A COMPARISON OF THE INFERENTIAL, COMPUTATIONAL, AND PREDICTIVE PERFORMANCE OF JOINT SPECIES DISTRIBUTION MODELS

By

David Peter Wilkinson

ORCID: 0000-0002-9560-6499

SUBMITTED IN TOTAL FULFILMENT
OF THE REQUIREMENTS OF THE DEGREE OF
DOCTORATE OF PHILOSOPHY

NOVEMBER 2019

SCHOOL OF BIOSCIENCES
FACULTY OF SCIENCE
THE UNIVERSITY OF MELBOURNE

© David Peter Wilkinson, 2019.

ABSTRACT

The standard correlative species distribution model accounts for the effect of the environment on species distributions without explicitly accounting for the effect of species interactions. Community ecology, in contrast, studies species co-occurrence patterns without accounting for the co-occurrence that could be explained by the shared response of species to the environment. Joint species distribution models (JSDMs) are a relatively recent development in the ecological literature that extend the single species distribution model framework to model multiple species simultaneously while accounting for species co-occurrence, and are an exciting prospect for bridging the gap between these two disciplines. Research into JSDMs is still in its early stages as a field and we have identified six outstanding problems in the literature that are potentially hampering the wider uptake of JSDMs by the wider ecological community.

The confluence of statistical and computational advancements has led to the rapid development of JSDMs, and the proposal of several models and implementations. These models are statistically complex but, while well defined as stand-alone studies, are described using a diverse array of model notation, terminology, and symbols. This often conflicting notation makes comparing the different JSDM implementations a daunting task to most ecologists who are trying to identify which model is best suited to their purpose. In this thesis I develop and present a clearly-defined, singular notation for the description of JSDMs and use it to elucidate the similarities and differences between seven different JSDM implementations.

All of the newly proposed JSDM implementations have been presented on completely different datasets to each other and thus any measure of inferential performance is not directly comparable. Without a direct comparison potential practitioners are unable to

make an informed decision about the performance of these models. In this thesis I conduct a direct comparison of seven JSDM implementations on six datasets to assess their inferential performance. I found that all JSDMs identified similar species-environment relationships to each other. All JSDMs identified species co-occurrence patterns in the same direction but with different strengths and uncertainties.

The newly proposed JSDMs have been implemented using a variety of software with different default or suggested parameters for model fitting. In addition to being fit to datasets of a varying size it is difficult to directly compare the computational performance of different implementations. Potential practitioners are going to be influenced by factors such as how fast a model can run and the size of dataset it is capable of feasible fitting to. In this thesis I perform a direct comparison of seven JSDMs on six datasets to directly assess their computational performance. I found a greater difference in computational performance between JSDMs than for inferential performance. I found evidence that scaling issues due to the size of datasets, and thus the number of parameters to estimate, was more prevalent for some implementations than others. A comparison of effective sample size in their respective Markov chain Monte Carlo regimes also indicated that some implementations are far more computationally efficient than others.

The use of JSDMs for prediction has only recently begun to be addressed in the literature, but has almost exclusively focussed on environment-only prediction types akin to those of single species models. The multivariate nature of JSDMs, however, allows for new, community-level predictions to be used. To date none of these prediction types have a formal definition for JSDMs. In this thesis I present four types of prediction that can be applied with JSDMs. Marginal predictions are environment-only predictions that do not account for the residual associations between species. Joint

predictions are a community-level prediction that can return either a probability of observing a particular assemblage or generate a series of plausible assemblages at a site. These two prediction types can then be conditioned upon the known occurrence states of one or more species at a site, and thus present conditional marginal and conditional prediction types respectively.

Single-species models have been subject to substantial research into the evaluation of their prediction, but this has seen minimal to date for JSDMs. In this thesis I present five classes of evaluation metrics that assess different aspects of species distributions and the community assemblage process. Threshold-independent and threshold-dependent metrics operate at the species-level, community dissimilarity and species richness metrics operate at the community-level, and likelihood-based metrics which assess model fit.

The literature on JSDM prediction has to date been limited in its comparisons of predictive performance. Studies have either compared a limited number of models, a subset of evaluation metrics, and/or generally only considered environment-only prediction types. In this thesis I present a comparison of predictive performance for six JSDMs and two stacked single-species models. I fit the models to 22 real and simulated datasets, predict with the four prediction types I have defined for JSDMs, and evaluate these predictions with up to 32 metrics from five classes. I found that likelihood-based metrics suggested the JSDMs were better fit to the data but that the other metric classes generally showed them to underperform compared to the standard stacked single species model for all prediction types. The stacked single species model with the SESAM constraint was found to consistently outperform all other models. This suggests an interesting avenue of future research by applying this constraint to JSDM predictions.

DECLARATION

This is to certify that:

i) the thesis only comprises of my own original work towards the PhD, except where indicated in the Preface,

ii) due acknowledgement has been made in the text to all other materials used,

iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies, and appendices.

David Peter Wilkinson

November, 2019

PREFACE

The work presented in this thesis is predominately my own. It was conducted under the supervision of Prof Michael McCarthy, Dr Reid Tingley, Dr Nick Golding, and Dr Gurutzeta Guillera-Arroita in the Quantitative and Applied Ecology Group in the School of BioSciences, Faculty of Science, The University of Melbourne. In order to facilitate journal publication both throughout and subsequent to the PhD candidature, Chapters 2-4 have been written has independent, stand-alone studies and include introduction and discussion sections of their own. This will lead to some unavoidable repetition between chapters. To be consistent with academic publishing convention, all chapters have been written using the collective first person plural "we".

Research is a collaborative process, and the chapter contributions of collaborators have been outlined below:

Chapter 2

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution*, **10**, 198–211.

Contributions: DW, NG, GGA, RT, and MM conceived the research ideas and designed the methodology. DW led the formation of the statistical notation, implemented the analysis, and led the writing of the manuscript. All authors contributed critically to manuscript revisions.

Chapter 3

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (in review). Defining and evaluating predictions of joint species distribution models.

Contributions: DW, NG, GGA, RT, and MM conceived the research ideas and designed the methodology. DW implemented the analysis, and led the writing of the manuscript. All authors contributed critically to manuscript revisions.

Chapter 4

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (in prep). A comparison of predictive performance of joint species distribution models for presence-absence data.

Contributions: DW, NG, GGA, RT, and MM conceived the research ideas and designed the methodology. DW implemented the analysis, and led the writing of the manuscript. All authors contributed critically to manuscript revisions.

Funding Support

My PhD candidature has been supported by an Australian Government Research Training Program Scholarship and an ARC Centre of Excellence for Environmental Decisions top-up scholarship. Conference travel has been supported by a Science Abroad Travelling Scholarship and a School of BioSciences Travel Grant.

Reproducibility

The code and data required to replicate the studies performed in this thesis can be found in the GitHub repositories: Doi90/JSDM_Inference (Chapter Two) and Doi90/JSDM_Prediction (Chapters Three and Four).

DEDICATION

For the two masked men, the man with the painted face, the man in the hat, the man in the DeLorean, the man with the chin, the man with the voice, the irreverent six, and those who revere the mountain of fate.

We have not, more than likely will not, and in some cases cannot meet, but without your influence I would not be who I am today. A different mindset begets a different thesis, and so your mark upon this one is undeniable. For one of you this mark is literal and is there to see for those who know where to look.

ACKNOWLEDGEMENTS

First, it would be prudent to thank my supervisors without whom I wouldn't have made it through the PhD: Michael McCarthy, Reid Tingley, Nick Golding, and Gurutzeta Guillera-Arroita. You took on the daunting task of a prospective student deciding to throw themselves into the deep end of statistical ecology without a background in it and taught them more than they ever expected. Thanks to my PhD committee: Peter Vesk, Brendan Wintle, and Jian Yen, whose insightful discussions helped shape this thesis throughout the whole process. And I would be remiss to not also thank Jose-Lahoz Monfort, Brendan (again!), and Don Driscoll whose supervision throughout my Masters put me in a position to be able to tackle a PhD. To QAECO at large, thanks for being a fantastic group to have been in for the past two degrees.

To everyone that thought highly enough of me abilities to pay me to work for them, you gave me the confidence that I actually might know what I'm talking about at least some of the time. Short-term or long-term, code monkey or tutoring, it meant a lot that you thought I could do it. So thanks to Nick, Guru, Nat, Terry, Matt, Jose, and Flanders.

I've been generously funded throughout the degree from a variety of sources that have allowed for national and international conference travel, and to just general exist during a PhD. Thanks to the Australian Government for the APA turned RTS scholarship, CEED for the top-up scholarship, and the Faculty of Science, School of BioSciences, Mick, and Guru for travel support.

I've been lucky enough to share my corner of the Old Library with an excellent group of other PhDs throughout the years. I might keep to myself most days (which is probably to your benefit) but I have always enjoyed our discussions. So, thanks to Linda, Manisha, Payal, Erica, Matthew, and Vishnu for the company. This isn't to say the rest of the Old Library crew isn't to be mentioned, but this thesis is long enough as it is...

I've also been lucky enough to be welcomed into a second group of excellent people while I've done the PhD in the Research Platforms Services gang. The work you do is highly appreciated by researchers all of the uni. So, thanks to Brendan, Lucy, Pablo, and Errol for the opportunity to help run workshops, and I need to specifically call out Nikki for bringing me into the fold. Most importantly, although he won't admit it, I need to

thank Flanders for effectively taking on the unofficial role as a mentor. I know I'm not the only one that has valued your advice over the years.

On the topic of unofficial mentors, I've got to thank Jian and Darren. Our chats over the years about my projects, your projects, and academia in general (and cricket with Darren) have been extremely valuable over the years.

Life supposedly also exists outside of a PhD (although I don't know when) and for which I owe thanks to my family, Andrew, Brad, Al, Jeremy, the past and present Costa's crew (because I've been there *way* too long to list you all), and Teh Wastoids from all over the world.

Finally, I have to thank Els, Saras, Kristina, and Lee for the friendships that fostered here in QAECO. I 100% would not have come out of the past twelve months or so in once piece (or as close as) without you. As colleagues I look up to you, but as friends I could not be more proud of what you've all accomplished. You've set a high bar for me to follow. Els, consider the favour re-payed a hundred fold. Our fortnightly Zoom chats have been the highlight of my week every time since you moved back home.

"Much is demanded of those to whom much is given"
First Canticle of the Hallowed Knights

TABLE OF CONTENTS

Abstract iii
Declaration vii
Preface ix
Dedication xi
Acknowledgements xiii
List of Figures xix
List of Tables xxiii

Chapter One: General Introduction 1
1.1. What is joint species distribution modelling? 3
1.2. What challenges are associated with selecting a joint species distribution model? 5
1.2.1. Model descriptions 5
1.2.2. Inferential performance 6
1.2.3. Computational performance 6
1.3. Prediction with joint species distribution models 7
1.3.1. Defining prediction 7
1.3.2. Evaluating predictions 8
1.4. Statement of the problem 9
1.5. Outstanding problems 9
1.6. Thesis outline 11
Reference list 15

Chapter Two: A Comparison of Joint Species Distribution Models for Presence-Absence Data 23
Preface 25
Abstract 27
2.1. Introduction 29
2.2. Materials & Methods 31
2.2.1. Approach 31
2.2.2. Datasets 31
2.2.3. The models 32
2.2.3.1. Multivariate probit regression (MPR) 34
2.2.3.2. Multivariate logistic regression (MLR) 34
2.2.3.2. Hierarchical multivariate probit regression (HPR) 35
2.2.3.4. Multivariate probit regression with latent factors (LPR) 36
2.2.3.5. Dimension reduction model (DPR) 37
2.2.3.6. Hierarchical multivariate probit regression with latent factors (HLR-S and HLR-NS) 39
2.2.4. Model fitting 40

2.2.5. Simulation comparison 43
2.3. Results 44
2.3.1. Computational comparison 44
2.3.2. Similarity of parameter estimates 46
2.4. Discussion 50
2.5. Conclusion 58
Reference list 59

Chapter Three: Defining and Evaluating Predictions of Joint Species Distribution Models 67
Preface 69
Abstract 71
3.1. Introduction 73
3.2. Materials And Methods 75
3.2.1. Types of probability 75
3.2.2. Joint SDMs 76
3.2.3. Prediction types 81
3.2.3.1. Marginal prediction 81
3.2.3.2. Joint prediction 82
3.2.3.3. Conditional prediction 83
3.2.3.4. Conditional marginal prediction 85
3.2.4. Evaluation metrics 86
3.2.5. Case study 89
3.2.6. Output analysis 91
3.3. Results 92
3.4. Discussion 95
3.5. Conclusion 99
Reference list 101

Chapter Four: A Comparison of Predictive Performance of Joint Species Distribution Models for Presence-Absence Data 107
Abstract 109
4.1. Introduction 111
4.2. Materials and methods 114
4.2.1. Models 114
4.2.2. Prediction methods 115
4.2.3. Datasets 118
4.2.4. Evaluation metrics 119
4.2.5. Model prediction comparisons 121
4.2.6. Analysis of results 122
4.2.7. Software 123
4.3. Results 123
4.3.1. Marginal prediction types 124

4.3.1.1. Threshold-independent and threshold-dependent metrics 124
4.3.1.2. Community dissimilarity metrics 126
4.3.1.3. Species richness 127
4.3.2. Joint prediction types 128
4.3.2.1. Threshold-independent and threshold-dependent metrics 128
4.3.2.2. Community dissimilarity metrics 129
4.3.2.3. Species richness 130
4.3.3. Likelihood-based metrics 130
4.3.4. SESAM performance 131
4.4. Discussion 131
4.4.1. Threshold-independent and threshold-dependent metrics . . 131
4.4.2. Community dissimilarity metrics 132
4.4.3. Likelihood 133
4.4.4 Species richness 134
4.4.5 SESAM performance 134
4.5. Conclusion 136
Reference list 137

Chapter Five: General Discussion 145
5.1. Overview 147
5.2. The six outstanding problems 147
5.2.1. P1: Model descriptions 147
5.2.2. P2: Inferential comparison 148
5.2.3. P3: Computational comparison 150
5.2.4. P4: Defining prediction 151
5.2.5. P5: Evaluating prediction 152
5.2.6. P6: Prediction comparison 153
5.3. Future work 155
5.4. Conclusion 156
Reference list 159

APPENDIX A: Published version of Chapter One 163
APPENDIX B: Forest plots of regression coefficients 177
APPENDIX C: LPR fitting issue 651
APPENDIX D: Regression coefficient uncertainty metrics 669
APPENDIX E: Correlation in regression and correlation coefficient estimates between JSDMs 699
APPENDIX F: Simulated dataset comparison 715
APPENDIX G: Evaluation metrics - Chapter Two 725
APPENDIX H: Forest plots for evaluation metric absolute values 731
APPENDIX I: Forest plots of the relative performance difference between the JSDM and SSDM 971

APPENDIX J: Model definitions 1021
APPENDIX K: Prediction definitions 1031
APPENDIX L: Evaluation metrics 1042
APPENDIX M: Results: Absolute value forest plots 1047
APPENDIX N: Results: Relative performance heatmaps 1273
APPENDIX O: Testing results for errors 1285

LIST OF FIGURES

2.1 Effective sample size of regression (a) correlation (b) estimates for each JSDM's respective MCMC samplers on the mosquito dataset. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively, and the black dots represent outliers. 45

2.2 Mean regression coefficients (dots) and 95% credible intervals (bars) for damselflies in the mosquito dataset, estimated by different JSDMs. 46

2.3 Correlation of mean regression coefficient estimates between JSDMs, averaged across all datasets. 47

2.4 Absolute value of the Gini coefficient of the variation in the regression coefficient estimates in the eucalypt dataset for all JSDMs. 48

2.5 Estimated correlation coefficients (top row) and the uncertainty of those estimates defined as the width of their 95% credible interval (bottom row) of all JSDMs on the frog dataset. 49

2.6 Average correlation of mean correlation coefficients between JSDMs across all datasets. 50

2.7 Relative correlation strength of all JSDMs compared to a baseline of the MPR model. Plot a) is for the four smallest datasets, and plot b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively. 51

2.8 Relative correlation uncertainty (95% credible interval width) of all JSDMs compared to a baseline of the MPR model. Plot a) is for the four smallest datasets, and plot b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively. HPR and HLR-S are missing from 8b as they did not complete within the time limit. 52

3.1 Visualisations of the normally-distributed latent variable, z_i, for two species (A and B). The mean of the distribution, μ_i, is 0.5 for species A and -1 for species B. The probability of occurrence of a species is equal to the area under the curve 78

where $z_i > 0$, shown here in grey. . . . . .

3.2 Visualisation of the multivariate, normally-distributed latent variable, z_{ij}, for a two species scenario. The mean of the distribution, μ_{ij}, on each species' respective axis is the same as their independent distributions from Figure 1 ($\mu_{iA} = 0.5$, $\mu_{iB} = -1$) and there is positive correlation of 0.75 between them. The contours of the probability distribution, the grey ellipses, indicate probability density values of 0.1, 0.3, 0.5, 0.7, and 0.9. The numbers in the four corners are the probabilities of the multivariate latent variable integrated in that quadrant, e.g. in the upper right quadrant; there is a probability of 0.16 that both species will occur at the site. 80

3.3. Five different prediction types possible with JSDMs. The boxes show the occurrence states of species *A*, *B*, and *C*. Question marks denote the species being predicted, empty boxes are species whose occurrence state is not being considered in that prediction type, and 1/0 a known presence/absence state of species being considered for prediction. *Prediction Type* is the name of the prediction methodology, and *Notation* shows the corresponding probabilistic notation, following the definitions in the text. 81

3.4 The conditional probability distribution of species B given species A is known to be present. This probability distribution is identical to that of Figure 3.2 except that it is truncated such that $z_A > 0$. The contours of the probability distribution, the grey ellipses, are the probability densities for 0.1, 0.3, 0.5, 0.7, and 0.9 proportions of the distribution's volume. The values in the four corners are the probabilities of the multivariate latent variable occurring in that quadrant. Since species A is known to be present, there is zero probability that the value of the latent variable falls in either of the left-hand quadrants. $Pr(z_B|z_A) = 0.23$, which is higher than the marginal distribution, $Pr(B) = 0.16$, of the un-truncated distributions seen in Figures 3.1b and 3.2. 85

3.5 Relative performance difference between the JSDM joint predictions and binary SSDM predictions for the community dissimilarity metrics. The forest plots show the mean and 95% credible interval for the difference between the JSDM and SSDM estimates from the MEMs. The difference is calculated as $(JSDM - SSDM/|SSDM|) * 100$. 94

4.1 Heat maps showing the relative performance of the different JSDMs compared to the SSDM for all metrics applicable to the marginal prediction types. Each ring on the y-axis represents a particular model, while each ray on the x-axis represents a different evaluation metric, 125

which have been clustered my class. Values in blue indicate better performance by the model compared to the SSDM, while red values indicate worse performance by the model compared to the SSDM. The four prediction types shown are a) probabilistic marginal, b) low-prevalence conditional marginal, c) medium-prevalence conditional marginal, and d) high-prevalence conditional marginal prediction. These heat maps have been generated using Circos (Krzywinski et al. 2009). . . .

4.2 Model performance for the bias (mean error) evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% quantiles of metric values, over species in each datasets, after accounting for dataset and fold using a linear mixed-effects model. The dashed dark grey line corresponds to the SSDM mean bias, and the dashed light grey line corresponds to 95% quantile from SSDM predictions. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs. This figure is illustrative of the absolute metric plots provided for all metrics and prediction types as supplementary information in Appendix M. SESAM is not plotted here as it does not make predictions for individual species. 127

4.3 Heat maps showing the relative performance of the different JSDMs and SESAM compared to the SSDM for all metrics applicable to the joint prediction types. Each ring on the y-axis represents a particular model, while each ray on the x-axis represents a different evaluation metric, which have been clustered my class. Values in blue indicate better performance by the model compared to the SSDM, while red values indicate worse performance by the model compared to the SSDM. The four prediction types shown are a) joint, b) low-prevalence conditional, c) medium-prevalence conditional, and d) high-prevalence conditional prediction. These heat maps have been generated using Circos (Krzywinski et al. 2009). 129

LIST OF TABLES

2.1 Dataset summary 32

2.2 Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an accompanying R package have code scripts available in the supplementary material of their source paper. . 41

2.3 Symbology. Matrix dimensions supplied in brackets 42

2.4 Model compatibility with datasets. Pairs marked with X were compatible. 44

2.5 Model runtimes (in hours). >168 corresponds to the cut-off time of seven days. . 45

3.1 Summary of evaluation metrics for JSDM predictions. 90

3.2 Summary of model performance for the different metric classes in each prediction type. ↑ = JSDM performed better,↓ = SSDM performed better, ↑↓ = mixed results, - = no difference between SSDM and JSDM, greyed out cells show where a metric class is not applicable for a prediction type. 93

4.1 Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an accompanying R package have code scripts available in the supplementary material of their source paper. 116

CHAPTER ONE

GENERAL INTRODUCTION

1.1. WHAT IS JOINT SPECIES DISTRIBUTION MODELLING?

Why species occur in certain places and not in others remains a fundamental question in ecology. A species geographic distribution is driven by abiotic factors, such as temperature or rainfall, and biotic factors (Hutchinson 1957), like species interactions, yet the ubiquitous correlative species distribution model (SDM) typically explores a species' relationship with the environment without explicitly considering the effect of species' interactions (Dormann *et al.* 2012; Wisz *et al.* 2013). Community ecology, in contrast, generally studies species co-occurrence patterns without accounting for the co-occurrence that can be explained by the shared response of species to the environment (Hardy 2008).

There have been several approaches to quantifying both species-environment relationships and species interactions in a single model. Early examples of this are found as extensions to the occupancy modelling framework (Guillera-Arroita 2017), and include explicitly modelling interaction terms in a linear model (MacKenzie, Bailey & Nichols 2004), or dominate/subordinate species relationships (Richmond, Hines & Beissinger 2010; Waddle *et al.* 2010). However, these approaches face practical and computational limits that restrict their application to only a handful of species simultaneously.

Another avenue of research into tackling this problem has involved incorporating species co-occurrence into the single-species SDM framework. These approaches include using distribution estimates of additional species alongside abiotic factors as predictor variables (Leathwick & Austin 2001; Araújo & Luoto 2007; Meier *et al.* 2010;

Pellissier *et al.* 2010), or restricting the predicted distribution of the target species to that of another it depends on (Schweiger *et al.* 2012). However, these approaches and the dominate/subordinate relationship extension of the occupancy model framework only incorporate uni-directional interactions in which one species affects the other but not vice-versa (Kissling *et al.* 2012; Pollock *et al.* 2014).

Recent years have seen the development of joint species distribution models (JSDMs) as an extension to the single-species SDM framework to incorporate the effect of biotic interactions in species communities (Ovaskainen, Hottola & Siitonen 2010; Kissling *et al.* 2012; Clark *et al.* 2014a, 2017; Pollock *et al.* 2014; Warton *et al.* 2015; Golding, Nunn & Purse 2015; Letten *et al.* 2015; Harris 2015; Thorson *et al.* 2016; Ovaskainen *et al.* 2016b,a; Hui 2016; Nieto-Lugilde *et al.* 2017). JSDMs bridge the gap between the distribution modelling and community ecology fields by simultaneously modelling multiple species in a single model while accounting for the effect of both species-environment relationships and residual associations between species on species co-occurrence patterns (Kissling *et al.* 2012; Pollock *et al.* 2014). The distinction between residual associations and biotic interactions is important as correlative models are not able to tease apart the potential drivers of species co-occurrence. Correlations between species estimated by a JSDM could potentially be explained by species interactions, or they could be driven by shared responses to unmeasured variables, dispersal, or other drivers of species' distributions (Dormann *et al.* 2018a; Zurell, Pollock & Thuiller 2018; Barner *et al.* 2018). Thus a JSDM informs on both the biotic and abiotic constraints on a species distribution, and allows inference into both the drivers of distributions and the community assemblage process.

1.2. WHAT CHALLENGES ARE ASSOCIATED WITH SELECTING A JOINT SPECIES DISTRIBUTION MODEL?

1.2.1. MODEL DESCRIPTIONS

The confluence of statistical advances and increased computational resource availability in recent years has led to a surge in publications on JSDMs. As a result, there have been several statistical models and computational implementations of JSDMs proposed in the literature. Each of these models is statistically complex and defined by their respective authors using different, and potentially conflicting, notation and terminology. Some models are described using the same symbol for different features (Pollock *et al.* 2014; Hui 2016) and/or different symbols for the same feature (Pollock *et al.* 2014 & Ovaskainen *et al.* 2016b; Pollock *et al.* 2014 & Hui 2016) and/or use a completely different model notation format to describe the same type of model (Pollock *et al.* 2014; Ovaskainen *et al.* 2016b; Hui 2016). Any symbols used in a model must be defined and thus, with significant effort, a reader can persevere and identify the similarities and differences between models. This becomes exceedingly difficult, however, when the terminology being used is conflicting. For JSDMs this is particularly apparent in the usage of the terms “latent variable” and “latent factor” which have been used interchangeably despite occurring simultaneously in some models (Pollock *et al.* 2014; Warton *et al.* 2015; Ovaskainen *et al.* 2016b). The diversity of methods and their presentation in the literature potentially hampers wider uptake of JSDMs by ecologists. A clear definition of statistical models and their notation and terminology, and adhering to some general convention, can facilitate the communication and understanding of

models presented in the literature (Lowry 1959; Elith, Leathwick & Hastie 2008; Elith *et al.* 2011; Edwards & Auger-Méthé 2019).

1.2.2. INFERENTIAL PERFORMANCE

All of the JSDMs that have been presented in the literature have been fit to datasets with different numbers of species, sites, or covariates, which occur in different geographic locations, and/or have different spatial scales. As such, the reported performance of the different models cannot be directly compared. The similarities or differences in the ecological relationships inferred by a model are going to be influenced by the dataset. The impact of model choice on the direction and/or strength of these relationships cannot be teased apart from the effect of the dataset. To directly compare the relationships estimated by different models, such as co-occurrence patterns, they must be fit to identical datasets (Barner *et al.* 2018).

1.2.3. COMPUTATIONAL PERFORMANCE

Each of the JSDM alternatives presented in the literature has been implemented with different computational infrastructure. These include Mathematica (Ovaskainen, Hottola & Siitonen 2010), MATLAB (Ovaskainen *et al.* 2016b; and re-implemented in R: Blanchet, Tikhonov & Norberg 2019), R in conjunction with C++ (Golding & Harris 2015), and R in conjunction with JAGS (Pollock *et al.* 2014; Hui 2016). In addition to the underlying software infrastructure, these models also differ in their default or suggested settings/parameters for using them. Add on to this an inherent difference in how well a model can scale with increasing dataset size, as the number of parameters to be estimated is dependent on the model, and a researcher may be left wondering which approaches are even practical for their application.

1.3. PREDICTION WITH JOINT SPECIES DISTRIBUTION MODELS

1.3.1. DEFINING PREDICTION

Arguably, the most common use for SDMs is to predict into un-sampled areas or new environmental/climate conditions (Elith & Leathwick 2009). Attempts at multi-species predictions have historically been limited to overlaying, or stacking, the predictions from multiple single-species models to create some measure of species richness or community assemblage predictions (Guisan & Rahbek 2011; Thuiller *et al.* 2015). However, doing so implicitly assumes that species occurrence probabilities are independent of each other (Calabrese *et al.* 2014).

As JSDMs are multi-species models that account for both species-environment relationships and residual associations between species, they open up several new prediction methods unavailable to single-species models. Early literature on the topic of using JSDMs for prediction were initially restricted to performing environment-only predictions akin to those of single-species models (Harris 2015; Zhang *et al.* 2018). By leveraging the additional information found in the residual correlations between species, JSDMs are able to perform community-level predictions that involve predicting the entire species assemblages.

An advantage JSDMs hold over single-species models is the ability to perform community-level predictions, but they have only recently been explored in the literature (Norberg *et al.* 2019). By simultaneously predicting all species while accounting for any residual correlations between them, it is purported that community-level predictions

should be more accurate than stacked single-species predictions that assume species' independence. Recently there has been a call for research into conditioning these predictions using the known occurrence state of one or more species in the community (Zurell *et al.* 2019; Norberg *et al.* 2019) but these prediction types have yet to be implemented. These prediction types allow for the exploration of ecological hypotheses, such as how community composition changes under the presence/absence of species archetypes (e.g. predators, competitors), or how easy-to-detect indicator species could be used to inform predictions of rare or hard-to-detect species in the community.

1.3.2. EVALUATING PREDICTIONS

There is abundant literature on the evaluation of single-species SDMs (Fielding & Bell 1997; Liu, White & Newell 2009; Lawson *et al.* 2014) but there has been minimal research into the evaluation of JSDM predictions. To date, studies have generally relied on the species-level metrics common to SDMs, such as AUC, bias-based metrics (e.g. root-mean-square error), and correlation-based metrics (e.g. Pearson's correlation coefficient) (Zhang *et al.* 2018; Zurell *et al.* 2019). Community-level predictions open up a suite of potential evaluation metrics in the form of the community dissimilarity indices widely used in community ecology (Legendre & De Cáceres 2013). Where the species-level metrics are assessed independently for each species and then averaged for a single model performance estimate, community-level metrics are assessed independently for each site and then averaged for a single model performance estimate.

To date there has not been any discussion in the literature about how JSDM predictions should be evaluated. Do species-level metrics still apply in the multi-species context of JSDMs? What insight can be gained from access to community dissimilarity indices? Are certain metrics only applicable to certain prediction types?

1.4. STATEMENT OF THE PROBLEM

Joint species distribution models are a promising development in ecology that has seen increased adoption in the literature in recent years. While the majority of the initial literature unsurprisingly focussed on the development of models and implementations (Pollock *et al.* 2014; Thorson *et al.* 2015; Ovaskainen *et al.* 2016b; Hui 2016; Harris 2016; Clark *et al.* 2017), we have also seen the development of extensions to the model framework to account for the influence of traits and phylogenies (Ovaskainen *et al.* 2017), different spatial scales (Ovaskainen *et al.* 2016a), the effect of environment on species associations (Tikhonov *et al.* 2017), and mixed data types (Clark *et al.* 2017). More recently we have begun to see research into using JSDMs for prediction (Zhang *et al.* 2018; Zurell *et al.* 2019; Norberg *et al.* 2019) and for applied purposes (Stjernman *et al.* 2019). However, research into JSDMs is still in relatively early stages as a field.

1.5. OUTSTANDING PROBLEMS

I have identified six outstanding problems that are potentially hampering the wider uptake of JSDMs by the wider ecological community:

P1: The rapid development of JSDMs has led to several proposed models and implementations that, while well-defined as stand-alone studies, have been presented using a diverse array of model notation, terminology, and symbols. It is unsurprising that early research into statistically complex methods has been led by statistical ecologists who have the aptitude to navigate the at times conflicting model definitions. For JSDMs to see wider uptake amongst practitioners, there needs to be a clearer definition of the similarities and differences inherent in the proposed JSDMs. How distinct are the different proposed modelling methods?

P2: Each newly proposed JSDM has been presented on completely different datasets to any other implementation and as such any inferred ecological relationships, or patterns thereof, cannot be directly compared. Potential practitioners are not able to make educated decisions about which JSDM to select based on inferential performance without a direct comparison on identical datasets. Are models performing similarly or do some outperform the rest? Is there a performance difference between estimates of species-environment relationships compared to the residual associations between species?

P3: Newly proposed JSDMs have been implemented in a variety of software with different default or suggested parameters for model fitting. JSDMs have also been presented on datasets that vary widely in their number of species, sites, and environmental variables. As such it is unclear which model, if any, is suitable for potential practitioners. Which software is free and which requires not insubstantial licensing fees? Are some models more computationally efficient than the others? Do some models scale better than others as the size of the dataset increases?

P4: There are several new prediction methods available to JSDMs but they do not have formal definitions. Early work on prediction with JSDMs has focussed on the environment-only predictions akin to those of single-species models, and only recently have some of these prediction types been implemented in practice. How do these new prediction types differ from each other in definition? How can these prediction types be implemented in practice?

P5: Once prediction has been performed using JSDMs, how should it be evaluated? There has been substantial research into the evaluation of single-species models but so far minimal into the multi-species predictions of JSDMs. Are the species-level metrics

of the single-species SDMs still applicable? What about the community dissimilarity metrics common in community ecology? Can some metrics only be evaluated against certain prediction types?

P6: There have been several comparisons of model predictive performance for single-species SDMs, but to date only one study doing so for JSDMs. However, this study does not utilise all of the prediction types available to JSDMs and only a select subset of potential evaluation metrics. A large scale comparison of JSDM predictive performance for all prediction types is currently lacking. Do some JSDMs outperform the others? How do JSDM prediction compare to those of stacked single-species models?

1.6. THESIS OUTLINE

In this thesis, I aim to address the aforementioned outstanding problems in joint species distribution modelling. By resolving these problems, I hope to facilitate wider uptake of JSDMs by the ecological community to better make use of large, multi-species datasets and recent developments in the availability of computational resources for conservation goals.

In the following section I outline the structure of this thesis and the aims of the chapters contained within.

Chapter 2: A comparison of joint species distribution models for presence-absence data

In this chapter, I address the outstanding problems P1, P2, and P3. To address P1 I present seven different JSDMs that have been proposed in the literature using a singular, clearly defined notation to elucidate the structural comparison between

models. I highlight that each of the JSDMs share the same core model, upon which they have built their own unique extensions. To address P2 and P3 I then fit all seven JSDMs to the six datasets that they were independently presented with in the literature to perform a large scale comparison of inferential and computational performance of the different implementations.

Chapter 3: Defining and evaluating predictions of joint species distribution models

In this chapter I address outstanding problems P4 and P5. To address P4 I present a formal definition of the different prediction types available to JSDMs including environment-only marginal predictions, full species assemblage joint predictions, and the conditional and conditional marginal predictions that leverage information on the known occurrence of species to improve predictions. To address P5 I outline five different classes of evaluation metric that can be used to evaluate JSDM predictions. These different classes analyse different aspects of predictive performance and are variously assessed at either the species- or community-level. I also present a case study comparing the performance of a standard SSDM against a simple JSDM on a single dataset to highlight both these new prediction methods and the evaluation of these predictions.

Chapter 4: A comparison of predictive performance of joint species distribution models for presence-absence data

In this chapter, I address the outstanding P6 problem. I perform a large scale comparison of predictive performance wherein I compare six JSDMs and two stacked single-species SDMs. I fit each model to 22 datasets (two real, twenty simulated) using either five-fold random or five-fold spatial block cross-validation depending on the dataset. For the JSDMs I then perform a total of nine different prediction types:

probabilistic and binary marginal predictions, binary joint predictions, binary conditional predictions under three scenarios (low-, medium-, and high-prevalence of the known species), and probabilistic conditional marginal predictions under three scenarios (low-, medium-, and high-prevalence of the known species). I perform both binary and probabilistic predictions, where appropriate, for the two single-species models. I then evaluate a suite of 32 evaluation metrics over five classes for all predictions. To compare model performance I fit a mixed effects model to each combination of prediction type and evaluation metric to evaluate the mean performance of each model

Chapter 5: General Discussion

In this chapter I synthesise my findings in Chapters 2-4 and present general conclusions. I discuss the inferential, predictive, and computational performance of the different JSDM implementations and the implications for their uptake by the wider ecological community. I also present some avenues for future work.

Appendices

For ease of navigation I present the supplementary material associated with each chapter in a single section at the end of the thesis.

REFERENCE LIST

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Barner, A.K., Coblentz, K.E., Hacker, S.D. & Menge, B.A. (2018). Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, **99**, 557–566.

Blanchet, F.G., Tikhonov, G. & Norberg, A. (2019). HMSC: Hierarchical Modelling of Species Community.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.

Clark, J.S., Gelfand, A.E., Woodall, C.W. & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, **24**, 990–999.

Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, **87**, 34–56.

Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., Moretti, M.D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S.I., Sheppard, C.S., Steinbauer, M.J., Zeuss, D. & Kraan, C. (2018a). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false

conclusions. *Global Ecology and Biogeography*.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.

Edwards, A.M. & Auger-Méthé, M. (2019). Some guidance on using mathematical notation in ecology (S. McMahon, Ed.). *Methods in Ecology and Evolution*, **10**, 92–99.

Elith, J. & Leathwick, J.J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, …*, **40**, 677–697.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.

Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which

drive the spatial distribution of a mosquito community. *Parasites & Vectors*, **8**, 367.

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.

Guisan, A. & Rahbek, C. (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

Hardy, O.J. (2008). Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, **96**, 914–926.

Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model (D. Warton, Ed.). *Methods in Ecology and Evolution*, **6**, 465–473.

Harris, D.J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, **97**, 3308–3314.

Hui, F.K.C. (2016). boral - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, **7**, 744–750.

Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to

modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54–64.

Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology*, **82**, 2560–2573.

Legendre, P. & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

Letten, A.D., Keith, D.A., Tozer, M.G. & Hui, F.K.C. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, **103**, 1264–1275.

Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution models: A review. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*, 4241–4247.

Lowry, W.P. (1959). On the Use of Mathematical Notation in Ecological Literature. *Ecology*, **40**, 492–492.

MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B.,

Guisan, A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.

Nieto-Lugilde, D., Maguire, K.C., Blois, J.L., Williams, J.W. & Fitzpatrick, M.C. (2017). Multiresponse algorithms for community-level modeling: review of theory, applications, and comparison to species distribution models. *Methods in Ecology and Evolution*, **12**, 3218–3221.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., Foster, S.D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N.A., Holt, R.D., Hui, F.K.C., Husby, M., Kålås, J.A., Lehikoinen, A., Luoto, M., Mod, H.K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N.E., Gravel, D. & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, **89**, 1–24.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, **7**, 549–555.

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514–2521.

Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, **7**, 428–436.

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software (J. Chave, Ed.). *Ecology Letters*, **20**, 561–576.

Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM) (J. McPherson, Ed.). *Methods in Ecology and Evolution*, **5**, 397–406.

Richmond, O.M.W., Hines, J.E. & Beissinger, S.R. (2010). Two-species occupancy models: a new paramaterization applied to co-occurence of secretive rails. *Ecological Applications*, **20**, 2036–2046.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, **21**, 88–99.

Stjernman, M., Sahlin, U., Olsson, O. & Smith, H.G. (2019). Estimating effects of arable land use intensity on farmland birds using joint species modeling. *Ecological Applications*, **29**, e01875.

Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C. &

Zipkin, E.F. (2016). Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, **25**, 1144–1158.

Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J. & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, **6**, 627–637.

Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015). From species distributions to meta-communities. *Ecology Letters*, **18**, 1321–1328.

Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context (D. Warton, Ed.). *Methods in Ecology and Evolution*, **8**, 443–452.

Waddle, J.H., Dorazio, R.M., Walls, S.C., Rice, K.G., Beauchamp, J., Schuman, M.J. & Mazzotti, F.J. (2010). A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, **20**, 1467–1475.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, **30**, 766–779.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping

distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Zhang, C., Chen, Y., Xu, B., Xue, Y. & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, **41**, 1876–1887.

Zurell, D., Pollock, L.J. & Thuiller, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co- occurrence data in homogenous environments? *Ecography*, 0–2.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, **00**, 1–13.

CHAPTER TWO

A COMPARISON OF JOINT SPECIES DISTRIBUTION MODELS FOR PRESENCE-ABSENCE DATA

PREFACE

The material presented in Chapter One has been published in *Methods in Ecology and Evolution*. The full citation is as follows:

> Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution*, 10, 198–211.

As per University regulations the author-accepted version of the manuscript is the format presented as Chapter One in the thesis. The published version of the manuscript, as presented in *Methods in Ecology and Evolution*, can be found in Appendix A.

The author-accepted version of the manuscript has only been modified to adhere to the section numbering and style formatting of the thesis at large. Content remains otherwise unmodified.

ABSTRACT

Joint species distribution models (JSDMs) account for biotic interactions and missing environmental predictors in correlative species distribution models. Several different JSDMs have been proposed in the literature, but the use of different or conflicting nomenclature and statistical notation potentially obscures similarities and differences among them. Furthermore, new JSDM implementations have been illustrated with different case studies, preventing direct comparisons of computational and statistical performance. We aim to resolve these outstanding issues by (i) highlighting similarities among seven presence-absence JSDMs using a clearly-defined, singular notation; and (ii) evaluating the computational and statistical performance of each JSDM using six datasets that vary widely in numbers of sites, species, and environmental covariates considered. Our singular notation shows that many of the JSDMs are very similar, and in turn parameter estimates of different JSDMs are moderate to strongly, positively-correlated. In contrast, the different JSDMs clearly differ in computational efficiency and memory limitations. Our framework will allow ecologists to make educated decisions about the JSDM that best suits their objective, and enable wider uptake of JSDM methods among the ecological community.

2.1. INTRODUCTION

Understanding the factors underlying species' geographic distributions is a fundamental goal of ecology. Abiotic factors, such as temperature or rainfall, and biotic factors, such as species interactions or population dynamics, ultimately drive species distributions (Hutchinson 1957). Yet correlative species distribution models (SDMs), the most common tool for predicting species distributions, typically quantify species' environmental relationships without explicitly considering effects of species interactions (Dormann *et al.* 2012; Wisz *et al.* 2013). In contrast, community ecology generally studies co-occurrence patterns and community structure without accounting for co-occurrence due to shared environmental responses among species (Hardy 2008).

Several statistical approaches have been proposed to quantify species' environmental relationships and interactions in a single model. Within the occupancy-detection modelling framework (Guillera-Arroita 2017), early examples explicitly modelled effects of biotic interactions via linear models, including pair-wise and higher-order effects (MacKenzie, Bailey & Nichols 2004), or dominate/subordinate species relationships, where the dominant species' occupancy state influences that of the subordinate species, but not vice versa (Richmond, Hines & Beissinger 2010; Waddle *et al.* 2010). Practical and computational limits restrict the application of these methods to communities of only a handful of species (Richmond, Hines & Beissinger 2010).

Other methods incorporate species co-occurrence data into the classical SDM framework. Distribution estimates of additional species can be used alongside abiotic variables as predictors (Leathwick & Austin 2001; Araújo & Luoto 2007; Meier *et al.* 2010; Pellissier *et al.* 2010), or a species' predicted distribution can be restricted to that of another it depends on (Schweiger *et al.* 2012). Like dominate/subordinate models,

however, these approaches only incorporate unidirectional interactions (Kissling *et al.* 2012; Pollock *et al.* 2014).

Joint species distribution models (JSDMs) have emerged as extensions of SDMs to capture the effects of biotic interactions in communities (Ovaskainen, Hottola & Siitonen 2010; Kissling *et al.* 2012; Clark *et al.* 2014b, 2017; Pollock *et al.* 2014; Warton *et al.* 2015; Golding, Nunn & Purse 2015; Letten *et al.* 2015; Harris 2015; Thorson *et al.* 2016; Ovaskainen *et al.* 2016b,a; Hui 2016; Nieto-Lugilde *et al.* 2017). JSDMs simultaneously model multiple species' distributions, accounting for both environmental relationships and residual associations (that might arise from species interactions) on species co-occurrence (Pollock *et al.* 2014). Thus, JSDMs inform about biotic and abiotic constraints on species' distributions (Pollock *et al.* 2014), and can improve predictions of community composition (Harris 2015). Promising advances in JSDMs have led to the development of several statistical models and computational implementations. This diversity of methods and their presentation might hamper wider use of JSDMs by ecologists, since prospective users have little guidance on which implementation best suits their objectives.

Our study formally compares seven presence-absence JSDMs that consider species interactions by characterizing the residual correlation among species not captured by environmental predictors. These models are statistically complex, and defined in their respective papers using different, potentially conflicting, notation and terminology. In addition, each JSDM was evaluated on a different dataset; any ecological relationships inferred are therefore not directly comparable, nor are reports of an implementation's computational efficiency. To help ecologists better understand the different models, we define them using a consistent notation to elucidate their similarities and differences, and compare their parameter estimates and computational performance against six

datasets. The results can help ecologists make informed decisions about appropriate JSDM selection.

2.2. MATERIALS & METHODS

2.2.1. APPROACH

Each JSDM considered here consists of three components: a Bayesian statistical model; software to fit the model to data; and default or suggested settings for model fitting (e.g., choice of priors and Markov chain Monte Carlo (MCMC) sampling regime). A JSDM's computational efficiency and parameter estimates will undoubtedly be affected by all of these components. Our aim was not to disentangle the effects of each of these components, but to highlight the differences and similarities between the overall methodologies, as applied in their respective papers. The term JSDM hereafter refers to these overall methodologies, unless otherwise stated.

2.2.2. DATASETS

Each JSDM was fit to six presence-absence datasets from recent JSDM papers (Ovaskainen, Hottola & Siitonen 2010; Pollock *et al.* 2014; Golding, Nunn & Purse 2015; Harris 2015; Ovaskainen *et al.* 2016b) to compare parameter estimates and computational efficiency across a range of datasets. Datasets not made publicly available were sourced from the authors. These datasets covered a broad range of taxa, geographic locations, and numbers of sites and species (Table 2.1**)**. Datasets were used as originally published with these exceptions: testing and training splits were merged; the Breeding Bird Survey online repository had been updated resulting in more sites; and we only used one of the eighteen fungi datasets from Ovaskainen, Hottola &

Siitonen (2010). Geographic coordinates were rescaled to have a maximum distance between sites of 1 to avoid issues of numerical instability in the spatial model.

Dataset	Source	Geographic Location	Species	Sites	Covariates
Birds	Harris (2015)	North America	370	2752	8
Butterflies	Ovaskainen *et al*. (2016a)	Great Britain	55	2609	4
Eucalypts	Pollock *et al*. (2014)	Grampians National Park, Australia	12	458	7
Frogs	Pollock *et al*. (2014)	Melbourne, Australia	9	104	3
Fungi	Ovaskainen *et al*. (2010)	Southern Finland	11	800	15
Mosquitoes	Golding *et al*. (2015)	South-East England	16	167	13

Table 2.1: Dataset summary

2.2.3. THE MODELS

We restrict our evaluation of JSDMs to seven models for species presence-absence data that account for species interactions via residual correlations. These JSDMs are extensions of the generalised linear modelling (GLM) framework, which is widely used for modelling species distribution data (Gelfand *et al.* 2006). The statistical models are defined below using a common notation. The following terms are consistent across all models: $\boldsymbol{y}$, the response variable; $1(.)$, an indicator function that returns 1 when the expression in brackets is true and 0 otherwise; $\boldsymbol{z}$, a normally-distributed latent variable; $\boldsymbol{\mu}$, the linear predictor for the measured covariates; **X**, the matrix of measured covariates; **β**, the matrix of regression coefficients; and **I**, the identity matrix. Subscript notation for sites is $i = 1, \dots, n$; for species $j = 1, \dots, J$; and for predictors $k = 1, \dots, K$.

Each JSDM is built on the foundation of Chib & Greenberg's (1998) multivariate probit regression model (hereafter, the core model). This model uses a latent variable

parameterisation of a probit model rather than the probit link directly. The probability of species presence is modelled as the probability of a latent multivariate normally-distributed variable exceeding a threshold, such that $y_{ij} = 1$ if $z_{ij} > 0$, and $y_{ij} = 0$ otherwise. The probability of species presence at a site is represented by a one-dimensional latent variable, while the community present at a site is represented by a multi-dimensional latent variable (see Pollock *et al* (2014) for a visual representation). This core model is described as follows:

$$y_{ij} = 1(z_{ij} > 0) \quad (1)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

The probability of species j being present at site i is the probability that latent variable z_{ij} is greater than zero. The latent variable is the sum of the linear predictor μ_{ij} and the correlated residual error e_{ij}. The linear predictor is the product of the measured environmental variables $\boldsymbol{X}_{i,.}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{.,j}$, as in generalised linear models. Correlations in the residual error $\boldsymbol{e}_i$ are captured in **R**, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements are restricted between -1 and 1. Standard deviations are constrained to equal to 1 in probit regression, thus covariance and correlation matrices are equivalent. The elements of **R** reflect species co-occurrence patterns not described by the environmental predictors (i.e. species interactions, or missing predictors). A limitation of this model is that it does not specify overdispersion in species occurrences like other extensions of the GLM framework (e.g., generalised linear mixed models), but it could be extended to do so.

The following sections define statistical models and describe the software and MCMC regimes of the JSDMs. All models were fit using their default setting for priors and MCMC regimes. Table 2.2 provides an overview of the statistical models' respective features, and Table 2.3 provides a glossary of all introduced notation.

2.2.3.1. MULTIVARIATE PROBIT REGRESSION (MPR)

The multivariate probit regression (MPR) model used by Golding *et al* (2015), *BayesComm*, is identical to the core model. The regression coefficients have a normal prior, $\beta \sim N(0, 10)$, and the correlation coefficients an inverse Wishart prior with $n + 2J$ degrees of freedom and scale matrix **I**. MPR was fit in R (R Core Team 2016) by MCMC using a Gibbs sampler implemented in R and C++. A single MCMC chain of 11,000 samples, discarding the first 1,000 as burn-in, sampled the posterior distribution.

2.2.3.2. MULTIVARIATE LOGISTIC REGRESSION (MLR)

Ovaskainen, Hottola & Siitonen (2010) introduced a logistic regression version of the core model based on O'Brien & Dunson (2004). The model (hereafter MLR) is as follows:

$$
\begin{aligned}
y_{ij} &= 1\left(z_{ij} > 0\right) \\
z_{ij} &= \mu_{ij} + logit\left(\Phi\left[e_{ij}\right]\right) \\
\mu_{ij} &= \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j} \\
\boldsymbol{e}_i &\sim MVN(\mathbf{0}, \mathbf{R})
\end{aligned}
\tag{2}
$$

Here, Φ is the cumulative density function of the standard normal distribution. The regression coefficients have a normal prior, $\beta \sim N(0{,}10)$, and the correlation coefficients a uniform prior, $U(-1{,}1)$. MLR was fit in Mathematica (Wolfram Research

Inc 2016), and the posterior distributions sampled via MCMC using a Metropolis-within-Gibbs sampler with 100,000 samples (thinned to keep 1 every 10) discarding the first 5,000 as burn in

2.2.3.3. HIERARCHICAL MULTIVARIATE PROBIT REGRESSION (HPR)

The hierarchical multivariate probit regression model of Pollock *et al* (2014) (hereby HPR) models the regression coefficients hierarchically such that β_{jk} is drawn from a normal distribution with mean ω_k and standard deviation σ_k.

$$y_{ij} = 1(z_{ij} > 0) \tag{3}$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\beta_{jk} \sim N(\omega_k, \sigma_k)$$

$$e_i \sim \boldsymbol{MVN}(\mathbf{0}, \mathbf{R})$$

Regression coefficients have a normal prior, $N(0, 100)$, on ω and a uniform prior, $U(0{,}100)$, on σ. Correlation coefficients have an inverse Wishart prior with $J + 1$ degrees of freedom and an **I** scale matrix. HPR was fit in R, and the posterior distributions sampled via MCMC using Gibbs sampling in JAGS (Plummer 2003), with three chains of 1,000,000 samples (thinned to keep 1 every 1,000) with the first 15,000 discarded as burn in.

2.2.3.4. MULTIVARIATE PROBIT REGRESSION WITH LATENT FACTORS (LPR)

The *boral* JSDM (Hui 2016) is a multivariate probit regression model with latent factors (hereafter LPR).

$$y_{ij} = 1(z_{ij} > 0) \quad (4)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$$

$$\varepsilon_i \sim N(\mathbf{0}, \mathbf{I})$$

This differs from the core model by using latent factors ("hypothetical" unmeasured variables) to explain any residual covariation that is not accounted for by the measured variables. This explained variation is the product of H unmeasured latent factors, $\boldsymbol{\eta}_{i,.}$ ($h = 1, \ldots, H$), and the factor loadings, $\boldsymbol{\lambda}_{.,j}$, of species j to latent factor h. We can interpret $\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$ and $\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$ as linear predictors for measured and unmeasured variables respectively. Now z_{ij} is the sum of the linear predictor of measured covariates, μ_{ij}, the linear predictor of unmeasured covariates, ν_{ij}, and uncorrelated residual error ε_{ij}. This model can be defined equivalently as an extension of MPR where the correlation matrix has low-rank structure, modelled as: $\mathbf{R} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{I}$ (Warton *et al.* 2015). Species interactions can be estimated by converting the factor loadings to correlation coefficients using this expression.

Latent factors bring considerable computational benefits by reducing the number of coefficients to estimate. A full-rank correlation matrix has $J(J-1)/2$ parameters, while

a factor loadings matrix only JH. The parameter ratio between latent factor and correlation matrix models is therefore $2H/(J-1)$, so the parameter reduction increases with J. Where $J = 10$ and $H = 2$, the latent factor model has 20 parameters rather than 45 (44%), whilst for $J = 100$ and $H = 5$ the latent factor model has 500 parameters rather than 4,950 (10%). This reduction does not necessarily correspond to an increase in parsimony, as the choice of priors for these matrices also impact model complexity. As latent factors are trying to approximate a fully unstructured correlation matrix, a greater H provides a better approximation, while a smaller H reduces computational requirements. The number of latent factors balances model simplicity and model fit, and 2-8 latent factors has been suggested as suitable in practice (Warton *et al.* 2015; Hui 2017).

The regression coefficients in the LPR implementation have a normal prior, $\beta \sim N(0, 10)$, and the latent factors a normal prior, $\eta \sim N(0,4.47)$, constrained such that upper diagonal elements are 0, and the diagonal elements are positive, $U(0,20)$. LPR was fit in R using the *boral* package. The posterior distribution was sampled via MCMC, with a Gibbs sampler using JAGS, in a single chain of 60,000 samples (thinned to keep 1 every 50) discarding the first 10,000 as burn in.

2.2.3.5. DIMENSION REDUCTION MODEL (DPR)

The multivariate generalised regression model of Clark *et al* (2017), *gjam* (hereby DPR), fits various types of response data. For presence-absence data it is a multivariate probit regression model that takes on two different forms depending on the size of the dataset. The small dataset form is equivalent to the core model. For datasets above a size threshold ($J > 100$ or $J > \frac{2}{3} * n - 1$) , it is similar to LPR, but with an

additional dimension reduction step (with respect to species) using the Dirichlet process (Taylor-Rodríguez *et al.* 2017):

$$
\begin{aligned}
y_{ij} &= 1\left(z_{ij} > 0\right) \\
z_{ij} &= \mu_{ij} + \nu_{ij} + \varepsilon_{ij} \\
\mu_{ij} &= \boldsymbol{X}_{i,\cdot}\boldsymbol{\beta}_{\cdot,j} \\
\nu_{ij} &= \boldsymbol{\eta}_{i,\cdot}\boldsymbol{\lambda}_{\cdot,j} \\
\boldsymbol{\lambda}_{\cdot,j} &= \mathbf{A}_{\boldsymbol{l}_j} \\
\varepsilon_i &\sim N(\mathbf{0}, \mathbf{I})
\end{aligned}
\tag{5}
$$

The Dirichlet process assumes that groups of species will have the same response to the unmeasured variables, so *a priori* reduces $\boldsymbol{\lambda}$ along the j-dimension to a matrix, $\mathbf{A}$, for a fixed number, L, of species archetypes (rather than J species). The Dirichlet process can be seen as a distribution of distributions, and is used to cluster species into archetypes, via an index variable $\boldsymbol{l}$. Using $\mathbf{A}$ as a look-up table for the low-rank $\boldsymbol{\lambda}$ allows conversion back to full-rank $\boldsymbol{\lambda}$, and then estimate correlation coefficients via the factor loadings. The parameter ratio of DPR to a full-rank correlation matrix is $2LH/(J(J-1))$, which is typically much smaller than 1. This decouples the number of parameters to estimate from J, though adequate L and H must be used to accurately model interactions.

The regression coefficients in the DPR implementation have an improper uniform prior, $U(-Inf, Inf)$, correlation coefficients use an inverse-Wishart prior with $n - K + J - 1$ degrees of freedom and an $\mathbf{I}$ scale matrix, and latent factors are normally distributed as $\eta_{ih} \sim N(0,1)$. DPR was fit in R using the *gjam* package. The posterior distribution was sampled via MCMC using a Gibbs sampler in R of a single chain of 60,000 samples (thinned to keep 1 every 50), discarding the first 10,000 as burn in.

2.2.3.6. HIERARCHICAL MULTIVARIATE PROBIT REGRESSION WITH LATENT FACTORS (HLR-S AND HLR-NS)

The last two JSDMs are spatially-explicit (HLR-S) and non-spatially-explicit (HLR-NS) versions of the hierarchical multivariate probit regression with latent factors used by Ovaskainen *et al*. (2016b):

$$
\begin{aligned}
y_{ij} &= 1\left(z_{ij} > 0\right) \\
z_{ij} &= \mu_{ij} + \nu_{ij} + \varepsilon_{ij} \\
\mu_{ij} &= \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j} \\
\beta_j &\sim MVN\left(\boldsymbol{\omega}_{\boldsymbol{l}_j}, \boldsymbol{\sigma}\right) \\
\nu_{ij} &= \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j} \\
\varepsilon_i &\sim N(0,1)
\end{aligned}
\tag{6}
$$

The HLR models make use of latent factors like LPR and DPR, but also use hierarchical regression coefficients like HPR. However, instead of independent normal distributions unique to the K covariates, $\beta_{.,j}$ is drawn from a multivariate normal distribution, allowing for correlation among coefficients. The mean of this distribution, $\boldsymbol{\omega}_{\boldsymbol{l}_j}$, is unique to species' archetype, $\boldsymbol{l}$. Archetype is specified *a priori* in HLR models using trait data, assuming that species with similar traits exhibit similar responses. Here, $\boldsymbol{l}_j$ is set to 1 as trait data was unavailable for most datasets. Thus, each species' $\beta_{.,j}$ are independently drawn from a shared distribution. The hierarchical regression coefficients have a normal prior, $N(0,1)$ for $\boldsymbol{\omega}$, and an inverse Wishart prior with five degrees of freedom and a **I** scale matrix for $\boldsymbol{\sigma}$. The latent factors are normally distributed as $\eta_{ih} \sim N(0,1)$ in HLR-NS, and as independent spatially homogenous Gaussian processes with exponential covariance in HLR-S:

$$Cov(\eta_{ih}, \eta_{i'h}) = exp(-d_{ii'}/\alpha_h) \quad (7)$$

Here, d is the spatial distance between sampling units, and α is a positive parameter controlling decay in correlation with distance. At zero distance the covariance function is 1, so the latent factors have zero mean and unit variance like in the non-spatial model, and this declines towards zero with increasing α.

These JSDMs were fit in MATLAB (The MathWorks Inc 2016). The posterior distribution was sampled via MCMC using a Gibbs sampler in a single chain of 50,000 samples (thinned to keep 1 every 50) with the first 10,000 discarded as burn in.

2.2.4. Model Fitting

Model fitting was carried out on desktop computers with identical specifications (Windows 7 OS, i7-6700 CPU @ 3.40GHz, 16GB RAM) to compare computational performance. JSDM software that explicitly enabled model fitting in parallel was run so (HPR), while those that did not were run sequentially. An upper time limit for runtime was set at seven days, after which the JSDM was deemed incompatible with that dataset on a desktop computer. JSDMs that reached this limit were then run on high performance computing infrastructure with increased memory availability (Boab: Dual Intel Xeon E5-2699 CPUs @ 2.30GHz, ~40GB RAM per model. Spartan: Intel(R) Xeon(R) CPU E5-2643 v3 @ 3.40GHz, 200GB RAM (University of Melbourne 2017)) with the same time limit. JSDMs were fit using: Mathematica 10.4 (Wolfram Research Inc 2016) (MLR), MATLAB R2016a (The MathWorks Inc 2016) (HLR models), and R 3.2.5 (R Core Team 2016) (all others). Model-specific R packages were *BayesComm* 0.1-2 (Golding & Harris 2015) (MPR), *boral* 1.3.1 (Hui 2017) (LPR), and *gjam* 2.1.1

	Source Paper	R Package	Hierarchical Regression Coefficients	Dirichlet Process	Latent Factors	Multivariate	Probit	Logistic	Regression	Spatially-Explicit	Other Compatible Data Types
MPR	Golding *et al*. (2015)	*BayesComm*				X	X		X		
MLR	Ovaskainen *et al*. (2010)	NA				X		X	X		
HPR	Pollock *et al*. (2014)	NA	X			X	X		X		
LPR	Hui (2016)	*boral*			X	X	X		X		Abundance
DPR	Clark *et al*. (2017)	*gjam*		o	o	X	X		X		Multiple including continuous and discrete abundance, and ordinal counts
HLR-S	Ovaskainen *et al*. (2016a)	NA	X		X	X	X		X	X	
HLR-NS	Ovaskainen *et al*. (2016a)	NA	X		X	X	X		X		

Table 2.2: Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an accompanying R package have code scripts available in the supplementary material of their source paper.

Symbol	Definition
Subscripts	
i	Site. $i = 1, \dots, n$
j	Species. $j = 1, \dots, J$
k	Measured covariate. $k = 1, \dots, K$
h	Latent factor / Unmeasured covariate. $h = 1, .., H$
l	Species archetype. $l = 1, \dots, L$
Main Terms	
y	Binary response variable (species presence/absence)
$1(.)$	Indicator function
z	Normally-distributed latent variable
μ	Linear predictor for the measured covariates
ν	Linear predictor for the unmeasured covariates
$\mathbf{X}$	Matrix of measured covariates $(n \cdot K)$
$\boldsymbol{\beta}$	Matrix of regression coefficients $(K \cdot J)$
$\boldsymbol{\eta}$	Matrix of latent factors $(n \cdot H)$
$\boldsymbol{\lambda}$	Matrix of factor loadings $(H \cdot J)$
$\mathbf{I}$	Identity matrix $(J \cdot J)$
$\mathbf{A}$	Archetype-reduced factor loadings matrix $(H \cdot J)$
$\boldsymbol{\Sigma}$	Covariance matrix $(J \cdot J)$
$\mathbf{R}$	Symmetric, positive-definite correlation matrix $(J \cdot J)$
e	Correlated residual error
ε	Uncorrelated residual error
Φ	Cumulative density function of $N(0,1)$
ω	Mean of the normal distribution for hierarchical β coefficients
σ	Standard deviation of the normal distribution for hierarchical β coefficients
d	Spatial distance
α	Positive parameter controlling decay in correlation with distance for latent factors

Table 2.3: Symbology. Matrix dimensions supplied in brackets

(Clark *et al.* 2017) (DPR). JAGS 4.3.0 (Plummer 2003) was used for some MCMC sampling.

ESS estimates are not directly comparable unless each JSDM was fit using identical MCMC regimes, so each JSDM was fit to the mosquito dataset a second time with identical MCMC regimes, consisting of a single chain of 11,000 iterations with no thinning and the first 1,000 discarded as burn-in. ESS was calculated separately for

regression and correlation coefficients by dividing the mean effective number of samples of each coefficient in the posterior distribution by the run-time.

Statistical inference was assessed by comparing the estimated regression and correlation coefficients. For latent factor JSDMs, the residual covariance matrix was recovered from the factor loadings ($\mathbf{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{I}$) and scaled to the correlation matrix $\mathbf{R}$. To evaluate the similarity of JSDM regression coefficient estimates, we compared posterior mean values and their 95% credible intervals, and calculated the Pearson correlation coefficient between mean estimates of each coefficient of each JSDM. Variation in coefficient estimates was evaluated using four uncertainty metrics calculated from post-hoc standardised posterior distributions: coefficient of variation, Gini coefficient, quartile coefficient of dispersion, and 90% quantile coefficient of dispersion. Similarity of correlation coefficient estimates was evaluated by calculating the Pearson correlation coefficient between mean estimates for each JSDM, and the magnitude and uncertainty (95% credible interval width) of estimates was calculated relative to the simplest JSDM (MPR) as the baseline.

2.2.5. Simulation Comparison

We also compared a subset of these models (MPR, HPR, LPR, DPR, and HLR-NS) when fit to datasets simulated under different modelling component assumptions. See Appendix S5.

2.3. RESULTS

2.3.1. COMPUTATIONAL COMPARISON

Only four JSDMs ran to completion within the time limit for all datasets: MPR, LPR, DPR, and HLR-NS (Table 2.4). MLR did not complete for four datasets, took 142 hours for the eucalypt dataset (nearly three times longer than HLR-S, and eight times longer than the closest non-spatial model, HPR), and 14 hours for the small frog data set; it was therefore deemed too slow for practical use and not considered further. Both HPR and HLR-S failed to complete within the time limit for the bird and butterfly datasets on desktop or high-performance computing resources. HLR-S was incompatible with the non-spatial fungi dataset.

	MPR	**MLR**	**HPR**	**LPR**	**DPR**	**HLR-S**	**HLR-NS**
Birds	X			X	X		X
Butterflies	X			X	X		X
Eucalypts	X	X	X	X	X	X	X
Frogs	X	X	X	X	X	X	X
Fungi	X		X	X	X		X
Mosquitoes	X		X	X	X	X	X

Table 2.4: Model compatibility with datasets. Pairs marked with X were compatible.

MLR, HLR-S, and HLR-NS were the slowest for converting the datasets into the required format (40 +/- 29 minutes). These were the only JSDMs implemented in programming languages other than R, the language with which the experimenter was most familiar. Datasets for all other JSDMs were prepared within an average of five minutes.

MPR was the fastest to run to completion for all datasets, followed by LPR for the two smallest datasets, and HLR-NS was fastest for the four largest (Table 2.5**)**. HPR and HLR-S were considerably slower for all datasets. MPR and HLR-NS had substantially greater sampler efficiency than the other JSDMs; MPR was most efficient for the regression coefficients (Figure 2.1a), whereas HLR-NS was most efficient for correlation coefficients (Figure 2.1b).

	MPR	**MLR**	**HPR**	**LPR**	**DPR**	**HLR-S**	**HLR-NS**
Birds	3.8	>168	NA	120.4	27.3	>168	15.2
Butterflies	0.23	>168	>168	13.9	6.5	>168	2.1
Eucalypts	<0.02	142.1	7.6	0.33	0.25	50.0	0.21
Frogs	<0.02	14.1	0.94	0.04	0.06	1.4	0.13
Fungi	<0.02	>168	15.8	0.62	0.67	NA	0.26
Mosquitoes	<0.02	>168	6.4	0.14	0.73	2.0	0.2

Table 2.5: Model runtimes (in hours). >168 corresponds to the cut-off time of seven days.

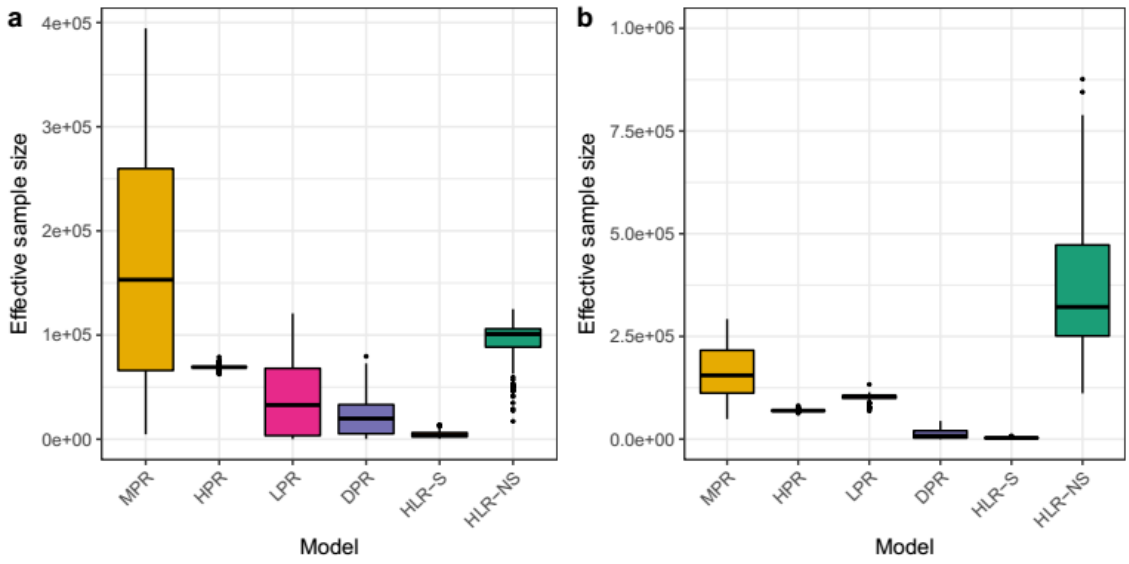


Figure 2.1: Effective sample size of regression (a) correlation (b) estimates for each JSDM's respective MCMC samplers on the mosquito dataset. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively, and the black dots represent outliers.

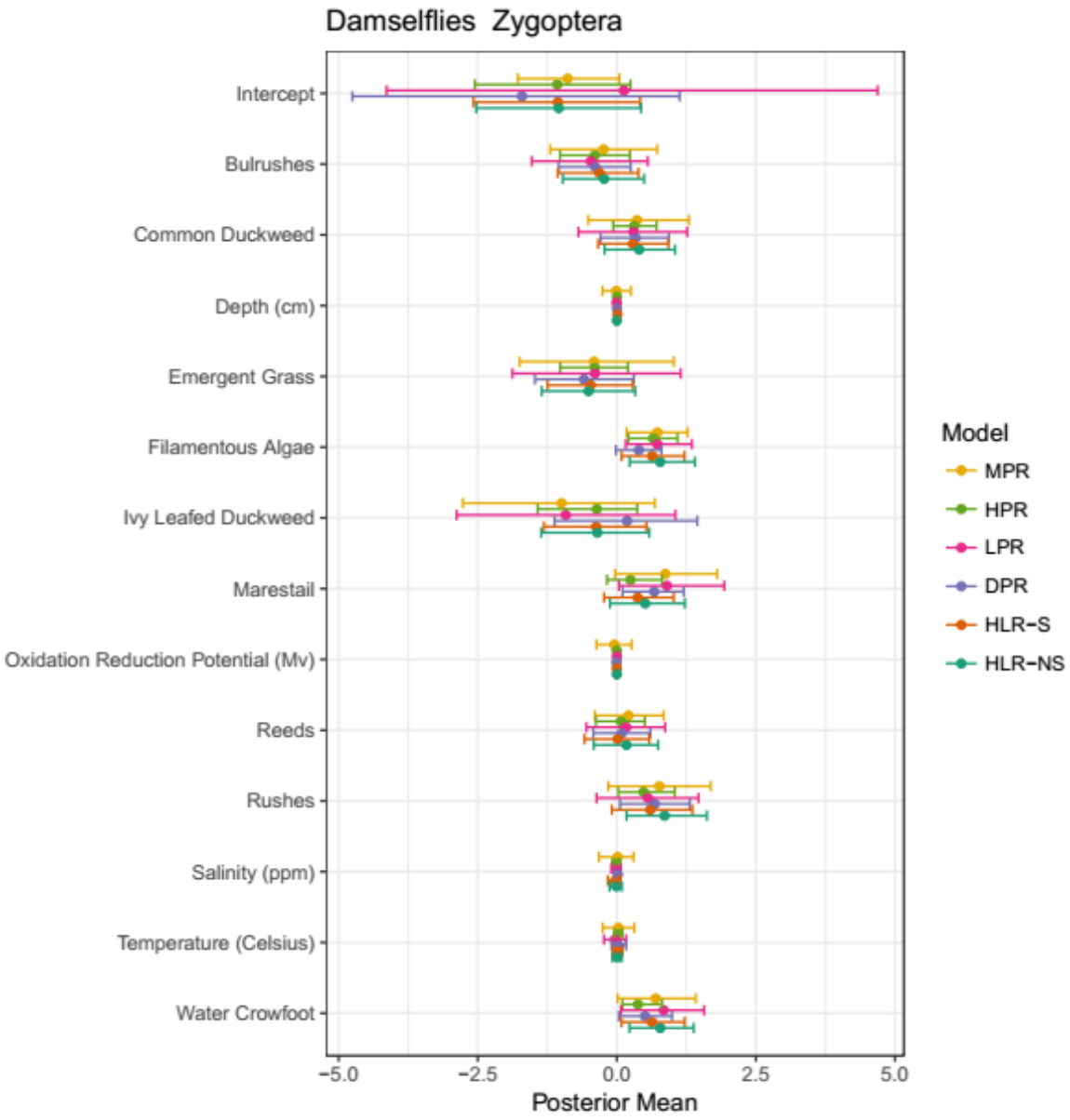


Figure 2.2: Mean regression coefficients (dots) and 95% credible intervals (bars) for damselflies in the mosquito dataset, estimated by different JSDMs.

2.3.2. Similarity of Parameter Estimates

All JSDMs yielded similar regression coefficients across all datasets (see Figure 2.2 for one species; all species in Appendix B). LPR parameter estimates for binary variables in the fungi and mosquito datasets were markedly more uncertain and frequently stronger than those estimated by other JSDMs for species-environment pairs with complete

separation (see Appendix B, Figure 458). Results presented here are based on default priors, but the impact of complete separation is alleviated with more informative priors (Appendix C).

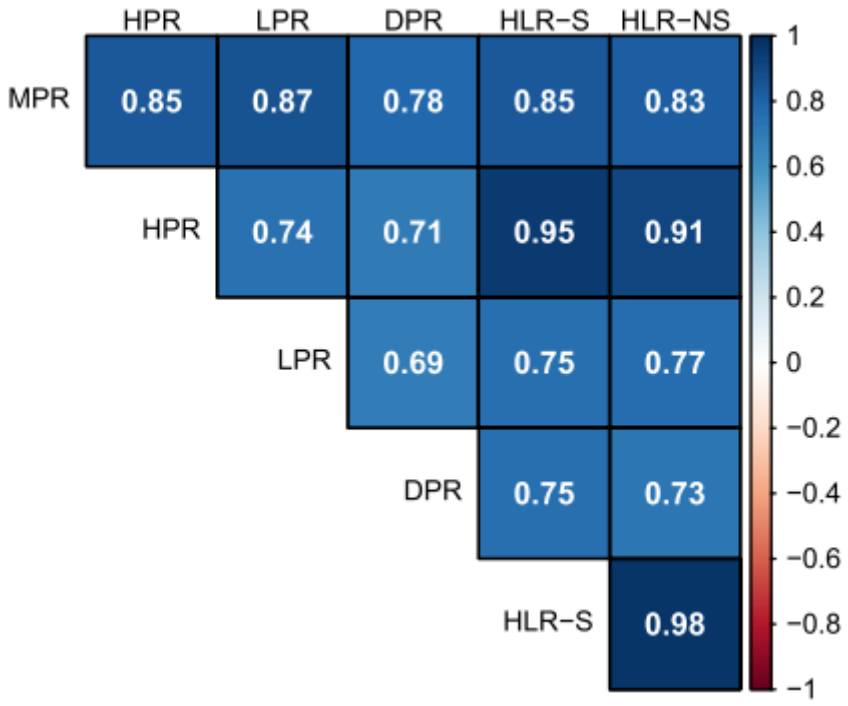


Figure 2.3: Correlation of mean regression coefficient estimates between JSDMs, averaged across all datasets.

Figure 2.3 shows the correlation of mean regression coefficient estimates between JSDMs, averaged across all datasets (individual datasets in Appendix E). There is strong correlation (>0.69) between all JSDM estimates, with the strongest correlations between HPR, HLR-S and HLR-NS (>0.9), and between MPR and all models (>0.75).

There was no consistent, substantial difference in the variation of regression coefficient estimates between JSDMs (see Figure 2.4 for the Gini coefficient for all species in the Eucalypt dataset). In a small number of instances, there were differences between JSDMs for particular coefficients (Appendix D, Figure 21), but there was no apparent

pattern to these occurrences, and Appendix D Figures 25-28 show only minor differences between JSDMs for each uncertainty metric across all datasets. There was no difference in observed patterns between different uncertainty metrics.

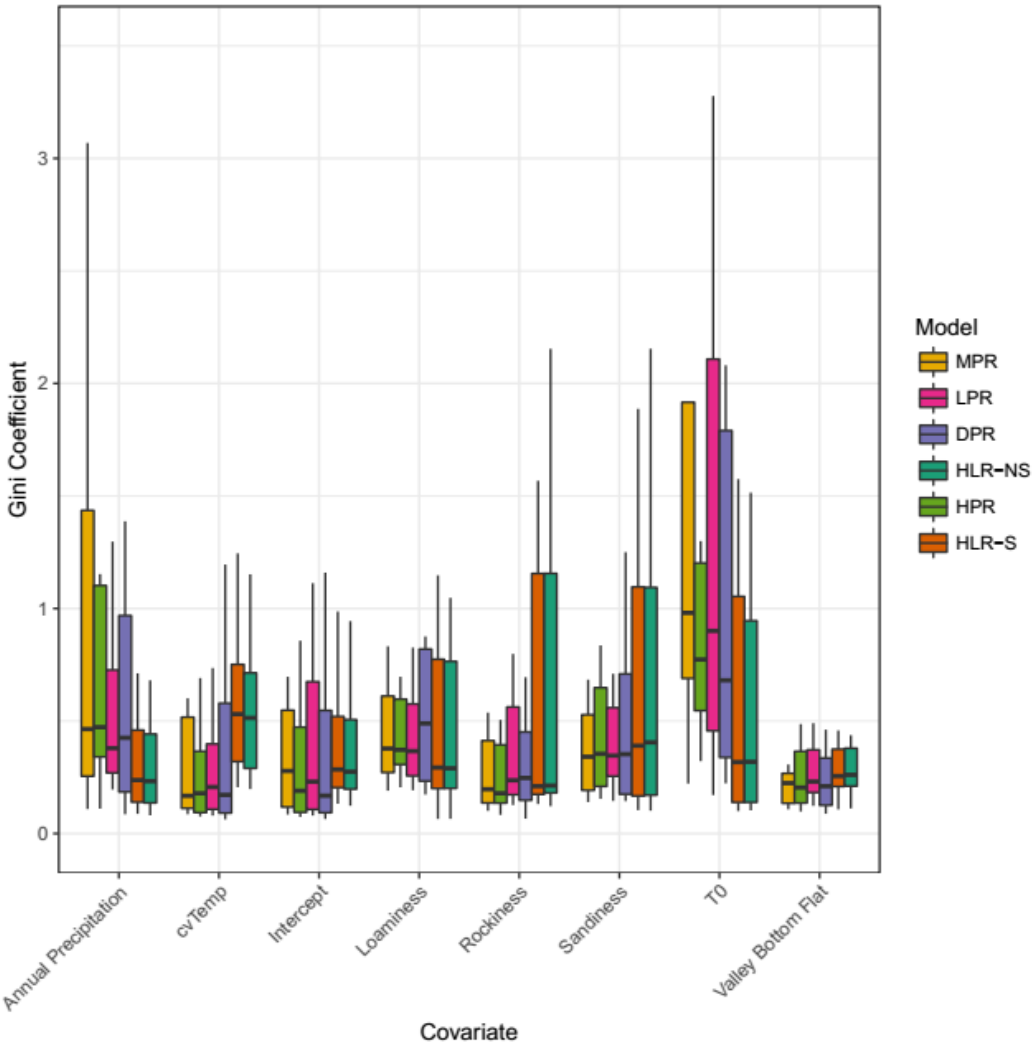


Figure 2.4: Absolute value of the Gini coefficient of the variation in the regression coefficient estimates in the eucalypt dataset for all JSDMs.

All JSDMs identified similar patterns in estimated correlation coefficients (Figure 2.5). The direction of species interactions was largely the same among JSDMs, but estimates differed in strength and uncertainty. Observed differences in the direction of estimated correlations between JSDMs involved weak and uncertain coefficients.

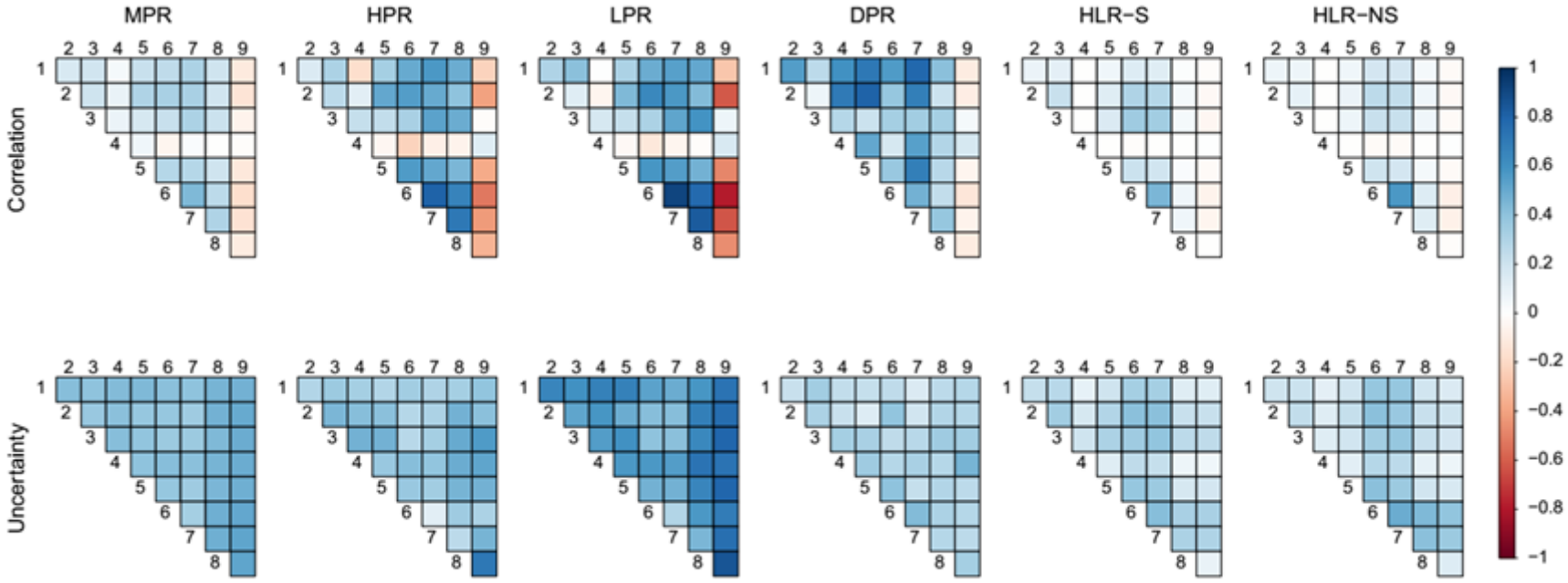


Figure 2.5: Estimated correlation coefficients (top row) and the uncertainty of those estimates defined as the width of their 95% credible interval (bottom row) of all JSDMs on the frog dataset.

Figure 2.6 shows the correlation of mean correlation estimates between JSDMs for all datasets (individual datasets in Appendix E). Strong, positive correlations (>0.7) were observed between most JSDMs. However, there was only moderate correlation (>0.49) between DPR and all JSDMs except HPR (0.77), and between LPR and the HLR-S (0.57) and HLR-NS (0.63).

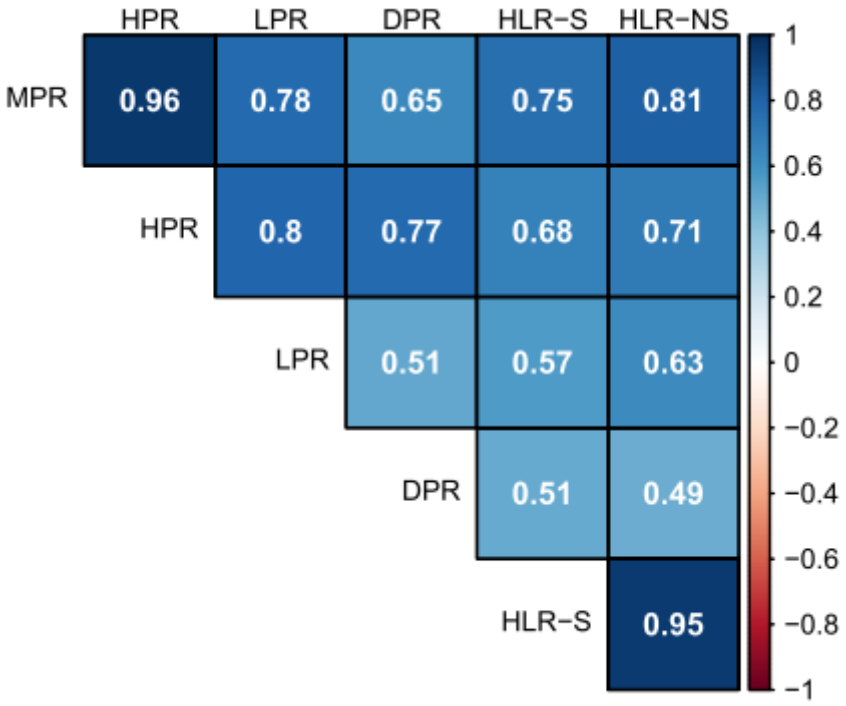


Figure 2.6: Average correlation of mean correlation coefficients between JSDMs across all datasets.

The strength of the JSDM correlation estimates is shown in Figures 2.7a and 2.7b. We split this comparison between the two largest (birds and butterflies) and four smallest datasets, as sample size imbalances between datasets masked observable trends. Averaged across the four smallest datasets, DPR, HLR-S, and HLR-NS estimated the weakest correlation strengths relative to MPR, yet for the two largest datasets HLR-NS

estimates were slightly stronger than MPR on average. HPR and LPR consistently estimated stronger correlations than MPR.

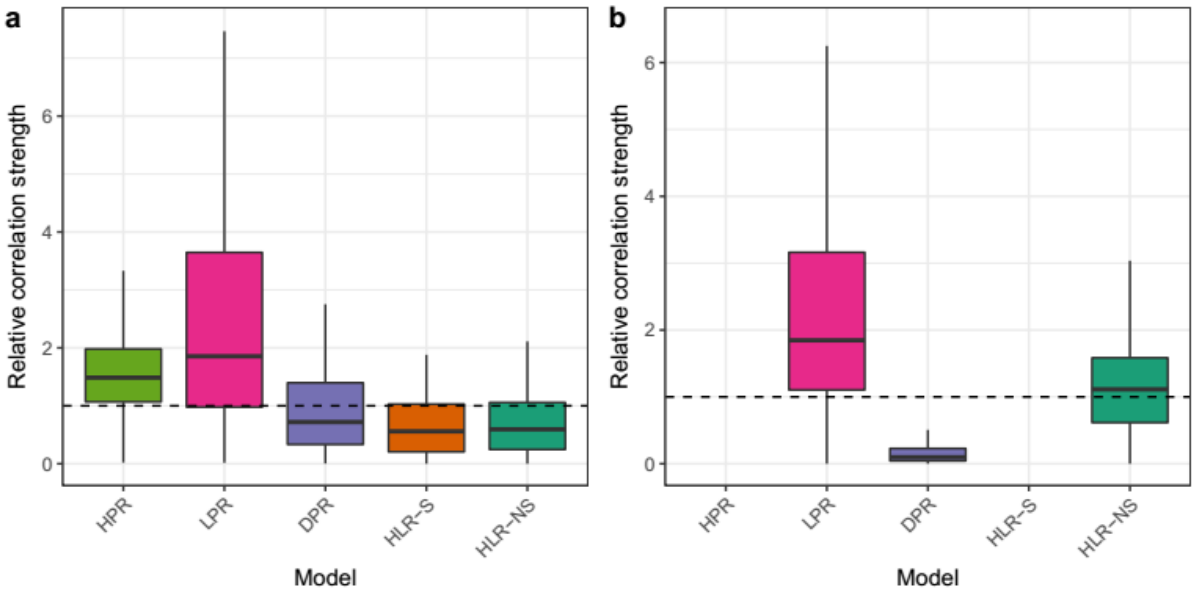


Figure 2.7: Relative correlation strength of all JSDMs compared to a baseline of the MPR model. Plot a) is for the four smallest datasets, and plot b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively.

For the smallest four datasets (Figure 2.8a**)** DPR, HLR-S, and HLR-NS had more certain estimates of correlation coefficients than MPR, HPR was relatively more uncertain, while MPR and LPR were generally equal. However, for the largest two datasets (Figure 2.8b), HLR-NS was relatively more uncertain than MPR; LPR and MPR were similarly uncertain, and DPR tended to be more certain than MPR.

For results of the simulation study see Appendix F.

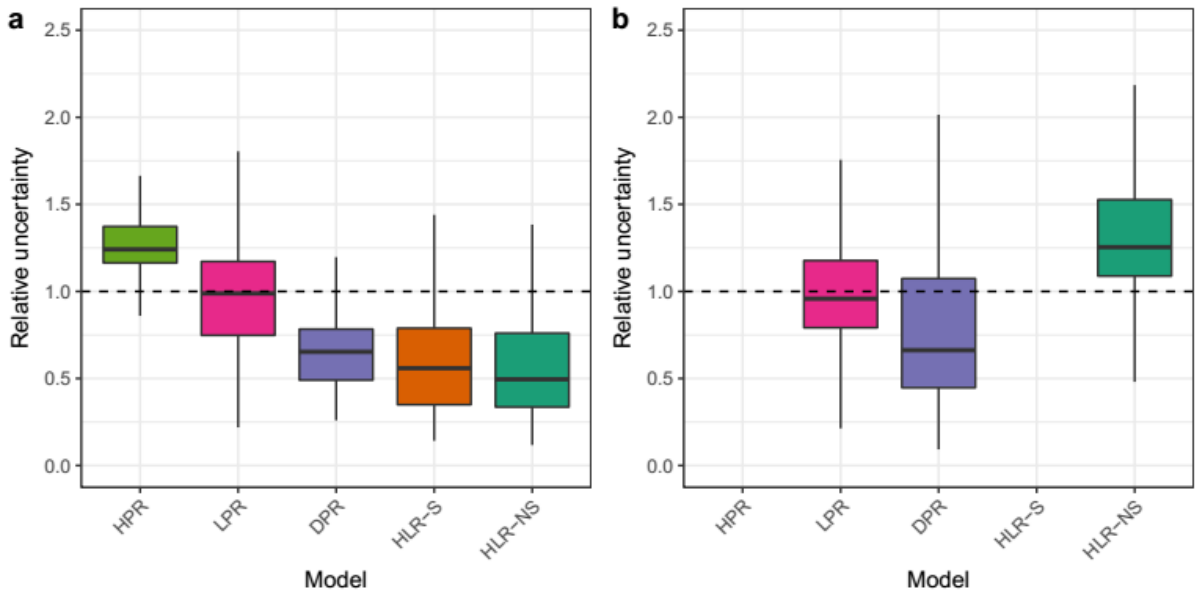


Figure 2.8: Relative correlation uncertainty (95% credible interval width) of all JSDMs compared to a baseline of the MPR model. Plot a) is for the four smallest datasets, and plot b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range respectively. HPR and HLR-S are missing from 8b as they did not complete within the time limit.

2.4. DISCUSSION

Only four JSDMs could be successfully fit to all six datasets, illustrating that different JSDMs scale differently with dataset size. That MLR could only be successfully fit to the two smallest datasets, and was significantly slower than the other JSDMs, highlights the computational benefits of the probit link over the logit link. Since probit models allow a latent variable parameterisation instead of using the link function directly, they can be fitted using a highly efficient Gibbs sampler (Chib & Greenberg 1998; O'Brien & Dunson 2004). Of the probit models, HPR and HLR-S scaled the worst, and LPR

exhibited signs of slowing down for the two largest datasets. The effect of scaling will only become more apparent as datasets become larger.

Estimates of runtime and sampler efficiency show that MPR and HLR-NS are markedly faster than the other JSDMs. HPR took longer to run than DPR, yet this difference is due to different MCMC regimes, since the HPR sampler was more efficient. HPR also exhibited the most consistent ESS across all coefficients. The MCMC sampling of HPR is implemented in JAGS, which prioritises ease of use for those with limited programming experience, rather than more efficient, customised samplers utilised in other JSDMs. HLR-S had the least efficient sampler as it was the only one that included spatial Gaussian process components. This requires inversion of an $n \times n$ covariance matrix per latent factor, which is computationally expensive and scales poorly with dataset size (Rasmussen & Williams 2006).

Datasets were coerced into the required formats for most JSDMs within five minutes, which suggests data formatting is not a barrier to JSDM usage. A vignette can be found in the code repository for this analysis (Wilkinson 2018) that explains data formatting. The HLR models took longer for dataset formatting but formatting time reduced markedly with experience. HLR models are fit in MATLAB not R, the language most ecologists are familiar with, so many ecologists might experience a slight learning curve. MATLAB also requires a software license.

All JSDMs estimated similar regression coefficients across all datasets (>0.7 correlation). The three JSDMs with hierarchically-drawn regression coefficients (HPR, HLR-S, HLR-NS) were strongly correlated (>0.9); however, this does not mean the additional hierarchical structure is necessarily beneficial. Minor differences between the uncertainty in JSDM coefficient estimates were observed but with no clear pattern. We

suggest that the JSDMs were as uncertain as each other when considered across all coefficients.

For this comparison we used the default priors for each JSDM, though these will not always be the ideal choice. The fungi and mosquito datasets exhibit complete separation for multiple species in the case of binary environmental variables. For maximum-likelihood estimation methods, a finite maximum likelihood would not exist and standard calculation of standard errors would fail. This is less of an issue for models using a Bayesian framework, as using prior information switches the focus of parameter estimation to posterior distribution summaries (Rainey 2016). LPR had convergence issues for these datasets with default priors, but the use of more informative priors led to performance similar to the other JSDMs (see Appendix C), highlighting the importance of careful consideration of priors.

Different JSDMs estimated residual correlations between species in the same direction (with some exceptions for near-zero strength interactions whose 95% credible intervals included zero), suggesting that all JSDMs lead to similar ecological inferences. Correlations between each JSDM's estimated correlation coefficients were moderate to very strong when averaged across all datasets. For the smaller datasets, DPR estimated weaker residual correlations than other JSDMs, and showed only weak correlation (~0.3) with the other JSDMs in the two largest datasets. The reason for this is unclear; the Dirichlet process was only implemented for the larger of the two, so it cannot be affecting both datasets.

Whilst the sign of residual correlations was the same across all JSDMs, the estimates differed in their relative strength and uncertainty. For the four smallest datasets, HLR-S and HLR-NS estimated relatively weaker correlations between species than the non-

latent factor JSDMs (e.g. MPR and HPR). This is because latent factor models account for correlations through shared responses to a small set of latent factors, which provides a form of regularisation on residual correlations. For the two largest datasets, however, HLR-NS estimated stronger correlations between species than MPR (HLR-S failed to complete within the time limit). LPR, despite being a latent factor model like HLR-S and HLR-NS, estimated stronger correlations between species. This is likely a result of the number of latent factors used to fit the model, as the use of more latent factors provides a better approximation of the full-rank correlation matrix. By default, LPR uses two latent factors, while HLR-S and HLR-NS estimate the number of latent factors during the model fitting process (here between 3 and 12). This further highlights the trade-off between model fit and model simplicity when using the latent factor approach. DPR estimated relatively weaker correlations than the other JSDMs in most cases (similar to HLR-S/HLR-NS in the small datasets); the reason for this is unclear.

Differences in the sizes of residual correlations between the smaller and larger datasets could indicate the effect of spatial scale on the implementation and interpretation of latent factors (Ovaskainen *et al.* 2016b,a). For datasets collected over small spatial scales (coincidently our smaller datasets) HLR-S estimated small α values: 0 (frogs/eucalypts) or 0.04 (mosquitoes), suggesting no benefit of a spatial implementation. Datasets collected at larger spatial scales, like the birds and butterflies (computationally incompatible with HLR-S in this study), would be more likely to exhibit spatial correlation (i.e., a moderate to large α). Indeed, Ovaskainen *et al*. (2016a,b) observed an α between 0.25-0.26 for the butterfly dataset (in which a training/testing split of the data reduced the computational requirements). We suggest the additional effort required for spatially-explicit JSDMs might only be necessary for large spatial scale datasets where spatial correlation could be reasonably expected, and

it becomes important to account for any decay in correlation between latent factors due to distance. Correlations between species at finer spatial scales might reflect species interactions, but at larger scales correlations are likely influenced by other factors like shared environmental responses and dispersal ability (Götzenberger *et al.* 2012; Tikhonov *et al.* 2017). This can be explored by defining latent factors at different spatial scales (Ovaskainen *et al.* 2016a).

In Appendix F we compared model performance for a subset of these models on simulated datasets. As for the real datasets included in the main analysis, we found that there were no significant differences in parameter estimation between JSDMs. We observed that models that accounted for hierarchical regression coefficients performed better than the other models for datasets generated hierarchically but only if they also matched the latent factor assumptions. Models also performed worse when fit to datasets that were simulated using unmeasured variables as well as the measured ones used to fit the models, but models that could account for that using latent factors (like HLR-NS) had smaller performance decreases. This shows that the added complexity of these models can provide a benefit to parameter estimation, where the ecological system that generates the data has these features.

Methodological differences between the different JSDMs beyond their core statistical framework might have biased our results. The program in which JSDMs are developed (e.g., R vs MATLAB) could influence computational efficiency, as could their respective MCMC samplers and regimes. The choice of priors and other default settings, such as the number of latent factors, could also impact statistical inference. However, since the goal of this study was to compare the default implementations of these different JSDMs and their underlying statistical frameworks, these differences are

unavoidable. Clearly, particular case studies should explore the appropriateness of default settings prior to implementation.

While we only considered presence-absence JSDMs in which species interactions are modelled via multivariate regression methods, other JSDMs have been proposed. *mistnet* (Harris 2015), a neural network JSDM, focuses on predicting community composition but does not provide coefficient estimates comparable to the JSDMs considered here. The JSDM of Johnson & Sinclair (2017) models species interactions by grouping them into guilds by environmental responses rather than via residual correlation. JSDMs for abundance data also exist (e.g.Dorazio, Connor & Askins 2015; Letten *et al.* 2015; Thorson *et al.* 2015; Warton *et al.* 2015).

We have compared the ability of JSDMs to make inferences about species' responses to biotic and abiotic factors. We have used both real and simulated datasets, but a more comprehensive simulation study accounting for all nuances of the model fitting process presents an interesting avenue for further research. The predictive ability of different JSDMs also warrants further study. Single species SDMs are commonly used to predict into un-sampled areas or under new environmental conditions, such as climate change scenarios (Elith & Leathwick 2009), and these are clearly areas where JSDMs are applicable. There is an open question, however, of how best to define prediction in multivariate space. JSDMs can predict community composition (joint prediction), distributions of individual species (marginal prediction), or distributions of individual species conditional on the presence or absence of other species (conditional prediction). Ecological interpretations, and the statistical and computational aspects of these different types of prediction should be considered.

Another important consideration is that all methods described here assume perfect detection (i.e. that absence records reflect true species absence). However, imperfect detection is often an issue in survey data. Extending these models to account for imperfect detection is an important avenue for future JSDM research (Beissinger *et al.* 2016; Warton *et al.* 2016), as is the evaluation of impacts of imperfect detection in the estimation of residual co-occurrence. Rota *et al* (2016) introduced a multispecies occupancy model accounting for imperfect detection, but in principle, any of the methods above could be extended to include a description of the detection process.

2.5. CONCLUSION

Akin to comparisons of single species SDMs (Elith *et al.* 2006a), this study serves as a guide for prospective JSDM users to make informed decisions about the JSDM, or JSDMs, that might be relevant to their objective. Expressing the different statistical JSDM frameworks with a common notation has helped to identify their differences and similarities. The core statistical differences between different JSDMs can be broadly defined by two methodological choices: the inclusion of a hierarchical structure on the regression coefficients, and the use of latent factors to account for shared species responses to "hypothetical" unmeasured environmental variables impacting species co-occurrence patterns. There are statistical and computational implications related to these methodological differences, but they also have ecological interpretations. As for all ecological modelling, the user should be aware of all ecological assumptions inherent in these models when making an informed selection of JSDM.

REFERENCE LIST

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Beissinger, S.R., Iknayan, K.J., Guillera-Arroita, G., Zipkin, E.F., Dorazio, R.M., Royle, J.A. & Kéry, M. (2016). Incorporating Imperfect Detection into Joint Models of Communities: A response to Warton et al. *Trends in Ecology and Evolution*, **31**, 736–737.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Clark, J.S., Gelfand, A.E., Woodall, C.W. & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, **24**, 990–999.

Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, **87**, 34–56.

Dorazio, R.M., Connor, E.F. & Askins, R.A. (2015). Estimating the effects of habitat and biological interactions in an avian community. *PLoS ONE*, **10**, 1–16.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Elith, J. & Leathwick, J.J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, …*, **40**, 677–697.

Gelfand, A.E., Silander, J.A., Wu, S., Latimer, A., Lewis, P.O., Rebelo, A.G. & Holder, M. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, **1**, 41–92.

Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.

Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites & Vectors*, **8**, 367.

Götzenberger, L., de Bello, F., Bråthen, K.A., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and

communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.

Hardy, O.J. (2008). Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, **96**, 914–926.

Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model (D. Warton, Ed.). *Methods in Ecology and Evolution*, **6**, 465–473.

Hui, F.K.C. (2017). boral: Bayesian Ordination and Regression AnaLysis.

Hui, F.K.C. (2016). boral - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, **7**, 744–750.

Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.

Johnson, D.S. & Sinclair, E.H. (2017). Modeling joint abundance of multiple species using Dirichlet process mixtures. *Environmetrics*, **28**, 1–13.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology*, **82**, 2560–2573.

Letten, A.D., Keith, D.A., Tozer, M.G. & Hui, F.K.C. (2015). Fine-scale hydrological

niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, **103**, 1264–1275.

MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.

Nieto-Lugilde, D., Maguire, K.C., Blois, J.L., Williams, J.W. & Fitzpatrick, M.C. (2017). Multiresponse algorithms for community-level modeling: review of theory, applications, and comparison to species distribution models. *Methods in Ecology and Evolution*, **12**, 3218–3221.

O'Brien, S.M. & Dunson, D.B. (2004). Bayesian multivariate logistic regression. *Biometrics*, **60**, 739–746.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, **7**, 549–555.

Ovaskainen, O., Hottola, J. & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514–2521.

Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models (D. Orme, Ed.). *Methods in Ecology and Evolution*, **7**, 428–

436.

Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM) (J. McPherson, Ed.). *Methods in Ecology and Evolution*, **5**, 397–406.

R Core Team. (2016). R: A Language and Environment for Statistical Computing.

Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis*, **24**, 339–355.

Rasmussen, C.E. & Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, Secondn. The MIT Press, Cambridge, Massachusetts.

Richmond, O.M.W., Hines, J.E. & Beissinger, S.R. (2010). Two-species occupancy models: a new paramaterization applied to co-occurence of secretive rails. *Ecological Applications*, **20**, 2036–2046.

Rota, C.T., Ferreira, M.A.R., Kays, R.W., Forrester, T.D., Kalies, E.L., McShea, W.J., Parsons, A.W. & Millspaugh, J.J. (2016). A multispecies occupancy model for two or more interacting species. *Methods in Ecology and Evolution*, **7**, 1164–1173.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, **21**, 88–99.

Taylor-Rodríguez, D., Kaufeld, K., Schliep, E.M., Clark, J.S. & Gelfand, A.E. (2017). Joint Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes. *Bayesian Analysis*, **12**, 939–967.

The MathWorks Inc. (2016). MATLAB.

Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C. & Zipkin, E.F. (2016). Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, **25**, 1144–1158.

Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J. & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, **6**, 627–637.

Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context (D. Warton, Ed.). *Methods in Ecology and Evolution*, **8**, 443–452.

University of Melbourne. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility.

Waddle, J.H., Dorazio, R.M., Walls, S.C., Rice, K.G., Beauchamp, J., Schuman, M.J. & Mazzotti, F.J. (2010). A new parameterization for estimating co-occurrence of

interacting species. *Ecological Applications*, **20**, 1467–1475.

Warton, D.I., Blanchet, F.G., O'Hara, R., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2016). Extending Joint Models in Community Ecology: A Response to Beissinger et al. *Trends in Ecology and Evolution*, **31**, 737–738.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, **30**, 766–779.

Wilkinson, D.P. (2018). JSDM_Inference v0.1.6. Zenodo, https://dx.doi.org/10.5281/zenodo.1452066.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Wolfram Research Inc. (2016). Mathematica.

CHAPTER THREE

DEFINING AND EVALUATING PREDICTIONS OF JOINT SPECIES DISTRIBUTION MODELS

PREFACE

The material presented in Chapter Two is currently under review in *Methods in Ecology and Evolution*.

As per University regulations the version of the manuscript presented as Chapter Two in the thesis is the same as that currently under review. The manuscript has only been modified to adhere to the section numbering and style formatting of the thesis at large. Content remains otherwise unmodified.

ABSTRACT

Joint species distribution models (JSDMs) simultaneously model the distributions of multiple species, while accounting for residual co-occurrence patterns. Despite increasing adoption of JSDMs in the literature, the question of how to define and evaluate JSDM predictions has only begun to be explored. We define four different JSDM prediction types that correspond to different aspects of species distribution and community assemblage processes. *Marginal* predictions are environment-only predictions akin to predictions from single-species models; *joint* predictions simultaneously predict entire community assemblages; and *conditional* and *conditional marginal* predictions are made at the species- or assemblage-level, conditional on the known occurrence state of one or more species at a site. We define five different classes of metrics that can be used to evaluate these types of prediction: threshold-dependent, threshold-independent, community dissimilarity, species richness, and likelihood metrics. We illustrate different prediction types and evaluation metric using a case study, in which we fit a stacked species distribution model (SSDM) and a JSDM to a frog occurrence dataset collected in Melbourne, Australia. Overall, neither model (SSDM vs JSDM) consistently outperformed the other when compared across prediction types; however, there was evidence of consistent trends within evaluation metric classes. Likelihood-based metrics indicate that the JSDM better fits the data, species richness metrics show that the JSDM overpredicts the number of species at a site compared to the SSDM for prediction types that account for residual correlations between species, while community dissimilarity metrics provide mixed results. JSDMs present opportunities to explore different facets of species distribution and community assemblage processes not possible with single-species models. We show that there are a

variety of different metrics available to evaluate JSDM predictions; the choice of prediction type and evaluation metrics should closely match the questions being investigated.

3.1. INTRODUCTION

Species distribution models (SDMs) are commonly used to predict into un-sampled areas or to new environmental conditions (Elith & Leathwick 2009). Joint species distribution models (JSDMs) are an extension of standard correlative SDMs that allow multiple species to be modelled simultaneously while accounting for species correlations not explained by available environmental predictors, e.g. due to species interactions or important missing covariates (Kissling *et al.* 2012; Pollock *et al.* 2014; Golding, Nunn & Purse 2015; Ovaskainen *et al.* 2016b; Clark *et al.* 2017). Despite increasing adoption of JSDMs in the literature, it remains unclear how predictions of JSDMs differ from those of standard SDMs, and how specifically JSDM predictions can be used to address different questions in ecology and conservation (Ovaskainen et al. 2016a; Zhang et al. 2018; but see: Norberg et al. 2019).

Modelling distributions of single species with correlative SDMs ignores the impacts of species interactions and missing environmental predictors on species distributions, which potentially biases estimated coefficients and resultant predictions (Kissling *et al.* 2012; Wisz *et al.* 2013). There have been attempts to account for biotic interactions by using other species' occurrence states as predictor variables alongside abiotic variables (Leathwick & Austin 2001; Araújo & Luoto 2007; Meier *et al.* 2010; Pellissier *et al.* 2010), or by constraining predicted distributions to distributions of species on which the target depends (Schweiger *et al.* 2012), but these approaches are restricted to unidirectional interactions (Kissling *et al.* 2012). Stacked species distribution models (SSDMs) combine, or stack, multiple single-species SDMs to estimate community structure and species richness (Gelfand *et al.* 2005; Parviainen *et al.* 2009; Mateo *et al.* 2012). Yet, it has been suggested that, because they do not account for species

interactions, SSDMs tend to over-predict species richness (Pineda & Lobo 2009; Guisan & Rahbek 2011; Calabrese *et al.* 2014; Thuiller *et al.* 2015). By accounting for interactions between multiple species, the expectation is that JSDMs might allow for more accurate predictions.

An important decision when aiming to obtain predictions from a JSDM is defining what to predict, based on the JSDM's multivariate output. In a single-species SDM, there is only one response variable to predict: some feature of the target species (presence/absence, abundance, etc). Accounting for biotic interactions by using other species as predictors, or by constraining a target species' distribution to that of another species, means that distributions of non-target species must be known *a priori* or estimated using analogous single-species SDMs. In the multivariate space of JSDMs, however, there are multiple response variables (one per species), all of which are potentially correlated. Prediction can therefore be approached in several ways. For instance, we may aim to predict community composition at sites where we have no knowledge of distributions, or, we may be interested in exploring how having data on the distributions of some species changes our prediction for a focal species.

Once we have a prediction, we need to evaluate it. How do we approach this for JSDMs? There is substantial literature on the evaluation of SDM predictions (Fielding & Bell 1997; Liu, White & Newell 2009; Lawson *et al.* 2014) that covers a wide variety of metrics, but are they appropriate in a multi-species context? Are the most common metrics used for single-species SDMs (such as AUC) still the most relevant? JSDMs can predict community assemblages, which opens up a suite of potential evaluation metrics in the form of the community dissimilarity indices widely used in community ecology (Legendre & De Cáceres 2013). What insight can be gained from these additional metrics?

In this article we outline the different ways that prediction can be performed using JSDMs. Marginal predictions are environment-only predictions that average over the occurrence and co-occurrence patterns of other species, and would be the most familiar to ecologists that use single-species SDMs. Joint predictions simultaneously predict the occurrence of multiple species while accounting for environmental responses and species correlations. Conditional predictions are akin to joint predictions but with additional information about the known occurrence state(s) of one or more species in the community. We outline which evaluation methods are appropriate for different prediction types. To illustrate these issues, we fit a standard SSDM and a JSDM to a dataset of frog occurrences in Melbourne, Australia.

3.2. MATERIALS AND METHODS

3.2.1. TYPES OF PROBABILITY

The first step when predicting with JSDMs is to select the type of probability distribution that corresponds to the desired prediction type, which then dictates the methodology for its calculation. In this section, we describe the types of probability distribution available. We define species occurrence as a binary response variable, y_{ij} (for species $j = 1, \dots, J$, at site $i = 1, \dots, n$), where species j is present ($y_{ij} = 1$) or absent ($y_{ij} = 0$) at site i, with some probability, $\Pr(y_{ij})$. JSDMs predict multiple species distributions simultaneously. To simplify notation in the following description, we consider predictions for two species, A and B, at a single site. The joint probability of both species being present is $\Pr(A, B)$, and is the product of a conditional probability, $\Pr(A|B)$, and a marginal probability, $\Pr(B)$ (or with A and B switched, which yields the same probability):

$$\Pr(A, B) = \Pr(A|B) \cdot \Pr(B) \quad (1)$$
$$= \Pr(B|A) \cdot \Pr(A)$$

The marginal probability of a single species' presence (i.e.$\Pr(A)$) is the probability of its presence *averaging over* the occupancy status of the other species. In contrast, a conditional probability, $\Pr(A|B)$, gives the probability that species A is present given the known presence or absence of species B. As equation 1 shows, we can define the probabilities in multiple, equivalent ways, by changing the order in which species are marginalised out. This can result in a species' presence being conditional on a joint probability distribution. For example, in a four-species scenario, we would marginalise out three species one by one and end up with three conditional probability distributions and one marginal probability distribution (equation 2):

$$\Pr(A, B, C, D) = \Pr(A|B, C, D) \cdot \Pr(B, C, D) \quad (2)$$
$$= \Pr(A|B, C, D) \cdot \Pr(B|C, D) \cdot \Pr(C, D)$$
$$= \Pr(A|B, C, D) \cdot \Pr(B|C, D) \cdot \Pr(C|D) \cdot \Pr(D)$$

3.2.2. JOINT SDMS

So far, we have expressed calculations as a function of discrete response variables (presence/absence), however, this is not always the case in JSDMs. Most occupancy-based JSDMs are built on the foundation of the Chib and Greenberg (1998) multivariate probit regression model that uses the latent variable parameterisation of a probit model instead of using the probit link directly. We focus on the multivariate probit model here for simplicity, though the same approach can be applied to the multivariate logistic, and latent factor models that have also been used as JSDMs. For these models, the

occurrence state, y, of a species, j, at a site, i, is modelled via a normally-distributed latent variable, z_{ij}. The latent occurrence state of a species (present or absent), y_{ij}, is equal to 1 when $z_{ij} > 0$, and 0 otherwise (see Figure 1 for a visual representation). This core model is as follows:

$$y_{ij} = 1(z_{ij} > 0) \quad (3)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

where the latent variable, z_{ij}, is the sum of the linear predictor, μ_{ij}, and the correlated residual error, e_{ij}. The linear predictor is the product of the measured environmental variables $\boldsymbol{X}_{i,.}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{.,j}$, as in standard generalised linear models. Correlations in the residual error $\boldsymbol{e}_i$ are captured in **R**, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements – the residual correlations between species – are restricted between -1 and 1. The elements of **R** reflect species co-occurrence patterns not described by the environmental predictors (i.e. species interactions, or missing predictors). Standard deviations, and in turn variances, are constrained to equal 1 in probit regression, thus covariance and correlation matrices are equivalent. Because the variance of the latent variable distribution remains constant, the probability of presence for a single species is controlled only by the mean value of the distribution, μ (compare Figures 3.1a and 3.1b).

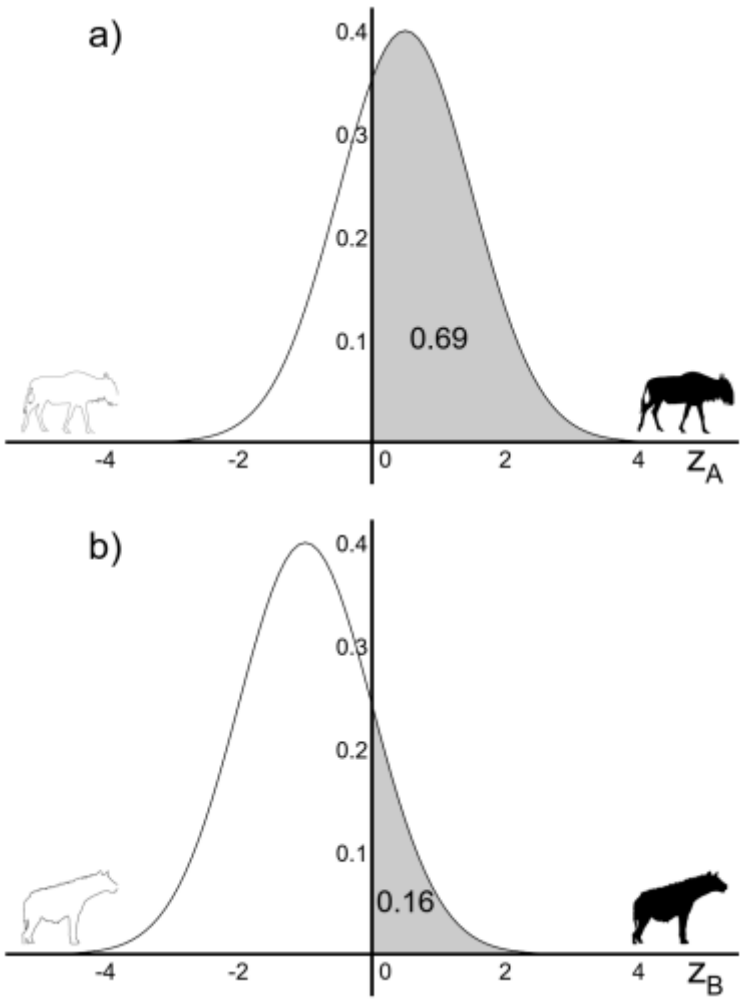


Figure 3.1: Visualisations of the normally-distributed latent variable, z_i, for two species (A and B). The mean of the distribution, μ_i, is 0.5 for species A and -1 for species B. The probability of occurrence of a species is equal to the area under the curve where $z_i > 0$, shown here in grey.

Since JSDMs simultaneously consider multiple species, the space of these latent variables is multivariate. We can consider each species as being represented by its own latent variable in a single dimension, and the whole species community at a site is represented by a multi-dimensional latent variable with as many dimensions as species. The JSDM therefore considers a joint probability distribution over these latent variables. With the latent variable parameterisation, the discrete response variables, e.g. $\Pr(A)$, correspond to the area under the curve of continuous response variables above or

below a threshold. Thus, a species is present with probability $\Pr(A) = \Pr(z_A > 0)$ and absent with probability $\Pr(A') = \Pr(z_A \leq 0)$. This probability can be computed as the integral of the probability density function $f()$ of the latent variablez_A over all values greater than zero: $\Pr(A) = \Pr(z_A > 0) = \int_0^\infty f(z_A)dz_A$. In a two-species scenario (Figure 3.2), the joint probability $\Pr(A, B)$ that both species A and B are present can be calculated as:

$$\Pr(A, B) = \Pr(z_A > 0, z_B > 0) \quad (4)$$

$$\Pr(A, B) = \int_0^\infty \int_0^\infty f(z_A|z_B) \cdot f(z_B)dz_A dz_B$$

$$= \int_0^\infty \int_0^\infty f(z_B|z_A) \cdot f(z_A)dz_B dz_A$$

That is, this joint probability $Pr(A, B)$ is the integral of the product of two density functions, in only the region where both z_A and z_B are greater than zero. In this two-species case this can be visualised as the volume under a two-dimensional surface (as in Figure 3.2). In Eq. 4, the conditional probability distribution $f(z_A|z_B)$ reflects probable values of z_A (and therefore $\Pr(A)$) *given* a fixed value of z_B. The marginal probability distribution $f(z_B)$ reflects probable values of z_B (and therefore $\Pr(B)$) *independent* of the latent occurrence state of species A. The order of items in the joint probability is inconsequential, so species can be marginalised out in any order (lines 2 and 3 in equation 4 are equivalent). Since the total area (or volume) under a probability distribution is always 1, the integral under a part of that distribution yields the probability that the value of the variable falls in that part of the space. This integration can be performed for any number of species, to compute joint probabilities for a community of any size.

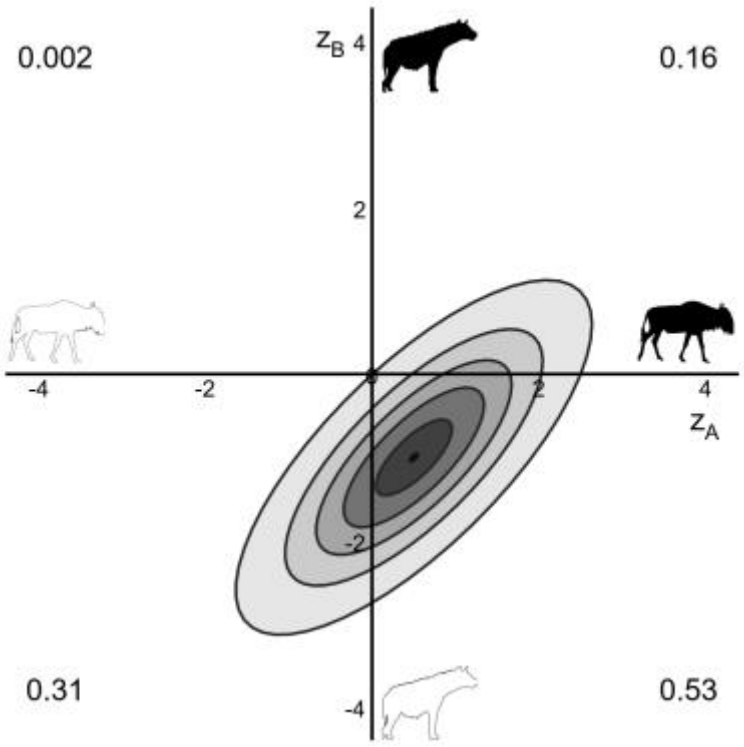


Figure 3.2: Visualisation of the multivariate, normally-distributed latent variable, z_{ij}, for a two species scenario. The mean of the distribution, μ_{ij}, on each species' respective axis is the same as their independent distributions from Figure 1 ($\mu_{iA} = 0.5$, $\mu_{iB} = -1$) and there is positive correlation of 0.75 between them. The contours of the probability distribution, the grey ellipses, indicate probability density values of 0.1, 0.3, 0.5, 0.7, and 0.9. The numbers in the four corners are the probabilities of the multivariate latent variable integrated in that quadrant, e.g. in the upper right quadrant; there is a probability of 0.16 that both species will occur at the site.

Integration over the relevant proportion of the multivariate latent variable space in a JSDM can therefore be used to compute the *joint* probability of observing a given configuration of presence/absence of all species at a site. These integrals can also be used to compute the *marginal* probability of species presence by averaging over (marginalising) presence of the other species. Multiple joint probabilities can be combined to compute the *conditional* probability of presence of a species given the presence (or absence) of another species, by restricting the space of the integration to

those indicating presence (or absence) of the other species. These predictive probability distributions can be combined in different ways to define a number of prediction types for JSDMs. We consider five types of JSDM prediction, each mapping onto ecological questions. These are illustrated in Figure 3.3 and detailed below.

3.2.3. PREDICTION TYPES

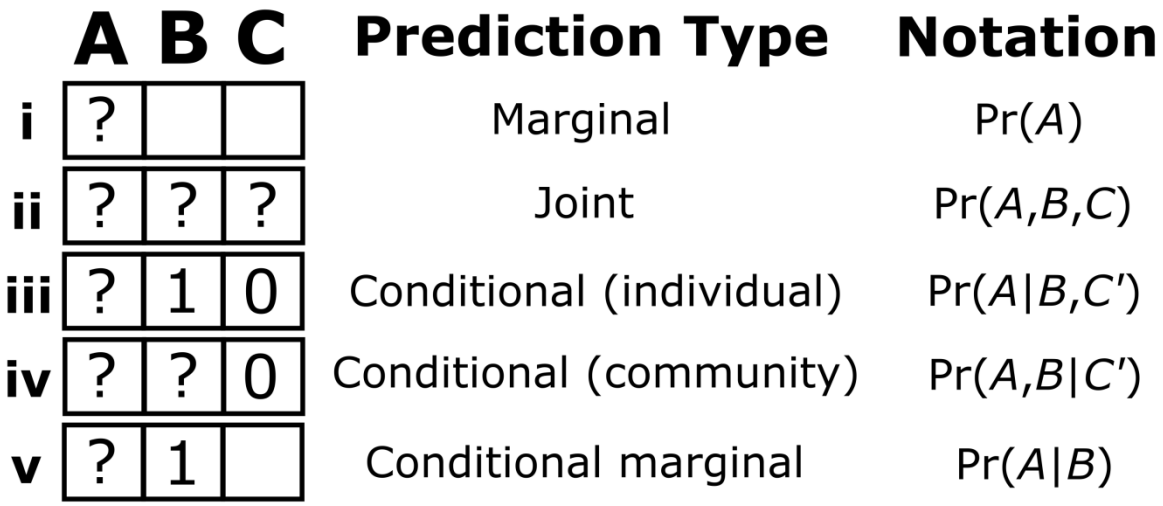


Figure 3.3: Five different prediction types possible with JSDMs. The boxes show the occurrence states of species *A*, *B*, and *C*. Question marks denote the species being predicted, empty boxes are species whose occurrence state is not being considered in that prediction type, and 1/0 a known presence/absence state of species being considered for prediction. *Prediction Type* is the name of the prediction methodology, and *Notation* shows the corresponding probabilistic notation, following the definitions in the text.

3.2.3.1. MARGINAL PREDICTION

Similar to single-species SDM predictions, marginal predictions are based solely on environmental attributes. The predicted probability of occurrence of each species can be calculated using the JSDM regression coefficients and the covariate values found at

each site, as in a standard generalised linear model. The sole difference between marginal predictions for JSDMs and those of single-species SDMs is that inter-species correlations are accounted for in the estimation of JSDM regression coefficients. Since a species' marginal probability is independent of the occurrence state of other species (see Figure 3.3i), one can sum the different community assemblage probabilities (that contain the focal species) at a site to get the marginal probability. In Figure 3.2, we can see that species A, on the x-axis, is predicted to be present in the top-right and bottom-right quadrants. Thus:

$$Pr(A) = Pr(A, B) + Pr(A, B') \quad (5)$$

$$Pr(A) = 0.16 + 0.53$$

$$Pr(A) = 0.69$$

We can see that $Pr(A)$ calculated from the multivariate distribution in Figure 3.2 is the same as that from the univariate distribution in Figure 3.1a; the same can be calculated for species B. This is because the correlation between species affects their co-occurrence rates and not the unconditional probability of each species' occurrence at a site. Consequently, the marginal prediction from a JSDM can be computed without having to explicitly compute the integrals discussed in the previous section.

3.2.3.2. JOINT PREDICTION

We can also make predictions of species community composition that account for both the environmental covariates *and* species co-occurrence, by using the joint probability distribution. Whereas the marginal prediction of a community at a given site can be represented by a single vector of probabilities (one for each species), the joint prediction instead yields a probability value for each possible realisation of the community

composition (a vector of 1s and 0s indicating presence/absence of each species; Figure 3.3ii).

However, with J species the number ofpossible community assemblages is 2^J. Whilst 3 species result in 8 probabilities (i.e. 8 possible species combinations), with 10 species that number rises to 1024, and with 20 species there are over a million possible assemblages. It is therefore generally infeasible to compute and store the probabilities of all assemblages, when using a JSDM for prediction in general.

Since the joint probability distribution is entirely defined by the JSDM parameters μ_{ij} and $\boldsymbol{R}$, these can instead be stored to represent the joint prediction. With these parameters, it then becomes straightforward to evaluate model predictions by computing the joint probability for the observed species assemblage in held-out data. These parameters can also be used to simulate plausible community assemblages by taking random draws from the joint probability distribution and these can then be summarised to compute other quantities of interest, such as a predictions of species richness.

Joint predictions cannot be represented via a single vector of probabilities of occurrence for individual species, since this discards crucial information about the jointness of the occurrence distribution. For example, computing an average over these simulated assemblages would simply result in (an approximation to) the marginal prediction. However, the joint prediction can be summarised in various ways that can be represented more concisely.

3.2.3.3. CONDITIONAL PREDICTION

Another way to define prediction in multivariate space is by using the conditional probability distribution of species, in which we estimate a species' occurrence

probability given a fixed occurrence state of other species. If we know the correlation between two species (estimated by the JSDM) and the occurrence state of one of those species, we can make a more informed prediction of the unknown occurrence state of the other species.

For each species with a known occurrence state, we can truncate the multivariate normal distribution over the latent variable in the dimension that represents that species, either to be positive if the species is present, or negative if it is absent (see Figure 3.4 for a visual representation). Because probability distributions integrate to one, truncating this distribution alters the probabilities of the remaining community assemblage possibilities. For instance, in our example, the marginal probability of species B, $\Pr(B)$, is 0.16 (Figure 3.1b), but if we know that species A is present, then the conditional probability of species B is $\Pr(B|A) = 0.23$ (Figure 3.4). This modification of the joint probability can be interpreted in another way: the conditional probability of species B being present is the joint probability of species A and B being present, normalised by the marginal probability of species A: $\Pr(B|A) = \Pr(B, A)/\Pr(A) = 0.16/0.69 = 0.23$. This follows from rearranging Eq 1.

Depending on the amount of information available to condition our joint distribution, there are two possible ways of approaching conditional predictions. In the scenario in which we know the occurrence state of all but one species at a site, we can calculate the probability of occurrence of that individual species conditional on the rest (see Figure 3.3iii). Alternatively, if we know the occurrence state of only a single, or small subset of, other species, we can calculate the joint probability of the remaining species conditional on the occurrence state of the known species (see Figure 3.3iv). This joint prediction has the same characteristics as the joint prediction described above, but for a subset of the species.

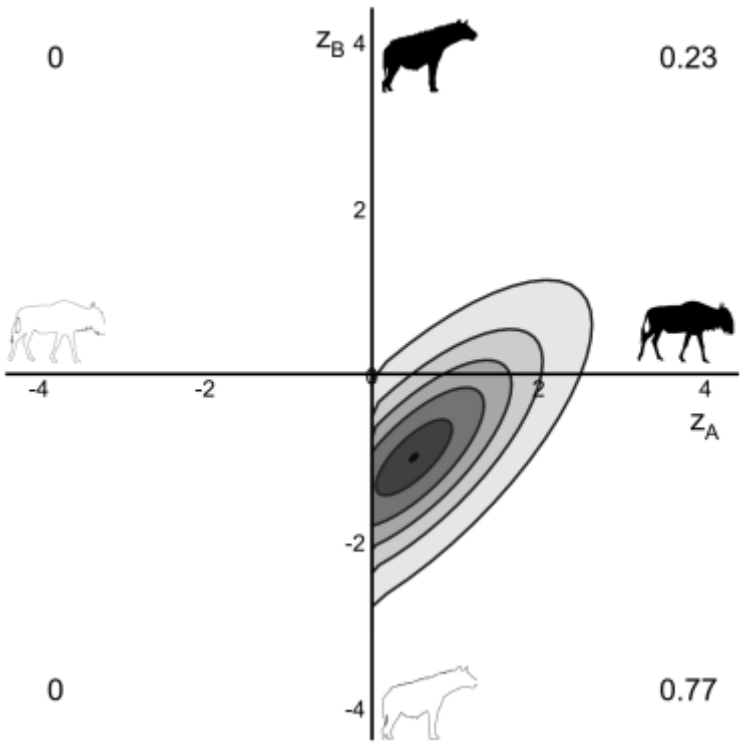


Figure 3.4: The conditional probability distribution of species B given species A is known to be present. This probability distribution is identical to that of Figure 3.2 except that it is truncated such that $z_A > 0$. The contours of the probability distribution, the grey ellipses, are the probability densities for 0.1, 0.3, 0.5, 0.7, and 0.9 proportions of the distribution's volume. The values in the four corners are the probabilities of the multivariate latent variable occurring in that quadrant. Since species A is known to be present, there is zero probability that the value of the latent variable falls in either of the left-hand quadrants. $Pr(z_B|z_A) = 0.23$, which is higher than the marginal distribution, $Pr(B) = 0.16$, of the un-truncated distributions seen in Figures 3.1b and 3.2.

3.2.3.4. CONDITIONAL MARGINAL PREDICTION

The final way of defining prediction using JSDMs is simultaneously conditional on and marginal to the occurrence states of the other species. As noted previously, conditional prediction can be defined in a variety of ways dependant on the number of species occurrence states that are known. This may result in the prediction of one species

conditional on the fixed presence/absence of all other species, or of $J-1$ species, conditioned on the presence/absence of one species (a joint, conditional prediction). We can also combine the conditional and marginal approaches to predict the presence of one or more species, conditional on some species, but marginal to others (see Figure 3.3v). This prediction type is also simple to represent; like marginal predictions it can be stored as a vector of probabilities for each species. It also accounts for residual correlations and can be matched onto ecological hypotheses. For example, predicting the marginal probability of presence of one set of species given the presence or absence of some other set of species - say predators/competitors, or perhaps some easy-to-detect indicator species.

3.2.4. Evaluation Metrics

Various metrics could be used to evaluate JSDM predictions. In a single-species context, choice of metric depends largely on factors such as data type (e.g., presence-absence vs. presence-only) and prediction format (binary or probabilistic) (Lawson *et al.* 2014). To date, there has not been a thorough exploration of the implications of the choice of evaluation metric in the multi-species context of JSDMs. Here we evaluate the performance of traditional single-species metrics for JSDMs as well as new metrics that may be suitable. We broadly classify metrics for evaluating JSDM predictions into five groups, in terms of the aspects of performance they focus on (see Table 3.1).

Threshold-independent metrics evaluate continuous predicted probabilities against observed presence-absence data. A widely used threshold-independent metric used in single-species SDMs is the Area Under the Receiver Operating Characteristic Curve (AUC), but other metrics include root mean square error (RMSE), the coefficient of determination (R^2), and the Pearson's or point-biserial correlation coefficient.

Threshold-dependent metrics compare binary predictions against observed presence-absence data. This is achieved by converting predicted values (which may be probabilities or some other metric) to presences if they exceed a set threshold value, or absences if they do not. The resulting confusion matrix contrasts observed and predicted species occurrence states. Example metrics include precision, sensitivity, and true/false positive/negative rates. The topic of thresholding predictions is contentious in the SDM literature (Liu *et al.* 2005; Freeman & Moisen 2008; Guillera-Arroita *et al.* 2015), as well as in the context of evaluation (Lawson *et al.* 2014). Aside from debates about whether we should threshold or not, there is also debate about how to determine the threshold value. It is common to set the threshold at an arbitrary value of 0.5 (Freeman & Moisen 2008), which provides an obvious decision threshold (i.e. the species is more likely to be present than absent) when predictions are calibrated estimates of probability of presence. A frequent suggestion is to set the threshold to the observed prevalence of the species in question (Hanberry & He 2013). In the multi-species context of JSDMs, a logical extension of this debate is community-wide or species-specific thresholds. However, Lawson *et al* (2014) showed that by using a probabilistic confusion matrix we can calculate threshold-dependent metrics without the need to threshold the probabilistic distributions. To avoid any issues with the choice of threshold impacting our analysis we used the probabilistic confusion matrix approach when calculating our threshold-dependent metrics.

Community dissimilarity indices are widely used in community ecology to quantify the dissimilarity between two realisations of species assemblages. Examples of common metrics are Bray-Curtis dissimilarity and Jaccard's distance. When used to compare predicted assemblages with observed assemblages, these metrics are restricted to evaluating binary predictions, representing a single hypothesised assemblage.

Species richness metrics consider the ability of models to predict a single (but widely studied) aspect of community composition – the number of species present. Two common examples are absolute species richness or species richness difference, defined as predicted minus observed richness. We only evaluated species richness difference, as it directly compares predictions with observed richness values.

Likelihood metrics assess model fit by computing the probability of observing a given community assemblage, assuming the model is 'correct' in its structure and parameter estimates. For reasons of numerical stability, it is common to work with the log of the likelihood. We use the term independent log-likelihood to represent the typical log-likelihood metric used in SSDMs. This assesses each species individually across all sites – computing the probability of observing that species' presence/absence observations– and then combining these into a single goodness of fit metric, assuming the species' distributions to be independent. For JSDMs, we can also define a joint log-likelihood that assesses all species simultaneously as an assemblage at each site and accounts for the correlation structure encoded in the JSDM formulation. Both metrics can be calculated for SSDMs and JSDMs. The method for the independent log-likelihood calculation is identical for both, while calculating the joint log-likelihood for SSDMs, which do not estimate correlations between species, assumes an identity matrix correlation structure. Consequently, both the independent and the joint log-likelihood metrics give the same overall value for SSDMs. Since the probit SSDM is nested within the JSDM (the JSDM can recover the SSDM exactly by setting correlations to zero), the joint likelihood metric can be compared between the two model types.

3.2.5. CASE STUDY

We illustrate the different JSDM prediction types and their evaluation metrics with a case study in which we compare performance of a SSDM and a JSDM. We used a presence-absence dataset of frog species in the Greater Melbourne area of Victoria, Australia. The dataset contains 9 species, 104 waterbodies (sites), and 3 measured covariates (area, road density, presence of vertical wall), and was previously analysed in Pollock *et al* (2014). In our analysis, we standardised the two continuous variables and used five-fold random cross validation for model evaluation.

In this case study we compared the performance of a standard SSDM with a standard multivariate probit regression JSDM (Chib & Greenberg 1998). Both models were fit in R v3.5.2 (R Core Team 2018). The SSDM was implemented by fitting separate generalised linear models (with probit links) to each species in the dataset using the *stats* package (R Core Team 2018). We fit the JSDM using *BayesComm* v0.1-2 (Golding & Harris 2015). The JSDM applies independent normal priors, $\beta_k \sim N(0, 100)$, on the regression coefficients and an inverse Wishart prior, with $n + 2J$ degrees of freedom and scale matrix **I**, on the correlation coefficients. The model was fit by MCMC using a Gibbs sampler implemented in R and C++. We used a single MCMC chain of 11,000 samples, discarding the first 1,000 as burn-in, to sample the posterior distribution. Model fitting and predictions were undertaken on The University of Melbourne's Spartan HPC infrastructure (Meade *et al.* 2017).

Table 3.1: Summary of evaluation metrics for JSDM predictions.

Name	Binary predictions	Probabilistic predictions	Threshold-dependent metric	Threshold-independent metric	Community dissimilarity metric	Species richness metric	Likelihood metric
Accuracy / True Skill Statistic	X	X	X				
Area under the Receiver Operating Characteristic curve (AUC)		X		X			
Bray-Curtis dissimilarity	X	X			X		
Canberra Index	X	X			X		
Cohen's Kappa	X		X				
Diagnostic odds ratio	X	X	X				
F_1 score	X	X	X				
False discovery ratio	X	X	X				
False negative rate	X	X	X				
False omission rate	X	X	X				
False positive rate	X	X	X				
Gower Index	X	X			X		
Gower Index (alternative)	X	X			X		
Jaccard distance	X	X			X		
Kendall rank correlation coefficient		X		X			
Kulczynski Index	X	X			X		
Log Likelihood – Independent		X					X
Log Likelihood – Joint	X						X
Mountford Index	X	X			X		
Mean error (bias)		X		X			
Mean square error		X		X			
Negative likelihood ratio	X	X	X				
Negative predictive performance	X	X	X				
Pearson correlation coefficient		X		X			
Positive likelihood ratio	X	X	X				
Positive predictive performance / Precision	X	X	X				
R^2/ Coefficient of determination	X	X		X			
Raup-Crick dissimilarity	X	X	X				
Root mean square error		X		X			
Spearman rank correlation coefficient		X		X			
Species richness difference	X	X				X	
Sum of squared errors		X		X			
True negative rate / Specificity	X	X	X				
True positive rate / Sensitivity	X	X	X				
Youden's J statistic	X	X	X				

Once fit to the training data, the models were used to predict to the held-out test data. The SSDM returned probabilistic predictions for each species at a given site. For comparability with JSDM joint predictions, we also generated 1000 binary predictions of each species at a given site under a Bernoulli distribution. The JSDM was used to return probabilistic and binary marginal predictions, binary joint predictions, binary conditional predictions under three scenarios (one known species occurrence state, where the known species has either low-, medium- or high-prevalence), and probabilistic conditional-marginal predictions under the same three scenarios. We evaluated the binary JSDM predictions against the binary SSDM predictions and the probabilistic JSDM predictions against the probabilistic SSDM predictions.

We evaluated a suite of 35 metrics in total, but only present subsets for each prediction type in the main text of this article. Some metrics are only applicable to binary predictions, others only to probabilistic predictions, while some can be calculated for both. More detail on the metrics including how they are calculated, which prediction types they are appropriate for, and how to interpret them can be found in Appendix G. Threshold-dependent and threshold-independent metrics are calculated on a per-species basis, whereas community dissimilarity metrics are calculated per-site. Most metrics can be calculated on either a per-species or per-site basis (as they are just comparing two binary vectors: observations and predictions) so this split is based on how they are historically used in ecological literature. Metrics were evaluated once for each of the 1000 posterior samples drawn.

3.2.6. Output Analysis

We analysed the performance of the two models for different metrics using linear mixed effects models (MEM). For each combination of metric and prediction type, we fit an

MEM to assess the relationship between the response variable (the value of the metric) and the predictor variables (model type and cross-validation fold), with either species or site modelled as a random intercept (depending on whether the metric was evaluated on sites or species). These MEMs explicitly considered different residual variances for the two species distribution model types to account for evident inhomogeneity of variance.

For each MEM, we used a Kolmogorov-Smirnov test (Massey Jr 1951) to assess whether the distribution of residuals was significantly different from a normal distribution with the same mean and standard deviation. As we performed many comparisons, each requiring a MEM and normality test, we used a Bonferroni-corrected p-value to determine failure of the test, so that the overall significance level remained at 0.05 (Dunn 1961). We re-fitted any model that failed the KS-test using an appropriate transformation of the response variable to account for the potential impact of bounding ranges of the metric in question. Metrics bound $0 - \infty$ were log-transformed and those bound $0 - 1$ were logit-transformed.

3.3. RESULTS

One type of model (SSDM vs JSDM) did not outperform the other overall prediction types (Table 3.2). The JSDM outperformed the SSDM in log-likelihood metrics, the SSDM had more accurate estimates of species richness than the JSDM, and the community dissimilarity indices had mixed results.

The independent log-likelihood metric indicated that the JSDM was better fit to the data than the SSDM across all probabilistic prediction types. The largest difference was for conditional marginal predictions conditioned on a high prevalence species, with an effect (and 95% credible interval) of -2.65 [-5.74, 0.45]. The relative difference between

the SSDMs and the JSDMs for the joint log-likelihood estimates, which are independent of prediction type, was -0.69 [-1.61, 0.24].

Table 3.2: Summary of model performance for the different metric classes in each prediction type. ↑ = JSDM performed better,↓ = SSDM performed better, ↑↓ = mixed results, - = no difference between SSDM and JSDM, greyed out cells show where a metric class is not applicable for a prediction type.

Prediction Type	Threshold-independent metrics	Threshold-dependent metrics	Community Dissimilarity Indices	Species richness metrics	Likelihood metrics
Marginal (binary)		↓ / -	-	-	
Marginal (probabilistic)	↓ / -	↓ / -	↓ / -	-	↑
Conditional (low-prevalence)		↓	↑↓	↓	
Conditional (medium-prevalence)		↓	↑↓	↓	
Conditional (high-prevalence)		↓	↑↓	↓	
Joint		↓	↑↓	↓	
Conditional marginal (low-prevalence)	↓ / -	-	↓ / -	↓	↑
Conditional marginal (medium-prevalence)	↓ / -	↓ / -	↓ / -	↓	↑
Conditional marginal (high-prevalence)	↑ / -	↑ / -	↓ / -	↓	↑

The predicted species richness was essentially unbiased for the JSDM marginal predictions and for both the binary and probabilistic predictions of the SSDM. All other JSDM prediction types, which account for correlations between species, overpredicted species richness. The magnitude of the relative over prediction varied from 0.48 [0.43, 0.54] species per site for conditional marginal (high-prevalence) predictions to 1.0 [0.86, 1.14] species per site for joint predictions. For binary JSDM prediction types this was accompanied by higher true and false positive rates of approximately 4% and 14% respectively, relative to the SSDM. For probabilistic predictions, there was no difference identified in true or false positive rates between JSDMs and SSDMs.

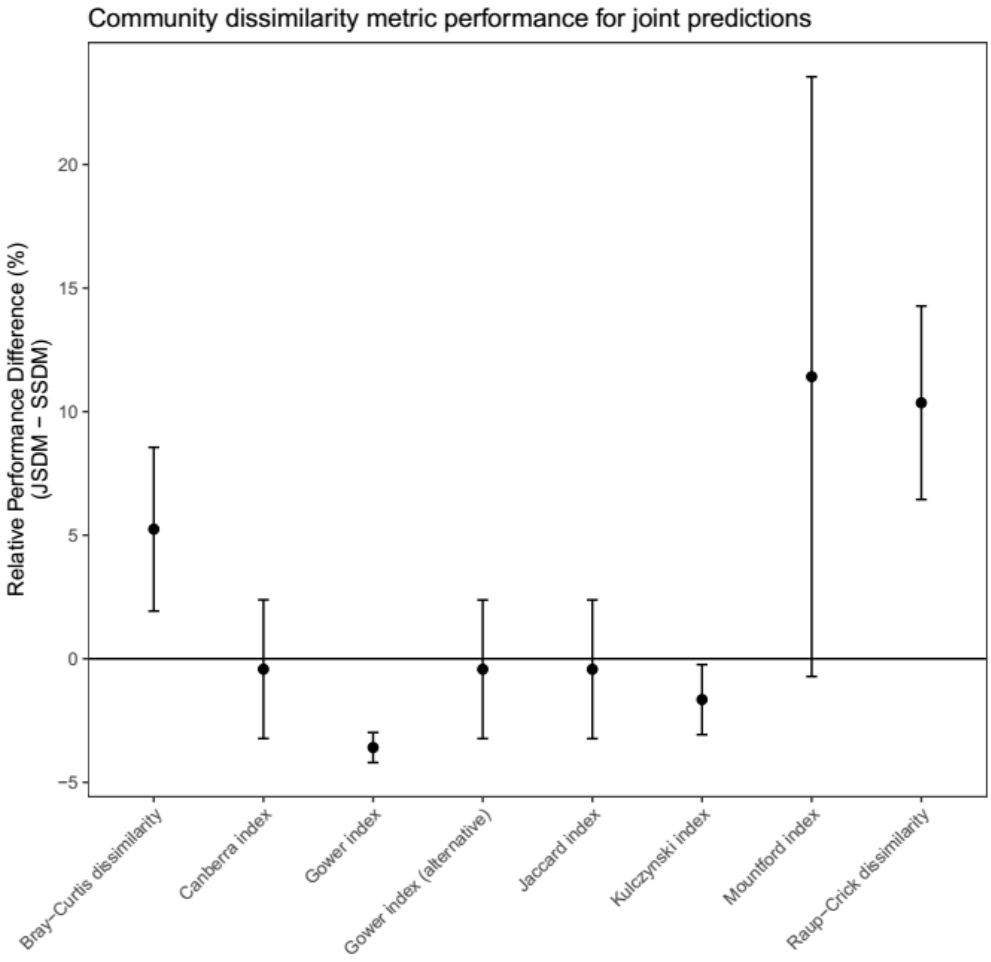


Figure 3.5: Relative performance difference between the JSDM joint predictions and binary SSDM predictions for the community dissimilarity metrics. The forest plots show the mean and 95% credible interval for the difference between the JSDM and SSDM estimates from the MEMs. The difference is calculated as $(JSDM - SSDM/|SSDM|) * 100$.

For both threshold-dependent and threshold-independent metric classes, the SSDM generally outperformed the JSDM. For binary prediction types that account for correlations between species, the SSDM outperformed the JSDM for all metrics. For the remaining prediction types, it was a mix of the SSDM performing better or no observed difference between the two models. Across all probabilistic prediction types, there was no difference in error rates between the SSDM and JSDM for mean error, mean square

error, and root mean square error metrics. The exception to this trend was for conditional marginal (high-prevalence) predictions, in which there was either no difference between the two models or the JSDM performed better.

The community dissimilarity metrics showed inconsistent trends (Figure 3.5). For non-marginal binary predictions, Bray-Curtis dissimilarity, Mountford index, and Raup-Crick dissimilarity showed more correct assemblage predictions for the JSDMs, while Gower index (and most metrics for conditional marginal (high-prevalence)) suggested JSDM assemblage predictions were more dissimilar to observed community data than SSDMs. For most probabilistic predictions, the majority of metrics showed no difference, but Gower index and Kulczynski index suggested worse assemblage predictions for JSDMs. A larger proportion of metrics suggested worse assemblage predictions for the JSDM when conditional on a high-prevalence species

Forest plots for all evaluation metric and prediction type combinations showing the mean value on the evaluation scale are available in Appendix H. Forest plots for the relative performance difference of the metrics are available in Appendix I.

3.4. DISCUSSION

We have reviewed and clarified the ways in which predictions of species distributions can be approached with JSDMs. Our review highlights the additional functionality that JSDMs enable relative to stacking single-species models. JSDMs have two main advantages in this regard: an ability to partition the effect of measured variables and residual correlations between species, and predictions of community assemblages that account for these correlations. By partitioning the effect of measured variables from

residual correlations, which may include species interactions, JSDMs potentially enable more accurate estimates of environmental drivers of species' distributions. By leveraging the information in the residual correlations, JSDMs can better predict community assemblages by simultaneously predicting all species rather than stacking multiple single-species predictions. Predictions are done for an entire assemblage using joint prediction, or for subsets of a community assemblage in conditional predictions. Both of these elements are integrated in the conditional marginal predictions, in which the measured variables/residual correlation partition is used in addition to the community assemblage process to predict single-species occurrences where some members of a community are known.

The different prediction types defined here for JSDMs correspond to different ecological aims. Marginal predictions correspond to the traditional single-species predictions; joint predictions can simultaneously predict entire assemblages; while the conditional and conditional marginal predictions let us explore ecological hypotheses about community assemblage processes, such as predator/prey interactions or how easy-to-detect indicator species can be used to help predict the rest of a community. Different aims also imply different evaluation metrics – community dissimilarity indices are suitable for community assemblage predictions (e.g. joint predictions) but may not always be appropriate for the single-species probabilities that can be returned via marginal or conditional marginal predictions. Each practitioner should determine an appropriate prediction type and evaluation metric, or metrics, for their ecological question.

We compared the predictive performance of JSDMs using different prediction types with a standard SSDM using a single dataset. We found no evidence to suggest either model outperformed the other in general. Some metric classes favoured one model, such

as species richness for SSDMs, while results for other metric classes, such as community dissimilarity indices, were mixed. Evaluations with a broader array of datasets and JSDMs seem warranted.

Both likelihood-based metrics indicated that the JSDM better fit the data than the SSDM, suggesting that accounting for residual correlations improved predictions of single species distributions. This is consistent with the results of Norberg *et al* (2019), who contrasted model pairs that were identical in all aspects except for residual correlations, although they did not use independent log-likelihood metrics directly. Our finding that the JSDM had a higher value of the joint likelihood metric suggests that accounting for residual correlations between species is more beneficial than assuming independence.

The mixed results for community dissimilarity metrics is due to their different formulas (see Appendix G for formulas). For example, Bray-Curtis dissimilarity and Gower index produced conflicting results regarding which model performed best (in favour of the JSDM and SSDM, respectively), which could result from their choice of denominator (they have identical numerators). For Bray-Curtis dissimilarity, the denominator is the sum of the number of observed and predicted presences, whereas the Gower index uses the total number of species. Thus, the denominator for the Gower index will be the same for SSDMs and JSDMs. In contrast, the Bray-Curtis index will have a larger denominator for JSDMs than SSDMs because JSDM predictions had higher true and false positive rates. The numerator would be higher for models overpredicting species richness, but the corresponding larger denominator of Bray-Curtis dissimilarity will balance this out. Therefore, we suggest careful consideration of the meaning of any given community dissimilarity metric in practice.

The JSDM tended to overpredict species richness for all prediction types that accounted for correlations between species. For our case study of nine frog species, the JSDM predicted approximately one extra species per site. In contrast, the mean species richness difference for both binary and probabilistic predictions of the SSDM was near zero. This finding contrasts with those of Guisan & Rahbek (2011) and Thuiller *et al* (2015), which suggested that SSDMs overpredict species richness. Thuiller *et al* (2015) suggested that SSDM overprediction will be more prevalent at finer spatial resolutions, such as that used in our case study. However, Zurell *et al* (2019) found results similar to ours, in which the JSDM overpredicted species richness compared to the SSDM.

As the JSDM only overpredicts species richness for prediction types that account for residual correlations, the estimated correlations could potentially explain why the JSDM tended to overpredict. The estimated species correlation coefficients for eight of the nine frog species were positive, and those eight species were negatively correlated with the ninth species. With the correlation matrix being positive for most species, the JSDM could predict likely assemblages in the absence of restrictions, such as site carrying capacities and/or dispersal limitations. In JSDMs the estimated species correlations could be explained by species interactions, but could also be driven by shared responses to unmeasured variables, dispersal, or other drivers of distributions (Barner *et al.* 2018). This overprediction could also explain the higher true and false positive rates of the JSDMs, as they predicted the presence of the correct species in addition to extra species that are positively correlated but absent.

A small body of literature focuses on larger-scale comparisons involving more species and model types (Zurell *et al.* 2019; Norberg *et al.* 2019), but without addressing all of the prediction types available to JSDMs. Norberg *et al* (2019) used joint predictions but used them differently than presented here. First, they generated binary assemblage

predictions but took the average of a large number of random draws to generate species-specific probabilities, which we have shown here to be an approximation of the marginal distribution. Second, they used joint predictions to generate species richness metrics equivalent to ours. Third, they used joint predictions and community dissimilarity metrics to test predictions of species turnover between sites rather than to test community composition at a site. To our knowledge, no JSDM studies have yet included conditional or conditional marginal distributions in this level of detail (but see Taylor-Rodriguez *et al* (2017) for a small scale exploration), although they have been identified as desirable avenues of research (Zurell *et al.* 2019; Norberg *et al.* 2019).

Both Norberg *et al* (2019) and Zurell *et al* (2019) included latent factor JSDMs in their comparisons, whereas we use a multivariate probit model. Norberg *et al* (2019) found that the HMSC JSDM (Ovaskainen *et al.* 2016b) outperformed both other JSDMs and SSDMs, while Zurell *et al* (2019) found that the *boral* JSDM (Hui 2016) had similar results to those described here. Zurell *et al* (2019) suggested that the poor performance of *boral* was a result of how the latent factor model is used when extrapolating predictions. Prediction using latent factor JSDMs can be performed in two ways: (1) marginalizing over the latent factors by assigning the mean value of modelled sites to prediction sites when performing regression-style predictions, and (2) defining the latent factor models in the same way as multivariate probit models (see Wilkinson *et al* (2019) for notation), which lets one use the prediction types defined in this paper. This second method may be a better alternative to latent factor model prediction than marginalizing over the latent factors.

3.5. CONCLUSION

JSDMs enable a variety of different ways to predict species distributions and community assemblages. Here we have defined environment-only marginal predictions, joint predictions for whole community assemblages, and conditional and conditional marginal predictions that can also leverage additional information on known species' occurrences. Previous studies have either not considered prediction with JSDMs, focussed on marginal predictions, or considered only limited aspects of joint prediction. We have also shown that there are several classes of evaluation metrics that can be applied to subsets of these predictions. Ecologists seeking to use these prediction methods and evaluation metrics should consider which method and metric are most closely linked to the ecological question they are investigating. A larger comparison of prediction types that considers different JSDM implementations and multiple datasets is required to evaluate general performance of these models and prediction types.

REFERENCE LIST

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Barner, A.K., Coblentz, K.E., Hacker, S.D. & Menge, B.A. (2018). Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, **99**, 557–566.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, **87**, 34–56.

Dunn, O.J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, **56**, 52–64.

Elith, J. & Leathwick, J.J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, …*, **40**, 677–697.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental*

Conservation, **24**, 38–49.

Freeman, E.A. & Moisen, G.G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48–58.

Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A., Latimer, A. & Rebelo, A.G. (2005). Modelling species diversity through species level hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 1–20.

Golding, N. (2015). Mosquito community data for Golding et al. 2015 (Parasites & Vectors) (Version 1). figshare. doi:10.6084/m9.figshare.1420528.v1. *figshare*.

Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.

Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites & Vectors*, **8**, 367.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., Tingley, R. & Wintle, B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292.

Guisan, A. & Rahbek, C. (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal

patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

Hanberry, B.B. & He, H.S. (2013). Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecol*, **13**, 13–19.

Hui, F.K.C. (2016). boral - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, **7**, 744–750.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54–64.

Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology*, **82**, 2560–2573.

Legendre, P. & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution models: A review. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with*

Mathematical and Computational Sciences, Proceedings, 4241–4247.

Massey Jr, F.J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, **46**, 68–78.

Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A. & Muñoz, J. (2012). Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, **7**.

Meade, B., Lafayette, L., Sauter, G. & Tosello, D. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., Foster, S.D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N.A., Holt, R.D., Hui, F.K.C., Husby, M., Kålås, J.A., Lehikoinen, A., Luoto, M., Mod, H.K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N.E., Gravel, D. & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, **89**, 1–24.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, **7**, 549–555.

Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species

distribution models. *Methods in Ecology and Evolution*, **7**, 428–436.

Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R.K. (2009). Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation*, **142**, 2501–2509.

Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.

Pineda, E. & Lobo, J.M. (2009). Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM) (J. McPherson, Ed.). *Methods in Ecology and Evolution*, **5**, 397–406.

R Core Team. (2018). R: A Language and Environment for Statistical Computing.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, **21**, 88–99.

Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015). From species distributions to meta-communities. *Ecology Letters*, **18**, 1321–1328.

Wilkinson, D.P. (2018). JSDM_Inference v0.1.6. Zenodo, https://dx.doi.org/10.5281/zenodo.1452066.

Wilkinson, D.P. (2019). JSDM_Prediction v0.1.0. Zenodo, https://dx.doi.org/10.5281/zenodo.3514766

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution*, **10**, 198–211.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Zhang, C., Chen, Y., Xu, B., Xue, Y. & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, **41**, 1876–1887.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, **00**, 1–13.

CHAPTER FOUR

A COMPARISON OF PREDICTIVE PERFORMANCE OF JOINT SPECIES DISTRIBUTION MODELS FOR PRESENCE-ABSENCE DATA

ABSTRACT

While there has been substantial literature on the evaluation of predictions from single species distribution models, the topic of prediction has only recently begun to be addressed for joint species distribution models (JSDMs). These studies have been limited in what aspects of prediction they cover: limited selection of models being compared, limited number of evaluation metrics, and/or not comparing the different prediction types available to JSDMs. In this study we perform a large scale comparison of the predictive performance of eight models: two stacked species distribution models (SSDMs) and six JSDMs. We fit these models to 22 real and simulated datasets, undertake four JSDM prediction types, and evaluate up to 32 metrics from five different classes that analyse different aspects of species distributions and the community assemblage process. We found that likelihood-based metrics indicated the JSDMs were better fit to the data than the SSDM, but that most other metric classes showed the SSDM outperforming the JSDMs by generally small amounts. The spatial and non-spatial implementations of the hierarchical multivariate probit regression model with latent factors typically performed better than the other JSDMs, but overall still performed worse than the SSDM. The SSDM predictions constrained with the spatially-explicit species assemblage modelling framework (SESAM) consistently outperformed both the SSDM and all JSDMs for both species- and community-level metrics. Our results indicate that despite the additional inference they provide about the community assemblage process by accounting for the residual association between species, JSDMs generally perform worse predictions than stacked single species models when evaluated at either the species or community level. The strong performance of the SESAM framework suggests that applying it to JSDM predictions is an interesting future avenue of research.

4.1. INTRODUCTION

Species distributions are influenced by the response of a species to both the abiotic and biotic conditions it encounters, but the ubiquitous single-species distribution model (SDM) has historically not accounted for the (biotic) effect of species interactions. Over the past decade, joint species distribution models (JSDMs) have seen rapid development for modelling multiple species simultaneously while accounting for both environmental responses and residual species associations that can be driven by factors including species interactions and shared responses to unmeasured environmental variables (Kissling *et al.* 2012; Wisz *et al.* 2013; Warton *et al.* 2015; Wilkinson *et al.* 2019). Research on JSDMs initially focussed on the development of different statistical approaches to implementing JSDMs (Pollock *et al.* 2014; Harris 2015; Golding & Purse 2016; Ovaskainen *et al.* 2016b; Hui 2016; Clark *et al.* 2017), then considered how to extend the framework to account for additional factors such as the effect of spatial scale or environmental gradients on species associations (Thorson *et al.* 2016; Ovaskainen *et al.* 2016a; Clark *et al.* 2017; Tikhonov *et al.* 2017). Only recently has the question of prediction using JSDMs begun to be explored in any detail (Zhang *et al.* 2018; Norberg *et al.* 2019; Zurell *et al.* 2019, Chapter Three).

SDMs are most often constructed as correlative statistical models that relate spatial data on environmental variables to data on a single species' distribution (Elith *et al.* 2006b). Species, however, exist as part of a community in which they directly and/or indirectly interact with other species. The community assemblage process can thus be considered in two parts: an abiotic filter where species-environment relationships influence which species can occur in a given environment, and a biotic filter where between-species relationships influence which species are more or less likely to co-occur (Götzenberger *et al.* 2012; Cornell & Harrison 2014). There have been attempts to account for species interactions inside the single-species

framework by using information on additional species as predictor variables (Araújo & Luoto 2007; Meier *et al.* 2010) or restricting predicted distributions to that of another species it depends on (Schweiger *et al.* 2012). Stacked species distribution models (SSDMs) attempt to model community assemblages by fitting multiple single-species models and overlaying, or stacking, the predictions (Guisan & Rahbek 2011; Thuiller *et al.* 2015). These approaches, however, either ignore species interactions or can only account for unidirectional interactions, where the direction of the interaction is already known (Kissling *et al.* 2012; Wisz *et al.* 2013; Pollock *et al.* 2014).

Joint species distribution models incorporate both the biotic and the abiotic factors impacting species and can thus link distribution modelling and community ecology. By accounting for the residual co-occurrence between species during model fitting, JSDMs might provide more accurate predictions than stacked SDMs which ignore potential interactions between species. As detailed in Chapter Three, there are four main ways in which JSDMs can be used to make predictions, each considering different aspects of the community assemblage process (see Figure 3.3 in Chapter Three). Marginal prediction types are species-level predictions and can be performed in two ways. Marginal predictions are environment-only predictions, akin to those of single species models, but by accounting for species co-occurrence during the model fitting process they should provide more accurate estimates of the species' response to the environment. Conditional marginal predictions are species-level predictions that account for the known occurrence state of some, but not all, other species at a site and are marginal to the rest. Joint prediction types are community-level predictions that leverage the additional information from the JSDM on residual species occurrence that are unavailable to single-species models. Joint predictions predict entire community assemblages simultaneously, while conditional predictions predict community assemblages where subsets of species have known occurrence states.

The ability of a model to accurately predict species distributions needs to be evaluated to be confident that the model will perform well in practice. While there is substantial literature on the evaluation of single species SDMs (Fielding & Bell 1997; Liu, White & Newell 2009; Lawson *et al.* 2014), there is minimal research into the evaluation of the multi-species predictions of JSDMs. Whilst the metrics commonly used to evaluate single species predictions can be applied to the predictions of multi-species models, the multivariate nature of these models also allows for evaluation of predictions using the community dissimilarity indices widely used in community ecology (Legendre & De Cáceres 2013). As detailed in Chapter Three, there are five classes of evaluation metrics that can be applied to the predictions of JSDMs: threshold-independent, threshold-dependent, community dissimilarity, likelihood-based, and species richness metrics. These metrics can be used to compare overall JSDM performance against SSDMs, as well as to compare the performance of different JSDM implementations against each other.

The use of JSDMs for prediction has only begun to be addressed in the literature (Ovaskainen *et al.* 2016b; Zhang *et al.* 2018; Zurell *et al.* 2019; Norberg *et al.* 2019). However, these studies have been limited in what aspects of JSDM predictive performance they have considered. Only one of these studies (Norberg *et al.* 2019) has considered any of the prediction types that account for the residual correlations between species, and no study has yet considered either of the conditional prediction types. Only one of these studies has compared more than one JSDM implementation, and they all have compared predictive performance for only a select few evaluation metrics. In this study we compare the predictive performance of six JSDMs and two SSDMs when fit to two real and twenty simulated datasets. We use up to 32 metrics over five metric classes to evaluate the SSDM predictions against the four different prediction types available to JSDMs.

4.2. MATERIALS AND METHODS

In this study we compare the predictions of six JSDMs and two SSDMs across 22 datasets. We utilise 32 metrics from five metric classes to evaluate the predictive performance of the SSDMs and the four different prediction types available to JSDMs. The different aspects of this workflow are described below.

4.2.1. MODELS

The two SSDM models are generalised linear models with a probit link, fit individually to each species and then stacked. The two models differ in their approach to stacking their predictions. The first is a standard stacked approach (SSDM) where the individual species predictions are summed together. The second approach is spatially explicit species assemblage modelling (SESAM: Guisan & Rahbek 2011) that combines SDMs with macroecological models to set an upper limit to the number of species predicted to occur at a site.

All six JSDMs are based on the multivariate probit regression model of Chib and Greenberg (1998). The first is the standard multivariate probit regression model (MPR) implemented in the R package *BayesComm* (Golding, Nunn & Purse 2015). Second, the hierarchical multivariate probit regression model (HPR) of Pollock *et al* (2014). Third, the multivariate probit regression with latent factors (LPR) implemented in the R package *boral* (Hui 2016). Fourth, the multivariate generalised regression model (DPR) implemented in the R package *gjam* (Clark *et al.* 2017). The fifth and sixth models are the spatial (HLR-S) and non-spatial (HLR-NS) implementations of the hierarchical multivariate probit regression model with latent factors implemented in the R package *HMSC* (Ovaskainen *et al.* 2016b). All of these models are implemented using a Bayesian framework and fit using Markoc chain Monte

Carlo (MCMC). A summary of the models and their structural components is found in Table 4.1.

The JSDMs are all fit using their default or suggested settings as defined in their source articles or, if not mentioned, the documentation of the software implementing the models. Model equations, default priors, and MCMC regimes are defined in greater detail in Wilkinson *et al* (2019) and in Appendix J.

4.2.2. PREDICTION METHODS

Predictions from single-species distribution models, or stacked single-species models, are environment-only predictions that ignore any possible effects of species interactions. For the SSDM approach a prediction is obtained individually for each species using the regression coefficients estimated in the model and the measured values of the corresponding variables at the sites being predicted to. This results in a predicted probability of presence for each species at each site. These predictions can also be used to generate binary predictions of presence of a given species at a given site by taking draws from a Bernoulli distribution with the corresponding predicted probability of presence as its parameter. The SESAM model constrains predictions to calibrate the predicted species richness at each site. The model uses an estimate of species richness at a site, either from a macroecological model or an alternative source, to constrain predictions using the probability rank rule such that it predicts species as present in decreasing order of probability of occurrence up to the maximum number permitted by the constraint. Thus SESAM can only provide binary predictions.

JSDMs, in contrast, provide information on correlations between species in their model outputs which can then be used to inform the prediction. Marginal JSDM predictions are environment-only predictions, similar to those of stacked single-species models,

	Source Paper	R Package	Hierarchical Regression Coefficients	Dirichlet Process	Latent Factors	Multivariate	Probit	Logistic	Regression	Spatially-Explicit
MPR	Golding *et al*. (2015)	*BayesComm*				X	X		X	
HPR	Pollock *et al*. (2014)	NA	X			X	X		X	
LPR	Hui (2016)	*boral*			X	X	X		X	
DPR	Clark *et al*. (2017)	*gjam*		o	o	X	X		X	
HLR-S	Ovaskainen *et al*. (2016a)	*HMSC*	X		X	X	X		X	X
HLR-NS	Ovaskainen *et al*. (2016a)	*HMSC*	X		X	X	X		X	

Table 4.1: Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an accompanying R package have code scripts available in the supplementary material of their source paper.

where the information about species correlations is ignored during the prediction process. This returns a probabilistic prediction for each species at a given site, and binary predictions can be generated by taking Bernoulli samples, as for the SSDM. To incorporate prediction uncertainty implicit in the posterior distribution over model parameters in these Bayesian models, Bernoulli samples can be drawn for each species, site, and each of a number of samples of predicted probability from the model posterior.

Joint predictions use the information on species co-occurrence to construct a joint probability distribution over plausible community assemblages. This probability distribution can be used either to evaluate the posterior probability of an observed or hypothesised assemblage, or can be used to simulate a number of binary community assemblage predictions from this joint distribution. This method uses the regression coefficients, measured variables at a site, and the species correlation matrix to generate a multivariate normal distribution. Random draws from this distribution represent a plausible community assemblage at a site. Joint predictions consider the distribution over the whole assemblage, and cannot be expressed in terms of probabilistic predictions for individual species at a site. Note that averaging the presence-absence of each species across simulated community assemblages (or otherwise marginalising the joint distribution) results in an approximation to the marginal distribution, and discards the information on species co-occurrence that the joint prediction encodes.

Conditional predictions are performed in the same manner as joint predictions but with the additional information of known occurrence states of one or more species in the community. The known occurrence state of a species truncates the multivariate normal distribution on one axis. As the total probability must sum to 1 this affects the probability of each of the remaining possible community assemblages.

Conditional marginal predictions similarly truncate the multivariate normal distribution based on the occurrence state of one or more species. Rather than representing the joint distribution of the remaining species (as in conditional prediction), the distribution over the remaining species is marginalised, to yield a single probability of presence of a species given both environmental conditions and the known presence/absence status of other species. Thus, this prediction method generates probabilistic predictions that are conditional on one or more species but marginal to the remainder.

An in-depth explanation of these prediction methods can be found in Chapter Three or Appendix K.

4.2.3. DATASETS

For this comparison we have used two real datasets, on frogs and eucalypts, and twenty simulated presence-absence datasets. The frog dataset comprises 9 species, 104 sites, and 3 covariates from Melbourne, Australia (Pollock *et al.* 2014). The eucalypt dataset comprises 12 species, 458 sites, and 7 covariates from Grampians National Park, Australia (Pollock 2014). The simulated datasets all have 10 species, 100 sites, and 5 covariates. Each simulated dataset has 3 continuous and 2 binary variables. The species correlation matrix was generated as a latent variable approximated correlation matrix using three latent factors. Species presence absence data was generated using the *communitySimul* function from the *HMSC* R package (Blanchet, Tikhonov & Norberg 2019).

For all datasets the continuous variables were standardised while binary variables were unmodified. We implemented two different cross-validation approaches depending on the dataset. For the frog and 5 of the 10 simulated datasets we implemented five-fold random cross-validation, where the sites were randomly assigned to each of the five

folds using the *createFolds* function from the *caret* R package (Kuhn *et al.* 2019). For the eucalypts and remaining 5 of the simulated datasets we used five-fold spatial block cross-validation using the *spatialBlock* function from the *blockCV* R package (Valavi *et al.* 2018, 2019).

4.2.4. Evaluation Metrics

The evaluation metrics available for evaluating JSDM predictions can be broadly classified into five groups based on the aspects of performance they consider. Threshold-independent metrics evaluate continuous predicted probabilities against observed presence-absence data. A common threshold-independent metric used for SDMs is the Area Under the Receiver Operating Characteristic Curve (AUC), but other examples include root mean square error (RMSE), and the coefficient of determination (R^2).

Threshold-dependent metrics compare binary predictions against observed presence-absence data. Predicted probabilities are converted to presences if they exceed a set threshold value or absences if they do not. From this we can generate a confusion matrix that contrasts our observed and predicted species occurrence states. Example metrics here include precision, sensitivity, and true/false positive/negative rates. However, thresholding predictions is a contentious topic in the SDM literature (Liu *et al.* 2005; Freeman & Moisen 2008; Guillera-Arroita *et al.* 2015). In addition to debates about the use of thresholds in general, there are also debates about how to determine the threshold value. Thresholds are commonly set to an arbitrary value of 0.5 (Freeman & Moisen 2008), but an alternative is to make the threshold equivalent to the observed prevalence of the species (Hanberry& He 2013). A logical extension of this debate for JSDMs is species-specific or community-wide thresholds. However, Lawson *et al*

(2014) showed that by calculating a probabilistic confusion matrix we can evaluate threshold-dependent metrics without thresholding predictions. We have used this probabilistic confusion matrix approach here to avoid any influence of the choice of threshold impacting our analysis.

As JSDMs are multi-species in nature we can use an additional set of evaluation metrics from community ecology in the form of community dissimilarity indices. These metrics compare how dissimilar our observed and predicted species assemblages are. Examples of these frequently seen in community ecology are Bray-Curtis dissimilarity and Jaccard distance. These metrics are restricted to evaluating binary predictions. To evaluate these metrics on probabilistic predictions, we simulated binary community assemblages from the appropriate probability distribution, as defined above. For Bayesian models, each of these communities was drawn using a different posterior prediction sample, for non-Bayesian models the same number of community assemblages were simulated for each site, but using the same parameters.

Species richness metrics consider a model's ability to predict a single aspect of community composition: the number of species present at a site. We consider species richness difference - the predicted richness minus the observed richness, which compares predictions against observed data.

The final set of evaluation metrics we have considered are likelihood-based metrics which assess model fit by computing the probability of observing a given community assemblage, assuming a particular model structure, and given a set of parameter estimates representing the prediction. It is common to work with the log of the likelihood for numerical stability reasons. The independent log-likelihood represents the typical log-likelihood metric used in SSDMs. This metric independently assesses each

species across all sites – computing the probability of observing a species' presence/absence observations– and combines them into a single goodness of fit metric, assuming the species' distributions to be independent. We can also define a joint log-likelihood that simultaneously assesses all species, as an assemblage, at each site and accounts for the correlation structure encoded in the core JSDM formulation – a multivariate probit model with a given mean and correlation. We can calculate both metrics for SSDMs and JSDMs. The independent log-likelihood calculation method is identical for both models, while the joint log-likelihood calculation for SSDMs assumes an identity matrix correlation structure to correspond to the model assumption of species independence.

More detail on the metrics including how they are calculated, which prediction types they are appropriate for, and how to interpret them can be found in Appendix L.

4.2.5. Model Prediction Comparisons

For the standard SSDM we made both binary and probabilistic predictions of community assemblages, and for SESAM we made binary predictions. For the JSDMs we evaluated nine prediction types: binary and probabilistic marginal predictions, binary joint predictions, binary conditional predictions for low-, middle-, and high-prevalence known species scenarios, and probabilistic conditional marginal predictions for low-, middle-, and high-prevalence known species scenarios. We evaluated a suite of 32 evaluation metrics in total, but some applied only to binary or probabilistic prediction types.

4.2.6. Analysis of Results

We used linear mixed effects models (MEMs) to evaluate the predictive performance of the eight models for each of the nine prediction methods. We fit an MEM to each combination of evaluation metric and prediction type to assess the relationship between the response variable, the test statistic, and the predictor variables: model, dataset, and cross-validation fold, with a random effect on the intercept of either species or site (depending on the test statistic). We included a partial interaction between model and dataset for the HPR and DPR models to account for observed patterns in the residuals. The MEMs explicitly considered different residual variances for the eight model types to account for evident inhomogeneity of variance. We assessed whether the model residuals met the model assumptions of being normally-distributed with homogenous variance (after accounting for inhomogeneity between model types) with a Kolmogorov-Smirnov test (Massey Jr 1951), hereafter referred to as a KS-test. The KS-test determines if the distribution of the residuals is significantly different from a normal distribution with the same mean and standard deviation. As we performed a large number of comparisons, we used a Bonferroni-corrected p-value, $\frac{0.05}{405} = 1.2 * 10^{-4}$ to consider whether normality assumptions were violated; to reduce the sensitivity of these tests (Dunn 1961). Models that failed the KS-test were refit using a transformation appropriate for the response variable to account for the potential impact of bounding ranges of the test statistic in question. We used a log transformation for strictly positive test statistics, and a logit transformation for those in the unit interval.

4.2.7. SOFTWARE

All models were fit using R v3.5.2 (R Core Team 2018). R packages for model fitting include *BayesComm* v0.1-2 (Golding & Harris 2015), *boral* v1.7 (Hui 2018), *gjam* v2.2.5 (Clark & Taylor-Rodríguez 2018), and *HMSC* v2.2-0 (Blanchet, Tikhonov & Norberg 2019). R packages required for prediction and prediction evaluation are *mvtnorm* v1.0-10 (Genz *et al.* 2019), *tmvtnorm* v1.4-10 (Wilhelm & B G 2015), *TruncatedNormal* v1.0 (Zdravko 2015), *Metrics* v0.1.4 (Hamner & Frasco 2018), *caret* 6.0-84 (Kuhn *et al.* 2019), *vegan* v2.5-5 (Oksanan *et al.* 2018), and *psych* v1.8.12 (Revelle 2018). Model fitting, prediction, and evaluation were all undertaken on The University of Melbourne's Spartan HPC infrastructure (Meade *et al.* 2017).

4.3. RESULTS

No model outperformed the others across all prediction types, but there were consistent trends within each class of validation metrics. The relative performance of all models compared to the SSDM are summarised in Figure 4.1 for the marginal prediction types and Figure 4.3 for the joint prediction types. Likelihood-based metrics showed that the JSDMs were better fit to the data than the SSDM. Both threshold-dependent and threshold-independent metrics indicated better performance by the SSDM, although the difference was small for the majority of metrics. The SSDM generally outperformed JSDMs for community dissimilarity metrics, and the difference was greater for joint prediction types than marginal ones. The JSDMs almost always overpredicted species richness compared to the SSDM for marginal predictions, but HLR-S and HLR-NS had more accurate estimates than the SSDM for most joint prediction types. The SESAM

model outperformed both the SSDM and JSDMs for threshold-dependent and community dissimilarity metrics for binary prediction types.

Only a selection of figures is presented in the main article. Forest plots for the absolute value of model performance for each evaluation metric are presented in Appendix M. Individual heatmaps for the different prediction types are presented in Appendix N.

4.3.1. Marginal Prediction Types

4.3.1.1. Threshold-independent and Threshold-dependent Metrics

The SSDM outperformed all JSDMs for almost every threshold-independent metric and marginal prediction type combination. The only exception to this was the bias metric for DPR in low-prevalence conditional marginal predictions. The relative performance of the JSDMs for the bias metric suggests very poor performance of the JSDMs, but on an absolute scale these differences are actually quite small. This is because the optimum value of the bias metric is 0, so as model performance increases the resultant small metric value leads to high relative differences between two values with little absolute difference (Figure 4.2). Across all probabilistic predictions the SSDM had a mean bias of $-2.7*10^{-4}$, while the JSDMs average $-1.6*10^{-3}$, i.e. all models were largely unbiased. For all other metrics the average JSDM performance across all marginal prediction types was -6.3% [-20.7, -0.3]. HPR was the worst performer for the error-based metrics with a mean relative difference of -18.3% [-61.8, -2.2]. In the low- and medium-prevalence case, there was a large relative difference between SSDM and the JSDMs in sum of squared errors from conditional marginal predictions, with a mean of -26.9% [-62.0, -13.0]. The most common single-species evaluation metric applied to SDMs is

AUC and the mean difference over all marginal prediction types between the SSDM and the worst performing JSDM is only 0.01.

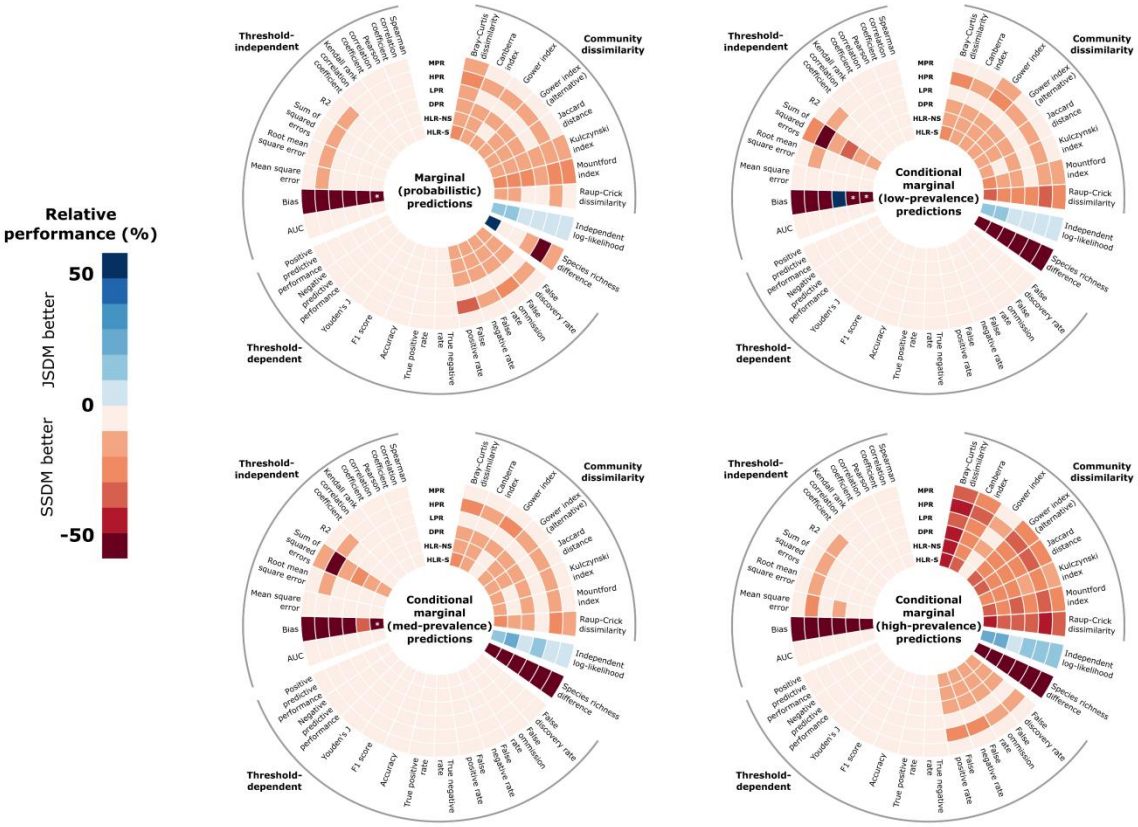


Figure 4.1: Heat maps showing the relative performance of the different JSDMs compared to the SSDM for all metrics applicable to the marginal prediction types. Each ring on the y-axis represents a particular model, while each ray on the x-axis represents a different evaluation metric, which have been clustered my class. Values in blue indicate better performance by the model compared to the SSDM, while red values indicate worse performance by the model compared to the SSDM. The four prediction types shown are a) probabilistic marginal, b) low-prevalence conditional marginal, c) medium-prevalence conditional marginal, and d) high-prevalence conditional marginal prediction. These heat maps have been generated using Circos (Krzywinski *et al.* 2009).

The JSDMs performed worse than the SSDM for threshold-dependent metrics but the difference was small. Across all probabilistic predictions the relative JSDM performance was -3.9% [-18.2, -0.2]. The only notable relative differences were the performance of the HPR, DPR, HLR-S, and HLR-NS models for the false rate metrics for probabilistic marginal predictions (-17.8% [-28.7, -13.6]) and high-prevalence conditional marginal predictions (-16.5% [-25.8, -13.5]).

4.3.1.2. COMMUNITY DISSIMILARITY METRICS

The SSDM outperformed the JSDMs for all community dissimilarity metrics on marginal prediction types. The mean JSDM performance was -14.5% [-25.5, -4.9] for probabilistic marginal predictions, -14.7% [-34.4, -5.7] for low-prevalence, -12.4% [-22.4, -6.1] for medium-prevalence, and -27.8% [-44.1, -6.1] for high prevalence conditional marginal predictions. In the high-prevalence predictions the largest differences between the SSDM and JSDMs were for Bray-Curtis dissimilarity (-40.4% [-46.9, -34.6]) and Raup-Crick dissimilarity (-36.2% [-42.5, -30.9]), while the smallest was for Gower index (-9.3% [-17.8, -5.15]). The HPR model was the worst performer overall with a mean relative difference of -23.1% [-43.7, -14.7] across all marginal prediction types. LPR and MPR were the best performers amongst the JSDMs with a mean relative performance for all marginal prediction types of -12.7% [-32.5, -4.8], compared to -19.7% [-41.2, -7.8] for the other JSDMs.

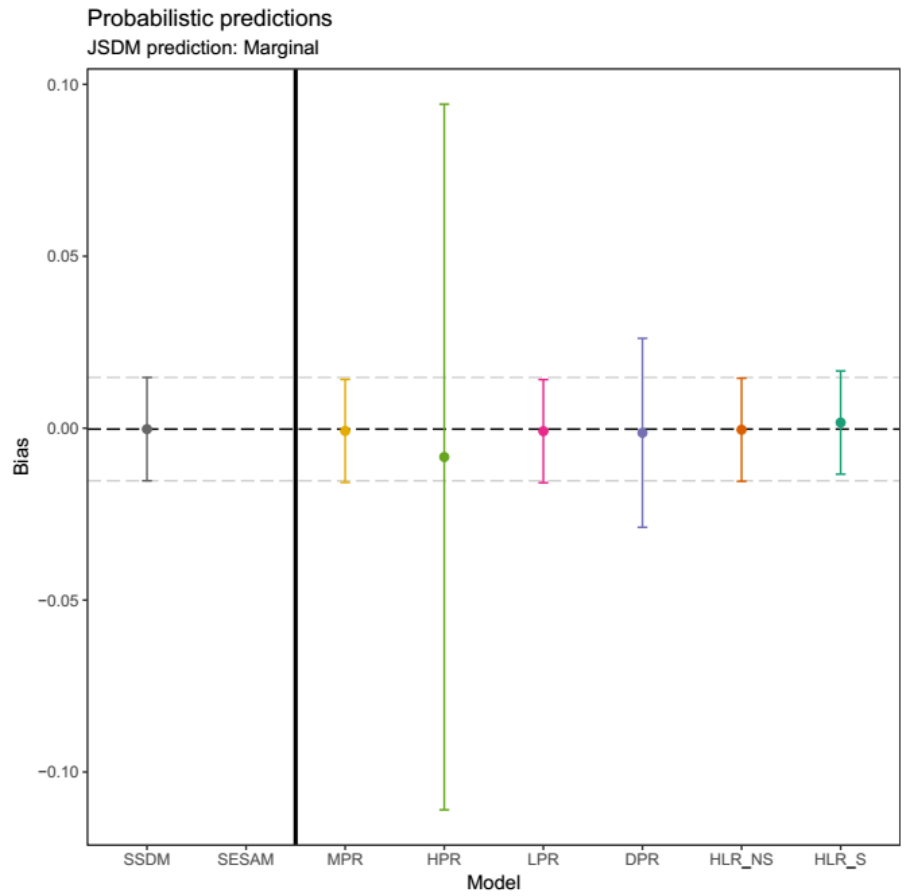


Figure 4.2: Model performance for the bias (mean error) evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% quantiles of metric values, over species in each datasets, after accounting for dataset and fold using a linear mixed-effects model. The dashed dark grey line corresponds to the SSDM mean bias, and the dashed light grey line corresponds to 95% quantile from SSDM predictions. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs. This figure is illustrative of the absolute metric plots provided for all metrics and prediction types as supplementary information in Appendix M. SESAM is not plotted here as it does not make predictions for individual species.

4.3.1.3. SPECIES RICHNESS

The SSDM generally outperformed the JSDMs in species richness difference estimates for marginal predictions, but the difference was not as severe as might be suggested by the relative performance results shown in Figure 4.1. As the optimum value of this metric is 0, differences relative to a small value, as found in the SSDM, can lead to large

relative percentage differences so we discuss the absolute differences here. The SSDM had a mean species richness difference of 0.06 [0.02, 0.08] across the MEMs for marginal prediction types. For probabilistic marginal predictions the JSDMs had a mean species richness difference of 0.09 [0.02, 0.15], with HLR-S outperforming the SSDM with a value of just 0.007. For low-, medium-, and high-prevalence conditional marginal predictions the JSDMs had a species richness difference of 0.86 [0.79, 0.93], 0.63 [0.57, 0.70], and 0.34 [0.28, 0.39] respectively. HLR-S exhibited the smallest differences of all of the JSDMs for each marginal prediction type.

4.3.2. Joint Prediction Types

4.3.2.1. Threshold-independent and Threshold-dependent Metrics

For threshold-dependent metrics evaluated on joint predictions the JSDMs performed worse than the SSDM. For binary marginal predictions the JSDMs had a mean relative difference of -3.4% [-6.9, -0.7] from the SSDM except for the false rate metrics for HPR, DPR, HLR-S, and HLR-NS with a mean of -17.9% [-29.0, -13.5]. The mean relative difference between JSDMs and SSDMs for conditional predictions was -4.1% [-15.0, -0.2] across all metrics, with a larger relative difference for the HPR and DPR models on the Cohen's Kappa and Youden's J metrics of -21.8 [-66.7, -11.9]. For joint predictions there was a large relative difference from SSDMs of -36.3 [-97.7, -3.5] across all JSDMs for false rate metrics. This was strongest for HPR and DPR with means of -70.4% and -88.7% respectively. LPR has a comparatively small relative difference of -4.0%. Across all threshold-dependent metrics for joint prediction HPR and DPR had a large relative difference of -35.7% [-82.2, -12.0] and -45.2% [-100.9, -15.0] respectively.

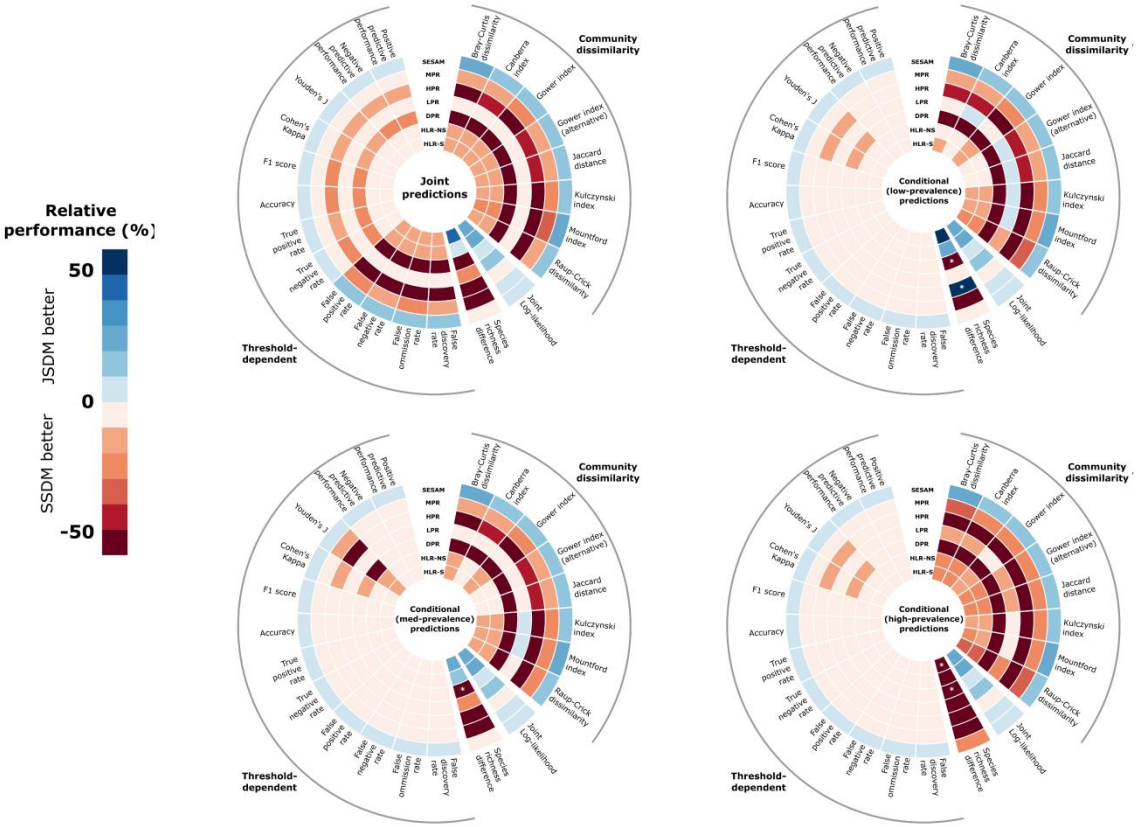


Figure 4.3: Heat maps showing the relative performance of the different JSDMs and SESAM compared to the SSDM for all metrics applicable to the joint prediction types. Each ring on the y-axis represents a particular model, while each ray on the x-axis represents a different evaluation metric, which have been clustered my class. Values in blue indicate better performance by the model compared to the SSDM, while red values indicate worse performance by the model compared to the SSDM. The four prediction types shown are a) joint, b) low-prevalence conditional, c) medium-prevalence conditional, and d) high-prevalence conditional prediction. These heat maps have been generated using Circos (Krzywinski *et al.* 2009).

4.3.2.2. COMMUNITY DISSIMILARITY METRICS

For community dissimilarity metrics on binary marginal predictions the SSDM outperformed the JSDMs. The relative performance difference was largest for HPR, DPR, HLR-S, and HLR-NS (-14.2% [-24.1, -10.0]), and smallest for MPR and LPR (-2.8% [-4.1, -2.0]). For joint prediction types the HPR and DPR models performed worst with a mean relative performance difference of -75.3% [-132.7, -44.7]. LPR performed

best with a mean difference of -4.3% [-20.5, 6.3], and performed better than the SSDM for several metrics in low- or medium-prevalence scenarios.

4.3.2.3. SPECIES RICHNESS

The results for species richness were mixed. Across all joint prediction types the mean difference between predicted and observed species richness was 0.06 [0.01, 0.09] for the SSDM. HLR-S had a small mean species richness difference of 0.02 [-0.01, 0.04] while HLR-NS performed similarly to the SSDM with a mean difference of 0.06 [0.03, 0.09]. LPR, HPR, and MPR exhibited slightly larger mean species richness differences than the SSDM with values of 0.07 [0.02, 0.1], 0.09 [-0.03, 0.17], and 0.14 [0.1, 0.17] respectively. DPR was the only JSDM to underpredict species richness with a mean difference value of -0.12 [-0.38, 0.14].

4.3.3. LIKELIHOOD-BASED METRICS

Both likelihood metrics indicate that the JSDMs were better fit to the out-of-sample data than the SSDMs, under the assumptions of the univariate and multivariate probit models. For the probabilistic marginal predictions the JSDMs on average performed 12.2% [7.9, 18.4] better than the SSDM, and for the low-, medium-, and high-prevalence conditional marginal predictions the JSDMs on average performed 12.5% [8.0, 19.0], 13.6% [8.5, 20.4], and 15.2% [10.4, 22.1] better than the SSDM. The HLR-S and HLR-NS models outperformed the other JSDMs across all probabilistic prediction types with a mean relative independent log likelihood of 19.3% [17.0, 22.0] compared to 9.4% [7.9, 12.0] for the remaining models. Results were similar for the joint log-likelihood except that HPR performed worse relative to the SSDM by 7.8%.

Relative to the SSDM, HLR-S and HLR-NS again performed better than the other JSDMs by an additional ~10%.

4.3.4. SESAM PERFORMANCE

SESAM outperformed the SSDM and all JSDMs for the threshold-dependent and community dissimilarity metrics. SESAM had a mean relative performance difference of 3.5% [0.7, 15.6] across threshold-dependent metrics for all joint prediction types. This performance is strongest in the false rate metric for binary marginal and joint predictions with a mean relative difference of 13.2% [10.0, 16.7]. For the community dissimilarity metrics, SESAM had a mean relative performance across all joint prediction types of 17.8% [14.4, 22.5]. The mean species richness difference of the SESAM predictions was 0.06 [0.01, 0.09] which is equivalent to the SSDM.

As some of these results are unexpected we performed checks to ensure the MEMs were not returning erroneous results. We found no evidence to suggest our results are an artefact of the model fitting process and present the result of these checks in Appendix O.

4.4. DISCUSSION

4.4.1. THRESHOLD-INDEPENDENT AND THRESHOLD-DEPENDENT METRICS

The SSDM routinely outperformed the JSDMs for all threshold-independent metrics, but the magnitude of the difference was generally small on the absolute scale of the metric. This is in contrast to the results of Zurell *et al* (2019) which found the JSDM performed worse than the stacked model. However that study used a stacked ensemble

SDM while we have just used a generalised linear model based SDM. Ensemble models have been shown to outperform single models (Dormann *et al.* 2018b), which will increase the difference in performance between the JSDM and SSDM. Similarly small differences on the absolute scale were seen for both common error-based metrics (e.g. bias, MSE, RMSE) or the correlation metrics. The only exception to this was the SSE metric for low- and medium-prevalence conditional marginal metrics, but all models, including the SSDM, had significantly larger errors here compared to the other marginal prediction types. This suggests that at the species-level the JSDM predictions are performing similarly to the SSDM for threshold-independent metrics.

The JSDMs performed worse than the SSDM for all threshold-dependent metrics in both the marginal and joint prediction types, but like the threshold-independent metrics the difference is negligible. The relative performance of HPR, DPR, HLR-S, and HLR-NS suggests a large difference between them and the SSDM for the false rate metrics in probabilistic marginal and high-prevalence conditional marginal predictions, but on the absolute scale the performance difference is ≤4% for all metrics. The same trend can be seen in the joint prediction type exhibit larger relative differences for the false rate metrics. On the absolute scale these differences are ≤4% for most metrics, and usually ≤1% for MPR and LPR. The exception to this is that both HPR and DPR perform poorly in the joint predictions, with difference on the absolute scale of 15.5% and 19.6% respectively. This suggests that at the species-level the JSDM predictions are generally performing similarly to the SSDM for threshold-dependent metrics.

4.4.2. Community Dissimilarity Metrics

One of the purported benefits of JSDMs over single species models is that by accounting for residual co-occurrences between species during the model fitting process

they are able to better inform on the community assemblage process. Whether it is due to factors like species interactions or shared responses to unmeasured environmental covariates, these identified species associations provide information on which species are more or less likely to co-occur outside of their response to the measured variables. However, our results have found that community dissimilarity metrics show the JSDMs to be predicting worse than the SSDM. The relative differences suggest that the difference is significant, but on the absolute scale the performance difference is fairly minor. Chapter Three evaluated these metrics on a single dataset to mixed results and found that MPR outperformed the SSDM for some metrics when MPR was over-predicting species richness. Here we have evaluated the metrics over 22 datasets and shown that the mean species richness difference is smaller, and all community dissimilarity metrics indicate worse performance for the JSDM. By accounting for the effect of dataset we have been able show that JSDMs are performing worse for all community dissimilarity metrics. However, LPR does outperform the SSDM for some conditional predictions, but only by a mean of 0.004, so performance is realistically near identical.

4.4.3. LIKELIHOOD

In almost all cases the JSDMs outperformed the SSDM for both of the likelihood-based metrics and were better fit to the out-of-sample data. This suggests that, by accounting for residual correlations in the model fitting process, JSDMs are able to estimate more accurate regression coefficients. This result is consistent with those of Norberg *et al* (2019), where they contrasted model pairs that were identical in all aspects except for accounting, or not, for residual correlations. However, more accurate coefficients should lead to more accurate predictions which is not what we observed in this study. This

suggests that while JSDMs are able to fit better to multi-species datasets than models that ignore the residual associations between species, there is some aspect of the prediction process causing them to predict relatively poorly.

4.4.4. Species Richness

The species richness difference metric presented mixed result where the SSDM did not continuously perform better or worse than the JSDMs. All models generally over-predicted species richness, with the exception of DPR for conditional predictions, but the mean difference was minimal at ~ ≤0.1 species per site. The HLR-S and HLR-NS models were the best performing JSDMs overall, and for all prediction types except the conditional marginal predictions HLR-S outperformed the SSDM and HLR-NS performed equivalently to it. As HLR-S was the only JSDM to regularly outperform the SSDM it suggests that the effect of spatial scale and/or spatially-driven unmeasured variables could be a potential driver of this result. Species co-occurrence can be driven by several factors, including species interactions or shared responses to environmental conditions, which operate at different scales. Two species can have the same environmental condition preferences but tend to rarely co-occur at a finer scale. This could lead to overpredicting species at a site if the effect of shared responses is stronger leading to higher co-occurrence rates than expected at the site level (Ovaskainen *et al.* 2016b). All JSDMs had much larger species richness difference estimates for the conditional marginal predictions but it is not clear why, and is in contrast to previous findings that suggest binary predictions are more prone to overprediction than probabilistic ones (Thuiller *et al.* 2015; Zurell *et al.* 2019).

4.4.5. SESAM Performance

The SESAM predictions outperformed the SSDM and each JSDMs for all evaluation metrics in the threshold-dependent and community dissimilarity metric classes. This suggests that the framework for constraining the species predicted at a site improves predictive performance at both the species level (threshold-dependent metrics) and community level (community dissimilarity metrics). SESAM did not outperform the SSDM for species richness difference, but the difference was only 0.001 species per site. SESAM uses external estimates of species richness at a site to constrain it's predictions: either from a macroecological model or the summed probabilistic predictions of an SSDM. As SESAM uses the predictions of the SSDM to set its species richness limits in our study this result is thus unsurprising. These results are consistent with those of Zurell *et al* (2019) who found that predictions constrained with the SESAM probability rank rule performed better for both species and community level metrics. Zurell *et al* (2019) found a small benefit to species richness metrics for predictions constrained with the SESAM framework but they used a macroecological model to set the species richness limit compared to using the SSDM predictions as we have in this study.

An interesting avenue of research not considered in this study is applying the SESAM framework to JSDM predictions. The high performance of SESAM in our results is in concordance with that of Zurell *et al* (2019) which suggested that the choice of how species predictions are combined into community-level predictions is potentially more important that the choice of underlying model used to generate them. Chapter Three suggested that JSDMs could be predicting likely community assemblages in the absence of limiting factors like site carrying capacities or dispersal limits, and thus they could potentially benefit from the application of a constraining framework. As it requires probabilistic prediction inputs, however, it can only be applied to marginal prediction

types in its current implementation. Research into how to incorporate this constraint into the prediction process itself rather than applying the constraint to predictions post-hoc is suggested.

4.5. CONCLUSION

While there were consistent trends within evaluation metric classes, we did not find evidence to suggest that any one model outperformed all of the others across all prediction types. The likelihood metrics indicated that the JSDMs were better fit to the data, while the other evaluation metric classes showed that the SSDM generally outperformed all of the JSDMs for all other metrics. On the absolute scale the difference in performance between models was generally small. The spatial and non-spatial HLR models were the best performing of the JSDMs and were able to outperform the SSDM for most species richness difference estimates in joint prediction types and generally had the smallest difference in performance from the SSDM when underperforming. The SESAM model consistently outperformed both the JSDMs and the SSDM for both species- and community-level metrics which suggests that the application of the SESAM framework to the JSDM prediction types should be evaluated in the future.

REFERENCE LIST

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Blanchet, F.G., Tikhonov, G. & Norberg, A. (2019). HMSC: Hierarchical Modelling of Species Community.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, **87**, 34–56.

Clark, J.S. & Taylor-Rodríguez, D. (2018). gjam: Generalised Joint Attribute Modelling.

Cornell, H. V. & Harrison, S.P. (2014). What Are Species Pools and When Are They Important? *Annual Review of Ecology, Evolution, and Systematics*, **45**, 45–67.

Dormann, C.F., Calabrese, J.M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C.M., Ciuti, S., Elith, J., Gerstner, K., Guelat, J., Keil, P., Lahoz-Monfort, J.J., Pollock, L.J., Reineking, B., Roberts, D.R., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Wood, S.N., Wüest, R.O. & Hartig, F. (2018). Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, **88**, 485–504.

Dunn, O.J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, **56**, 52–64.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Freeman, E.A. & Moisen, G.G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48–58.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. & Hothorn, T. (2019). mvtnorm: Multivariate Normal and t Distributions.

Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.

Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites & Vectors*, **8**, 367.

Golding, N. & Purse, B. V. (2016). Fast and flexible Bayesian species distribution

modelling using Gaussian processes. *Methods in Ecology and Evolution*, **7**, 598–608.

Götzenberger, L., de Bello, F., Bråthen, K.A., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.

Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., Tingley, R. & Wintle, B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292.

Guisan, A. & Rahbek, C. (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.

Hamner, B. & Frasco, M. (2018). Metrics: Evaluation Metrics for Machine Learning.

Hanberry, B.B. & He, H.S. (2013). Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecol*, **13**, 13–19.

Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model (D. Warton, Ed.). *Methods in Ecology and Evolution*, **6**, 465–473.

Hui, F.K.C. (2018). boral: Bayesian Ordination and Regression AnaLysis.

Hui, F.K.C. (2016). boral - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, **7**, 744–750.

Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J.,

Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. & Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, L., Candan, C. & Hunt, T. (2019). caret: Classification and Regression Training.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54–64.

Legendre, P. & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.

Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.

Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution models: A review. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*, 4241–4247.

Massey Jr, F.J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of*

the American Statistical Association, **46**, 68–78.

Meade, B., Lafayette, L., Sauter, G. & Tosello, D. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility.

Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., Foster, S.D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N.A., Holt, R.D., Hui, F.K.C., Husby, M., Kålås, J.A., Lehikoinen, A., Luoto, M., Mod, H.K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N.E., Gravel, D. & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, **89**, 1–24.

Oksanan, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2018). vegan: Community Ecology Package.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, **7**, 549–555.

Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, **7**, 428–436.

Pollock, L.J. (2014). Data from: Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Dryad Digital Repository*.

Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM) (J. McPherson, Ed.). *Methods in Ecology and Evolution*, **5**, 397–406.

R Core Team. (2018). R: A Language and Environment for Statistical Computing.

Revelle, W. (2018). psych: Procedures for Personality and Psychological Research.

Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, **21**, 88–99.

Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C. & Zipkin, E.F. (2016). Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, **25**, 1144–1158.

Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015). From species distributions to meta-communities. *Ecology Letters*, **18**, 1321–1328.

Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context (D. Warton, Ed.). *Methods in Ecology and Evolution*, **8**, 443–452.

Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, **10**, 225–232.

Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2018). blockCV: Spatial and environmental blocking for k-fold cross-validation.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, **30**, 766–779.

Wilhelm, S. & B G, M. (2015). tmvtnorm: Truncated Multivariate Normal and Student t Distribution.

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution*, **10**, 198–211.

Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K., Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Zdravko, B.I. (2015). TruncatedNormal: Truncated Multivariate Normal.

Zhang, C., Chen, Y., Xu, B., Xue, Y. & Ren, Y. (2018). Comparing the prediction of

joint species distribution models with respect to characteristics of sampling data. *Ecography*, **41**, 1876–1887.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, **00**, 1–13.

CHAPTER FIVE

GENERAL DISCUSSION

5.1. OVERVIEW

In this thesis I explored the six outstanding problems that I identified in the joint species distribution modelling literature that are potentially hampering their wider uptake by the ecological community. My aims were to i) describe several published JSDMs in a singular notation to elucidate their similarities and differences, ii) perform a comparison of the inferential performance of JSDMs, iii) perform a comparison of the computational performance of JSDMs, iv) define the different types of prediction methodologies available to JSDMs, v) define the methods for evaluating the predictions of JSDMs, and vi) perform a comparison of the predictive performance of JSDMs.

In this chapter I synthesise my findings and discuss their implications for resolving the six outstanding problems.

5.2. THE SIX OUTSTANDING PROBLEMS

5.2.1. P1: MODEL DESCRIPTIONS

The use of mathematical notation has increased in the ecological literature over time as statistical developments, and the computational developments that help facilitate them, occur and allow the tackling of more complex problems (Edwards & Auger-Méthé 2019). This provides more precise means of communicating the statistical models than allowed by text-based descriptions, and the increased reproducibility of both methods and results (Lowry 1959). However, to fully realise these advantages for joint species distribution models we need a singular, clearly defined set of notation and terminology in order to increase the accessibility of the literature (Lowry 1959; Elith, Leathwick & Hastie 2008; Elith *et al.* 2011; Edwards & Auger-Méthé 2019).

In Chapter Two I defined a set of notation and terminology in which to present JSDMs. Using this framework I presented seven presence-absence based JSDMs to elucidate their similarities and differences. This allowed me to show that all of these JSDMs share the same core model (Chib & Greenberg 1998) and differ only in how they extend the modelling framework. When using a shared notation we are able to show that the main structural differences between the JSDMs come down to the inclusion/exclusion of two extensions: hierarchical regression coefficients and latent factors. The use of hierarchical regression coefficients allows for sharing information between species to improve the estimates for rare species, and also reduces the overall number of parameters to be estimated during the model fitting process. Latent factors also provide both an ecological and computational benefit. Latent factors attempt to explain some of the variation in species occurrence as the response of species to hypothetical unmeasured environmental variables. This response can then be analysed, particularly in spatially-explicit models, to identify what additional environmental factors are affecting species distributions. The use of latent factors also provides a big computational benefit by reducing the number of parameters to estimate compared to an unstructured correlation matrix.

5.2.2. P2: Inferential Comparison

In order to directly compare the estimated relationships between different JSDMs they must be fit to identical datasets (Barner *et al.* 2018). The different JSDMs presented in the literature have all been published using different datasets as case studies, which does not allow us to compare the inferences they have made. In Chapter Two I performed a large scale comparison in which I fit seven published JSDMs for presence-absence data

to six datasets that cover a variety of taxa, geographic regions, and spatial scales. This allowed the direct comparison of model inferential performance.

I found that all of the models were generating similar inferences for both the species-environment relationships and species co-occurrence patterns. For the species-environment relationships the correlation between model estimates was highly positive, particularly between the models using hierarchical regression coefficients, which suggest all models are able to correctly identify the environmental effects. We also did not find any evidence to suggest that any single model was more or less uncertain in its estimates of the species-environment relationships. The only caveat to this finding is that the LPR model struggled to fit to datasets with a high number of binary predictor variables, particularly if there is an issue with complete separation, and occasionally returned stronger and more uncertain estimates. While this is usually less of an issue for models fit using a Bayesian framework than maximum likelihood estimation (Rainey 2016), the LPR model had convergence issues when fit to these datasets using its default priors. By using slightly more informative priors we were able to alleviate the problem and return similar parameter estimates as the other models. While I deliberately fit all of these JSDMs using their default settings to facilitate a comparison, this result serves as an important reminder to give considerable thought to all settings used during the model fitting process.

The different JSDMs also identified similar species co-occurrence patterns to each other. The correlation of the correlation coefficient estimates of the different models was highly positive but not as strongly as for the species-environment relationships. This suggests that the models are generating similar ecological inferences about the residual associations between species. The LPR and DPR models showed the weakest correlations with other model estimates, but were still moderately positive. This could

be explained in part by the convergence issues of LPR and the Dirichlet process used by DPR. By modelling the residual co-occurrence between species using species archetype groupings when using the latent factor approach, DPR generates a coarse approximation of the correlation matrices generated by other models. This will lead to weaker correlations with other model estimates as it cannot capture as much between species variation in coefficient estimates. However, DPR's weaker correlations with other models overall was not solely driven by the two larger datasets so there is still some unidentified factor having an effect.

While the JSDMs returned correlation coefficient estimates with the same sign, they did differ in the relative strength and uncertainties of their estimates. The spatial and non-spatial HLR models estimated weaker species co-occurrence patterns for the four datasets than the other JSDMs which is unsurprising when considering that the latent factor approach provides some measure of regularisation on the residual correlations. LPR, however, showed relatively stronger estimates despite being a latent factor model. This is most likely indicative of the number of latent factors used to fit the model as LPR uses a default of two while the HLR models estimate an appropriate amount during the model fitting process. A larger amount of latent factors generates a better approximation of the full-rank species correlation matrix, and despite LPR's default number falling in the suggested range in the literature (Warton *et al.* 2015; Hui 2017) it potentially leads to poorer estimates.

5.2.3. P3: Computational Comparison

Each of the JSDMs has been presented in the literature with alternative computational infrastructure implementations. This includes different software infrastructure that a user interacts with directly or is called indirectly during the model fitting process,

different MCMC regimes, different choice of priors, and other default settings. In conjunction with each model being fit to different datasets and thus rendering any measure of computational performance incomparable, a researcher may be left confused about which JSDMs is practical for their application.

In Chapter Two I fit seven JSDM implementations to six datasets on computers of identical specifications in order to assess their computational performance. I found that the difference in computational performance between models was significantly greater than any difference in inferential performance. The MLR model showed the worst performance of all models as it was only able to fit to the two smallest datasets, and did so far slower than all other models. HPR and HLR-S were the next slowest and were unable to fit to the two largest datasets as a result of scalability issues. This scalability issue was also apparent in all four models that successfully fit to all datasets. However, this shows that several JSDM implementations are capable of fitting to large datasets (the bird dataset containing 370 species at 2,752 sites), and for both MPR and HLR-NS they did so in relatively quick lengths of time. The comparison of effective sample size from the MCMC samplers of the different models also showed MPR and HLR-NS to have the most computationally efficient performance.

5.2.4. P4: Defining Prediction

Using JSDMs for prediction is a relatively nascent topic of discussion in the literature. One of the purported benefits of JSDMs over single species models is that by leveraging the additional information found in the estimated residual correlations they are able to perform both species- and community-level predictions. The majority of JSDM literature to date that have performed prediction have only addressed species-level predictions (Harris 2015; Zhang *et al.* 2018; Zurell *et al.* 2019) and only one has begun

to explore community-level predictions (Norberg *et al.* 2019). Recently there was been a call for further exploration of these community-level predictions (Zurell *et al.* 2019; Norberg *et al.* 2019).

In Chapter Three I defined four prediction types available to JSDMs. Marginal predictions are environment-only predictions akin to those of single species models most ecologists are familiar with. Joint predictions are performed at the community level and are able to generate realistic species assemblages at a site. These prediction types can then be conditioned on the known occurrence state of one or more species in the community. Conditional predictions occur at the community-level by generating realistic assemblages where one species is known to be present or absent. Conditional marginal predictions occur at the species level by calculating the probability of presence for a species conditional on the known occurrence state of one species but marginal to that of the remaining species in the community. Using these newly defined prediction types allows for the exploration of different ecological hypotheses related to species distributions and community composition.

5.2.5. P5: Evaluating Prediction

While there has been abundant literature on the evaluation of single-species distribution models (Fielding & Bell 1997; Liu, White & Newell 2009; Lawson *et al.* 2014) there has to date been minimal discussion about the evaluation of JSDM predictions. In addition to the single-species metrics traditionally used in distribution modelling, the multivariate nature of JSDMs opens up additional evaluation metrics that come from community ecology.

In Chapter Three I defined five classes of evaluation metrics that can be applied to the predictions of JSDMs and outline which prediction types that are appropriate for.

Threshold-independent metrics evaluate predicted probabilities against observed presence-absence data. Threshold-dependent metrics evaluate binary predictions against observed presence-absence data, but through the use of a probabilistic confusion matrix can be evaluated against probabilistic predictions (Lawson *et al.* 2014). Community dissimilarity metrics evaluate the dissimilarity between two realisations of a species assemblage (here, our observed and predicted assemblages). Species richness metrics consider a model's ability to predict the number of species present at a site. Likelihood metrics assess how well a model is fit to the data by computing the probability of observing a given community while assuming that the model is 'correct' in its structure and parameter estimates. By defining these different metrics classes and to which prediction types they apply I have outlined a suite of different evaluation approaches that can be used by ecologists. Thus a user is able to examine different aspects of the predictions, such as from a species or community perspective, as best matches their own objectives.

5.2.6. P6: PREDICTION COMPARISON

While there have been several comparisons of the predictive performance of different single-species distribution models there has to date only been one comparison between JSDMs (Norberg *et al.* 2019). In Chapter Four I present a thorough comparison of the predictive performance of six JSDMs and two SSDMs. In this comparison I fit the models to 22 datasets (two real and twenty simulated), use a mix of random and spatial-block five-fold cross validation, predict with the four prediction types available to JSDMs, and evaluate the predictions using 32 evaluation metrics across the five classes I have previously defined.

I did not find any evidence to suggest that a single model outperformed all of the others across all prediction types, but I did identify consistent trends within metric classes. The likelihood-based metrics showed the JSDMs to be better fit to the data than the SSDMs, which suggests that accounting for residual associations between species is more beneficial than not doing so, and is consistent with the results of Norberg *et al* (2019). However, most metrics in the other four metric classes generally showed the standard SSDM to outperform all JSDMs in all prediction types, although generally only by a small amount. The exceptions to this were HLR-S often having better species richness difference estimates than the SSDM, and LPR performing better for some community dissimilarity metrics in some prediction scenarios. Only HPR and DPR consistently showed a high relative difference between their performance of community dissimilarity metrics and the SSDM. This is somewhat consistent with the results of Norberg *et al*(2019) who found that the SSDM outperformed most JSDMs, but that the HLR models performed equally well or better than the SSDM for several scenarios.

I found that the SESAM framework when applied to SSDM predictions consistently outperformed all of the JSDMs and the standard SSDM for both the threshold-dependent and community dissimilarity metrics for the binary prediction scenarios. The only metric where SESAM did not outperform the standard SSDM is species richness difference where they both performed near identically, which is unsurprising when considering the predictions of the SSDM set the richness constraint in the SESAM predictions. This result suggests that using a framework that constrains the number of species predicted to occur at a site improves predictions at both the species- and community-levels, and is consistent with the results of Zurell *et al* (2019).

5.3. FUTURE WORK

The results of my comparison of predictive performance, in addition to the results of Zurell *et al* (2019), suggests that further work is required to investigate why the JSDMs are performing relatively poorly for prediction despite being shown to fit better to the data. It is possible that the JSDMs are predicting potential community assemblages in absence of limiting factors like site carrying capacity or dispersal limitations and would thus benefit from a constraining framework like that of SESAM. In its current form this framework can only be applied to probabilistic prediction outputs to generate binary community assemblages, and it is unclear how to appropriately apply it, or something similar, to the binary outputs of joint prediction types.

There are also several extensions to the JSDM framework that have only recently, or not at all, been explored that provide interesting avenues for future model development. The effect of imperfect detection on model performance has been well documented in the literature (Lahoz-Monfort, Guillera-Arroita & Wintle 2014; Guillera-Arroita 2017; Guillera-Arroita, Kéry & Lahoz-Monfort 2019), and while it has been applied multiple times to multi-species datasets from an occupancy modelling perspective there is only one study addressing its impact on JSDMs (Tobler *et al.* 2019). This study, however, found that to achieve satisfactory convergence during the model fitting process they had to use such a high number of latent factors that they lost most of the computational savings that approach usually provides.

In addition to accounting for imperfect detection it would be interesting to assess correlated detections in JSDMs. Two species may occur at the same sites due to shared environmental responses and thus appear to be positively correlated, whilst in reality they actively avoid each other at the fine scale due to factors like competition of

nocturnal/diurnal habits. Thus species may be using the same sites but never at the same time. This would be an excellent advantage for camera trap studies where they are able to identify shared usage of sites between species that directly compete with one another and thus don't use sites simultaneously, such as competing predator species like cats and foxes.

In light of the computational comparison results I presented in Chapter Two and the performance issues of Tobler *et al* (2019) it will be beneficial for JSDMs to be re-implemented using new, more efficient software. This will be a key development for both the application of JSDMs to larger datasets as well as for building model extensions that will significantly increase the number of parameters that need to be estimated (such as the additional correlation matrix in a correlated detection model). A prospective candidate for this would be the *greta* R package (Golding 2018) which has been shown to be faster and scale better than existing MCMC languages like BUGS and JAGS.

5.4. CONCLUSION

In this thesis I addressed the six outstanding issue I identified in the JSDM literature. By defining a singular set of notation and terminology in which to present JSDMs I was able to show that all of the presence-absence based models being considered shared the same core model. I was also able to elucidate the differences and similarities between models and show that the models could be broadly distinguished by their inclusion/exclusion of two structural components: hierarchical regression coefficients and latent factors. By taking the models published on separate datasets in the literature and performing a direct comparison of the performance of seven JSDM implementations across six datasets I was able to show that the models all performed

similarly for inferential performance. The different JSDMs identified the same species-environment relationships, and for the residual associations between species they identified correlations in the same direction but differed in their estimated strengths of those relationships. With this comparison I was also able to compare computational performance and found that there is a bigger difference between models here than inferential performance. There is an issue of scalability in all models, but some models are able to handle it much better than others and thus fit more easily to larger datasets. I defined four prediction types that are possible for JSDMs, three of which have been implemented once or not at all in the literature, and five classes of evaluation metrics that can be applied to them. These prediction types and metrics classes allow ecologists explore different hypothesis about species distributions and the community assemblage process. By performing a large scale comparison of predictive performance I showed that JSDMs are better fit to the data than models that do not account for the residual associations between species but perform slightly worse predictions. With the work presented in this thesis addressing these issues I hope to facilitate the wider uptake of joint species distribution modelling in the ecological community.

REFERENCE LIST

Barner, A.K., Coblentz, K.E., Hacker, S.D. & Menge, B.A. (2018). Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, **99**, 557–566.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Edwards, A.M. & Auger-Méthé, M. (2019). Some guidance on using mathematical notation in ecology (S. McMahon, Ed.). *Methods in Ecology and Evolution*, **10**, 92–99.

Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.

Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.

Golding, N. (2018). greta: Simple and Scalable Statistical Modelling in R.

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.

Guillera-Arroita, G., Kéry, M. & Lahoz-Monfort, J.J. (2019). Inferring species richness

using multispecies occupancy modeling: Estimation performance and interpretation. *Ecology and Evolution*, **9**, 780–792.

Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model (D. Warton, Ed.). *Methods in Ecology and Evolution*, **6**, 465–473.

Hui, F.K.C. (2017). boral: Bayesian Ordination and Regression AnaLysis.

Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504–515.

Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54–64.

Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution models: A review. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*, 4241–4247.

Lowry, W.P. (1959). On the Use of Mathematical Notation in Ecological Literature. *Ecology*, **40**, 492–492.

Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., Foster, S.D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N.A., Holt, R.D., Hui, F.K.C., Husby, M., Kålås, J.A., Lehikoinen, A., Luoto, M., Mod, H.K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N.E., Gravel, D. & Ovaskainen, O. (2019). A comprehensive

evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, **89**, 1–24.

Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis*, **24**, 339–355.

Tobler, M.W., Kéry, M., Hui, F.K.C., Guillera-Arroita, G., Knaus, P. & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, **100**, e02754.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, **30**, 766–779.

Zhang, C., Chen, Y., Xu, B., Xue, Y. & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, **41**, 1876–1887.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, **00**, 1–13.

APPENDIX A

PUBLISHED VERSION OF CHAPTER ONE

Received: 13 August 2018 | Accepted: 10 September 2018

DOI: 10.1111/2041-210X.13106

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

A comparison of joint species distribution models for presence–absence data

David P. Wilkinson | Nick Golding | Gurutzeta Guillera-Arroita | Reid Tingley | Michael A. McCarthy

School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

Correspondence
David P. Wilkinson
Email: davidpw@student.unimelb.edu.au

Funding information
Australian Government Research Training Program; University of Melbourne; Australian Research Council, Grant/Award Number: DE160100904 and DE170100601

Handling Editor: Pedro Peres-Neto

Abstract

1. Joint species distribution models (JSDMs) account for biotic interactions and missing environmental predictors in correlative species distribution models. Several different JSDMs have been proposed in the literature, but the use of different or conflicting nomenclature and statistical notation potentially obscures similarities and differences among them. Furthermore, new JSDM implementations have been illustrated with different case studies, preventing direct comparisons of computational and statistical performance.
2. We aim to resolve these outstanding issues by (a) highlighting similarities among seven presence–absence JSDMs using a clearly defined, singular notation; and (b) evaluating the computational and statistical performance of each JSDM using six datasets that vary widely in numbers of sites, species, and environmental covariates considered.
3. Our singular notation shows that many of the JSDMs are very similar, and in turn parameter estimates of different JSDMs are moderate to strongly, positively correlated. In contrast, the different JSDMs clearly differ in computational efficiency and memory limitations.
4. Our framework will allow ecologists to make educated decisions about the JSDM that best suits their objective, and enable wider uptake of JSDM methods among the ecological community.

KEYWORDS
biotic interactions, community assembly, hierarchical models, joint species distribution model, latent factors, presence–absence, residual correlation

1 | INTRODUCTION

Understanding the factors underlying species' geographic distributions is a fundamental goal of ecology. Abiotic factors, such as temperature or rainfall, and biotic factors, such as species interactions or population dynamics, ultimately drive species distributions (Hutchinson, 1957). Yet correlative species distribution models (SDMs), the most common tool for predicting species distributions, typically quantify species' environmental relationships without explicitly considering effects of species interactions (Dormann et al., 2012; Wisz et al., 2013). In contrast, community ecology generally studies co-occurrence patterns and community structure without accounting for co-occurrence due to shared environmental responses among species (Hardy, 2008).

Several statistical approaches have been proposed to quantify species' environmental relationships and interactions in a single model. Within the occupancy-detection modelling framework (Guillera-Arroita, 2017), early examples explicitly modelled effects of biotic interactions via linear models, including pair-wise and higher order effects (MacKenzie, Bailey, & Nichols, 2004), or

© 2018 The Authors. Methods in Ecology and Evolution © 2018 British Ecological Society

dominate/subordinate species relationships, where the dominant species' occupancy state influences that of the subordinate species, but not vice versa (Richmond, Hines, & Beissinger, 2010; Waddle et al., 2010). Practical and computational limits restrict the application of these methods to communities of only a handful of species (Richmond et al., 2010).

Other methods incorporate species co-occurrence data into the classical SDM framework. Distribution estimates of additional species can be used alongside abiotic variables as predictors (Araújo & Luoto, 2007; Leathwick & Austin, 2001; Meier et al., 2010; Pellissier et al., 2010), or a species' predicted distribution can be restricted to that of another it depends on (Schweiger et al., 2012). Like dominate/subordinate models, however, these approaches only incorporate unidirectional interactions (Kissling et al., 2012; Pollock et al., 2014).

Joint species distribution models (JSDMs) have emerged as extensions of SDMs to capture the effects of biotic interactions in communities (Ovaskainen, Hottola, & Siitonen, 2010; Kissling et al., 2012; Clark, Gelfand, Woodall, & Zhu, 2014; Clark, Nemergut, Seyednasrollah, Turner, & Zhang, 2017; Pollock et al., 2014; Warton et al., 2015; Golding, Nunn, & Purse, 2015; Letten, Keith, Tozer, & Hui, 2015; Harris, 2015; Thorson et al., 2016; Ovaskainen, Abrego, Halme, and Dunson, 2016; Ovaskainen, Roy, Fox, & Anderson, 2016; Hui, 2016; Nieto-Lugilde, Maguire, Blois, Williams, & Fitzpatrick, 2017). JSDMs simultaneously model multiple species' distributions, accounting for both environmental relationships and residual associations (that might arise from species interactions) on species co-occurrence (Pollock et al., 2014). Thus, JSDMs inform about biotic and abiotic constraints on species' distributions (Pollock et al., 2014), and can improve predictions of community composition (Harris, 2015). Promising advances in JSDMs have led to the development of several statistical models and computational implementations. This diversity of methods and their presentation might hamper wider use of JSDMs by ecologists, since prospective users have little guidance on which implementation best suits their objectives.

Our study formally compares seven presence–absence JSDMs that consider species interactions by characterising the residual correlation among species not captured by environmental predictors. These models are statistically complex, and defined in their respective papers using different, potentially conflicting, notation and terminology. In addition, each JSDM was evaluated on a different dataset; any ecological relationships inferred are therefore not directly comparable, nor are reports of an implementation's computational efficiency. To help ecologists better understand the different models, we define them using a consistent notation to elucidate their similarities and differences, and compare their parameter estimates and computational performance against six datasets. The results can help ecologists make informed decisions about appropriate JSDM selection.

2 | MATERIALS AND METHODS

2.1 | Approach

Each JSDM considered here consists of three components: a Bayesian statistical model; software to fit the model to data; and default or suggested settings for model fitting (e.g., choice of priors and Markov chain Monte Carlo (MCMC) sampling regime). A JSDM's computational efficiency and parameter estimates will undoubtedly be affected by all of these components. Our aim was not to disentangle the effects of each of these components, but to highlight the differences and similarities between the overall methodologies, as applied in their respective papers. The term JSDM hereafter refers to these overall methodologies, unless otherwise stated.

2.2 | Datasets

Each JSDM was fit to six presence–absence datasets from recent JSDM papers (Golding et al., 2015; Harris, 2015; Ovaskainen et al., 2010; Ovaskainen, Roy et al., 2016; Pollock et al., 2014) to compare parameter estimates and computational efficiency across a range of datasets. Datasets not made publicly available were sourced from the authors. These datasets covered a broad range of taxa, geographic locations, and numbers of sites and species (Table 1). Datasets were used as originally published with these exceptions: testing and training splits were merged; the Breeding Bird Survey online repository had been updated resulting in more sites; and we only used one of the 18 fungi datasets from Ovaskainen et al. (2010). Geographic coordinates were rescaled to have a maximum distance between sites of 1 to avoid issues of numerical instability in the spatial model.

TABLE 1 Dataset summary

Dataset	Source	Geographic location	Species	Sites	Covariates
Birds	Harris (2015)	North America	370	2,752	8
Butterflies	Ovaskainen, Roy et al. (2016)	Great Britain	55	2,609	4
Eucalypts	Pollock et al. (2014)	Grampians National Park, Australia	12	458	7
Frogs	Pollock et al. (2014)	Melbourne, Australia	9	104	3
Fungi	Ovaskainen et al. (2010)	Southern Finland	11	800	15
Mosquitoes	Golding et al. (2015)	South-East England	16	167	13

2.3 | The models

We restrict our evaluation of JSDMs to seven models for species presence–absence data that account for species interactions via residual correlations. These JSDMs are extensions of the generalised linear modelling (GLM) framework, which is widely used for modelling species distribution data (Gelfand et al., 2006). The statistical models are defined below using a common notation. The following terms are consistent across all models: $\boldsymbol{y}$, the response variable; 1(.), an indicator function that returns 1 when the expression in brackets is true and 0 otherwise; $\boldsymbol{z}$, a normally distributed latent variable; μ, the linear predictor for the measured covariates; $\boldsymbol{X}$, the matrix of measured covariates; $\boldsymbol{\beta}$, the matrix of regression coefficients; and $\mathbf{I}$, the identity matrix. Subscript notation for sites is $i = 1, \ldots, n$; for species $j = 1, \ldots, J$; and for predictors $k = 1, \ldots, K$.

Each JSDM is built on the foundation of Chib and Greenberg's (1998) multivariate probit regression model (hereafter, the core model). This model uses a latent variable parameterisation of a probit model rather than the probit link directly. The probability of species presence is modelled as the probability of a latent multivariate normally distributed variable exceeding a threshold, such that $y_{ij} = 1$ if $z_{ij} > 0$, and $y_{ij} = 0$ otherwise. The probability of species presence at a site is represented by a one-dimensional latent variable, while the community present at a site is represented by a multidimensional latent variable (see Pollock et al., 2014 for a visual representation). This core model is described as follows:

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

The probability of species j being present at site i is the probability that latent variable z_{ij} is greater than zero. The latent variable is the sum of the linear predictor μ_{ij} and the correlated residual error e_{ij}. The linear predictor is the product of the measured environmental variables $\boldsymbol{X}_{i,.}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{.,j}$, as in generalised linear models. Correlations in the residual error $\boldsymbol{e}_i$ are captured in $\mathbf{R}$, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements are restricted between −1 and 1. Standard deviations are constrained to equal to 1 in probit regression, thus covariance and correlation matrices are equivalent. The elements of $\mathbf{R}$ reflect species co-occurrence patterns not described by the environmental predictors (i.e., species interactions, or missing predictors). A limitation of this model is that it does not specify overdispersion in species occurrences like other extensions of the GLM framework (e.g., generalised linear mixed models), but it could be extended to do so.

The following sections define statistical models and describe the software and MCMC regimes of the JSDMs. All models were fit using their default setting for priors and MCMC regimes. Table 2

TABLE 2 Modelling method components. X denotes permanent features, o denotes features dependent on dataset size. Models without an accompanying R package have code scripts available in the supplementary material of their source paper

	Source paper	R package	Hierarchical regression coefficients	Dirichlet process	Latent factors	Multivariate	Probit	Logistic	Regression	Spatially explicit	Other compatible data types
MPR	Golding et al. (2015)	BayesComm				X	X		X		
MLR	Ovaskainen et al. (2010)	NA				X		X	X		
HPR	Pollock et al. (2014)	NA	X			X	X		X		
LPR	Hui (2016)	boral			X	X	X		X		Abundance
DPR	Clark et al. (2017)	gjam		o	o	X	X		X		Multiple including continuous and discrete abundance, and ordinal counts
HLR-S	Ovaskainen, Roy et al. (2016)	NA	X		X	X	X		X	X	
HLR-NS	Ovaskainen, Roy et al. (2016)	NA	X		X	X	X		X		

provides an overview of the statistical models'respective features, and Table 3 provides a glossary of all introduced notation.

2.3.1 | Multivariate probit regression

The multivariate probit regression (MPR) model used by Golding et al. (2015), BayesComm, is identical to the core model. The regression coefficients have a normal prior, $\beta \sim N(0, 10)$, and the correlation coefficients an inverse Wishart prior with $n + 2J$ *df* and scale matrix **I**. MPR was fit in R (R Core Team 2018) by MCMC using a Gibbs sampler implemented in R and C++. A single MCMC chain of 11,000 samples, discarding the first 1,000 as burn-in, sampled the posterior distribution.

TABLE 3 Symbology: matrix dimensions supplied in brackets

Symbol	Definition
Subscripts	
i	Site. $i = 1, \ldots, n$
j	Species. $j = 1, \ldots, J$
k	Measured covariate. $k = 1, \ldots, K$
h	Latent factor/Unmeasured covariate. $h = 1, .., H$
l	Species archetype. $l = 1, \ldots, L$
Main terms	
y	Binary response variable (species presence/absence)
$1(.)$	Indicator function
z	Normally distributed latent variable
μ	Linear predictor for the measured covariates
ν	Linear predictor for the unmeasured covariates
$\boldsymbol{X}$	Matrix of measured covariates $(n \cdot K)$
$\boldsymbol{\beta}$	Matrix of regression coefficients $(K \cdot J)$
$\boldsymbol{\eta}$	Matrix of latent factors $(n \cdot H)$
λ	Matrix of factor loadings $(H \cdot J)$
$\boldsymbol{I}$	Identity matrix $(J \cdot J)$
$\boldsymbol{A}$	Archetype-reduced factor loadings matrix $(H \cdot J)$
$\boldsymbol{\Sigma}$	Covariance matrix $(J \cdot J)$
$\mathbf{R}$	Symmetric, positive-definite correlation matrix $(J \cdot J)$
e	Correlated residual error
ε	Uncorrelated residual error
Φ	Cumulative density function of $N(0,1)$
ω	Mean of the normal distribution for hierarchical β coefficients
σ	Standard deviation of the normal distribution for hierarchical β coefficients
d	Spatial distance
α	Positive parameter controlling decay in correlation with distance for latent factors

2.3.2 | Multivariate logistic regression

Ovaskainen et al. (2010) introduced a logistic regression version of the core model based on O'Brien and Dunson (2004). The model (hereafter MLR) is as follows:

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mu_{ij} + logit(\Phi[e_{ij}])$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

Here, Φ is the cumulative density function of the standard normal distribution. The regression coefficients have a normal prior, $\beta \sim N(0, 10)$, and the correlation coefficients a uniform prior, $U(-1, 1)$. MLR was fit in Mathematica (Wolfram Research Inc, 2016), and the posterior distributions sampled via MCMC using a Metropolis-within-Gibbs sampler with 100,000 samples (thinned to keep 1 every 10) discarding the first 5,000 as burn-in.

2.3.3 | Hierarchical multivariate probit regression

The hierarchical multivariate probit regression model of Pollock et al. (2014) (hereby HPR) models the regression coefficients hierarchically such that β_{jk} is drawn from a normal distribution with mean ω_k and standard deviation σ_k.

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\beta_{jk} \sim N(\omega_k, \sigma_k)$$

$$e_i \sim \text{MVN}(\mathbf{0}, \mathbf{R})$$

Regression coefficients have a normal prior, $N(0, 100)$, on ω and a uniform prior, $U(0, 100)$, on σ. Correlation coefficients have an inverse Wishart prior with $J + 1$ *df* and an **I** scale matrix. HPR was fit in R, and the posterior distributions sampled via MCMC using Gibbs sampling in JAGS (Plummer, 2003), with three chains of 1,000,000 samples (thinned to keep 1 every 1,000) with the first 15,000 discarded as burn-in.

2.3.4 | Multivariate probit regression with latent factors

The boral JSDM (Hui, 2016) is a multivariate probit regression model with latent factors (hereafter LPR).

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.j}$$

$$\varepsilon_i \sim N\left(\mathbf{0}, \mathbf{I}\right)$$

This differs from the core model by using latent factors ("hypothetical" unmeasured variables) to explain any residual covariation that is not accounted for by the measured variables. This explained variation is the product of H unmeasured latent factors, $\eta_{i,.}$ ($h = 1, \ldots, H$), and the factor loadings, $\lambda_{.j}$, of species j to latent factor h. We can interpret $\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.j}$ and $\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.j}$ as linear predictors for measured and unmeasured variables, respectively. Now z_{ij} is the sum of the linear predictor of measured covariates, μ_{ij}, the linear predictor of unmeasured covariates, ν_{ij}, and uncorrelated residual error ε_{ij}. This model can be defined equivalently as an extension of MPR where the correlation matrix has low-rank structure, modelled as: $\mathbf{R} = \lambda\lambda' + \mathbf{I}$ (Warton et al., 2015). Species interactions can be estimated by converting the factor loadings to correlation coefficients using this expression.

Latent factors bring considerable computational benefits by reducing the number of coefficients to estimate. A full-rank correlation matrix has $J\left(J-1\right)/2$ parameters, while a factor loadings matrix only JH. The parameter ratio between latent factor and correlation matrix models is therefore 2H/(J − 1), so the parameter reduction increases with J. Where $J = 10$ and $H = 2$, the latent factor model has 20 parameters rather than 45 (44%), whilst for $J = 100$ and $H = 5$ the latent factor model has 500 parameters rather than 4,950 (10%). This reduction does not necessarily correspond to an increase in parsimony, as the choice of priors for these matrices also impact model complexity. As latent factors are trying to approximate a fully unstructured correlation matrix, a greater H provides a better approximation, while a smaller H reduces computational requirements. The number of latent factors balances model simplicity and model fit, and 2-8 latent factors has been suggested as suitable in practice (Hui, 2017; Warton et al., 2015).

The regression coefficients in the LPR implementation have a normal prior, $\beta \sim N\left(0, 10\right)$,and the latent factors a normal prior, $\eta \sim N\left(0, 4.47\right)$,constrained such that upper diagonal elements are 0, and the diagonal elements are positive, $U\left(0, 20\right)$. LPR was fit in R using the BORAL package. The posterior distribution was sampled via MCMC, with a Gibbs sampler using JAGS, in a single chain of 60,000 samples (thinned to keep 1 every 50) discarding the first 10,000 as burn-in.

2.3.5 | Dimension reduction model

The multivariate generalised regression model of Clark et al. (2017), GJAM (hereby DPR), fits various types of response data. For presence–absence data it is a multivariate probit regression model that takes on two different forms depending on the size of the dataset. The small dataset form is equivalent to the core model. For datasets above a size threshold ($J > 100$ or $J > \frac{2}{3} * n - 1$), it is similar to LPR, but with an additional dimension reduction step (with respect to species) using the Dirichlet process (Taylor-Rodríguez, Kaufeld, Schliep, Clark, & Gelfand, 2016):

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.j}$$

$$\boldsymbol{\lambda}_{.j} = \mathbf{A}_{l_j}$$

$$\varepsilon_i \sim N\left(\mathbf{0}, \mathbf{I}\right)$$

The Dirichlet process assumes that groups of species will have the same response to the unmeasured variables, so a priori reduces λ along the j-dimension to a matrix, $\mathbf{A}$, for a fixed number, L, of species archetypes (rather than J species). The Dirichlet process can be seen as a distribution of distributions, and is used to cluster species into archetypes, via an index variable $\boldsymbol{l}$. Using $\mathbf{A}$ as a look-up table for the low-rank λ allows conversion back to full-rank λ, and then estimate correlation coefficients via the factor loadings. The parameter ratio of DPR to a full-rank correlation matrix is $2LH/\left(J\left(J-1\right)\right)$, which is typically much smaller than 1. This decouples the number of parameters to estimate from J, though adequate L and H must be used to accurately model interactions.

The regression coefficients in the DPR implementation have an improper uniform prior, $U\left(-Inf, Inf\right)$, correlation coefficients use an inverse-Wishart prior with $n - K + J - 1$ df and an $\mathbf{I}$ scale matrix, and latent factors are normally distributed as $\eta_{ih} \sim N\left(0, 1\right)$. DPR was fit in R using the GJAM package. The posterior distribution was sampled via MCMC using a Gibbs sampler in R of a single chain of 60,000 samples (thinned to keep 1 every 50), discarding the first 10,000 as burn-in.

2.3.6 | Hierarchical multivariate probit regression with latent factors

The last two JSDMs are spatially explicit (HLR-S) and non-spatially explicit (HLR-NS) versions of the hierarchical multivariate probit regression with latent factors used by Ovaskainen, Roy et al. (2016):

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.j}$$

$$\beta_j \sim MVN\left(\boldsymbol{\omega}_{l_j}, \boldsymbol{\sigma}\right)$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.j}$$

$$\varepsilon_i \sim N\left(0, 1\right)$$

The HLR models make use of latent factors like LPR and DPR, but also use hierarchical regression coefficients like HPR. However, instead of independent normal distributions unique to the K covariates, $\beta_{.j}$ is drawn from a multivariate normal distribution, allowing for correlation among coefficients. The mean of this distribution, $\boldsymbol{\omega}_{l_j}$, is unique to species' archetype, $\boldsymbol{l}$. Archetype is specified a priori in HLR models using trait data, assuming that species with similar traits exhibit

similar responses. Here, l_j is set to 1 as trait data were unavailable for most datasets. Thus, each species' $\beta_{.j}$ are independently drawn from a shared distribution. The hierarchical regression coefficients have a normal prior, $N(0,1)$ for $\boldsymbol{\omega}$, and an inverse Wishart prior with five *df* and a **I** scale matrix for $\boldsymbol{\sigma}$. The latent factors are normally distributed as $\eta_{ih} \sim N(0,1)$ in HLR-NS, and as independent spatially homogenous Gaussian processes with exponential covariance in HLR-S:

$$\mathrm{Cov}\left(\eta_{ih}, \eta_{i'h}\right) = \exp\left(-d_{ii'}/\alpha_h\right)$$

Here, d is the spatial distance between sampling units, and α is a positive parameter controlling decay in correlation with distance. At zero distance the covariance function is 1, so the latent factors have zero mean and unit variance like in the non-spatial model, and this declines towards zero with increasing α.

These JSDMs were fit in MATLAB (The MathWorks Inc., 2016). The posterior distribution was sampled via MCMC using a Gibbs sampler in a single chain of 50,000 samples (thinned to keep 1 every 50) with the first 10,000 discarded as burn in.

2.4 | Model fitting

Model fitting was carried out on desktop computers with identical specifications (Windows 7 OS, i7-6700 CPU @ 3.40 GHz, 16GB RAM) to compare computational performance. JSDM software that explicitly enabled model fitting in parallel was run so (HPR), while those that did not were run sequentially. An upper time limit for runtime was set at 7 days, after which the JSDM was deemed incompatible with that dataset on a desktop computer. JSDMs that reached this limit were then run on high performance computing infrastructure with increased memory availability (Boab: Dual Intel Xeon E5-2699 CPUs @ 2.30 GHz, ~40GB RAM per model. Spartan: Intel(R) Xeon(R) CPU E5-2643 v3 @ 3.40 GHz, 200GB RAM (Lafayette, Sauter, Vu, & Meade 2016)) with the same time limit. JSDMs were fit using: Mathematica 10.4 (Wolfram Research Inc., 2016) (MLR), MATLAB R2016a (The MathWorks Inc., 2016) (HLR models), and R 3.2.5 (R Core Team 2018) (all others). Model-specific R packages were BayesComm 0.1-2 (Golding & Harris, 2015) (MPR), boral 1.3.1 (Hui, 2017) (LPR), and gjam 2.1.1 (Clark et al., 2017) (DPR). JAGS 4.3.0 (Plummer, 2003) was used for some MCMC sampling.

2.5 | Statistical and computational comparison

We compared the computational performance and parameter estimation of the different JSDMs. Computational metrics included: ease of use (time required for dataset formatting); dataset compatibility (model completion within 7 days on a desktop computer); runtime; and sampler efficiency (effective sample size (ESS) per unit CPU-time). To prevent bias in the *ease of use* metric due to user experience gained during the process, the JSDM/dataset combinations were fit by the same author (DW) in a random order, with two exceptions: (a) HLR-S and HLR-NS differed only by a single setting, so ease of use was measured for HLR-S only, and (b) LPR and DPR were not available when the study began and were included later.

ESS estimates are not directly comparable unless each JSDM was fit using identical MCMC regimes, so each JSDM was fit to the mosquito dataset a second time with identical MCMC regimes, consisting of a single chain of 11,000 iterations with no thinning and the first 1,000 discarded as burn-in. ESS was calculated separately for regression and correlation coefficients by dividing the mean effective number of samples of each coefficient in the posterior distribution by the runtime.

Statistical inference was assessed by comparing the estimated regression and correlation coefficients. For latent factor JSDMs, the residual covariance matrix was recovered from the factor loadings $(\boldsymbol{\Sigma} = \lambda\lambda' + \mathbf{I})$ and scaled to the correlation matrix **R**. To evaluate the similarity of JSDM regression coefficient estimates, we compared posterior mean values and their 95% credible intervals, and calculated the Pearson correlation coefficient between mean estimates of each coefficient of each JSDM. Variation in coefficient estimates was evaluated using four uncertainty metrics calculated from post hoc standardised posterior distributions: coefficient of variation, Gini coefficient, quartile coefficient of dispersion, and 90% quantile coefficient of dispersion. Similarity of correlation coefficient estimates was evaluated by calculating the Pearson correlation coefficient between mean estimates for each JSDM, and the magnitude and uncertainty (95% credible interval width) of estimates were calculated relative to the simplest JSDM (MPR) as the baseline.

2.6 | Simulation comparison

We also compared a subset of these models (MPR, HPR, LPR, DPR, and HLR-NS) when fit to datasets simulated under different modelling component assumptions. See Supporting Information Appendix S5.

3 | RESULTS

3.1 | Computational comparison

Only four JSDMs ran to completion within the time limit for all datasets: MPR, LPR, DPR, and HLR-NS (Table 4). MLR did not complete for four datasets, took 142 hours for the eucalypt dataset (nearly three times longer than HLR-S, and eight times longer than the closest non-spatial model, HPR), and 14 hours for the small frog data set; it was therefore deemed too slow for practical use and not considered further. Both HPR and HLR-S failed to complete within the time limit for the bird and butterfly datasets on desktop or high-performance computing resources. HLR-S was incompatible with the non-spatial fungi dataset.

MLR, HLR-S, and HLR-NS were the slowest for converting the datasets into the required format (40 ± 29 min). These were the only JSDMs implemented in programming languages other than R, the language with which the experimenter was most familiar. Datasets for all other JSDMs were prepared within an average of 5 min.

TABLE 4 Model compatibility with datasets. Pairs marked with X were compatible

	MPR	MLR	HPR	LPR	DPR	HLR-S	HLR-NS
Birds	X			X	X		X
Butterflies	X			X	X		X
Eucalypts	X	X	X	X	X	X	X
Frogs	X	X	X	X	X	X	X
Fungi	X		X	X	X		X
Mosquitoes	X		X	X	X	X	X

MPR was the fastest to run to completion for all datasets, followed by LPR for the two smallest datasets, and HLR-NS was fastest for the four largest (Table 5). HPR and HLR-S were considerably slower for all datasets. MPR and HLR-NS had substantially greater sampler efficiency than the other JSDMs; MPR was most efficient for the regression coefficients (Figure 1a), whereas HLR-NS was most efficient for correlation coefficients (Figure 1b).

3.2 | Similarity of parameter estimates

All JSDMs yielded similar regression coefficients across all datasets (see Figure 2 for one species; all species in Supporting Information Appendix S1). LPR parameter estimates for binary variables in the fungi and mosquito datasets were markedly more uncertain and frequently stronger than those estimated by other JSDMs for species-environment pairs with complete separation (see Supporting Information Appendix S1, Figure 458). Results presented here are based on default priors, but the impact of complete separation is alleviated with more informative priors (Supporting Information Appendix S2).

Figure 3 shows the correlation of mean regression coefficient estimates between JSDMs, averaged across all datasets (individual datasets in Supporting Information Appendix S4). There is strong correlation (>0.69) between all JSDM estimates, with the strongest correlations between HPR, HLR-S, and HLR-NS (>0.9), and between MPR and all models (>0.75).

There was no consistent, substantial difference in the variation of regression coefficient estimates between JSDMs (see Figure 4 for the Gini coefficient for all species in the Eucalypt dataset). In a small number of instances, there were differences between JSDMs for particular coefficients (Supporting Information Appendix S3, Figure 21), but there was no apparent pattern to these occurrences, and Supporting Information Appendix S3 Figures 25–28 show only minor differences between JSDMs for each uncertainty metric across all datasets. There was no difference in observed patterns between different uncertainty metrics.

All JSDMs identified similar patterns in estimated correlation coefficients (Figure 5). The direction of species interactions was largely the same among JSDMs, but estimates differed in strength and uncertainty. Observed differences in the direction of estimated correlations between JSDMs involved weak and uncertain coefficients.

Figure 6 shows the correlation of mean correlation estimates between JSDMs for all datasets (individual datasets in Supporting Information Appendix S4). Strong, positive correlations (>0.7) were observed between most JSDMs. However, there was only moderate correlation (>0.49) between DPR and all JSDMs except HPR (0.77), and between LPR and the HLR-S (0.57) and HLR-NS (0.63).

The strength of the JSDM correlation estimates is shown in Figure 7a,b. We split this comparison between the two largest (birds and butterflies) and four smallest datasets, as sample size imbalances between datasets masked observable trends. Averaged across the four smallest datasets, DPR, HLR-S, and HLR-NS estimated the weakest correlation strengths relative to MPR, yet for the two largest datasets HLR-NS estimates were slightly stronger than MPR on average. HPR and LPR consistently estimated stronger correlations than MPR.

For the smallest four datasets (Figure 8a) DPR, HLR-S, and HLR-NS had more certain estimates of correlation coefficients than MPR, HPR was relatively more uncertain, while MPR and LPR were generally equal. However, for the largest two datasets (Figure 8b), HLR-NS was relatively more uncertain than MPR; LPR, and MPR were similarly uncertain, and DPR tended to be more certain than MPR.

For results of the simulation study see Supporting Information Appendix S5.

4 | DISCUSSION

Only four JSDMs could be successfully fit to all six datasets, illustrating that different JSDMs scale differently with dataset size. That MLR could only be successfully fit to the two smallest datasets, and was significantly slower than the other JSDMs, highlights the computational benefits of the probit link over the logit link. Since probit models allow a latent variable parameterisation instead of using the link function directly, they can be fitted using a highly efficient Gibbs sampler (Chib & Greenberg, 1998; O'Brien & Dunson, 2004). Of the probit models, HPR and HLR-S scaled the worst, and LPR exhibited signs of slowing down for the two largest datasets. The effect of scaling will only become more apparent as datasets become larger.

Estimates of runtime and sampler efficiency show that MPR and HLR-NS are markedly faster than the other JSDMs. HPR took longer to run than DPR, yet this difference is due to different MCMC regimes, since the HPR sampler was more efficient. HPR also exhibited the most consistent ESS across all coefficients. The MCMC sampling of HPR is implemented in JAGS, which prioritises ease of use for those with limited programming experience, rather than more efficient, customised samplers utilised in other JSDMs. HLR-S had the

	MPR	MLR	HPR	LPR	DPR	HLR-S	HLR-NS
Birds	3.8	>168	NA	120.4	27.3	>168	15.2
Butterflies	0.23	>168	>168	13.9	6.5	>168	2.1
Eucalypts	<0.02	142.1	7.6	0.33	0.25	50.0	0.21
Frogs	<0.02	14.1	0.94	0.04	0.06	1.4	0.13
Fungi	<0.02	>168	15.8	0.62	0.67	NA	0.26
Mosquitoes	<0.02	>168	6.4	0.14	0.73	2.0	0.2

TABLE 5 Model runtimes (in hours). >168 corresponds to the cut-off time of 7 days

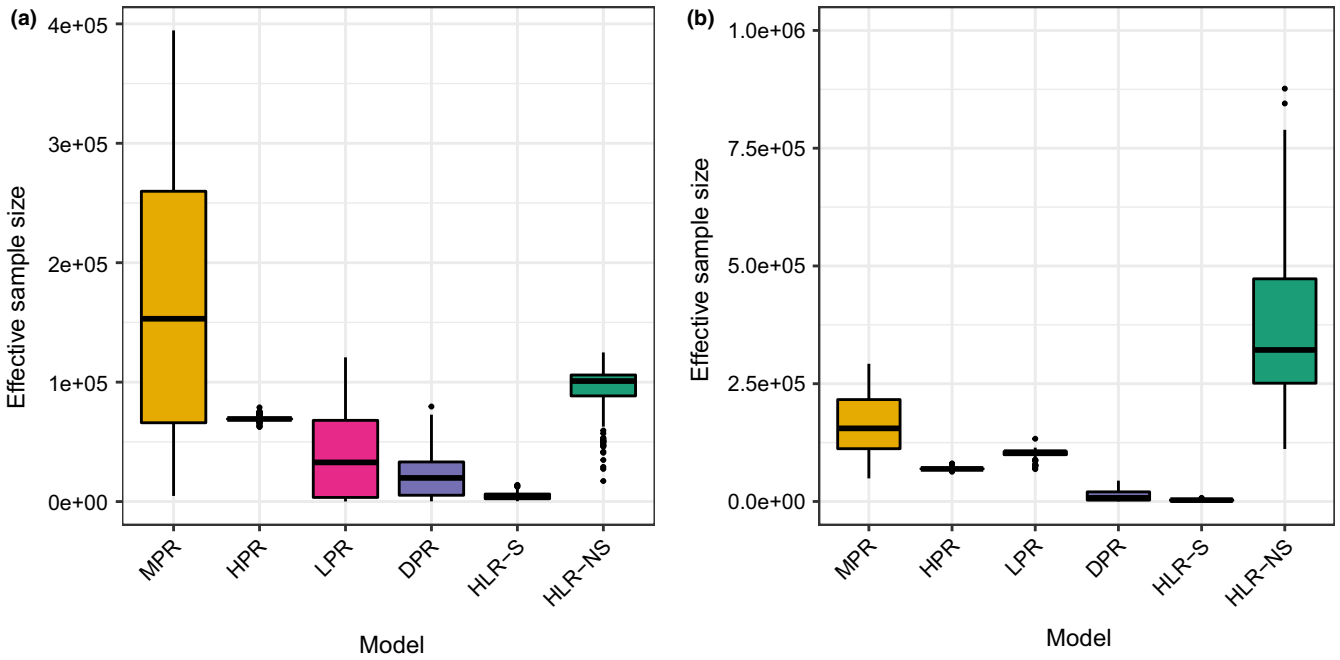


FIGURE 1 Effective sample size of regression (a) correlation (b) estimates for each JSDM's (joint species distribution models) respective MCMC (Markov chain Monte Carlo) samplers on the mosquito dataset. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles, respectively, the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range, respectively, and the black dots represent outliers

least efficient sampler as it was the only one that included spatial Gaussian process components. This requires inversion of an $n \times n$ covariance matrix per latent factor, which is computationally expensive and scales poorly with dataset size (Rasmussen & Williams, 2006).

Datasets were coerced into the required formats for most JSDMs within 5 min, which suggest data formatting is not a barrier to JSDM usage. A vignette can be found in the code repository for this analysis (Wilkinson, 2018) that explains data formatting. The HLR models took longer for dataset formatting but formatting time reduced markedly with experience. HLR models are fit in MATLAB not R, the language most ecologists are familiar with, so many ecologists might experience a slight learning curve. MATLAB also requires a software license.

All JSDMs estimated similar regression coefficients across all datasets (>0.7 correlation). The three JSDMs with hierarchically drawn regression coefficients (HPR, HLR-S, HLR-NS) were strongly correlated (>0.9); however, this does not mean the additional hierarchical structure is necessarily beneficial. Minor differences between the uncertainty in JSDM coefficient estimates were observed but with no clear pattern. We suggest that the JSDMs were as uncertain as each other when considered across all coefficients.

For this comparison we used the default priors for each JSDM, though these will not always be the ideal choice. The fungi and mosquito datasets exhibit complete separation for multiple species in the case of binary environmental variables. For maximum-likelihood estimation methods, a finite maximum likelihood would not exist and standard calculation of standard errors would fail. This is less of an issue for models using a Bayesian framework, as using prior information switches the focus of parameter estimation to posterior distribution summaries (Rainey, 2016). LPR had convergence issues for these datasets with default priors, but the use of more informative priors led to performance similar to the other JSDMs (see Supporting Information Appendix S2), highlighting the importance of careful consideration of priors.

Different JSDMs estimated residual correlations between species in the same direction (with some exceptions for near-zero strength interactions whose 95% credible intervals included zero), suggesting that all JSDMs lead to similar ecological

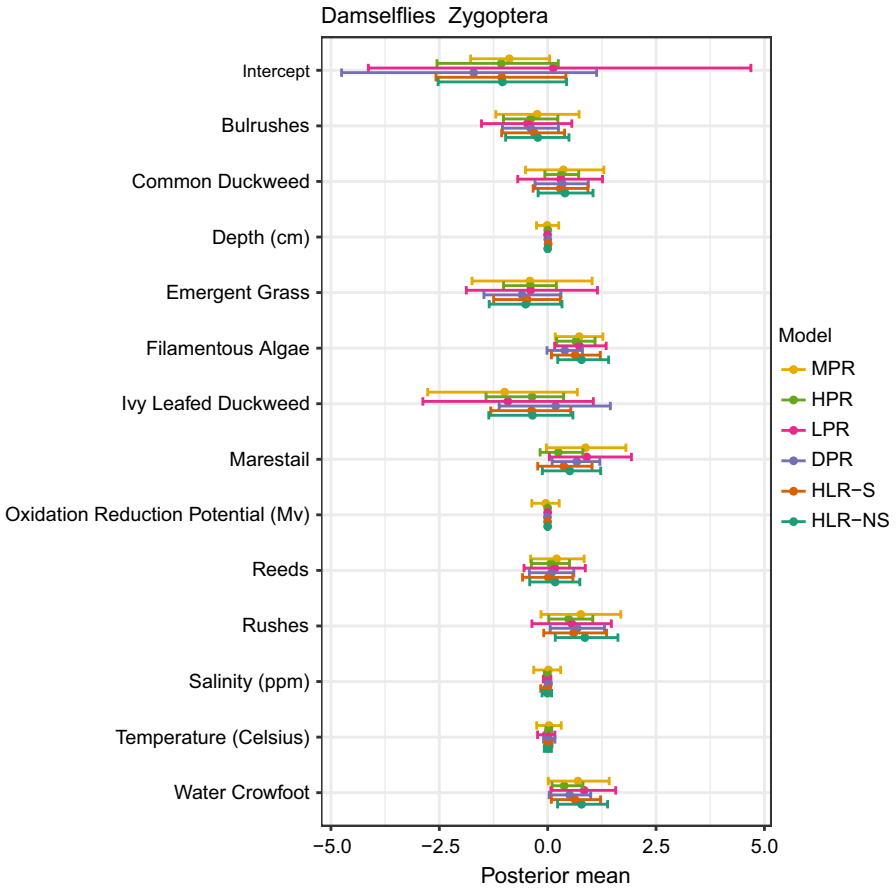


FIGURE 2 Mean regression coefficients (dots) and 95% credible intervals (bars) for damselflies in the mosquito dataset, estimated by different Joint species distribution models (JSDMs)

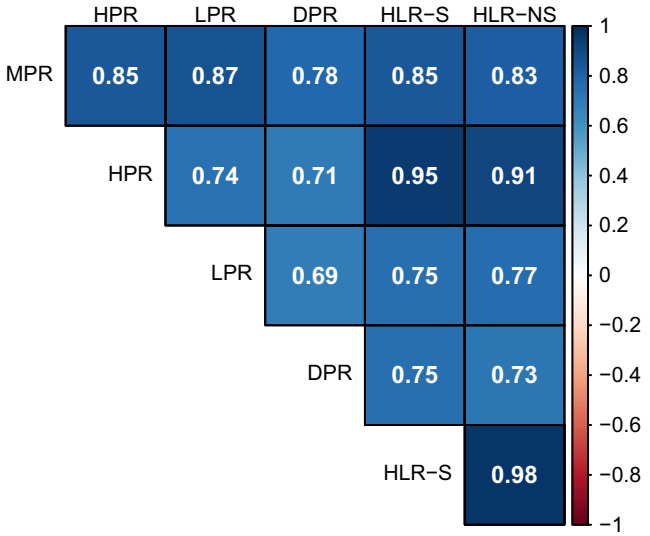


FIGURE 3 Correlation of mean regression coefficient estimates between Joint species distribution models (JSDMs), averaged across all datasets

inferences. Correlations between each JSDM's estimated correlation coefficients were moderate to very strong when averaged across all datasets. For the smaller datasets, DPR estimated weaker residual correlations than other JSDMs, and showed only

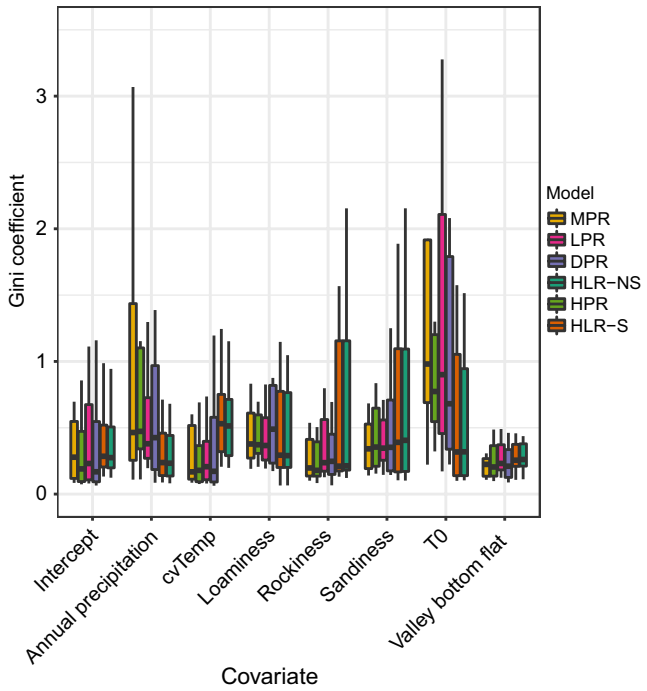


FIGURE 4 Absolute value of the Gini coefficient of the variation in the regression coefficient estimates in the eucalypt dataset for all Joint species distribution models (JSDMs)

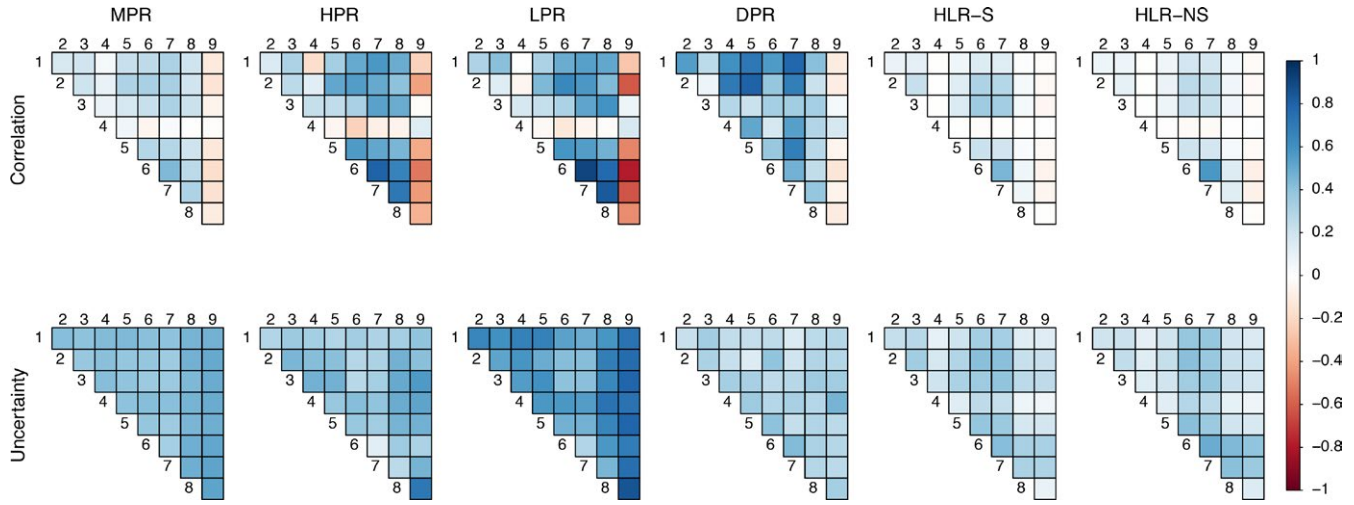


FIGURE 5 Estimated correlation coefficients (top row) and the uncertainty of those estimates defined as the width of their 95% credible interval (bottom row) of all Joint species distribution models (JSDMs)on the frog dataset

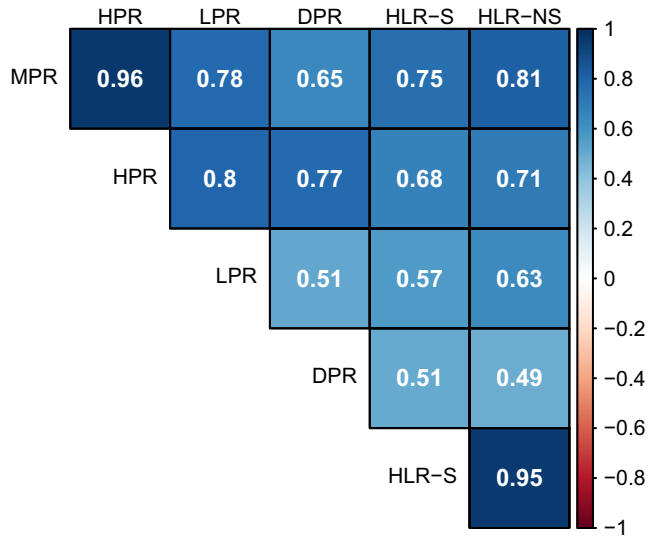


FIGURE 6 Average correlation of mean correlation coefficients between Joint species distribution models (JSDMs) across all datasets

weak correlation (~0.3) with the other JSDMs in the two largest datasets. The reason for this is unclear; the Dirichlet process was only implemented for the larger of the two, so it cannot be affecting both datasets.

Whilst the sign of residual correlations was the same across all JSDMs, the estimates differed in their relative strength and uncertainty. For the four smallest datasets, HLR-S and HLR-NS estimated relatively weaker correlations between species than the non-latent factor JSDMs (e.g., MPR and HPR). This is because latent factor models account for correlations through shared responses to a small set of latent factors, which provides a form of regularisation on residual correlations. For the two largest datasets, however, HLR-NS estimated stronger correlations between species than MPR (HLR-S failed to complete within the time limit). LPR, despite being a latent factor model like HLR-S and HLR-NS, estimated stronger correlations between species. This is likely a result of the number of latent factors used to fit the model, as the use of more latent factors provides a better approximation of the full-rank correlation matrix. By default, LPR uses two latent factors, while HLR-S and HLR-NS estimate the number of latent factors during the model fitting process (here between 3 and 12). This further highlights the trade-off between model fit and model simplicity when using the latent factor approach. DPR estimated relatively weaker correlations than the other JSDMs in most cases (similar to HLR-S/HLR-NS in the small datasets); the reason for this is unclear.

Differences in the sizes of residual correlations between the smaller and larger datasets could indicate the effect of spatial scale on the implementation and interpretation of latent factors (Ovaskainen, Abrego, et al., 2016; Ovaskainen, Roy et al., 2016). For datasets collected over small spatial scales (coincidently our smaller datasets) HLR-S estimated small α values: 0 (frogs/eucalypts) or 0.04 (mosquitoes), suggesting no benefit of a spatial implementation. Datasets collected at larger spatial scales, like the birds and butterflies (computationally incompatible with HLR-S in this study), would be more likely to exhibit spatial correlation (i.e., a moderate to large α).Indeed, Ovaskainen, Abrego, et al. (2016) and Ovaskainen, Roy et al. (2016) observed an α between 0.25 and 0.26 for the butterfly dataset (in which a training/testing split of the data reduced the computational requirements). We suggest the additional effort required for spatially explicit JSDMs might only be necessary for large spatial scale datasets where spatial correlation could be reasonably expected, and it becomes important to account for any decay in correlation between latent factors due to distance. Correlations between species at finer spatial scales might reflect species interactions, but at larger scales correlations are likely influenced by other factors like shared environmental responses and dispersal ability (Götzenberger et al.,

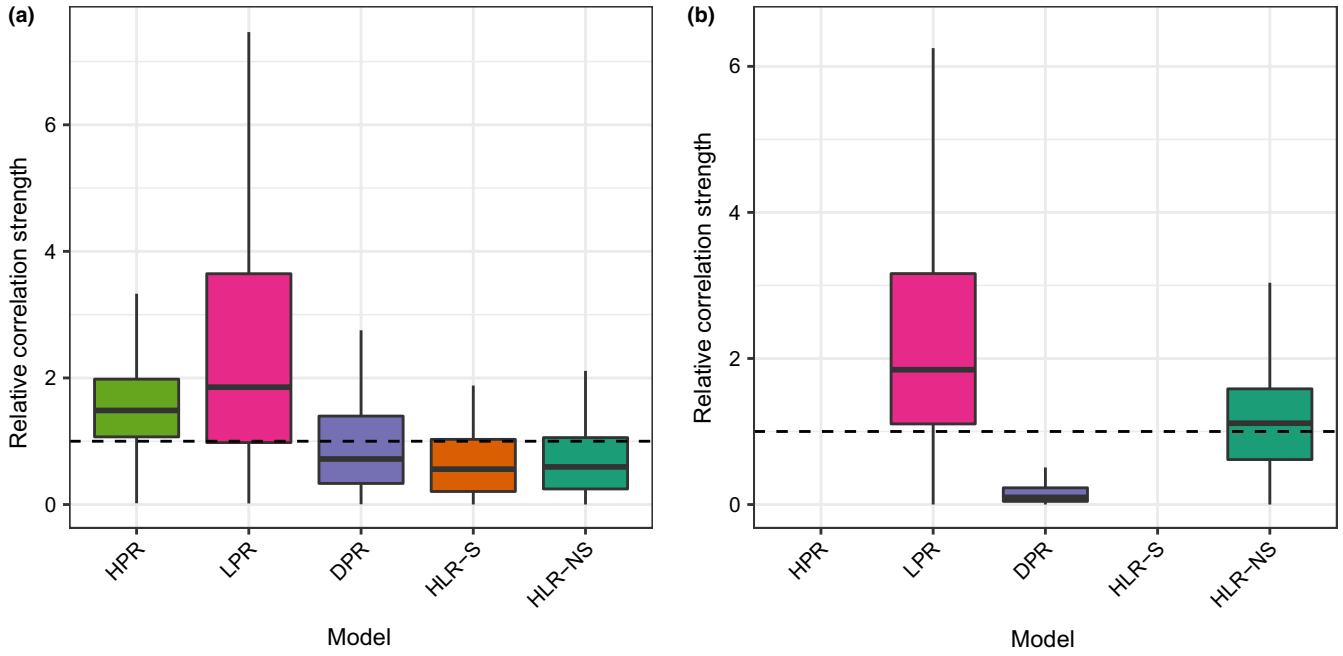


FIGURE 7 Relative correlation strength of all Joint species distribution models (JSDMs) compared to a baseline of the multivariate probit regression (MPR) model. Plot (a) is for the four smallest datasets, and plot (b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles, respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range, respectively. HPR and HLR-S are missing from 7b as they did not complete within the time limit

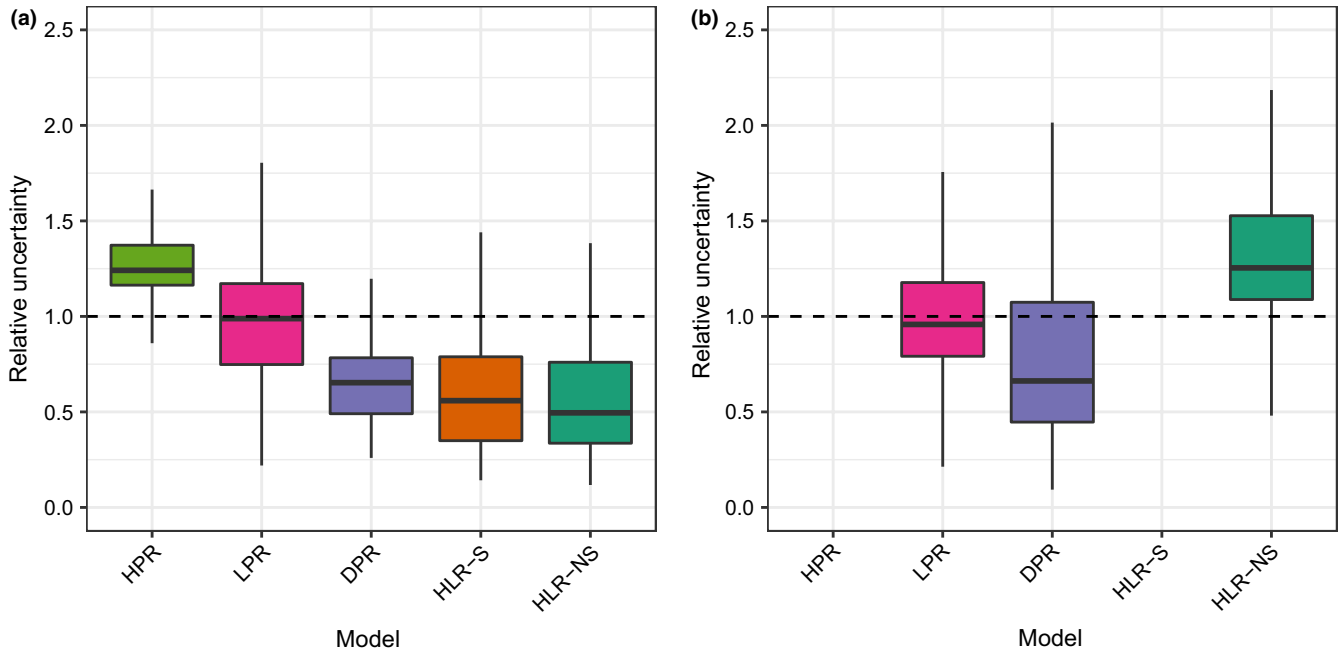


FIGURE 8 Relative correlation uncertainty (95% credible interval width) of all Joint species distribution models (JSDMs) compared to a baseline of the multivariate probit regression (MPR) model. Plot (a) is for the four smallest datasets, and plot (b) is for the two largest. The horizontal black lines represent the median value, the lower and upper hinges correspond to the 25th and 75th quantiles, respectively, and the lower and upper whiskers extend from the hinge to the smallest and largest value within 1.5 times the interquartile range, respectively. HPR and HLR-S are missing from 8b as they did not complete within the time limit

2012; Tikhonov, Abrego, Dunson, & Ovaskainen, 2017). This can be explored by defining latent factors at different spatial scales (Ovaskainen, Abrego et al., 2016).

In Supporting Information Appendix S5 we compared model performance for a subset of these models on simulated datasets. As for the real datasets included in the main analysis, we found

that there were no significant differences in parameter estimation between JSDMs. We observed that models that accounted for hierarchical regression coefficients performed better than the other models for datasets generated hierarchically but only if they also matched the latent factor assumptions. Models also performed worse when fit to datasets that were simulated using unmeasured variables as well as the measured ones used to fit the models, but models that could account for that using latent factors (like HLR-NS) had smaller performance decreases. This shows that the added complexity of these models can provide a benefit to parameter estimation, where the ecological system that generates the data has these features.

Methodological differences between the different JSDMs beyond their core statistical framework might have biased our results. The program in which JSDMs are developed (e.g., R vs MATLAB) could influence computational efficiency, as could their respective MCMC samplers and regimes. The choice of priors and other default settings, such as the number of latent factors, could also impact statistical inference. However, since the goal of this study was to compare the default implementations of these different JSDMs and their underlying statistical frameworks, these differences are unavoidable. Clearly, particular case studies should explore the appropriateness of default settings prior to implementation.

While we only considered presence–absence JSDMs in which species interactions are modelled via multivariate regression methods, other JSDMs have been proposed. MISTNET (Harris, 2015), a neural network JSDM, focuses on predicting community composition but does not provide coefficient estimates comparable to the JSDMs considered here. The JSDM of Johnson and Sinclair (2017) models species interactions by grouping them into guilds by environmental responses rather than via residual correlation. JSDMs for abundance data also exist (e.g., Dorazio, Connor, & Askins, 2015; Letten et al., 2015; Thorson et al., 2015; Warton et al., 2015).

We have compared the ability of JSDMs to make inferences about species' responses to biotic and abiotic factors. We have used both real and simulated datasets, but a more comprehensive simulation study accounting for all nuances of the model fitting process presents an interesting avenue for further research. The predictive ability of different JSDMs also warrants further study. Single species SDMs are commonly used to predict into un-sampled areas or under new environmental conditions, such as climate change scenarios (Elith & Leathwick, 2009), and these are clearly areas where JSDMs are applicable. There is an open question, however, of how best to define prediction in multivariate space. JSDMs can predict community composition (joint prediction), distributions of individual species (marginal prediction), or distributions of individual species conditional on the presence or absence of other species (conditional prediction). Ecological interpretations and the statistical and computational aspects of these different types of prediction should be considered.

Another important consideration is that all methods described here assume perfect detection (i.e., that absence records reflect true species absence). However, imperfect detection is often an issue in survey data. Extending these models to account for imperfect detection is an important avenue for future JSDM research (Beissinger et al., 2016; Warton et al., 2016), as is the evaluation of impacts of imperfect detection in the estimation of residual co-occurrence. Rota et al. (2016) introduced a multispecies occupancy model accounting for imperfect detection, but in principle, any of the methods above could be extended to include a description of the detection process.

5 | CONCLUSION

Akin to comparisons of single species SDMs (Elith et al., 2006), this study serves as a guide for prospective JSDM users to make informed decisions about the JSDM, or JSDMs, that might be relevant to their objective. Expressing the different statistical JSDM frameworks with a common notation has helped to identify their differences and similarities. The core statistical differences between different JSDMs can be broadly defined by two methodological choices: the inclusion of a hierarchical structure on the regression coefficients, and the use of latent factors to account for shared species responses to "hypothetical" unmeasured environmental variables impacting species co-occurrence patterns. There are statistical and computational implications related to these methodological differences, but they also have ecological interpretations. As for all ecological modelling, the user should be aware of all ecological assumptions inherent in these models when making an informed selection of JSDM.

ACKNOWLEDGEMENTS

We thank Peter Vesk, Brendan Wintle, and Jian Yen for insightful discussions, and Els Van Burm and Erica Marshall for early manuscript reviews. We thank Robert Dorazio, Devin Johnson, and one anonymous reviewer for their valuable input. D.P.W. is funded by an Australian Government Research Training Program Scholarship. N.G. is funded by a McKenzie fellowship from the University of Melbourne. G.G.-A. and R.T. are supported by Australian Research Council (ARC) Discovery Early Career Researcher Awards (DE160100904 and DE170100601).

CONFLICTS OF INTEREST

Some of the authors of this manuscript are authors on the JSDM papers being compared. N.G., R.T., and M.A.M. on Pollock et al. (2014), and N.G. on Golding et al. (2015). No other conflicts of interest are declared.

AUTHORS' CONTRIBUTIONS

All authors conceived the ideas and methodology. D.P.W. implemented the analysis. D.P.W. led writing the manuscript but all authors contributed significantly throughout and gave final approval before submission.

DATA ACCESSSIBILITY

- Script files to replicate the model running, data extraction, analysis, and plot generation are available in an online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018).
- The bird dataset is from the USGS Breeding Bird Survey and is available to download from https://www.mbr-pwrc.usgs.gov/bbs/bbs2011.html (Sauer et al., 2012). A version used in this analysis can also be found in our online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018).
- The butterfly dataset is available to download in the supplementary material of Ovaskainen, Roy et al. (2016). A version used in this analysis can also be found in our online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018).
- The eucalypt dataset is available to download from Pollock (2014) A version used in this analysis can also be found in our online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018).
- An anonymised version of the frog dataset is available in our online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018). This dataset will allow you to replicate our results without the ability to identify the particulars of the dataset.
- The fungi dataset is available in our online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018)
- The mosquito dataset is available to download from Golding (2015) https://doi.org/10.6084/m9.figshare.1420528.v1. A version used in this analysis can also be found in our online repository https://doi.org/10.5281/zenodo.1452066 (Wilkinson, 2018).

ORCID

David P. Wilkinson http://orcid.org/0000-0002-9560-6499
Nick Golding http://orcid.org/0000-0001-8916-5570
Gurutzeta Guillera-Arroita http://orcid.org/0000-0002-8387-5739
Reid Tingley http://orcid.org/0000-0002-7630-7434
Michael A. McCarthy http://orcid.org/0000-0003-1039-7980

REFERENCES

Araújo, M. B., & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, *16*, 743–753. https://doi.org/10.1111/j.1466-8238.2007.00359.x

Beissinger, S. R., Iknayan, K. J., Guillera-Arroita, G., Zipkin, E. F., Dorazio, R. M., Royle, J. A., & Kéry, M. (2016). Incorporating imperfect detection into joint models of communities: A response to Warton et al.. *Trends in Ecology and Evolution*, *31*, 736–737. https://doi.org/10.1016/j.tree.2016.07.009

Chib, S., & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, *85*, 347–361. https://doi.org/10.1093/biomet/85.2.347

Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, *24*, 990–999. https://doi.org/10.1890/13-1015.1

Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, *87*, 34–56. https://doi.org/10.1002/ecm.1241

Dorazio, R. M., Connor, E. F., & Askins, R. A. (2015). Estimating the effects of habitat and biological interactions in an avian community. *PLoS ONE*, *10*, 1–16.

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., … Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, *39*, 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., … Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., & Leathwick, J. J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution*, *40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Gelfand, A. E., Silander, J. A., Wu, S., Latimer, A., Lewis, P. O., Rebelo, A. G., & Holder, M. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, *1*, 41–92. https://doi.org/10.1214/06-BA102

Golding, N. (2015). Mosquito community data for Golding et al. 2015 (Parasites & Vectors) (Version 1). figshare. https://doi.org/10.6084/m9.figshare.1420528.v1 figshare.

Golding, N., & Harris, D. J. (2015). *BayesComm: Bayesian community ecology analysis*. R package version 0.1-2. https://CRAN.Rproject.org/package=BayesComm

Golding, N., Nunn, M. A., & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites & Vectors*, *8*, 367. https://doi.org/10.1186/s13071-015-0915-1

Götzenberger, L., de Bello, F., Bråthen, K. A., Davison, J., Dubuis, A., Guisan, A., … Zobel, M. (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, *87*, 111–127. https://doi.org/10.1111/j.1469-185X.2011.00187.x

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: Advances, challenges and opportunities. *Ecography*, *40*, 281–295. https://doi.org/10.1111/ecog.02445

Hardy, O. J. (2008). Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, *96*, 914–926. https://doi.org/10.1111/j.1365-2745.2008.01421.x

Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, *6*, 465–473. https://doi.org/10.1111/2041-210X.12332

Hui, F. K. C. (2016). boral - Bayesian ordination and regression analysis of multivariate abundance data in r. *Methods in Ecology and Evolution*, *7*, 744–750. https://doi.org/10.1111/2041-210X.12514

Hui, F. K. C. (2017). boral: Bayesian Ordination and Regression Analysis.

Hutchinson, G. E. (1957). Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, *22*, 415–427. https://doi.org/10.1101/SQB.1957.022.01.039

Johnson, D. S., & Sinclair, E. H. (2017). Modeling joint abundance of multiple species using Dirichlet process mixtures. *Environmetrics*, *28*, 1–13.

Kissling, W. D., Dormann, C. F., Groeneveld, J., Hickler, T., Kühn, I., Mcinerny, G. J., … O'Hara, R. B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, *39*, 2163–2178. https://doi.org/10.1111/j.1365-2699.2011.02663.x

Leathwick, J. R., & Austin, M. P. (2001). Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology*, *82*, 2560–2573. https://doi.org/10.1890/0012-9658(2001)082[2560:CIBTSI]2.0.CO;2

Letten, A. D., Keith, D. A., Tozer, M. G., & Hui, F. K. C. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, *103*, 1264–1275. https://doi.org/10.1111/1365-2745.12428

MacKenzie, D. I., Bailey, L. L., & Nichols, J. D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, *73*, 546–555. https://doi.org/10.1111/j.0021-8790.2004.00828.x

Meier, E. S., Kienast, F., Pearman, P. B., Svenning, J. C., Thuiller, W., Araújo, M. B., … Zimmermann, N. E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, *33*, 1038–1048. https://doi.org/10.1111/j.1600-0587.2010.06229.x

Nieto-Lugilde, D., Maguire, K. C., Blois, J. L., Williams, J. W., & Fitzpatrick, M. C. (2017). Multiresponse algorithms for community-level modeling: Review of theory, applications, and comparison to species distribution models. *Methods in Ecology and Evolution*, *12*, 3218–3221.

O'Brien, S. M., & Dunson, D. B. (2004). Bayesian multivariate logistic regression. *Biometrics*, *60*, 739–746. https://doi.org/10.1111/j.0006-341X.2004.00224.x

Ovaskainen, O., Abrego, N., Halme, P., & Dunson, D. (2016). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, *7*, 549–555. https://doi.org/10.1111/2041-210X.12501

Ovaskainen, O., Hottola, J., & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, *91*, 2514–2521. https://doi.org/10.1890/10-0173.1

Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, *7*, 428–436. https://doi.org/10.1111/2041-210X.12502

Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C. F., Vittoz, P., Dubuis, A., … Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, *33*, 1004–1014. https://doi.org/10.1111/j.1600-0587.2010.06386.x

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.3406

Pollock, L. J. (2014). Data from: Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Dryad Digital Repository*. https://doi.org/10.1111/2041-210x.12180

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., … Mccarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*, 397–406. https://doi.org/10.1111/2041-210X.12180

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis*, *24*, 339–355. https://doi.org/10.1093/pan/mpw014

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning* (2nd ed.). Cambridge, MA: The MIT Press.

Richmond, O. M. W., Hines, J. E., & Beissinger, S. R. (2010). Two-species occupancy models: A new parameterization applied to co-occurence of secretive rails. *Ecological Applications*, *20*, 2036–2046. https://doi.org/10.1890/09-0470.1

Rota, C. T., Ferreira, M. A. R., Kays, R. W., Forrester, T. D., Kalies, E. L., McShea, W. J., … Millspaugh, J. J. (2016). A multispecies occupancy model for two or more interacting species. *Methods in Ecology and Evolution*, *7*, 1164–1173. https://doi.org/10.1111/2041-210X.12587

Sauer, J. R., Hines, J. E., Fallon, J. E., Pardieck, K. L., Ziolkowski, Jr. D. J., & Link, W. A. (2012). The North American Breeding Bird Survey, Results and Analysis 1966 - 2011. Version 12.13.2011.

Schweiger, O., Heikkinen, R. K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., … Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global Ecology and Biogeography*, *21*, 88–99. https://doi.org/10.1111/j.1466-8238.2010.00607.x

Taylor-Rodríguez, D., Kaufeld, K., Schliep, E. M., Clark, J. S., & Gelfand, A. E. (2016). Joint species distribution modeling: Dimension reduction using Dirichlet processes. *Bayesian Analysis*, *12*(4), 939–967.

The MathWorks Inc (2016). *MATLAB R2016a*. MA, USA: Natick.

Thorson, J. T., Ianelli, J. N., Larsen, E. A., Ries, L., Scheuerell, M. D., Szuwalski, C., & Zipkin, E. F. (2016). Joint dynamic species distribution models: A tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, *25*, 1144–1158. https://doi.org/10.1111/geb.12464

Thorson, J. T., Scheuerell, M. D., Shelton, A. O., See, K. E., Skaug, H. J., & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, *6*, 627–637. https://doi.org/10.1111/2041-210X.12359

Tikhonov, G., Abrego, N., Dunson, D., & Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods in Ecology and Evolution*, *8*, 443–452. https://doi.org/10.1111/2041-210X.12723

Lafayette, L., Sauter, G., Vu, L., & Meade, B. (2016). *Spartan performance and flexibility: An HPC-cloud chimera*. Barcelona: OpenStack Summit.

Waddle, J. H., Dorazio, R. M., Walls, S. C., Rice, K. G., Beauchamp, J., Schuman, M. J., & Mazzotti, F. J. (2010). A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, *20*, 1467–1475. https://doi.org/10.1890/09-0850.1

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2015). So many variables: Joint modeling in community ecology. *Trends in Ecology and Evolution*, *30*, 766–779. https://doi.org/10.1016/j.tree.2015.09.007

Warton, D. I., Blanchet, F. G., O'Hara, R., Ovaskainen, O., Taskinen, S., Walker, S. C., & Hui, F. K. C. (2016). Extending joint models in community ecology: A response to Beissinger et al.. *Trends in Ecology and Evolution*, *31*, 737–738. https://doi.org/10.1016/j.tree.2016.07.007

Wilkinson, D. (2018). JSDM_Inference v0.1.6. Zenodo, https://doi.org/10.5281/zenodo.1452066

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., … Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, *88*, 15–30. https://doi.org/10.1111/j.1469-185X.2012.00235.x

Wolfram Research Inc. (2016). *Mathematica*, *Version 10.4*. Champaign, IL, USA.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Wilkinson DP, Golding N, Guillera-Arroita G, Tingley R, McCarthy MA. A comparison of joint species distribution models for presence–absence data. *Methods Ecol Evol*. 2018;00:1–14. https://doi.org/10.1111/2041-210X.13106

APPENDIX B

FOREST PLOTS OF REGRESSION COEFFICIENTS

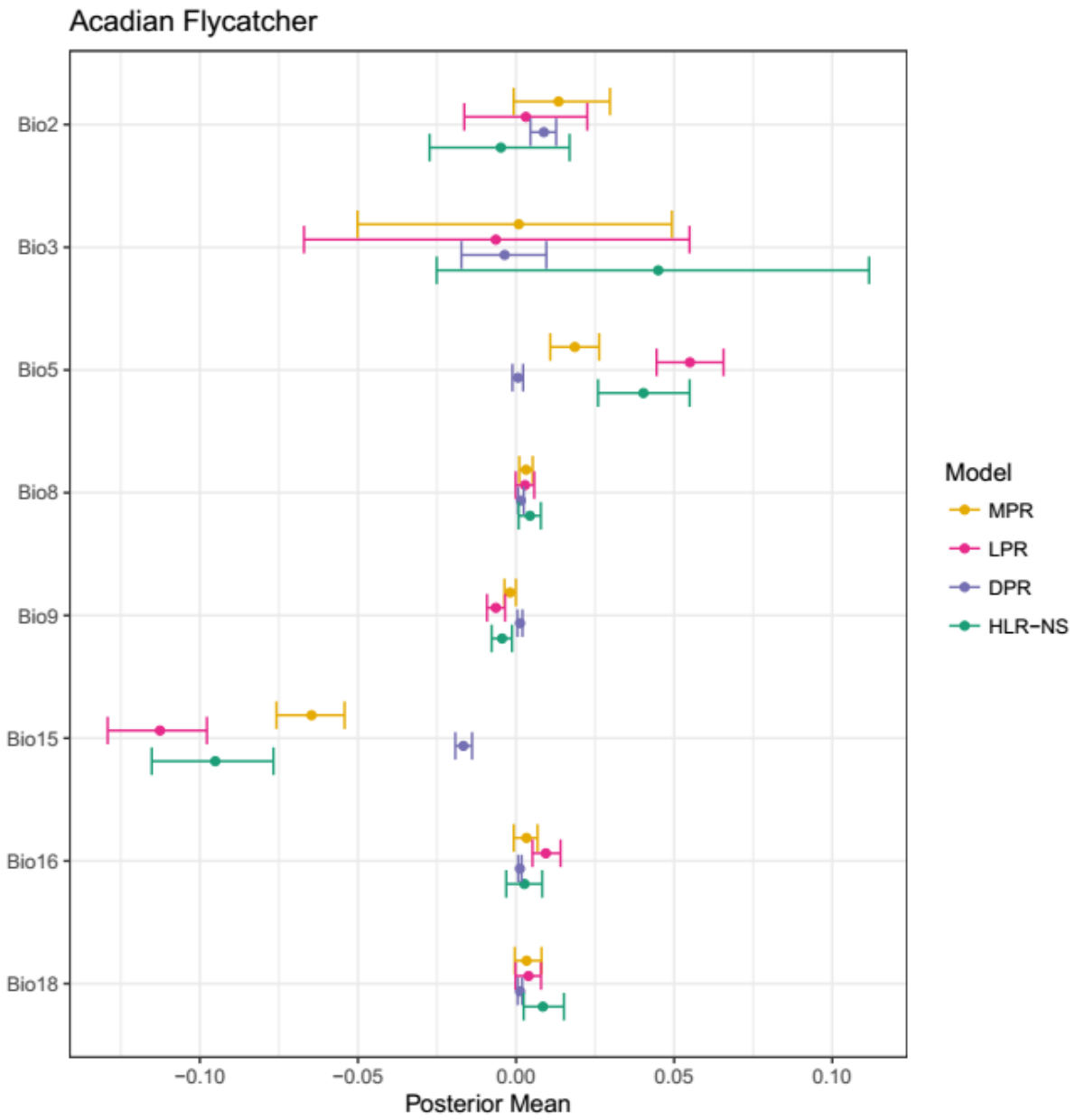


Figure 6: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Acadian flycatcher, *Empidonax virescens*.

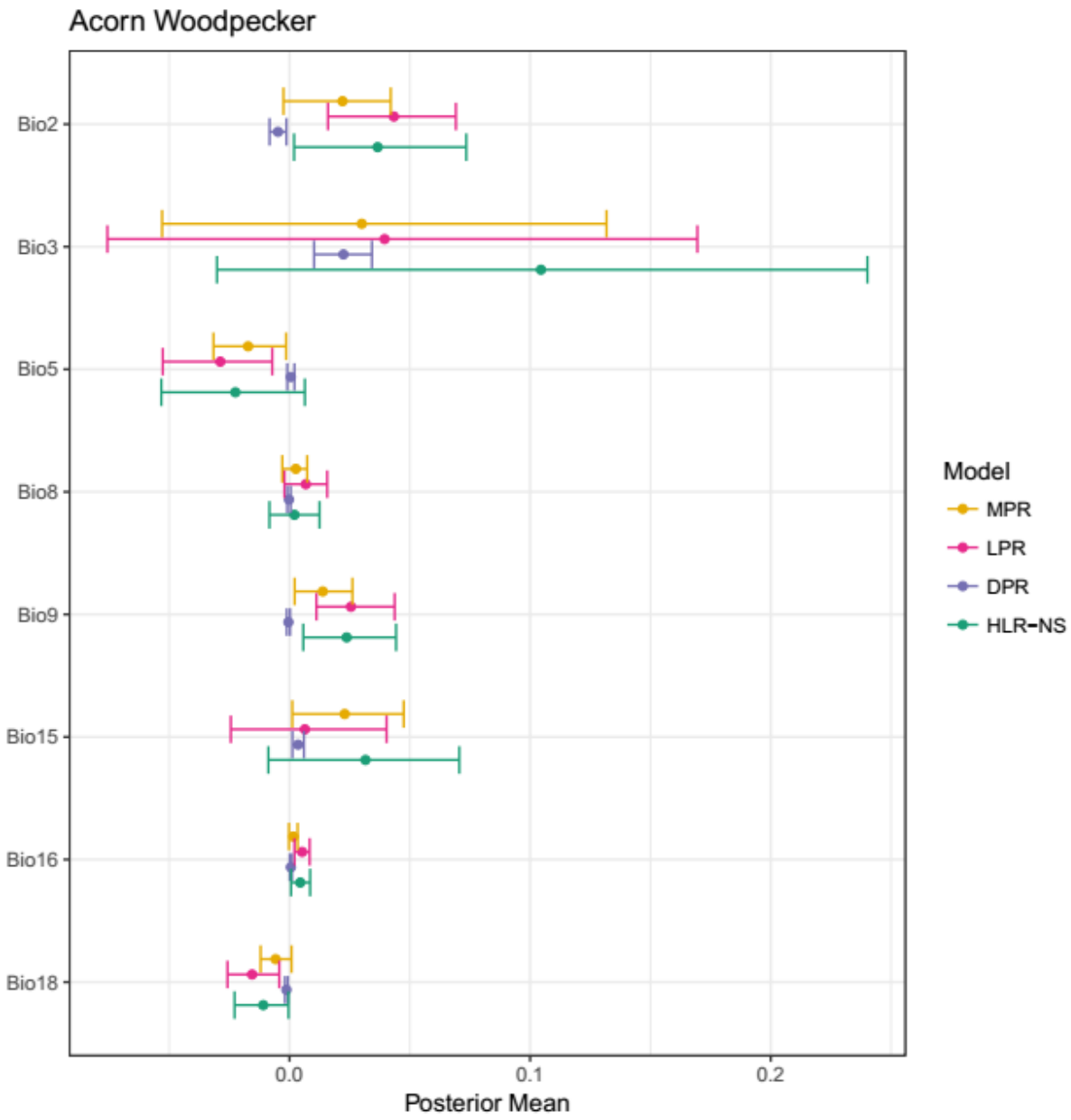


Figure 7: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the acorn woodpecker, *Melanerpes formicivorus*.

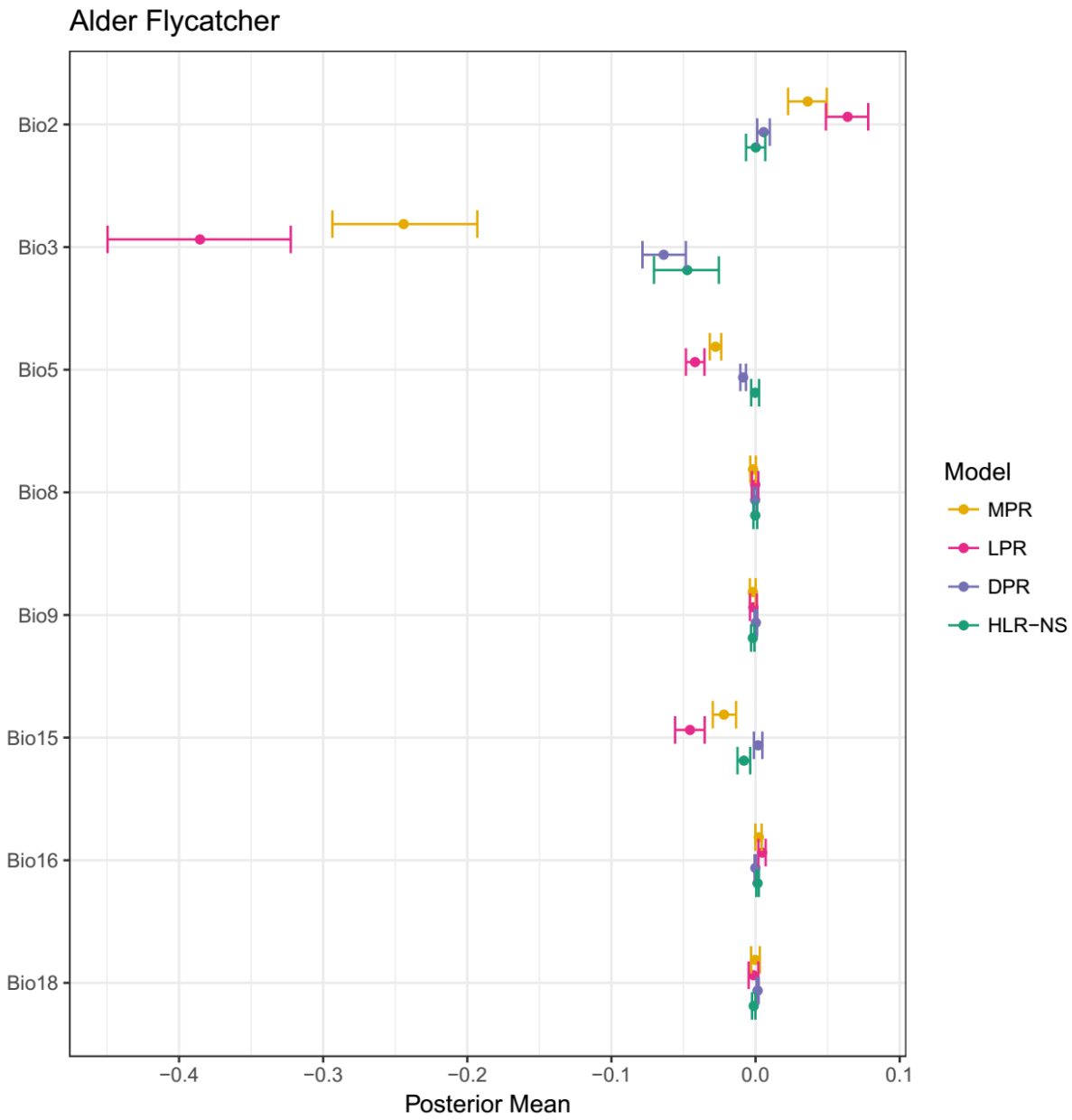


Figure 8: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Alder flycatcher, *Empidonax alnorum*.

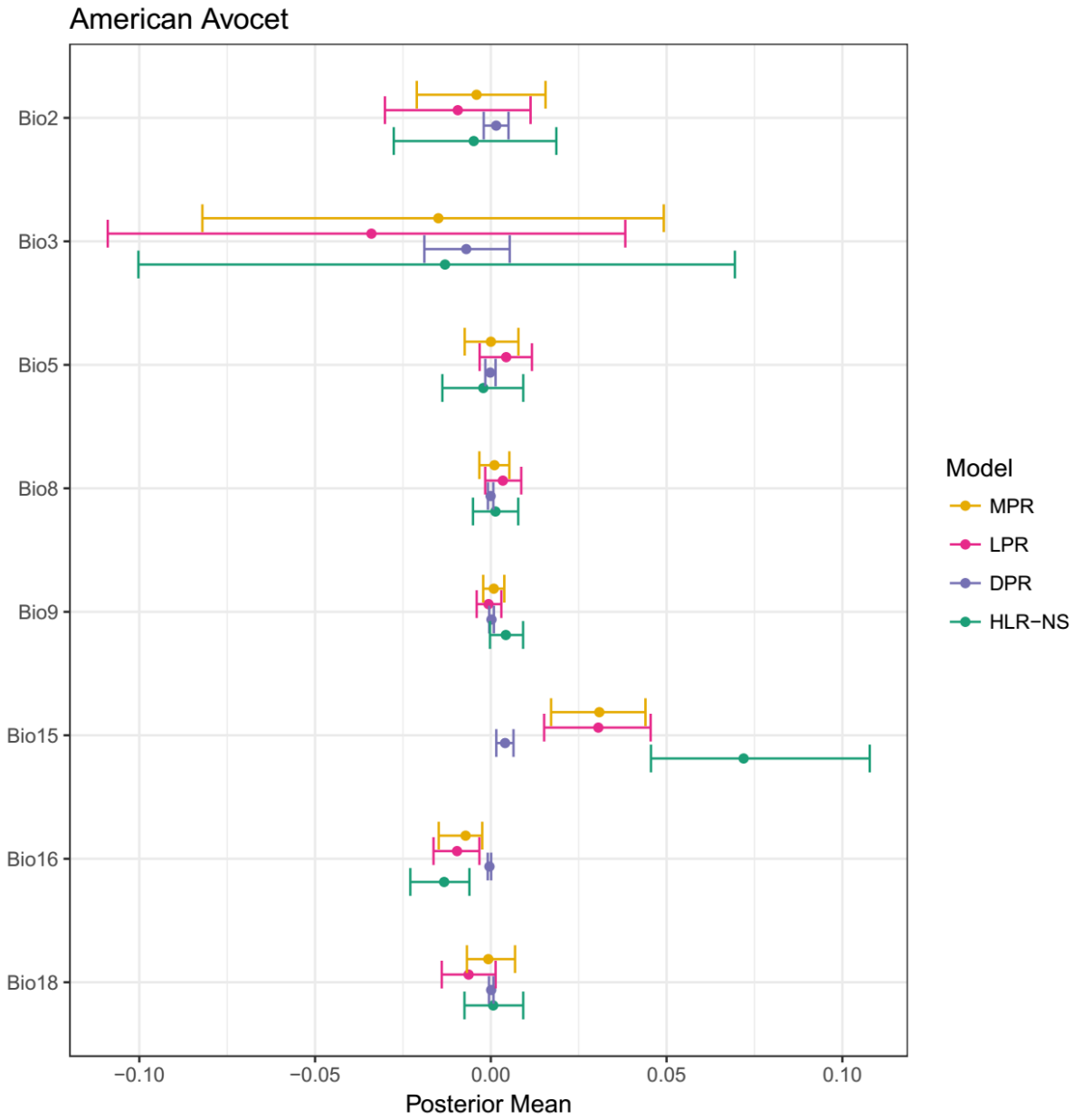


Figure 9: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American avocet, *Melanerpes formicivorus*.

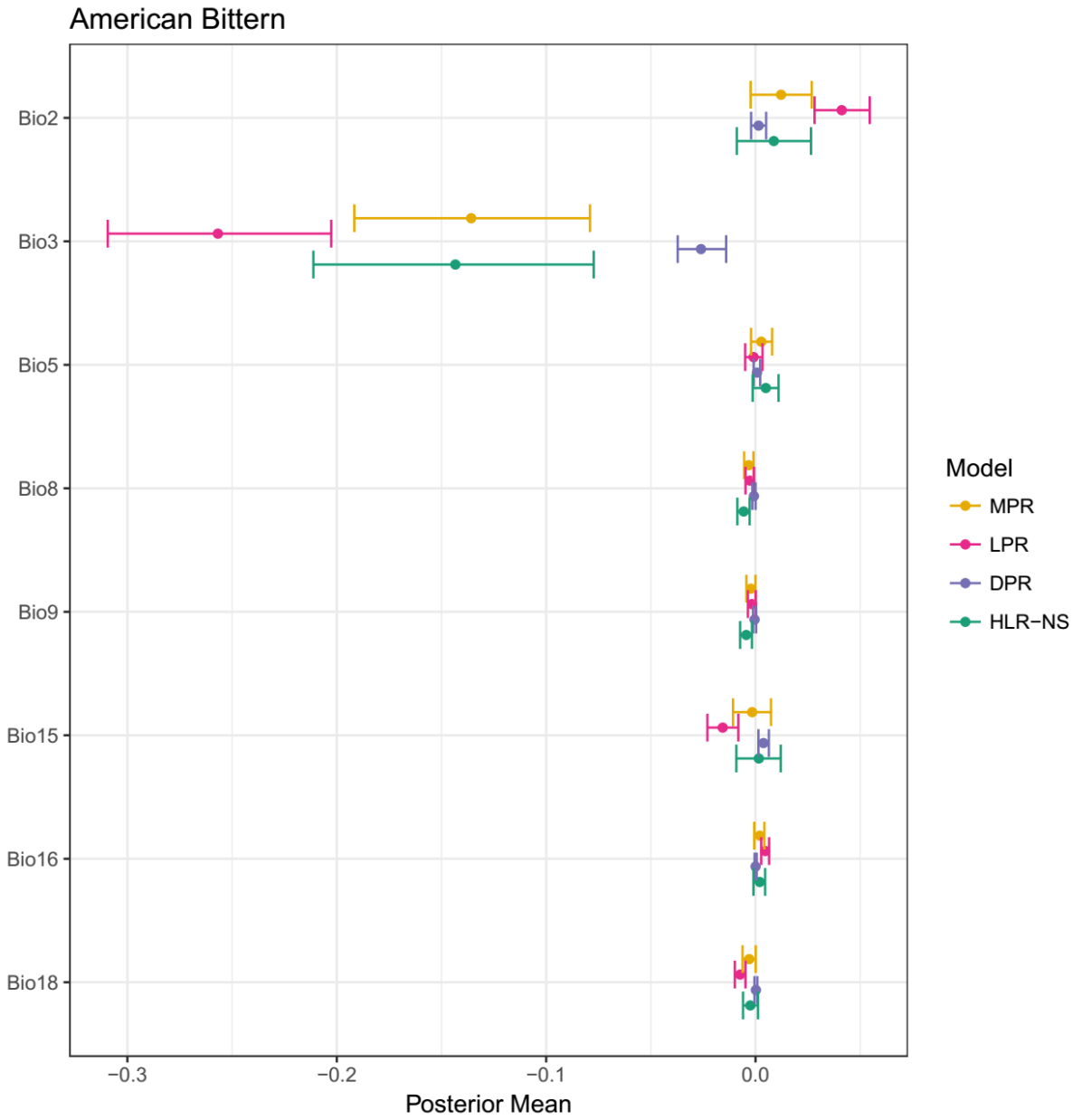


Figure 10: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American bittern, *Botaurus lentiginosus*.

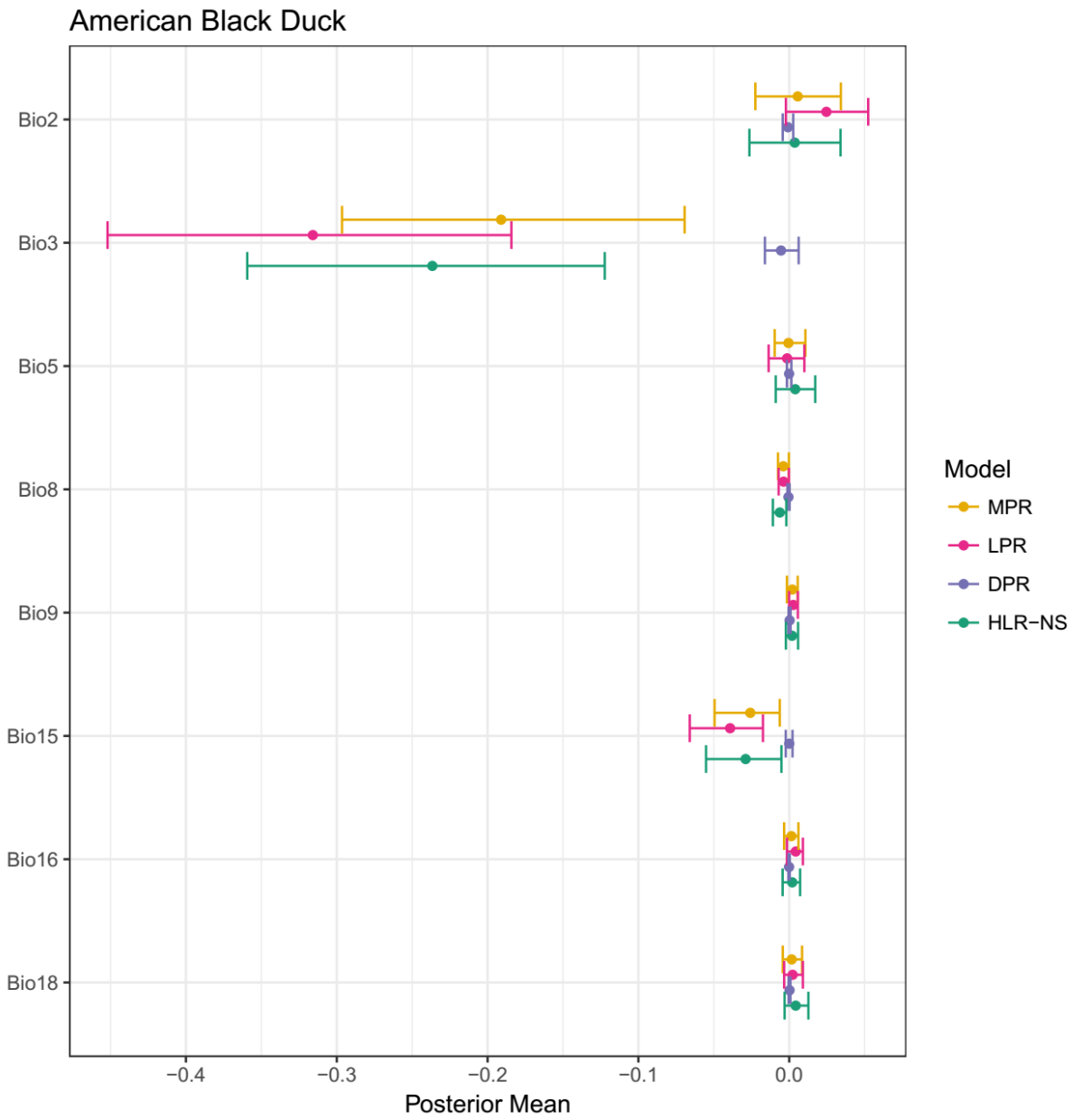


Figure 11: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American black duck, *Anas rubripes*.

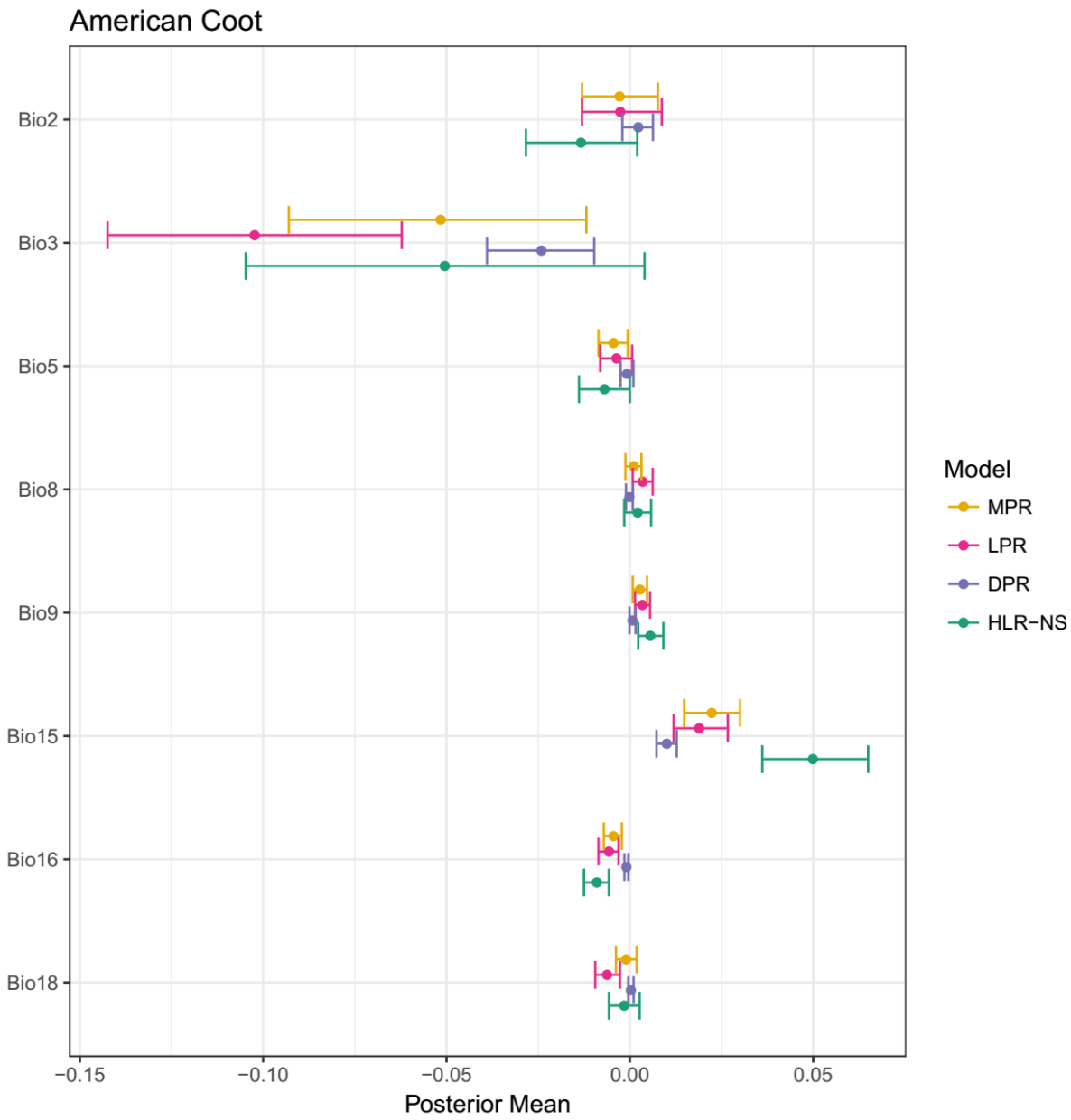


Figure 12: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American coot, *Fulica americana*.

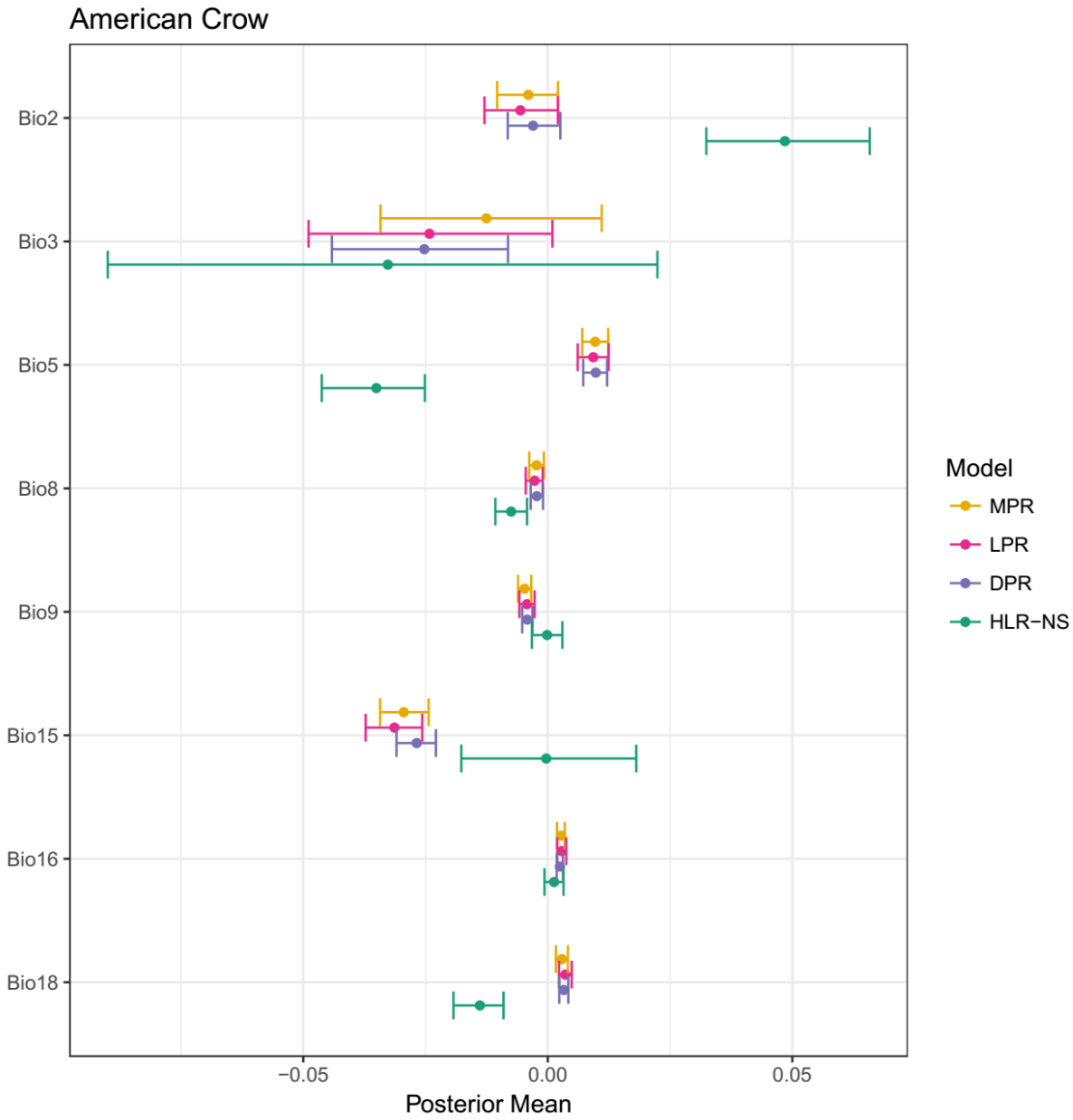


Figure 13: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American crow, *Corvus brachyrhynchos*.

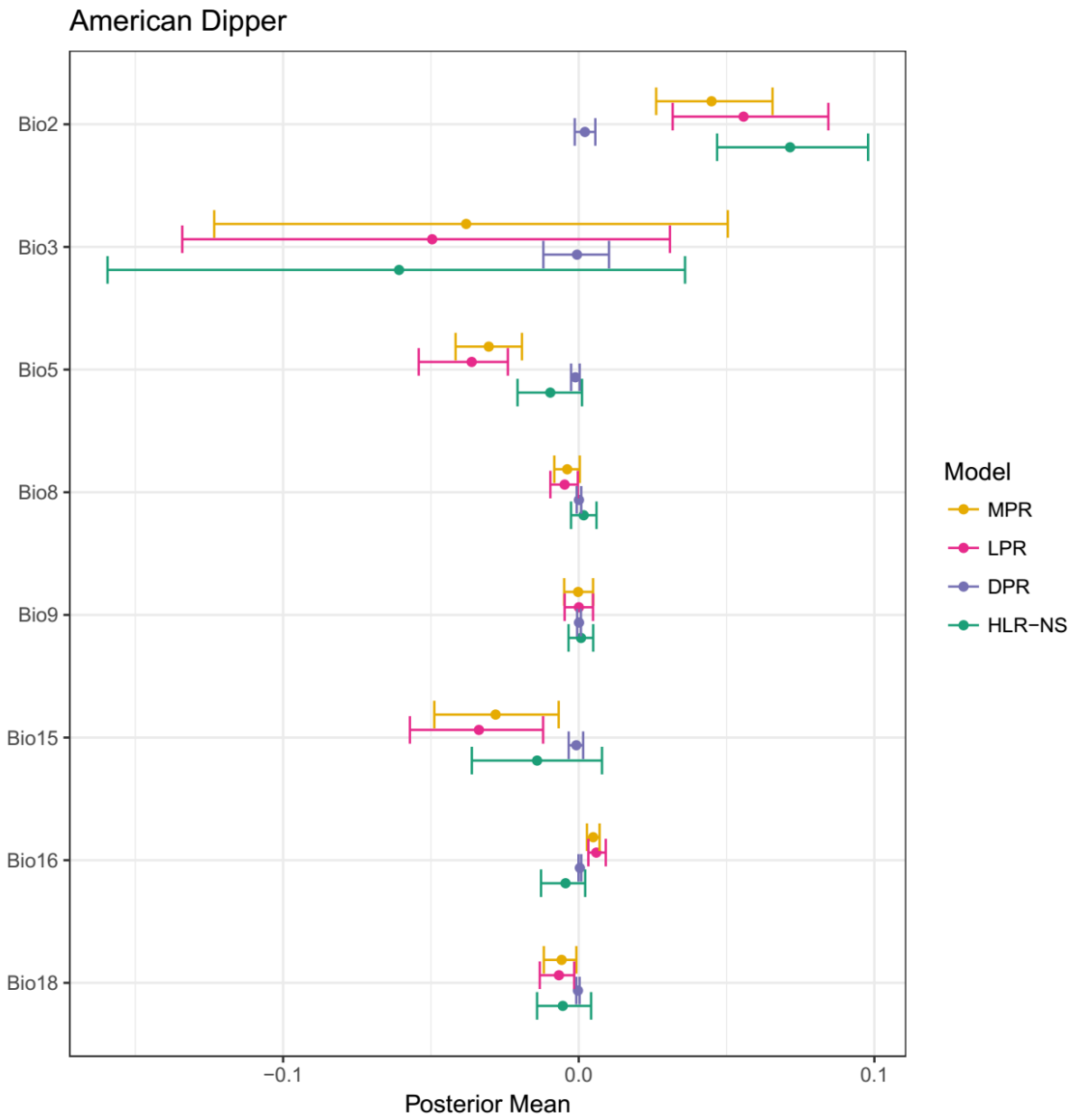


Figure 14: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American dipper, *Cinclus mexicanus*.

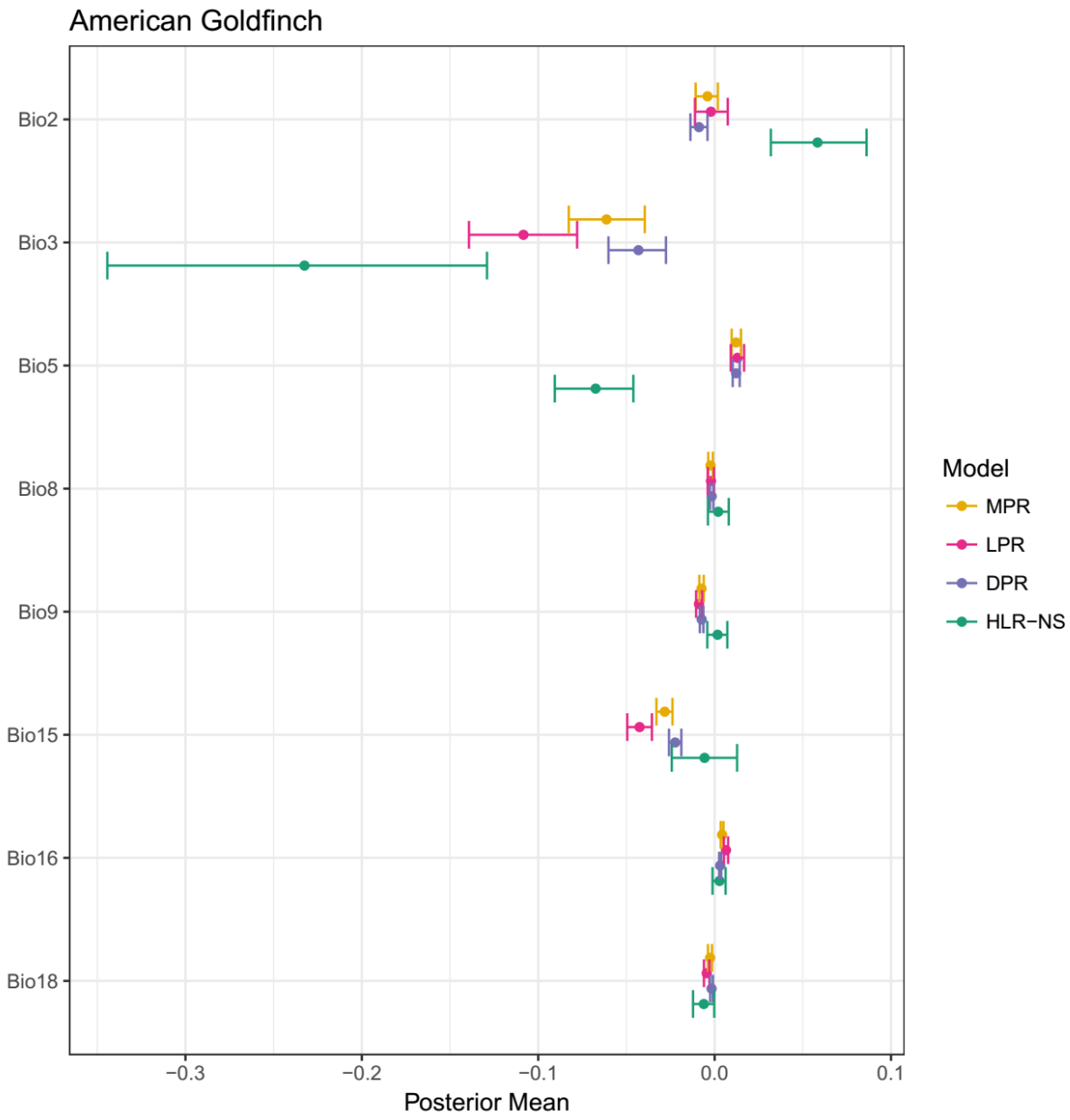


Figure 15: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American goldfinch, *Spinus tristis*.

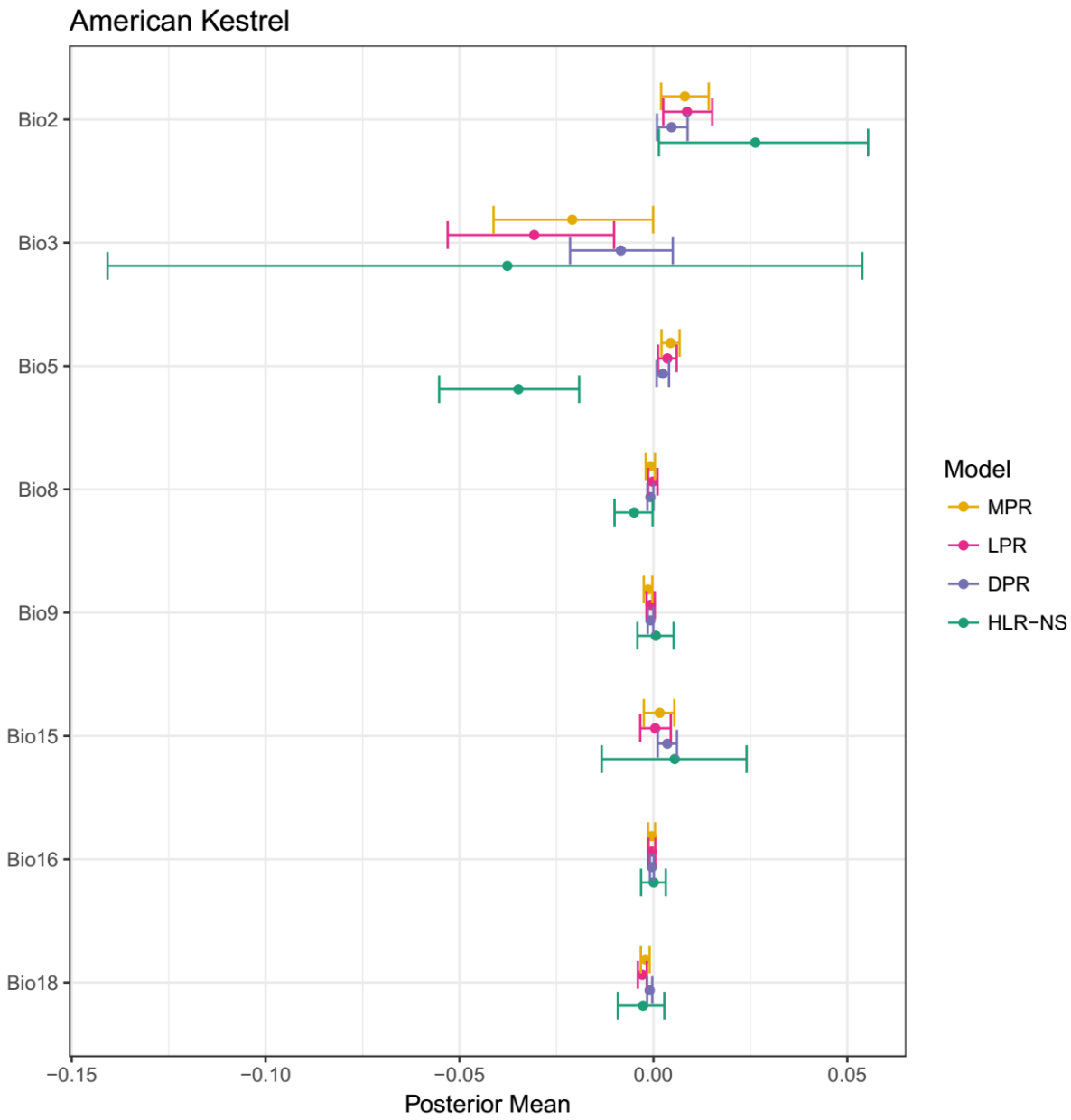


Figure 16: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American kestrel, *Falco sparverius*.

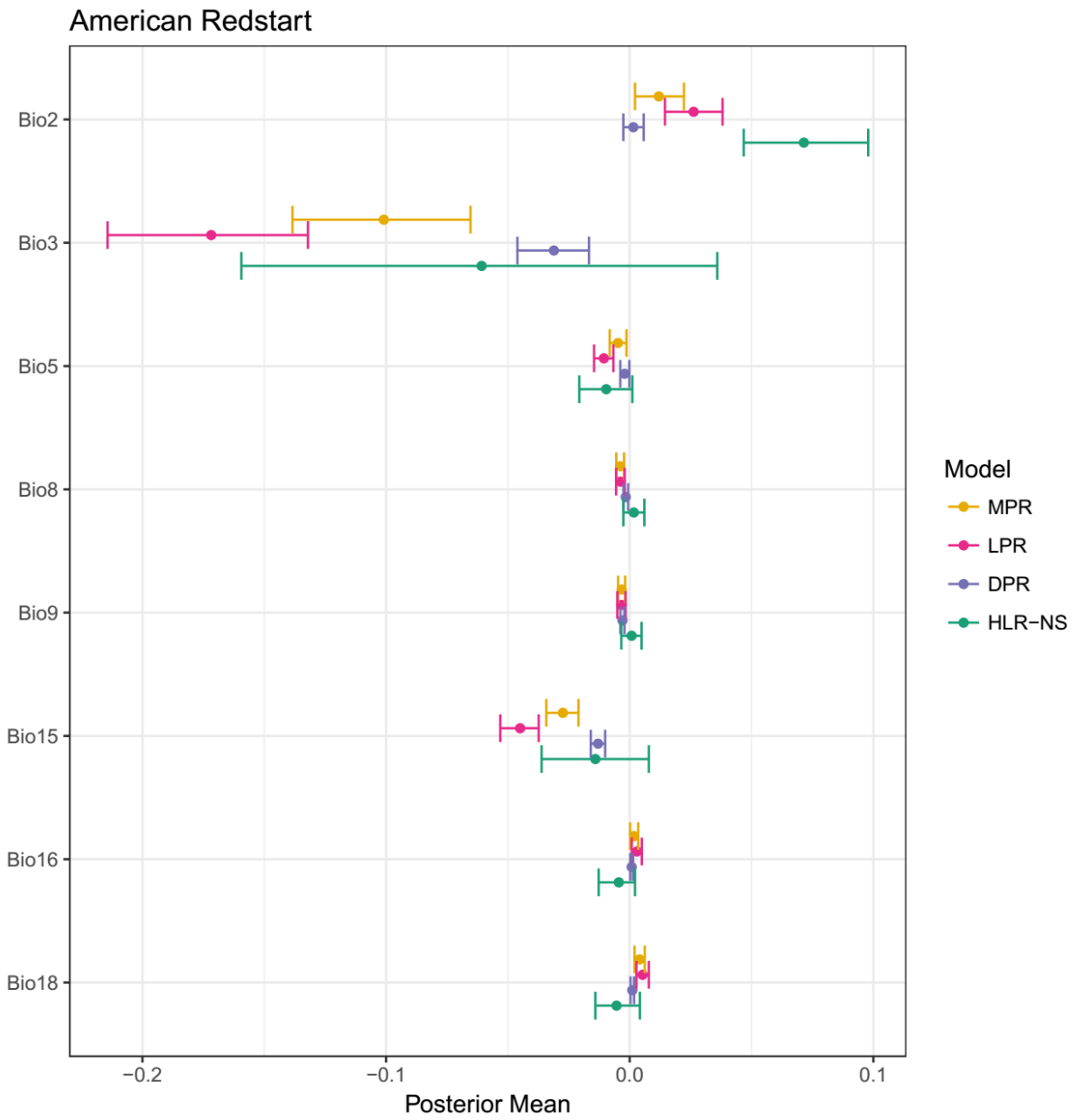


Figure 17: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American redstart, *Setophaga ruticilla*.

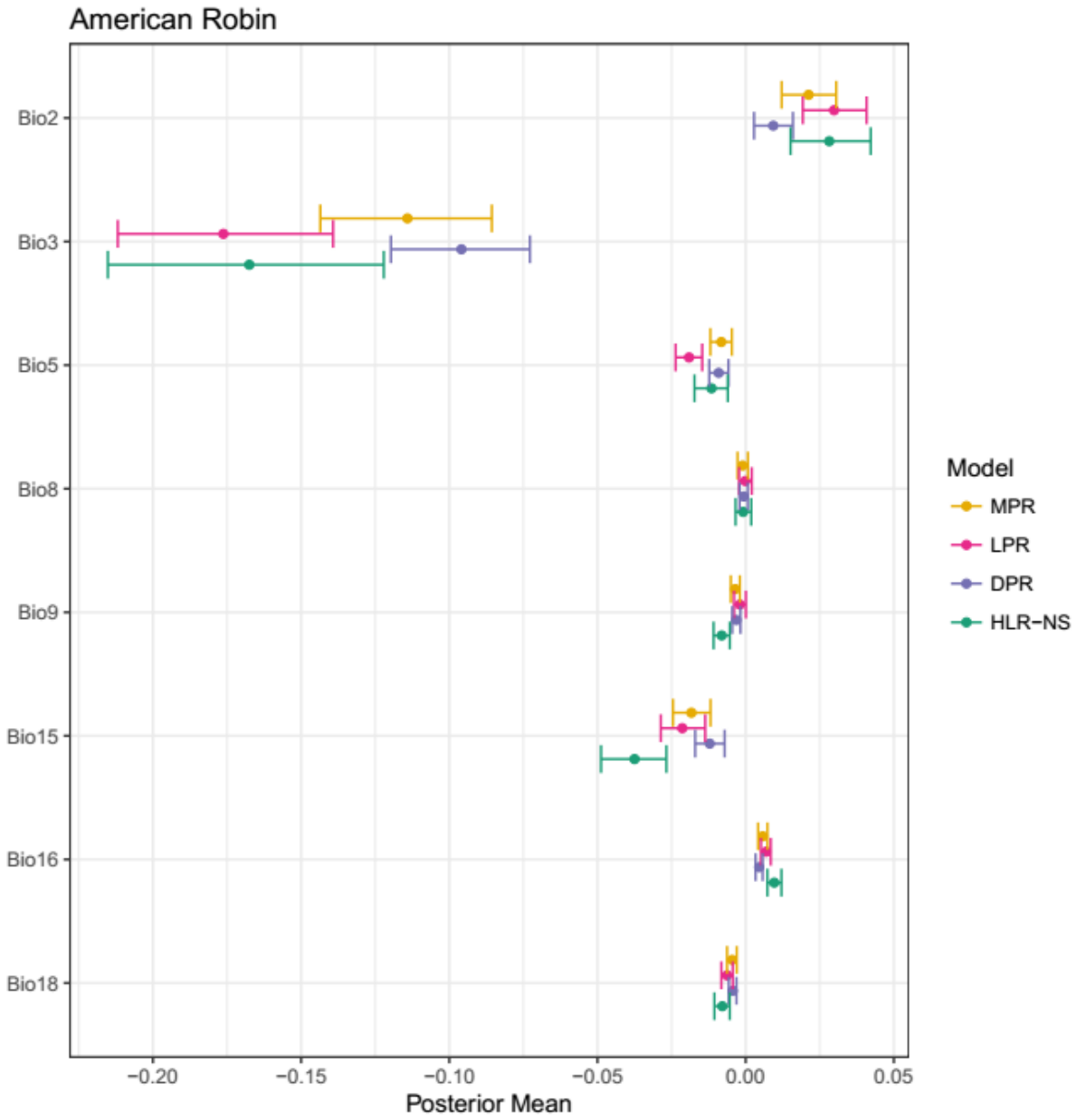


Figure 18: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American robin, *Turdus migratorius*.

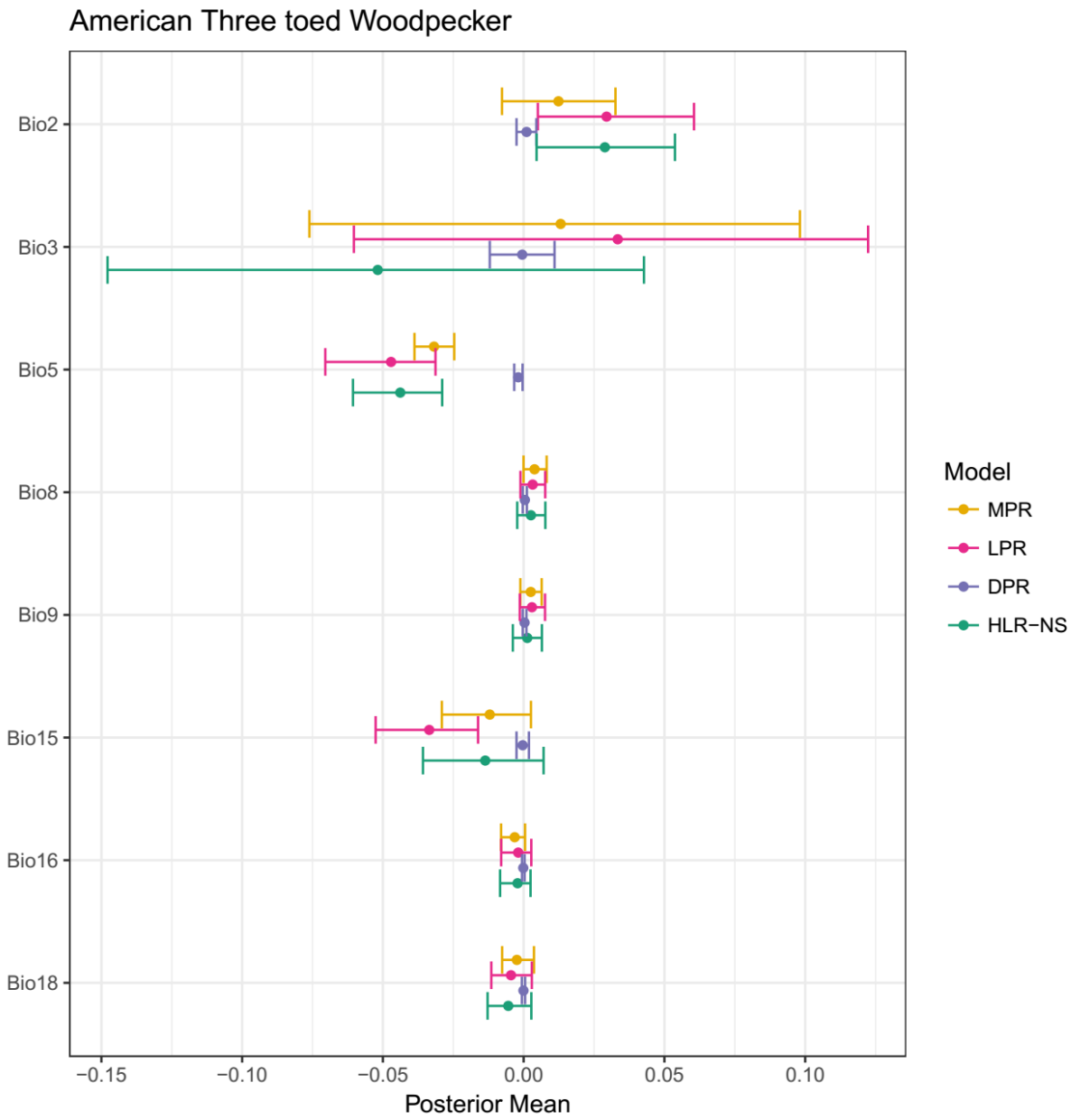


Figure 19: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American three-toed woodpecker, *Picoides dorsalis*.

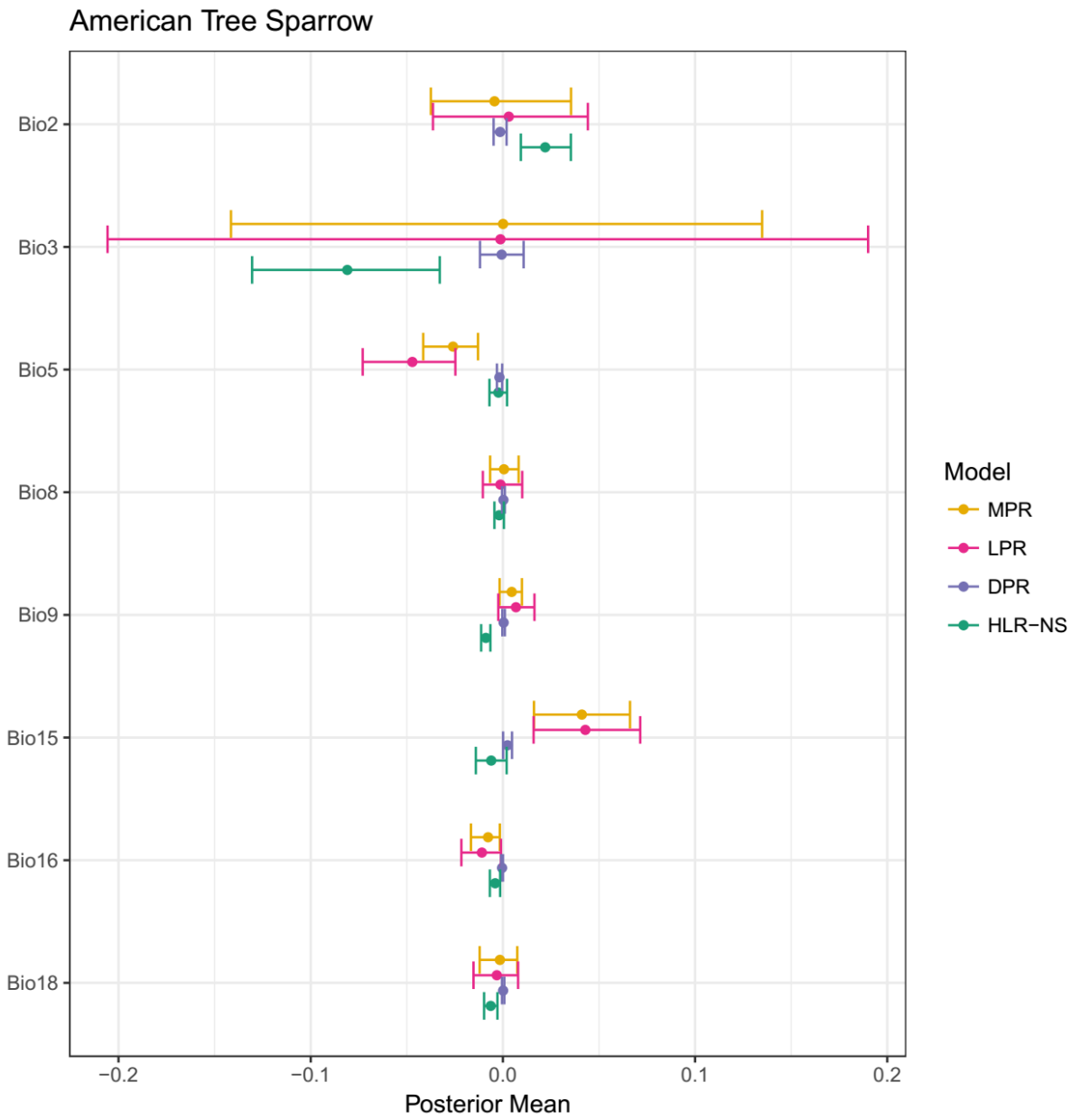


Figure 20: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American tree sparrow, *Spizella arborea*.

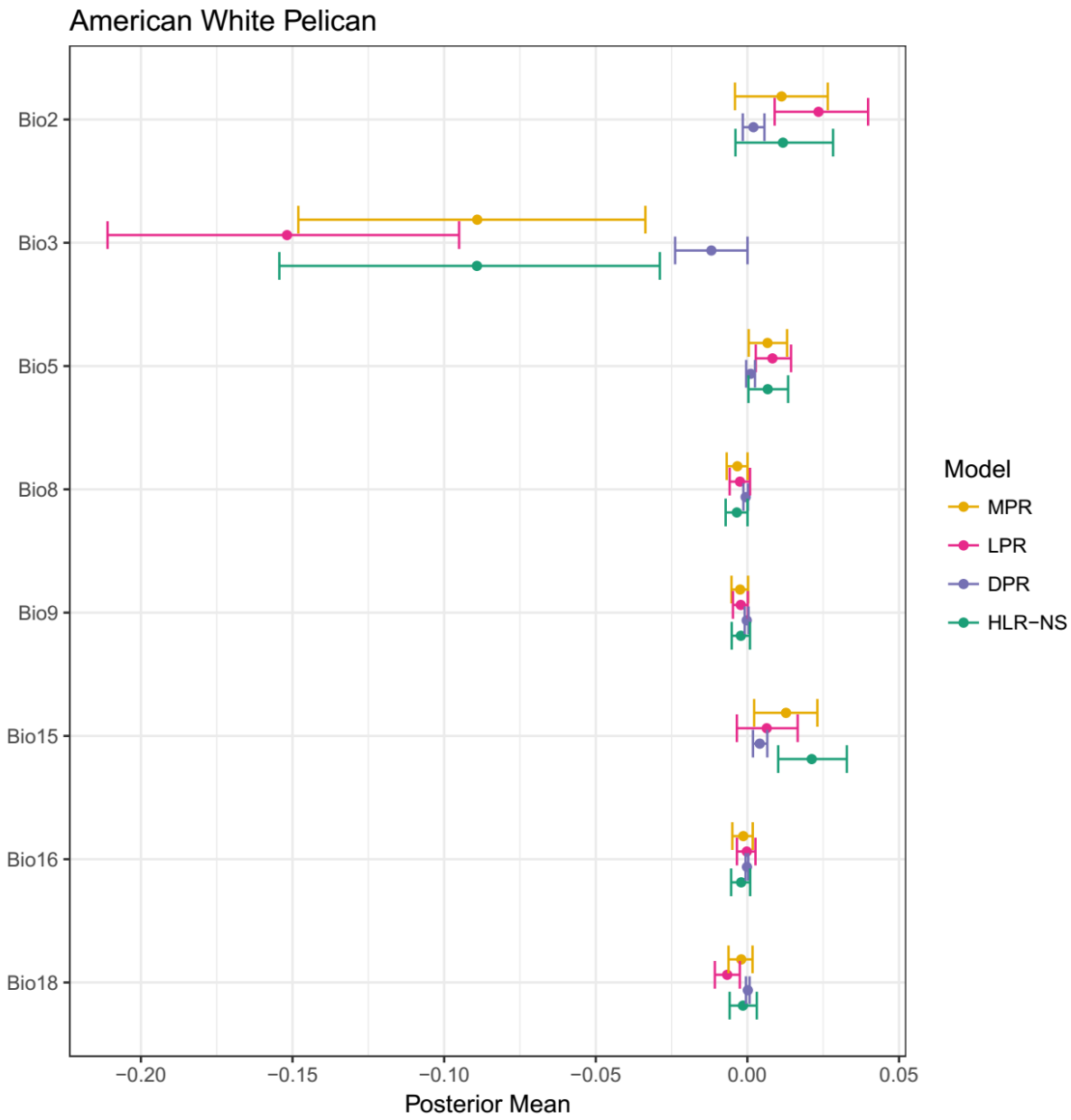


Figure 21: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American white pelican, *Pelecanus erythrorhynchos*.

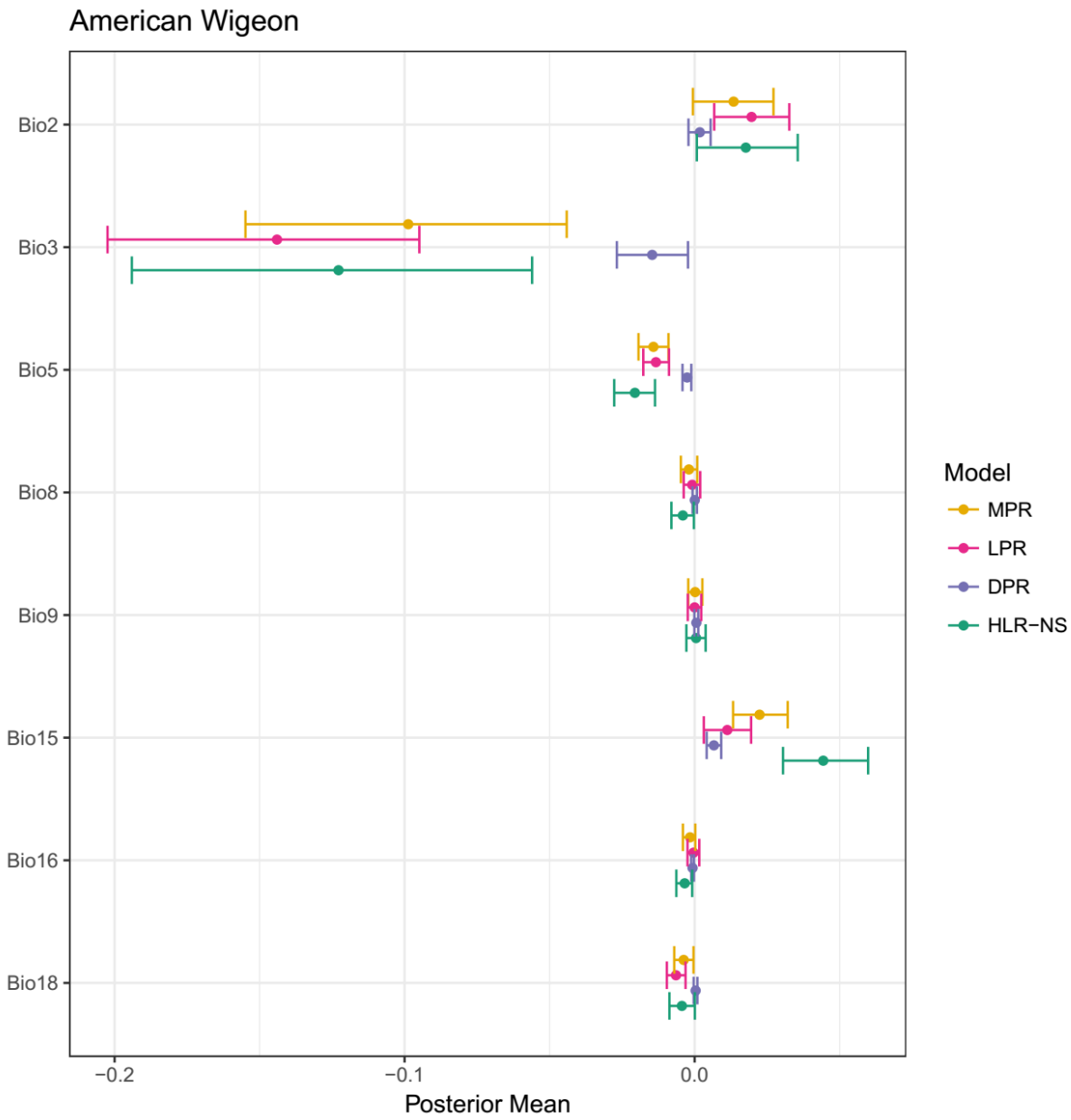


Figure 22: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American wigeon, *Anas americana*.

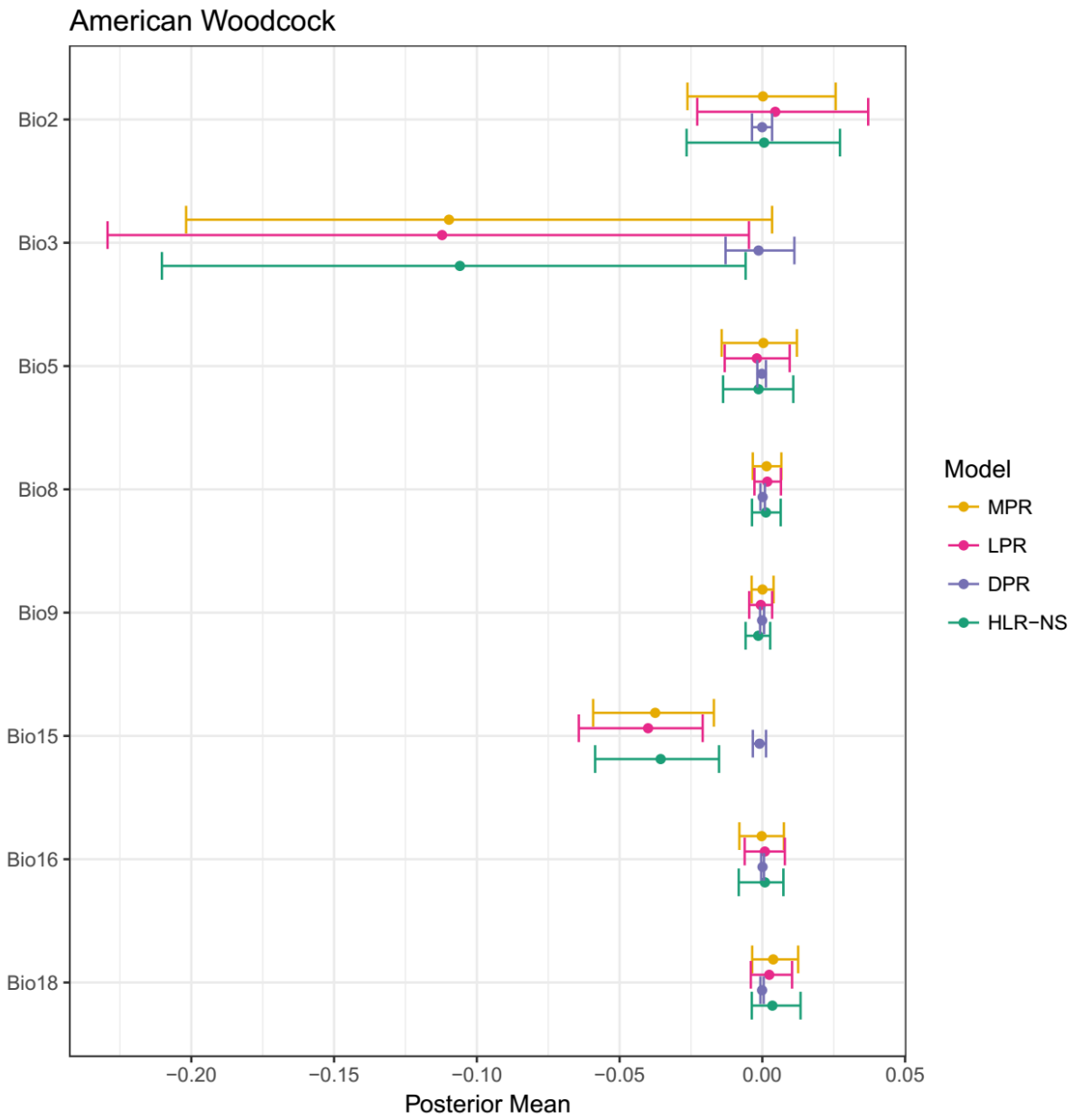


Figure 23: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the American woodcock, *Scolopax minor*.

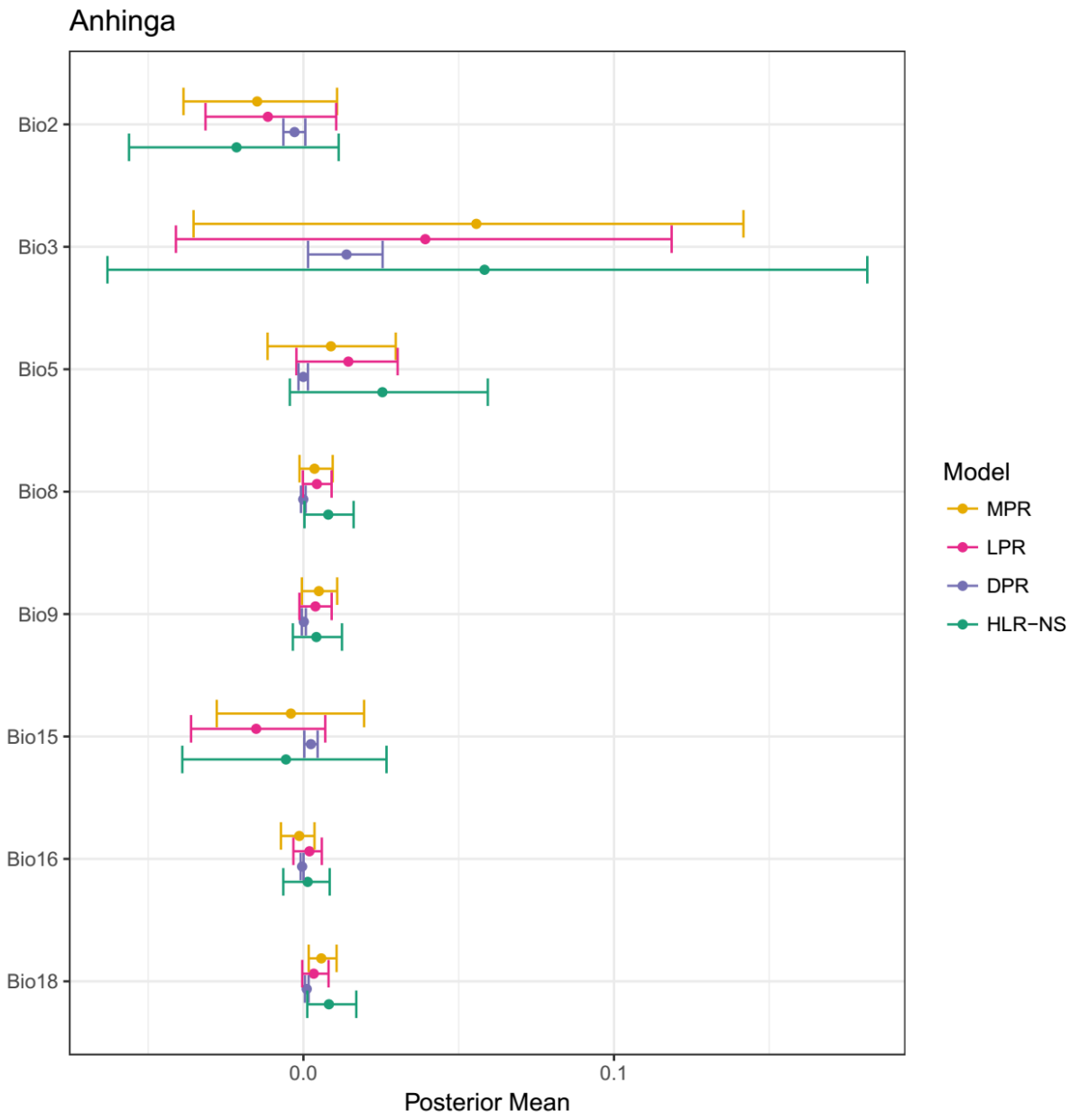


Figure 24: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Anhinga, *Anhinga anhinga*.

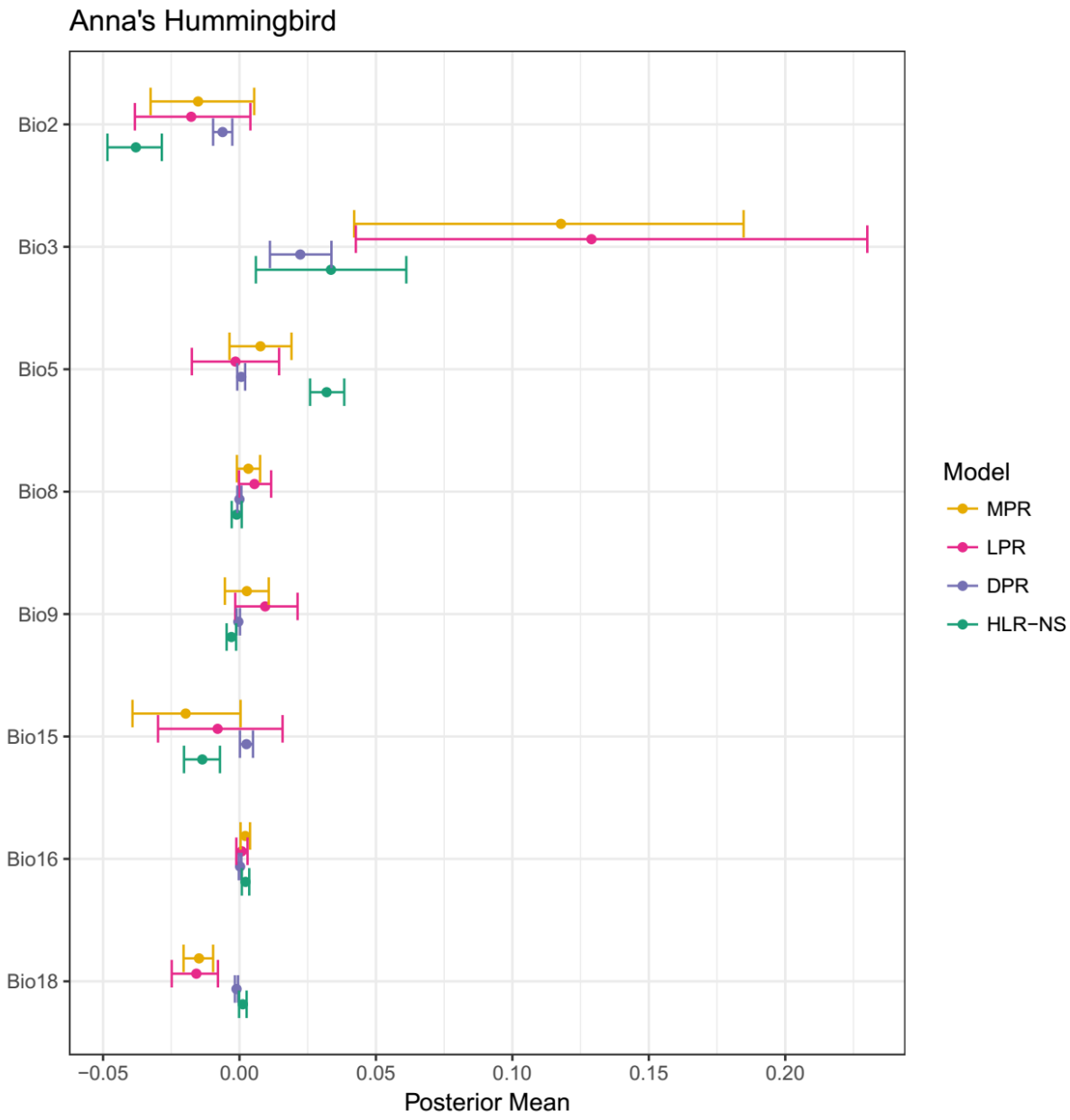


Figure 25: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Anna's hummingbird, *Calypte anna*.

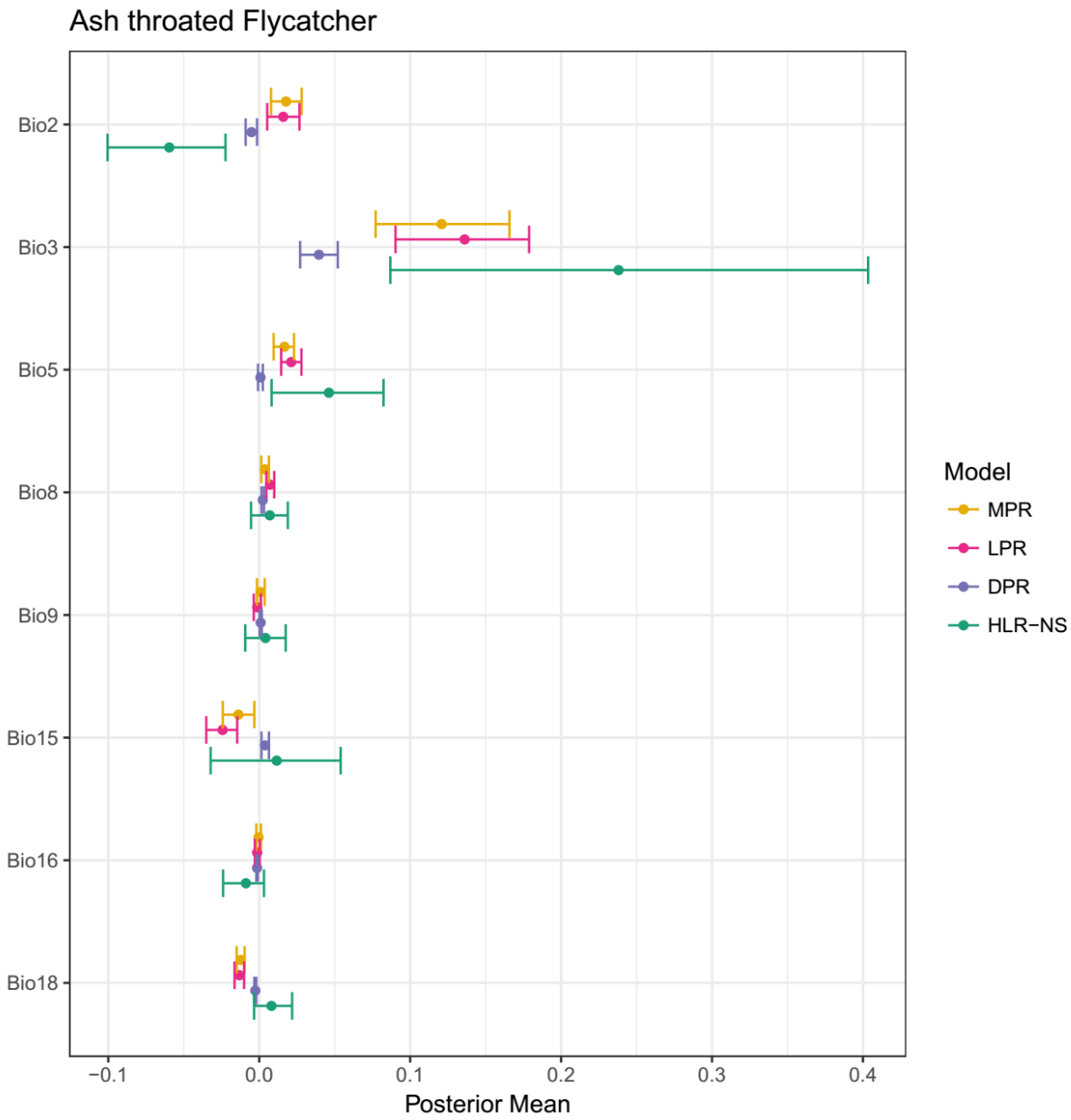


Figure 26: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ash-throated flycatcher, *Myiarchus cinerascens*.

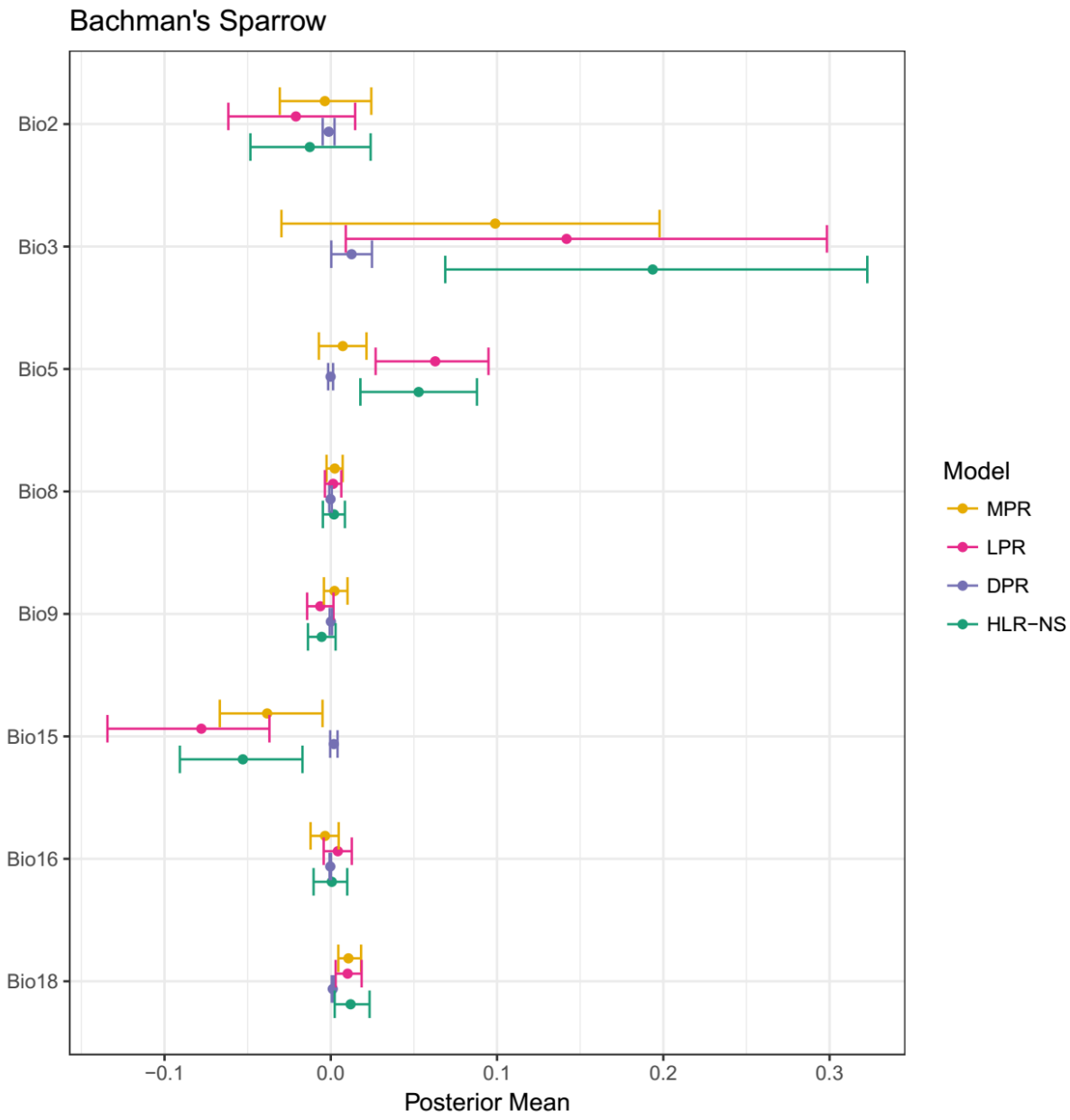


Figure 27: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Bachmann's sparrow, *Peucaea aestivalis*.

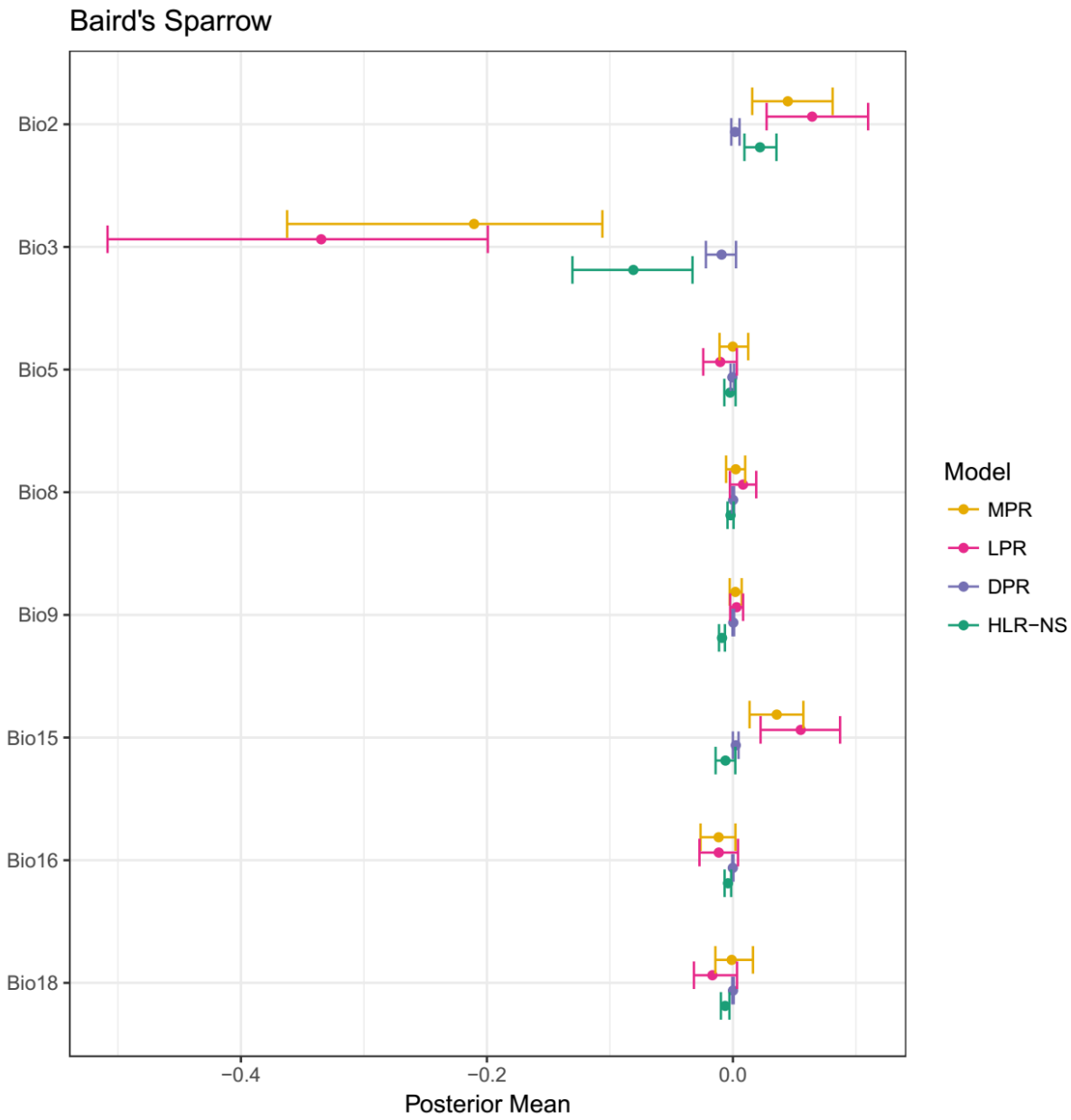


Figure 28: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Baird's sparrow, *Ammodramus bairdii*.

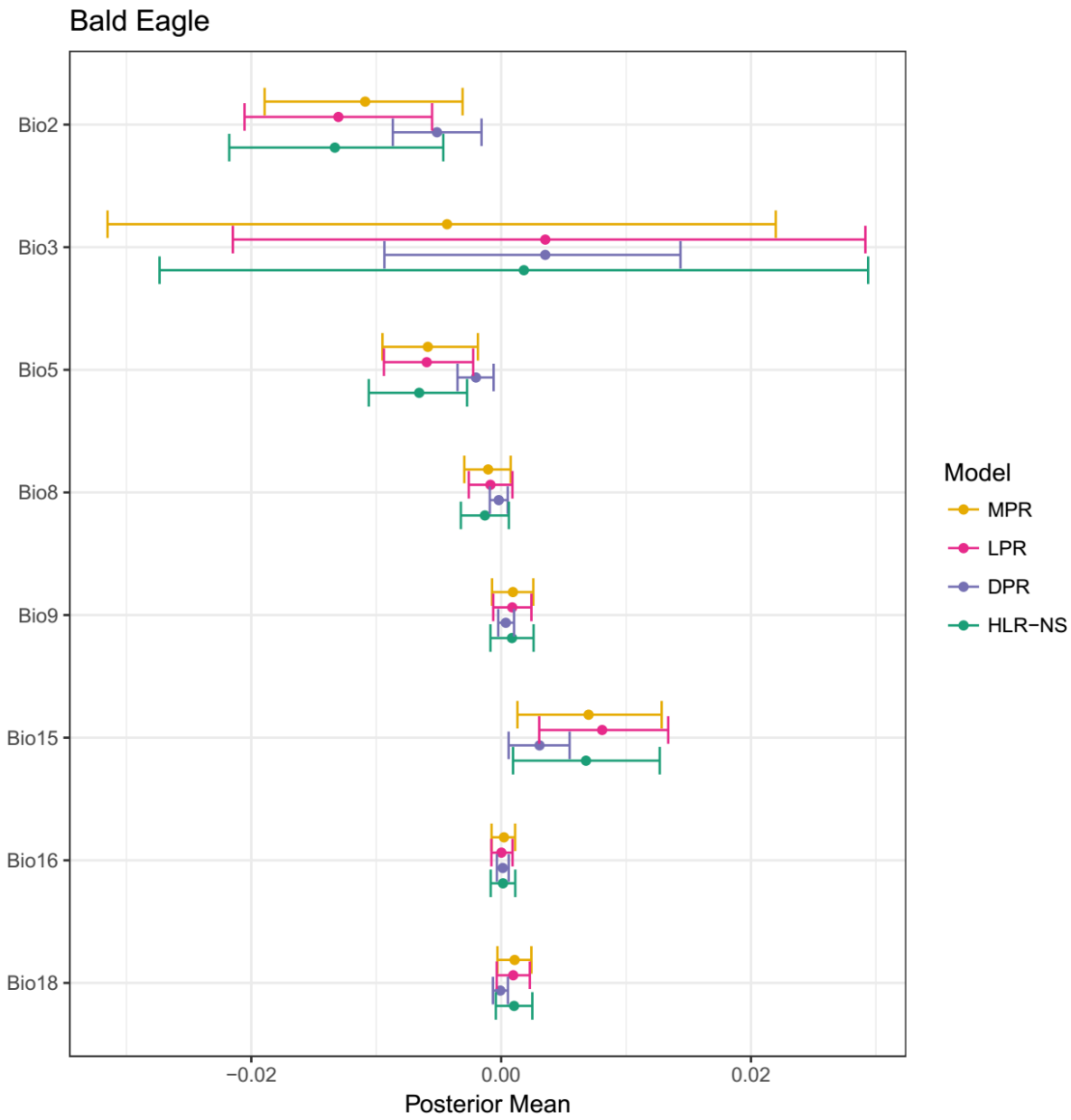


Figure 29: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bald eagle, *Haliaeetus leucocephalus*.

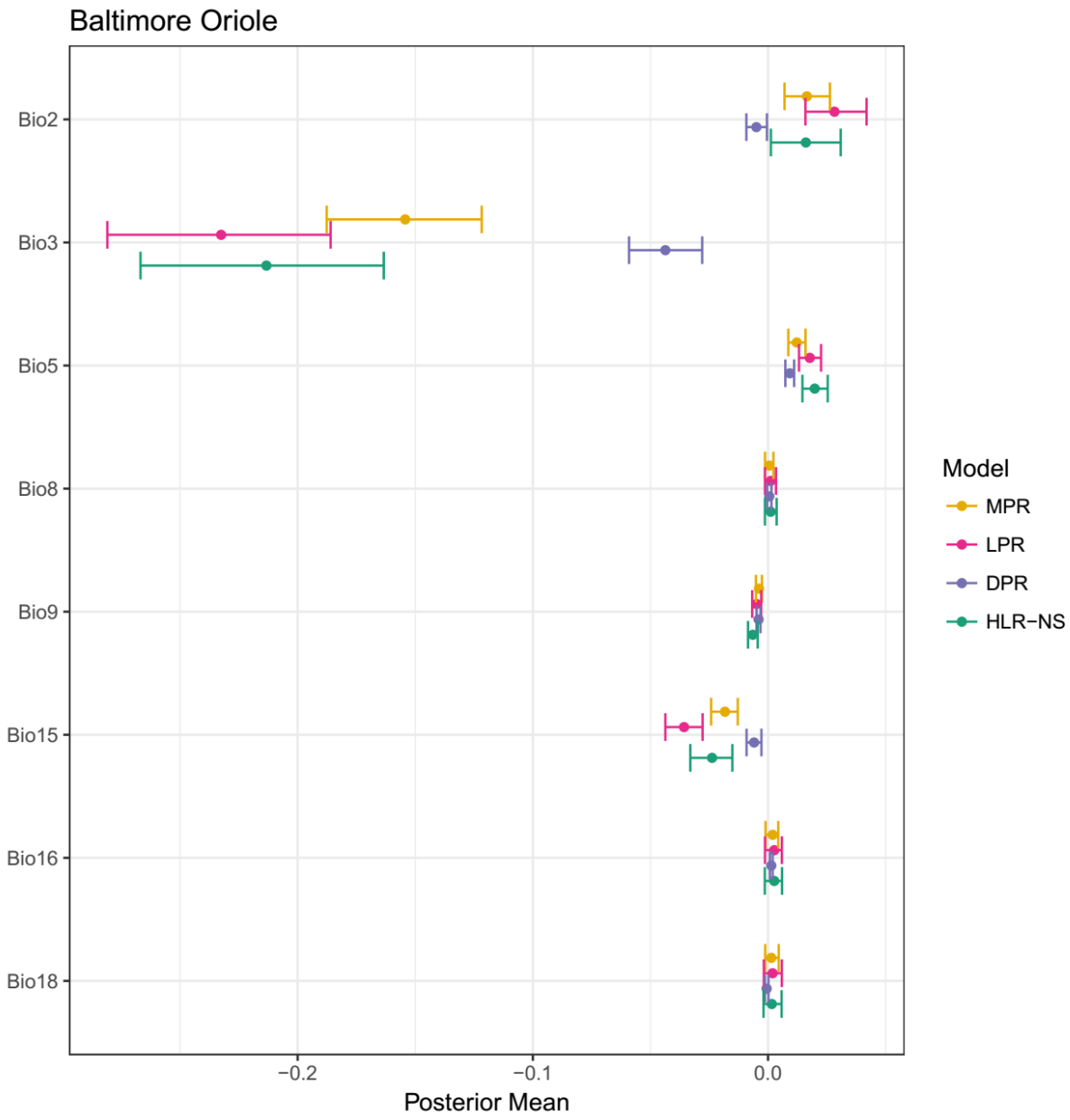


Figure 30: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Baltimore oriole, *Icterus galbula*.

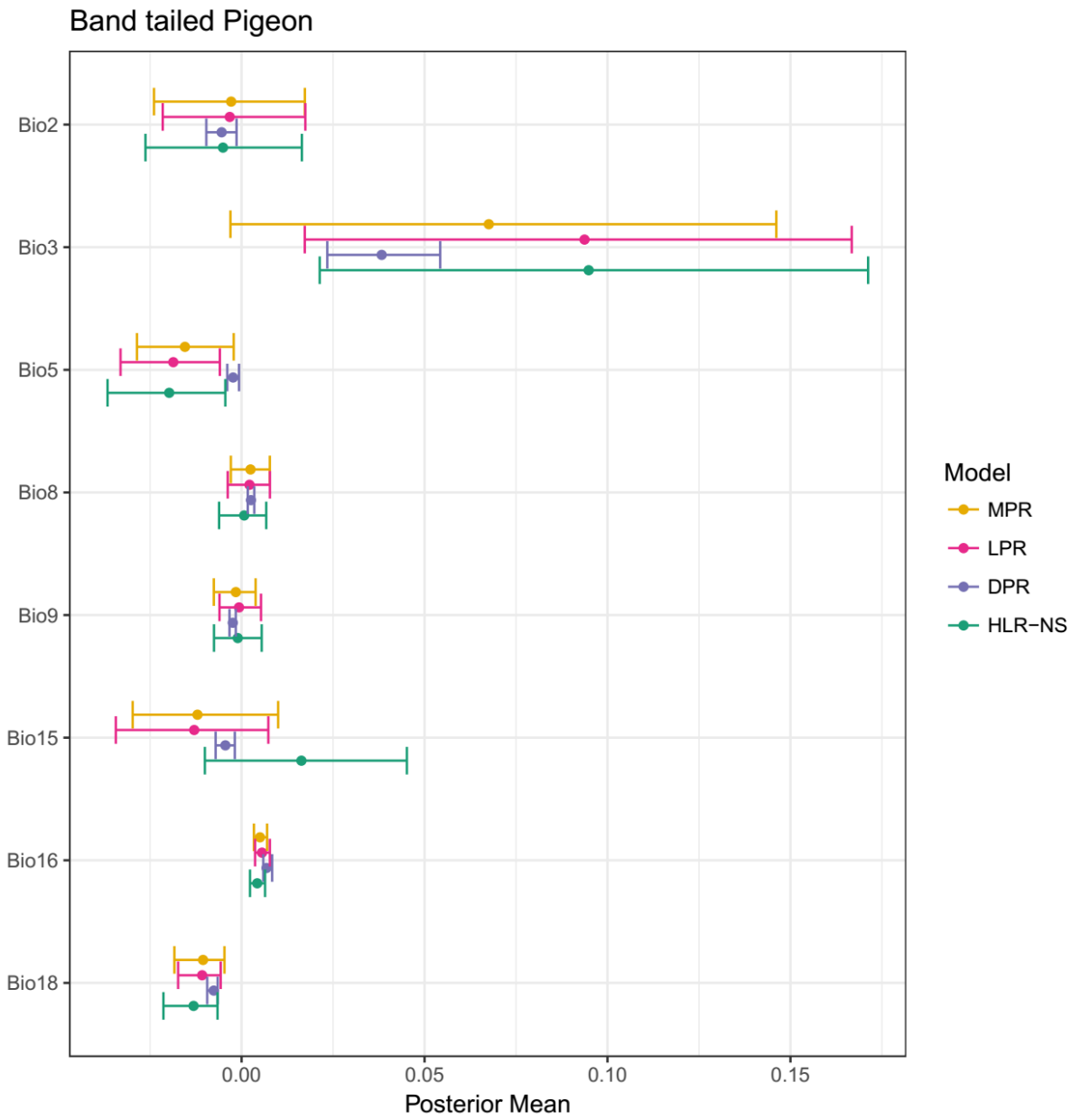


Figure 31: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Band-tailed pigeon, *Patagioenas fasciata*.

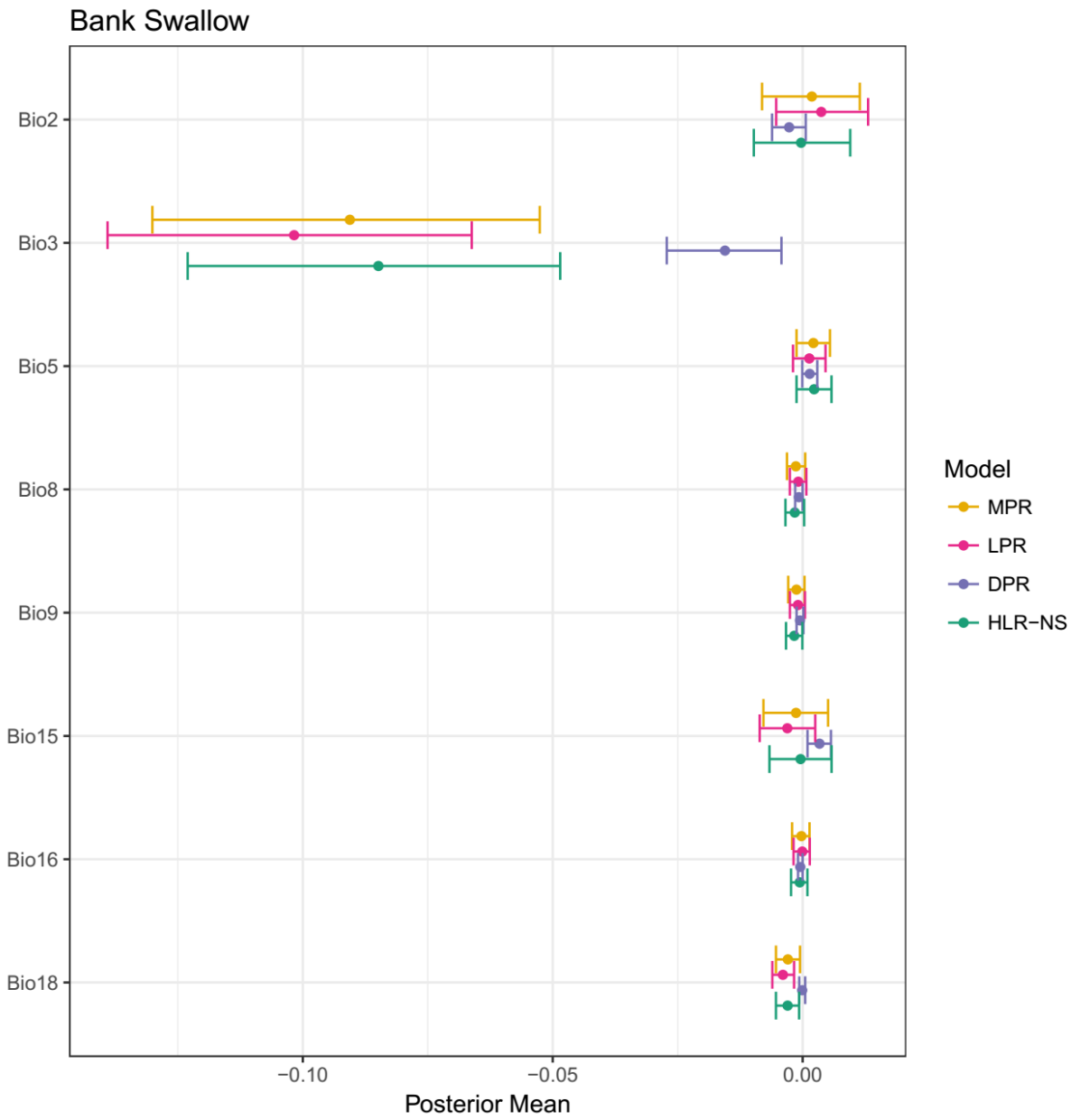


Figure 32: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bank swallow, *Riparia riparia*.

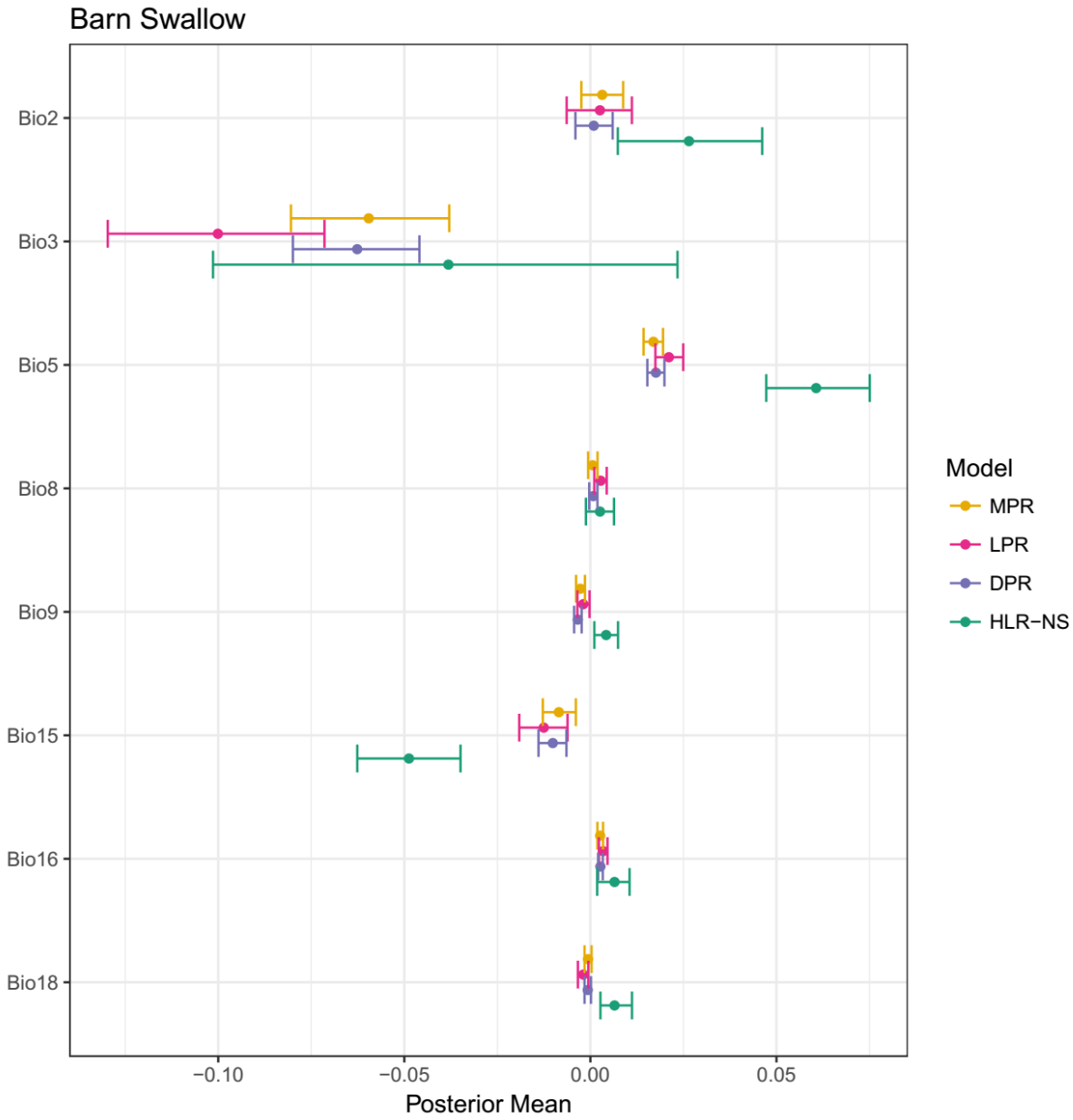


Figure 33: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the barn swallow, *Hirundo rustica*.

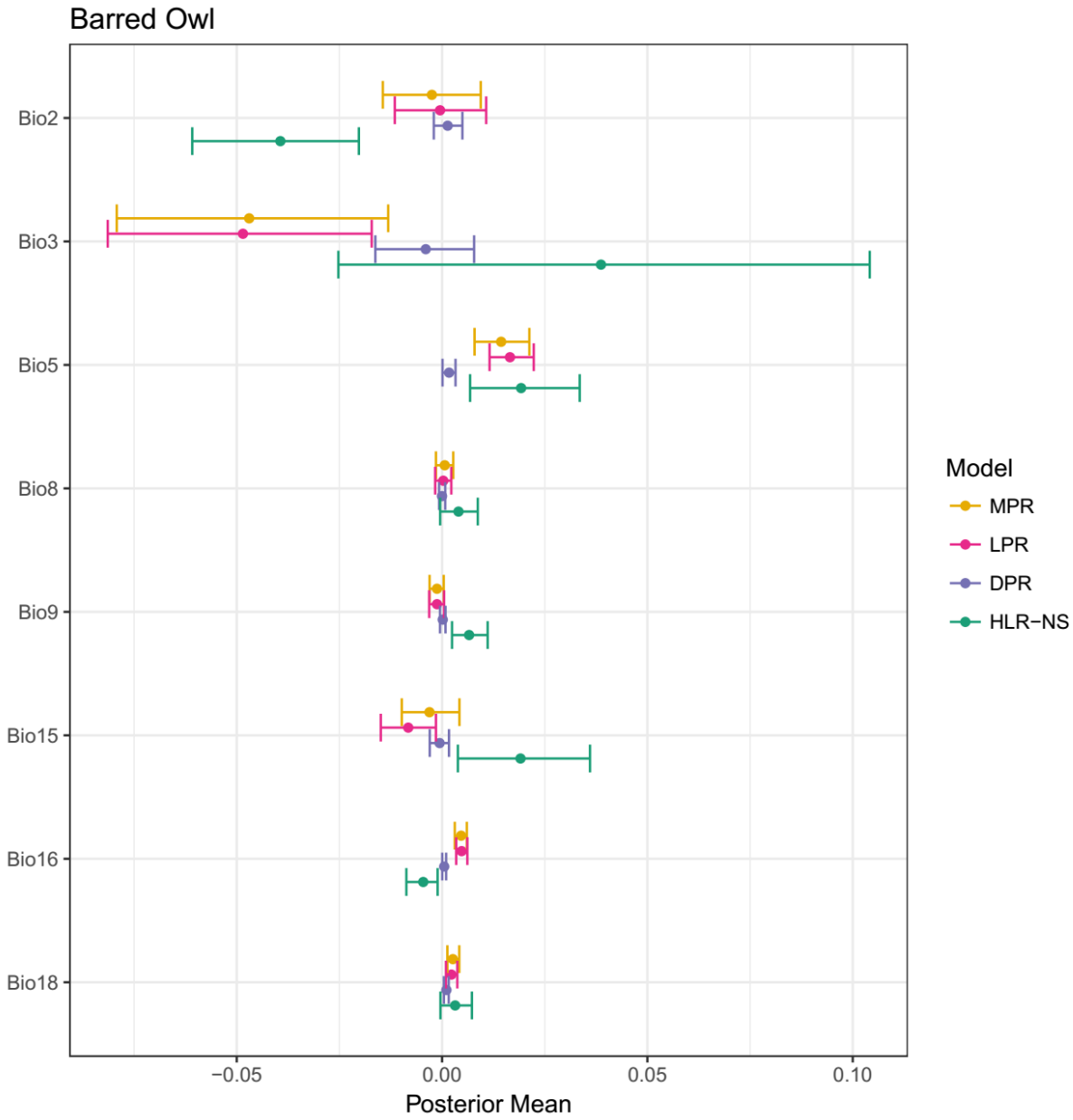


Figure 34: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the barred owl, *Strix varia*.

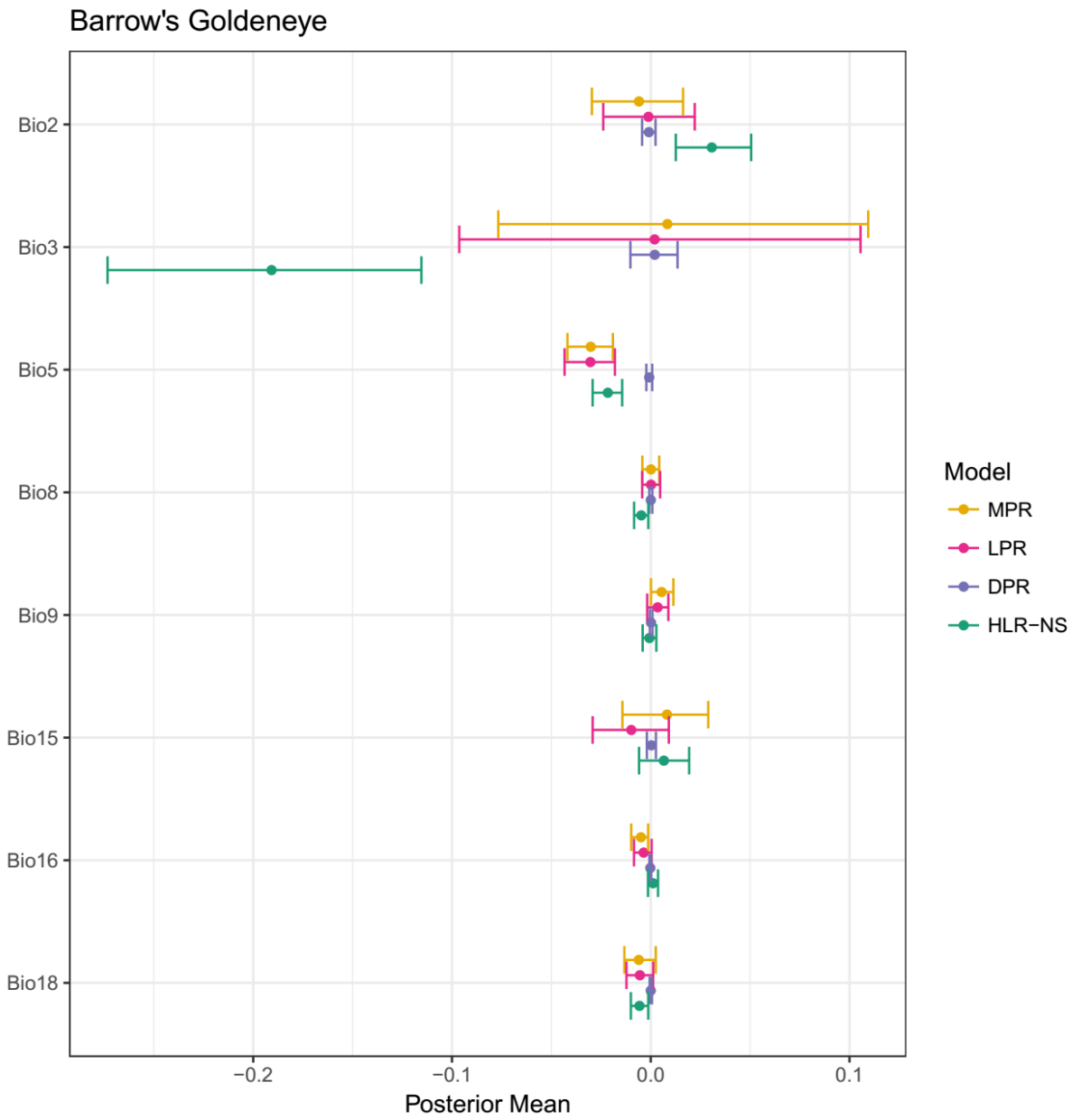


Figure 35: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Barrow's goldeneye, *Bucephala islandica*.

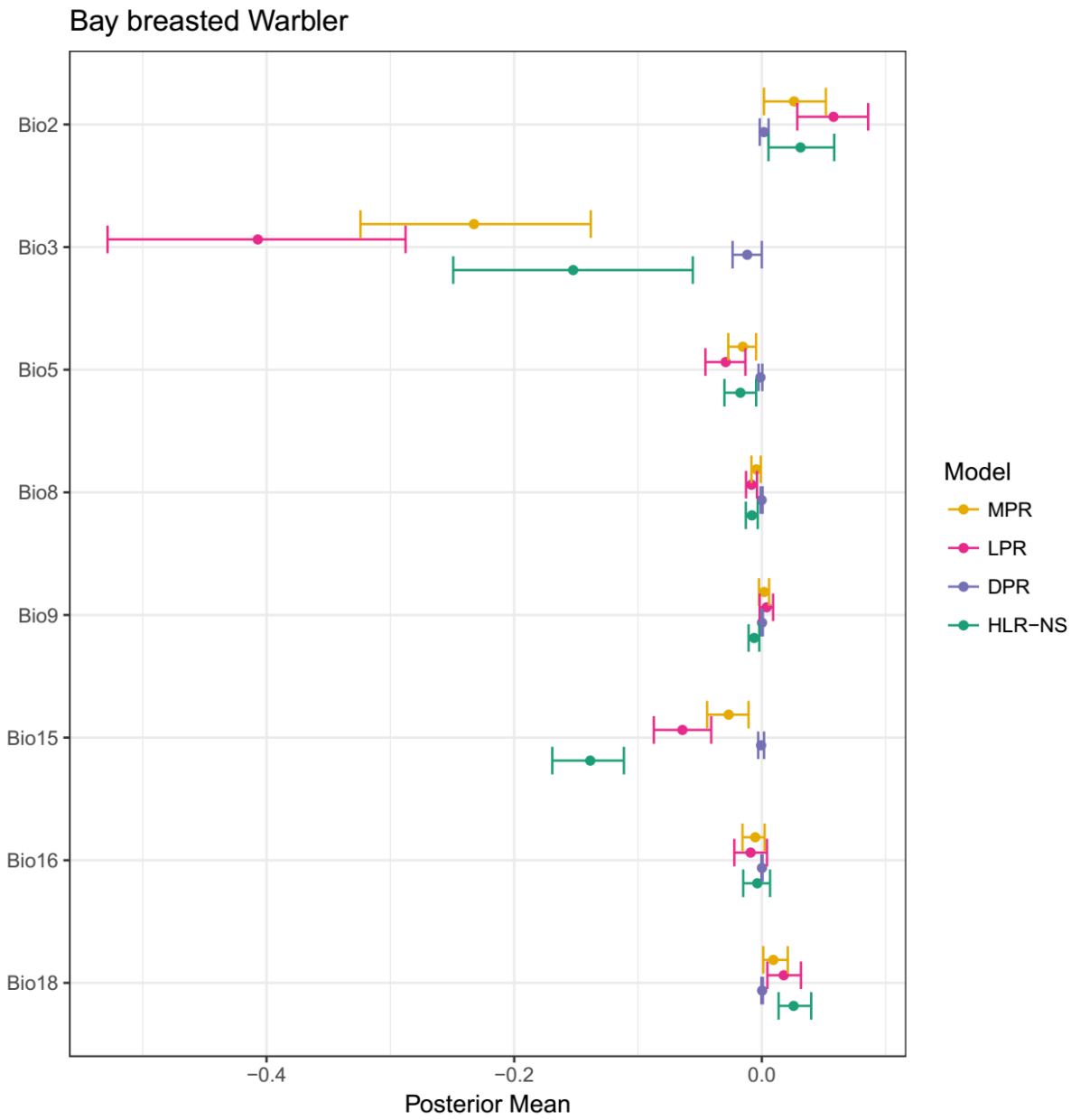


Figure 36: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bay-breasted warbler, *Setophaga castanea*.

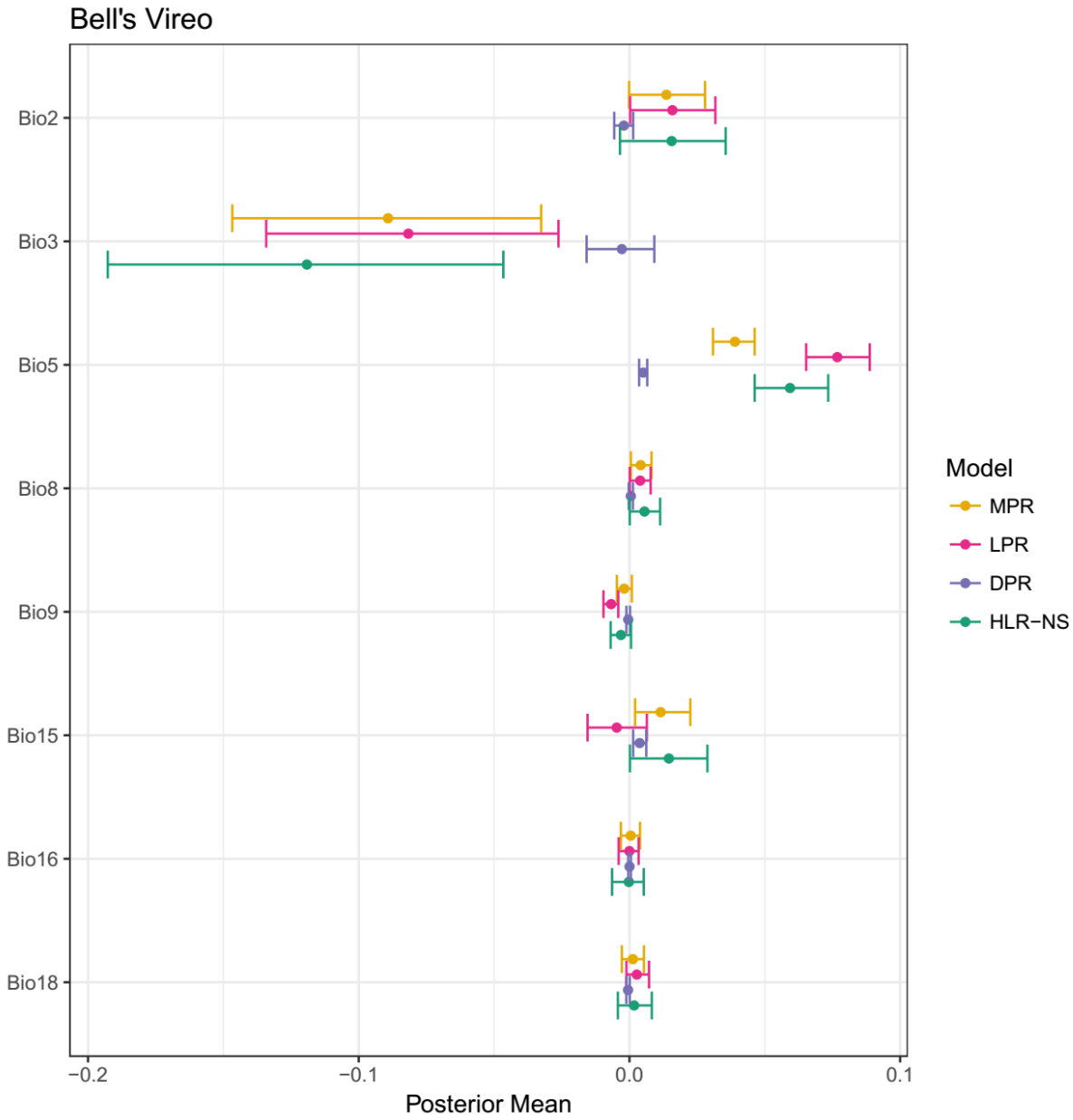


Figure 37: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Bell's vireo, *Vireo bellii*.

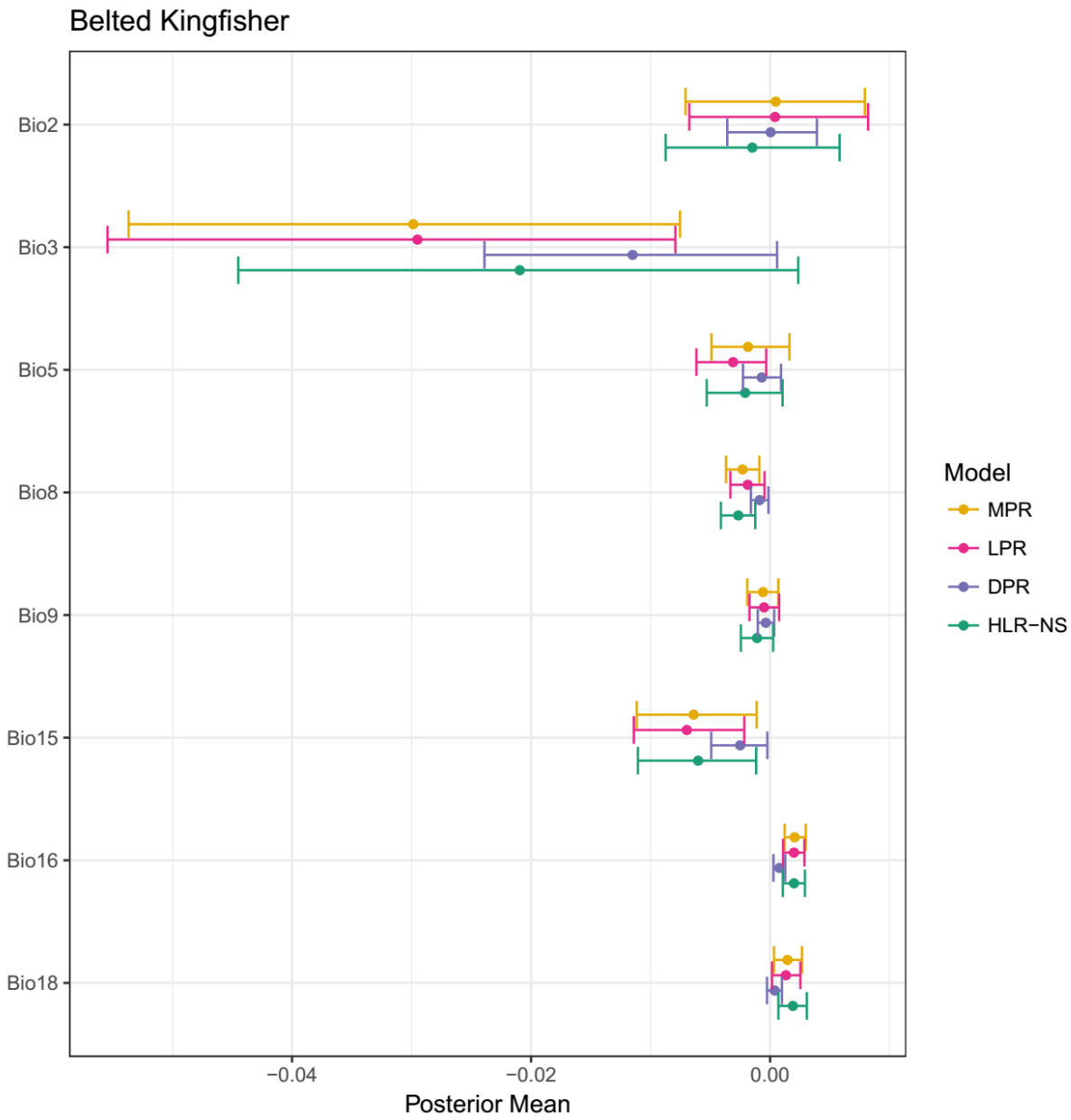


Figure 38: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the belted kingfisher, *Megaceryle alcyon*.

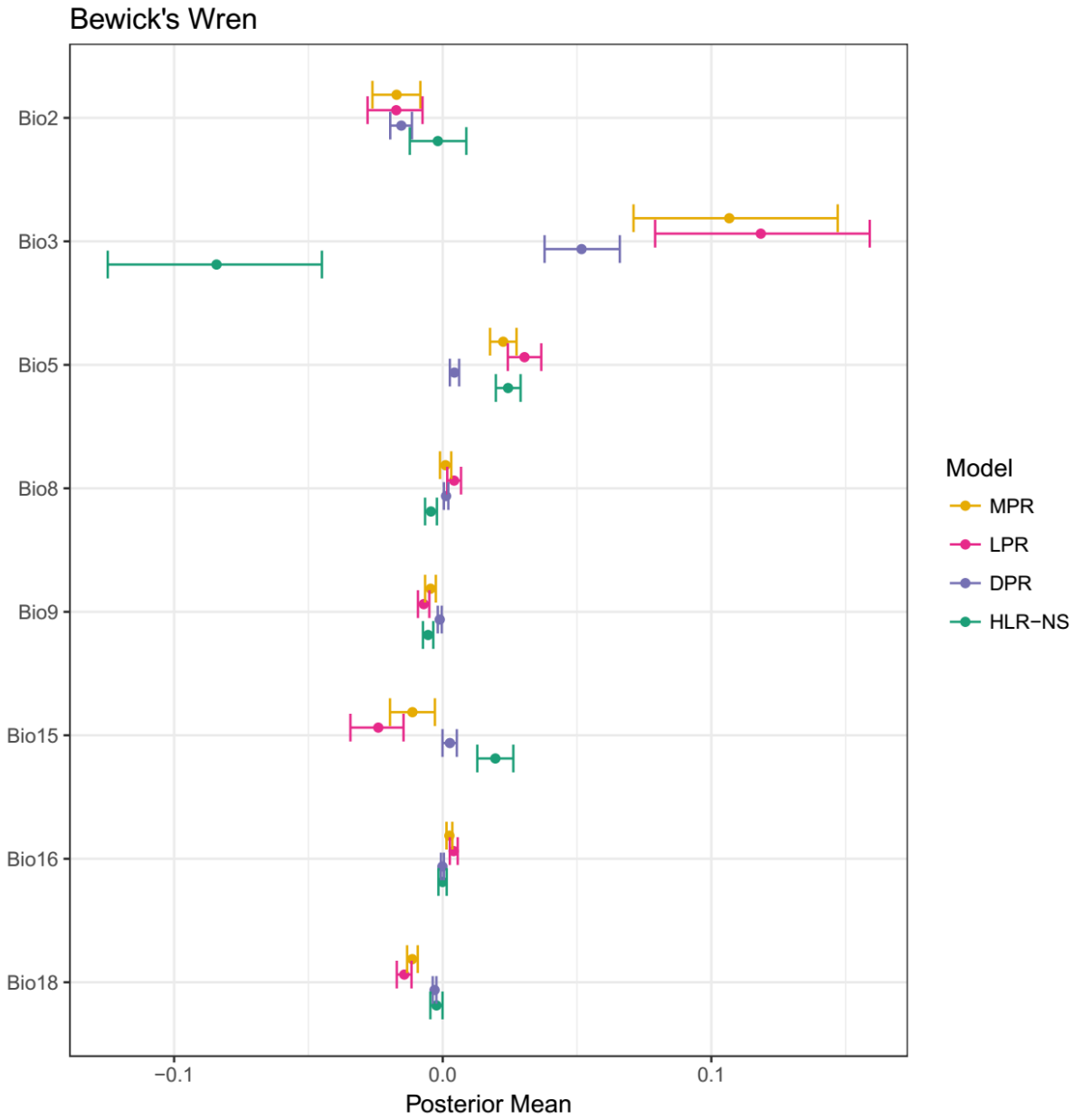


Figure 39: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Bewick's wren, *Thryomanes bewickii*.

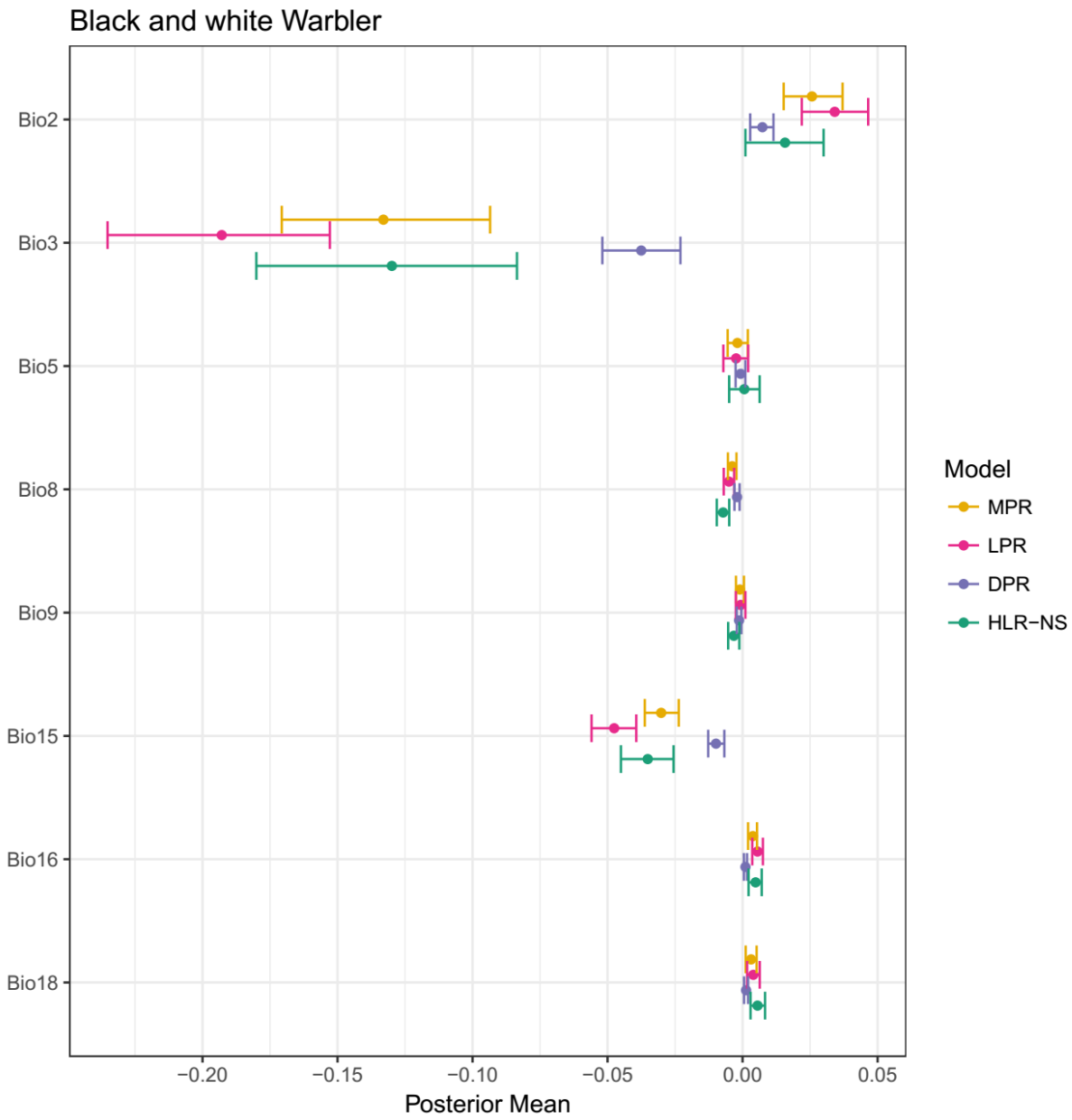


Figure 40: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-and-white warbler, *Mniotilta varia*.

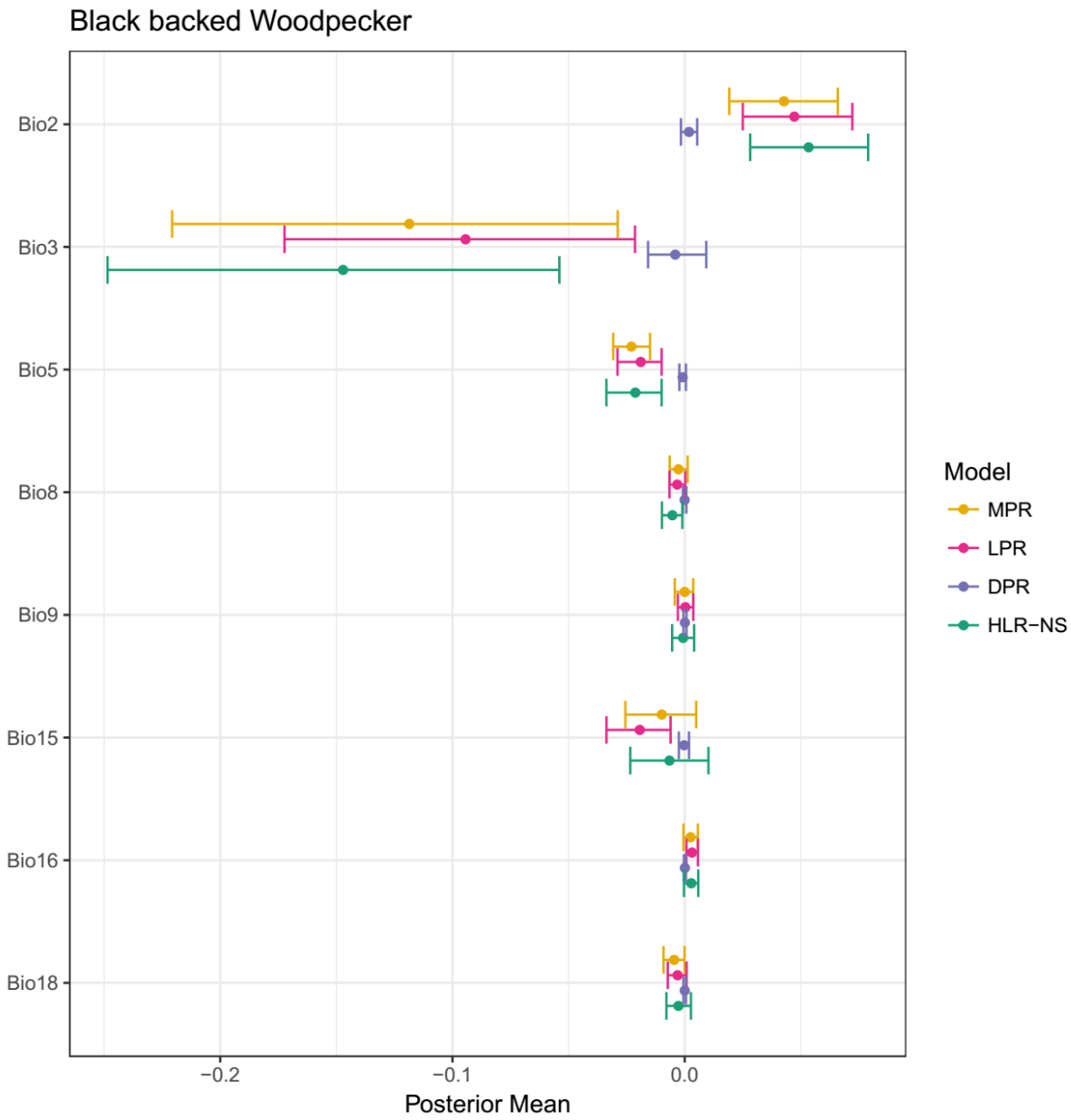


Figure 41: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-backed woodpecker, *Picoides arcticus*.

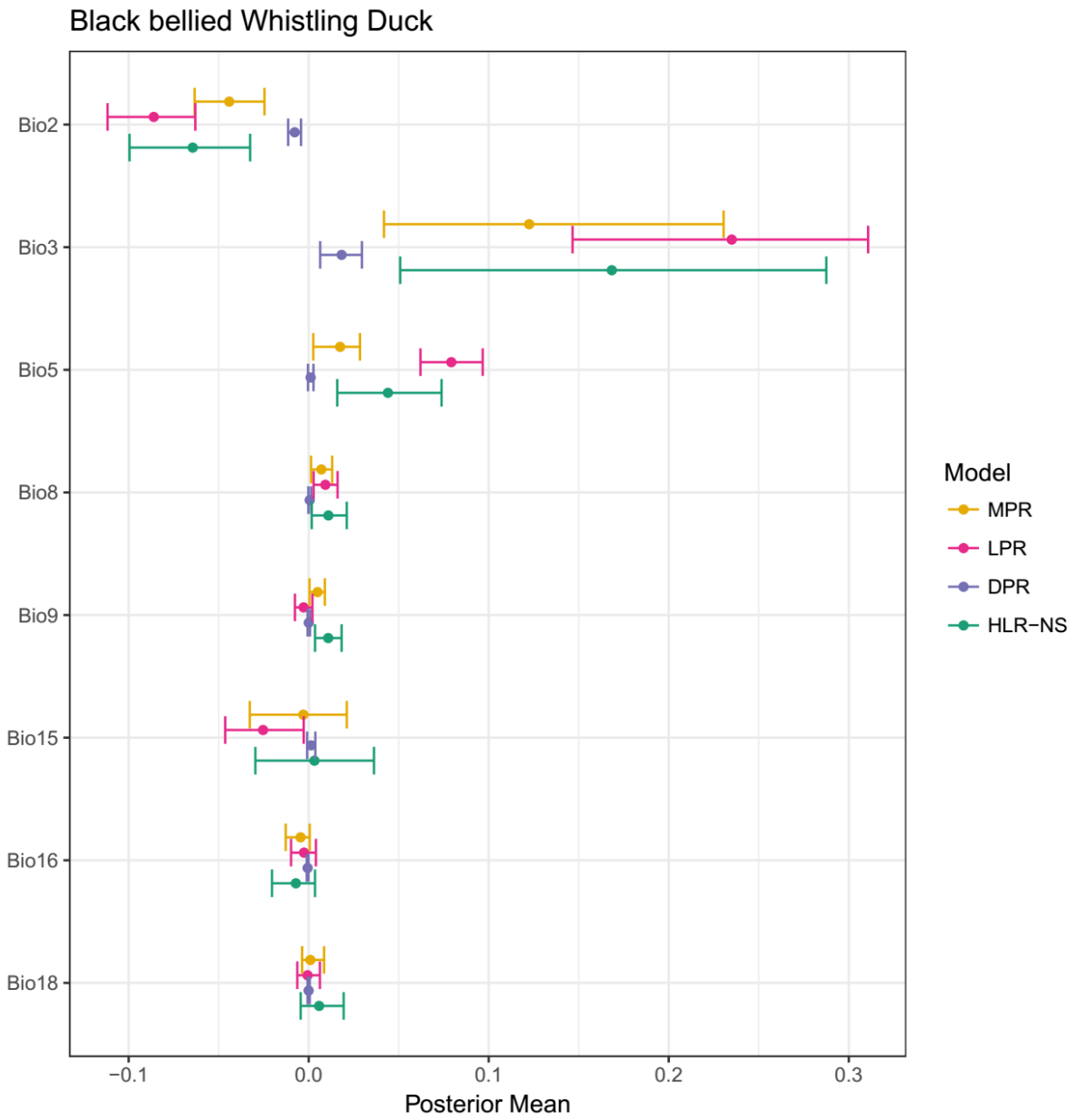


Figure 42: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-bellied whistling duck, *Dendrocygna autumnalis*.

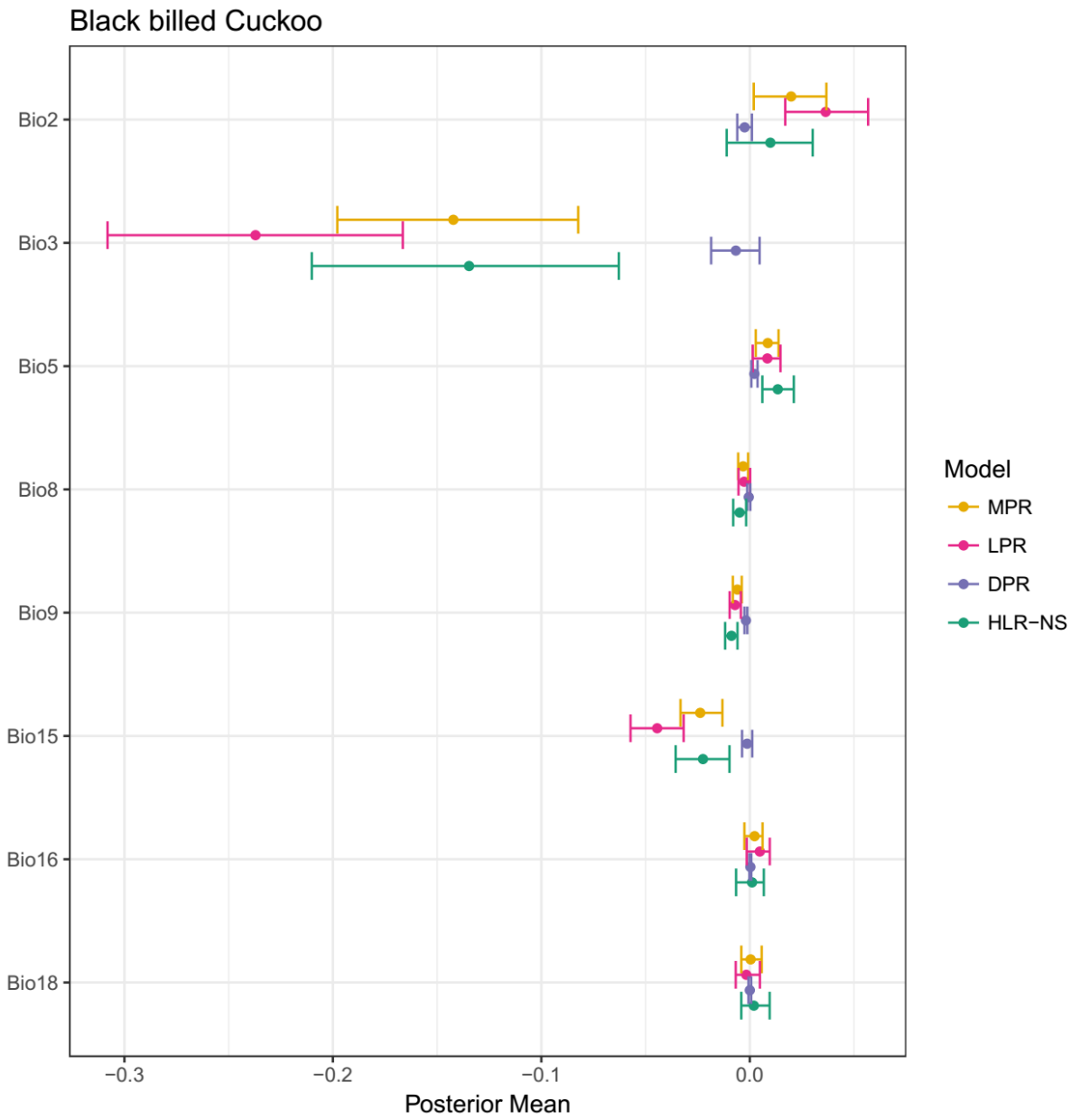


Figure 43: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-billed cuckoo, *Coccyzus erythropthalmus*.

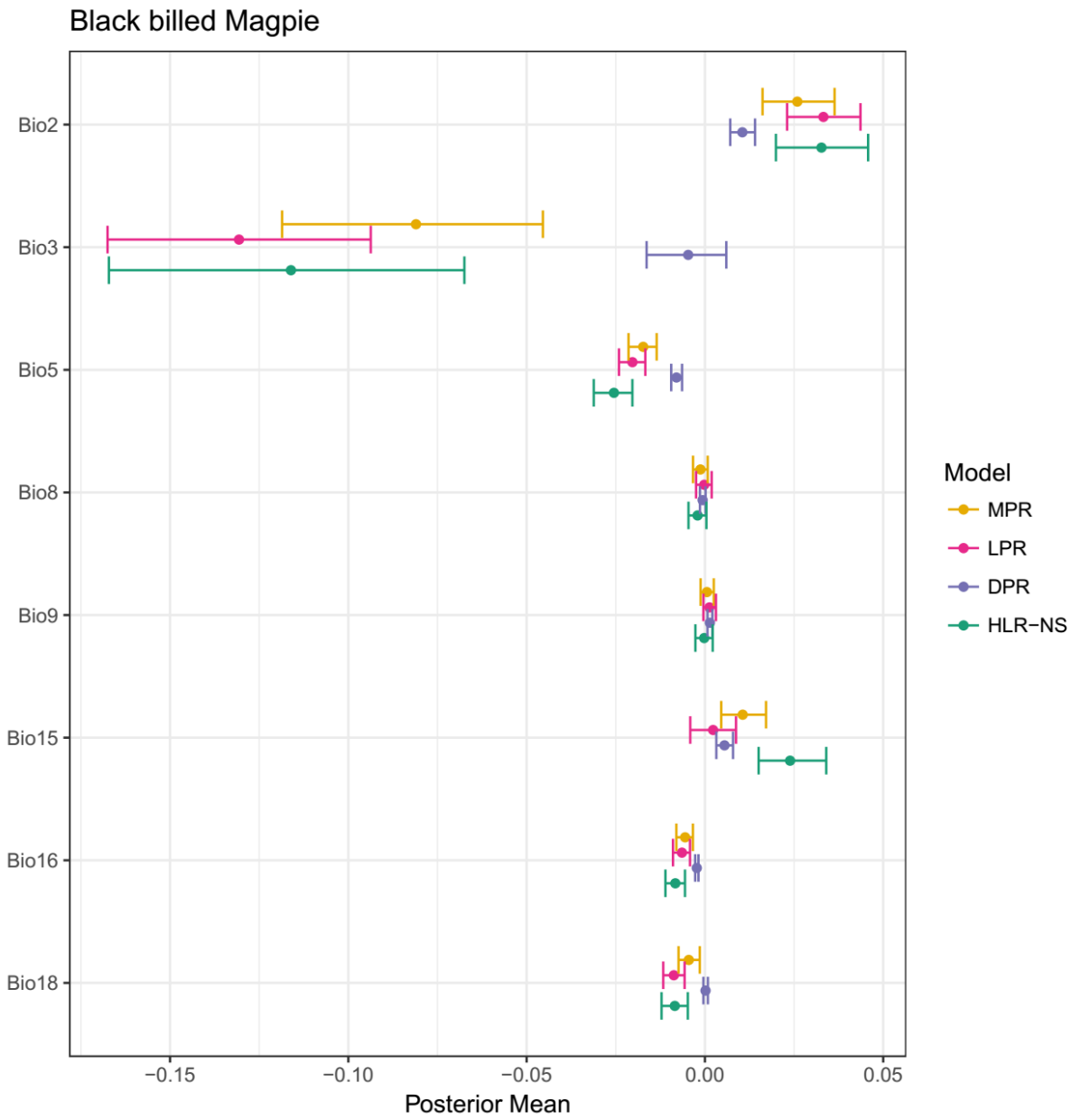


Figure 44: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-billed magpie, *Pica hudsonia*.

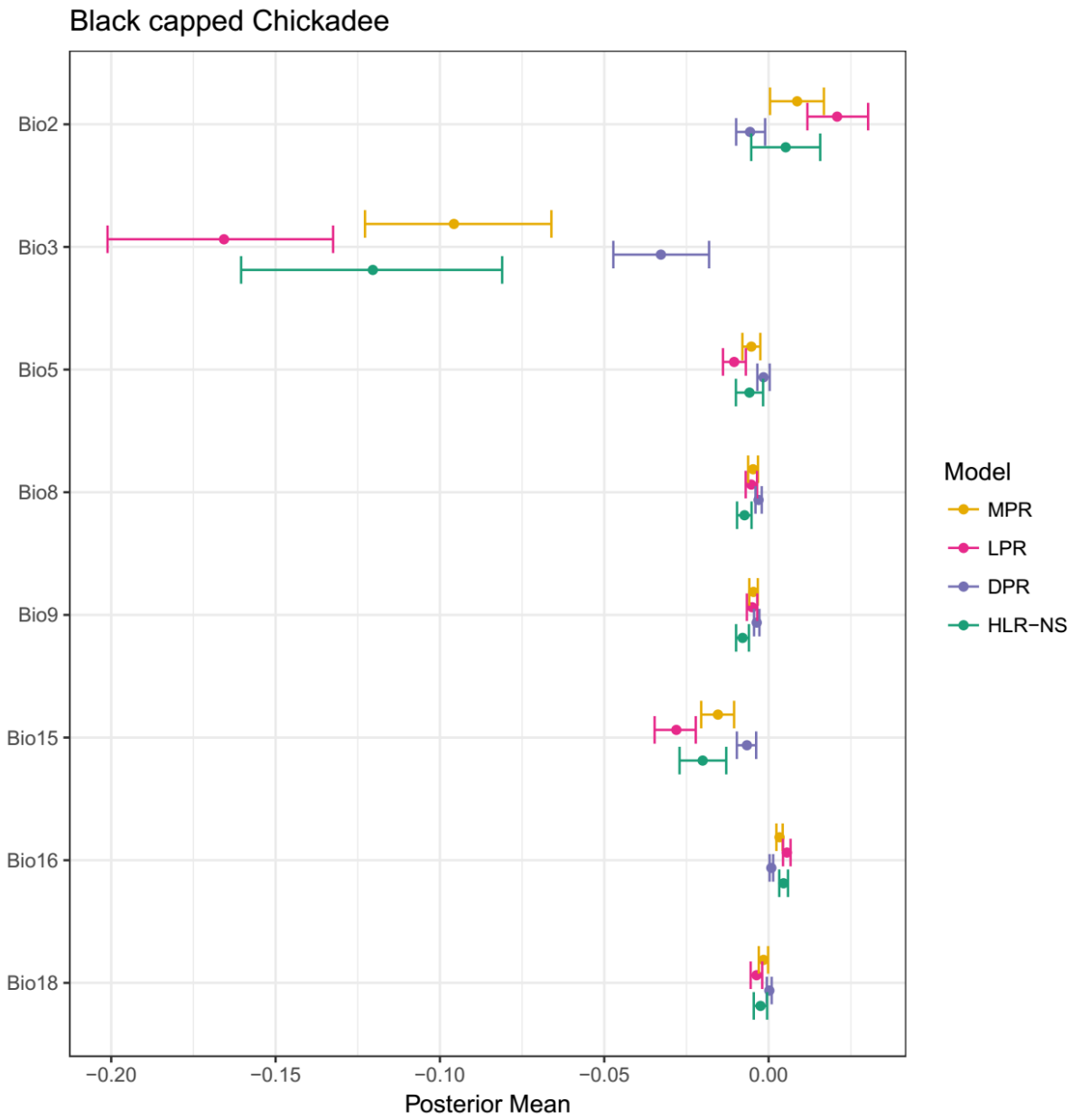


Figure 45: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-capped chicadee, *Poecile atricapillus*.

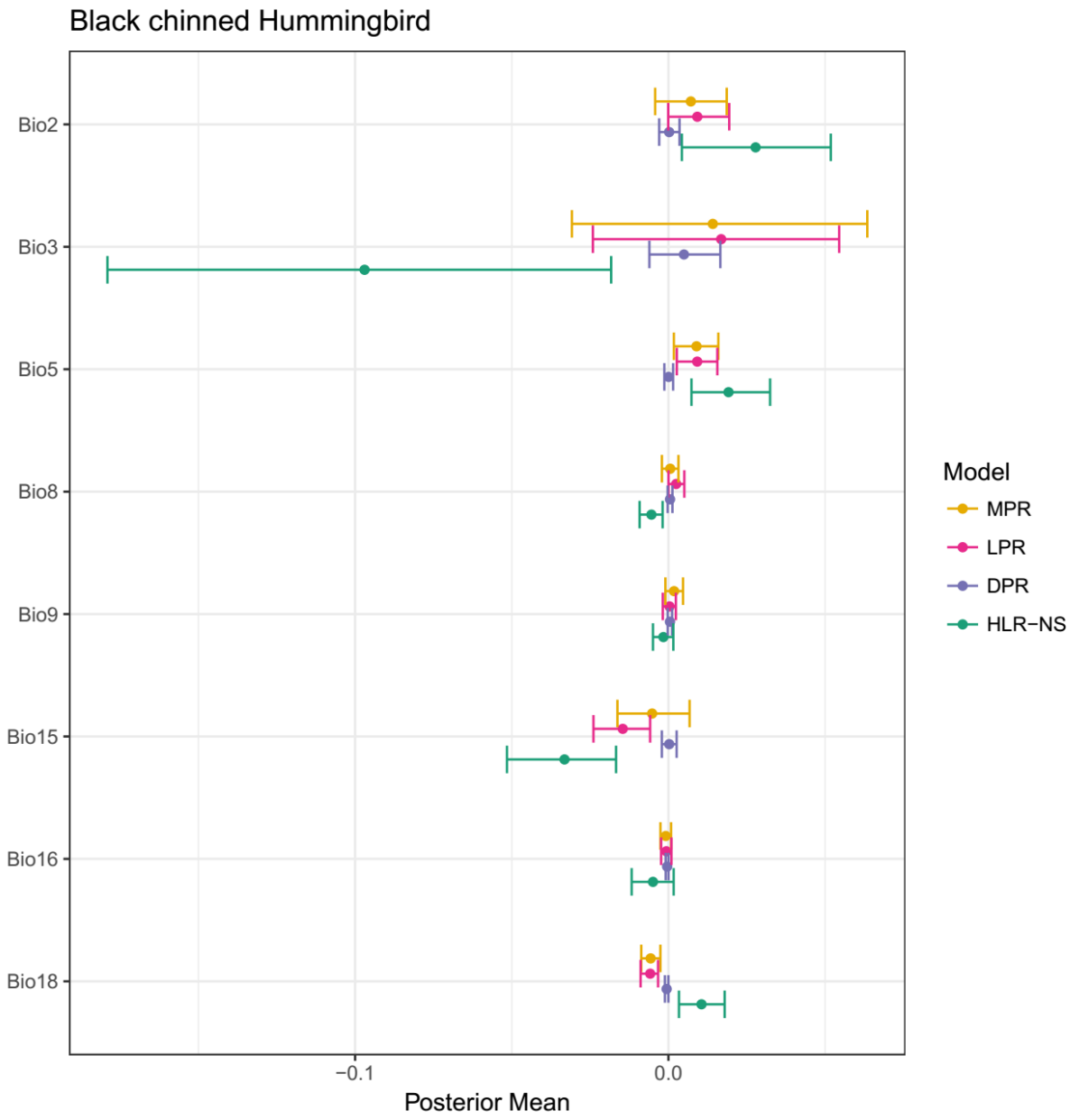


Figure 46: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-chinned hummingbird, *Archilochus alexandri*.

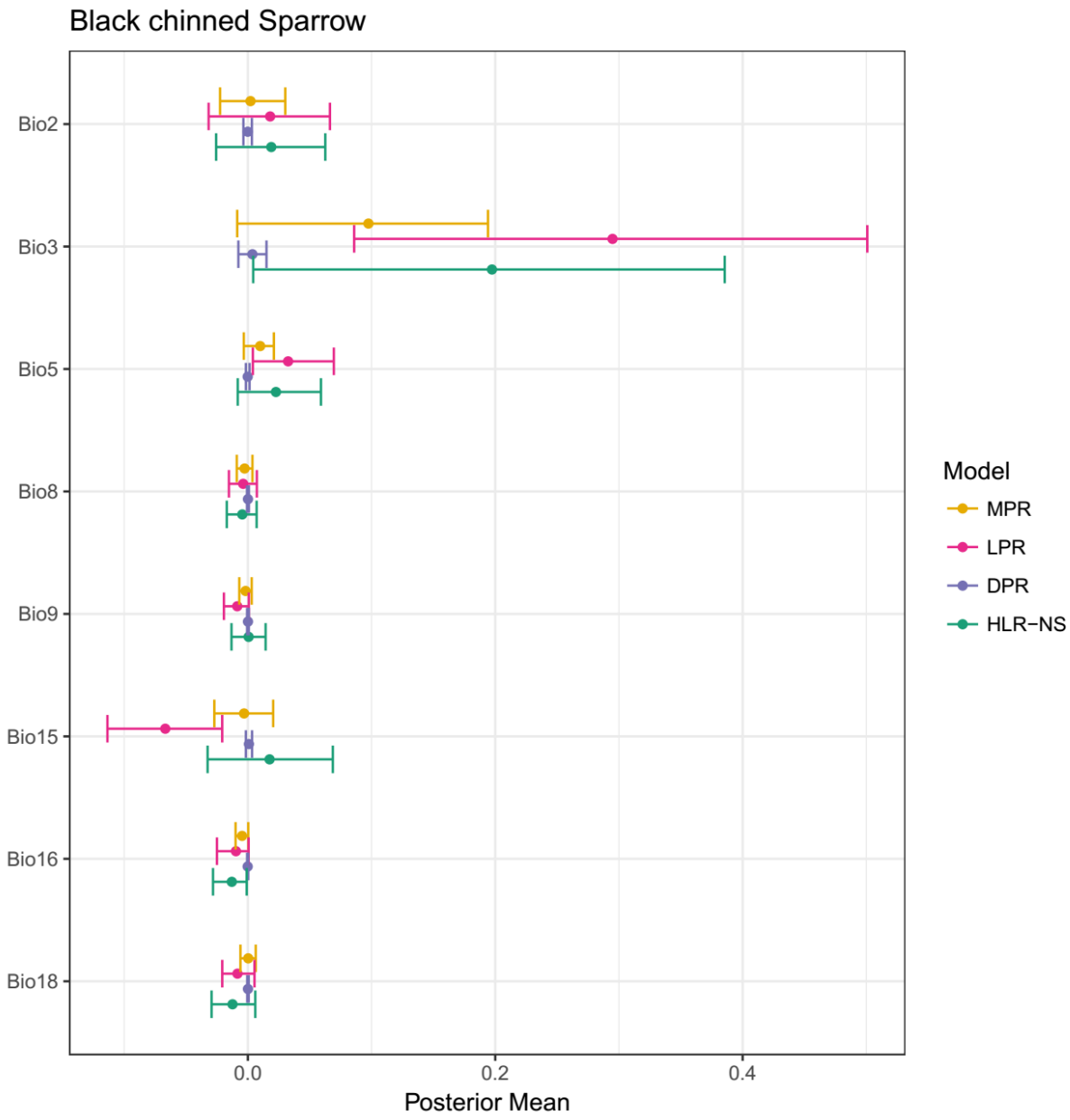


Figure 47: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-chinned sparrow, *Spizella atrogularis*.

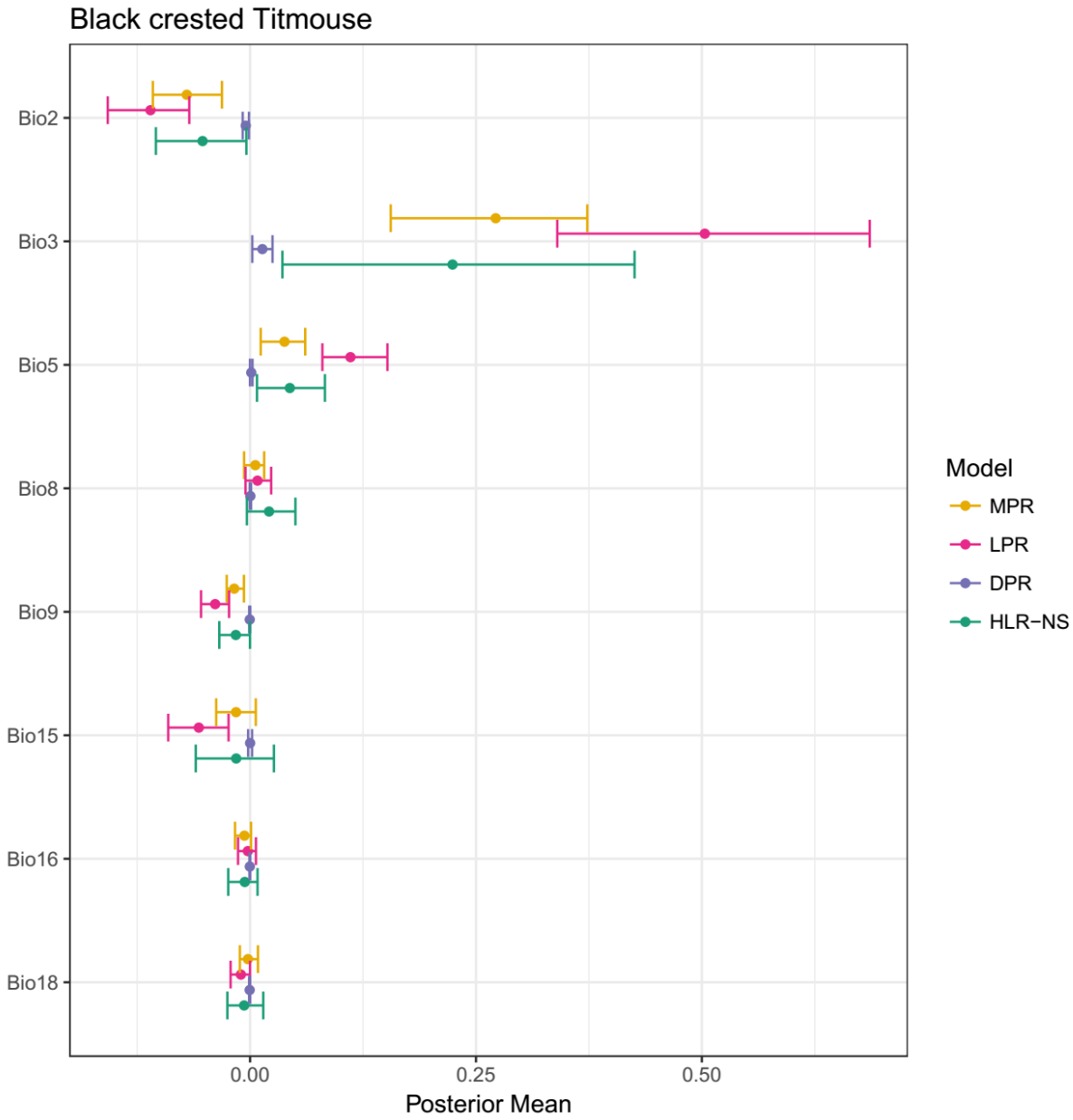


Figure 48: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-crested titmouse, *Baeolophus atricristatus*.

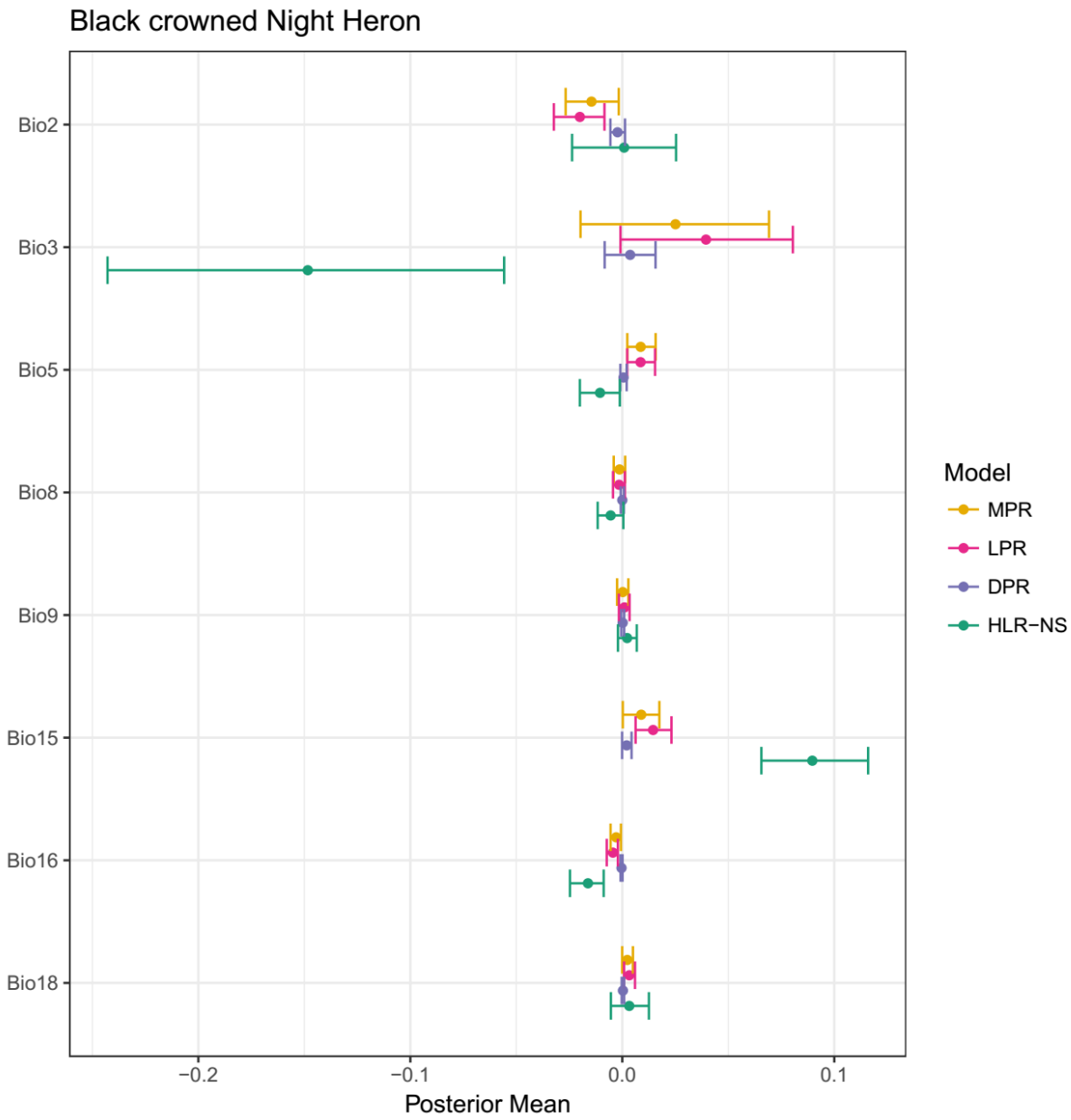


Figure 49: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-crowned night heron, *Nycticorax nycticorax*.

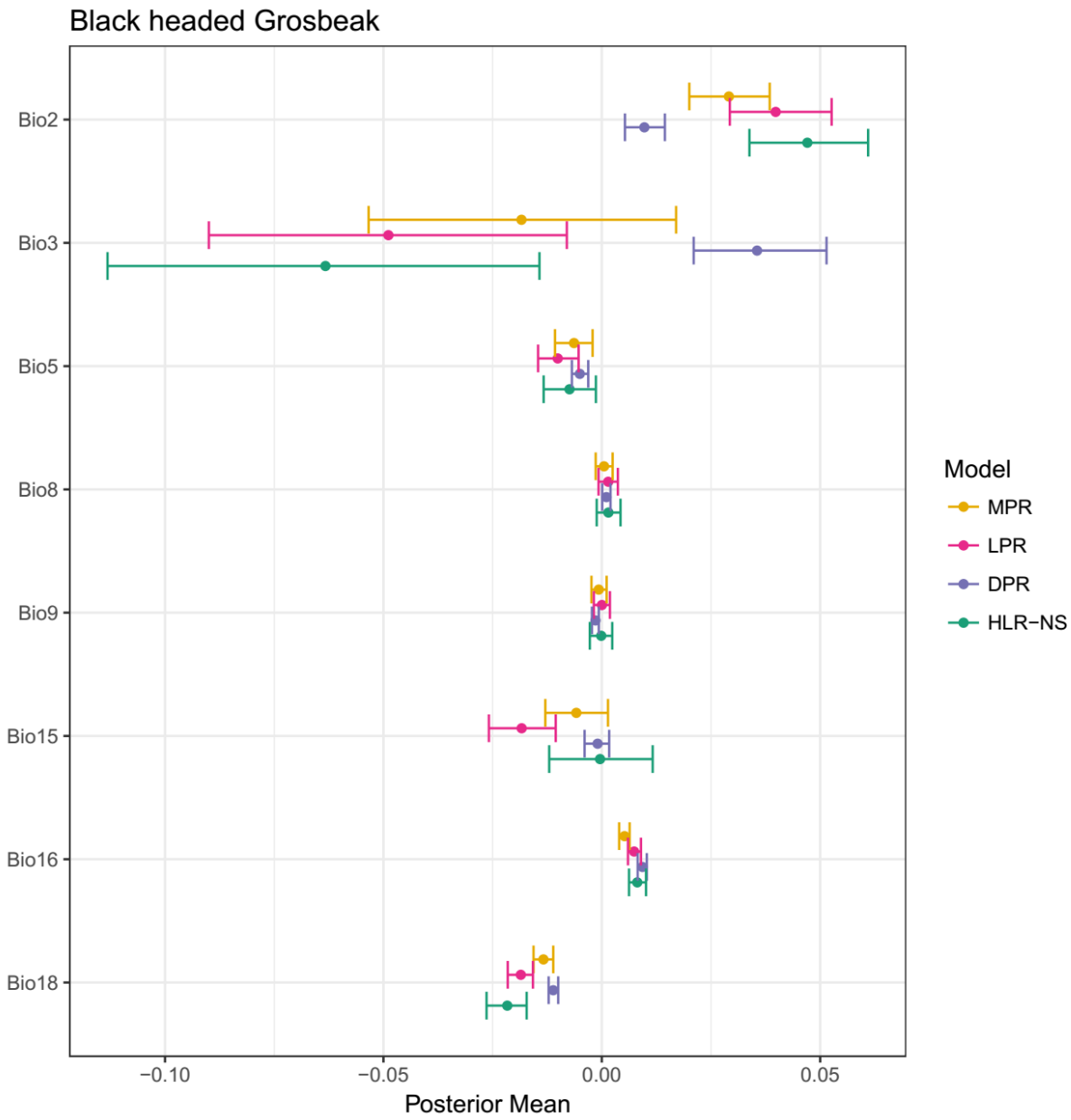


Figure 50: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-headed grosbeak, *Pheucticus melanocephalus*.

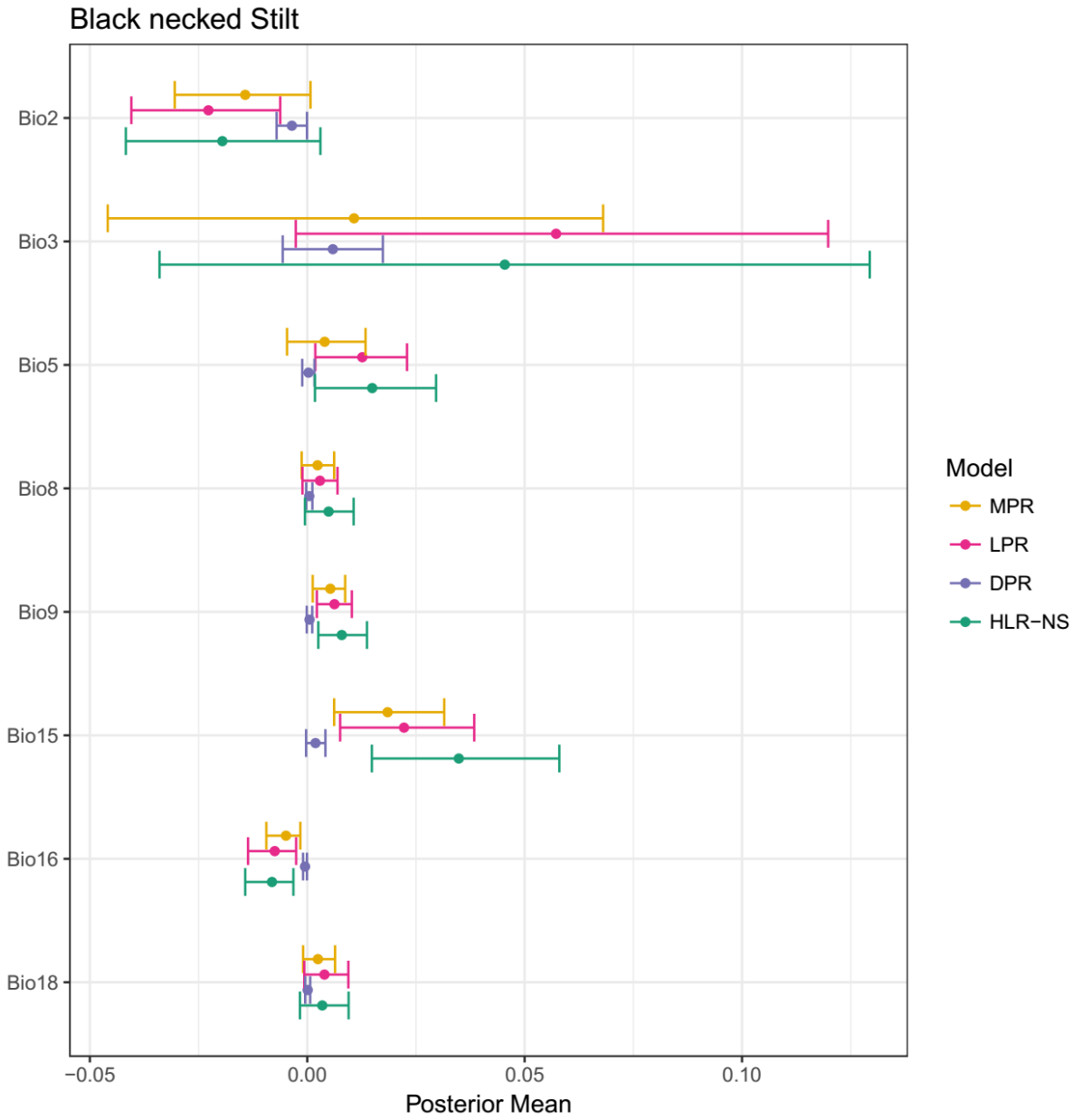


Figure 51: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-necked stilt, *Himantopus mexicanus*.

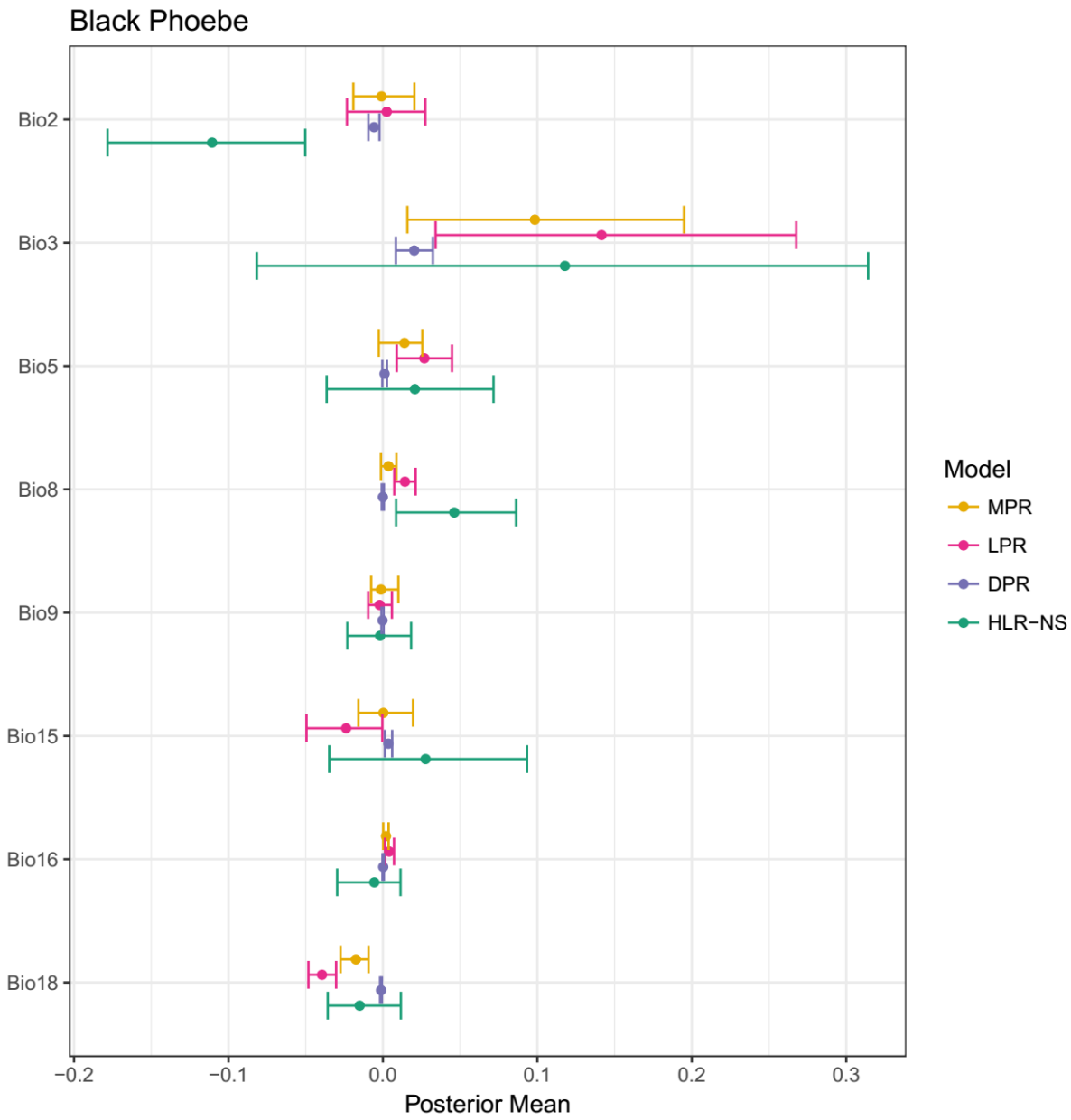


Figure 52: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black phoebe, *Sayornis nigricans*.

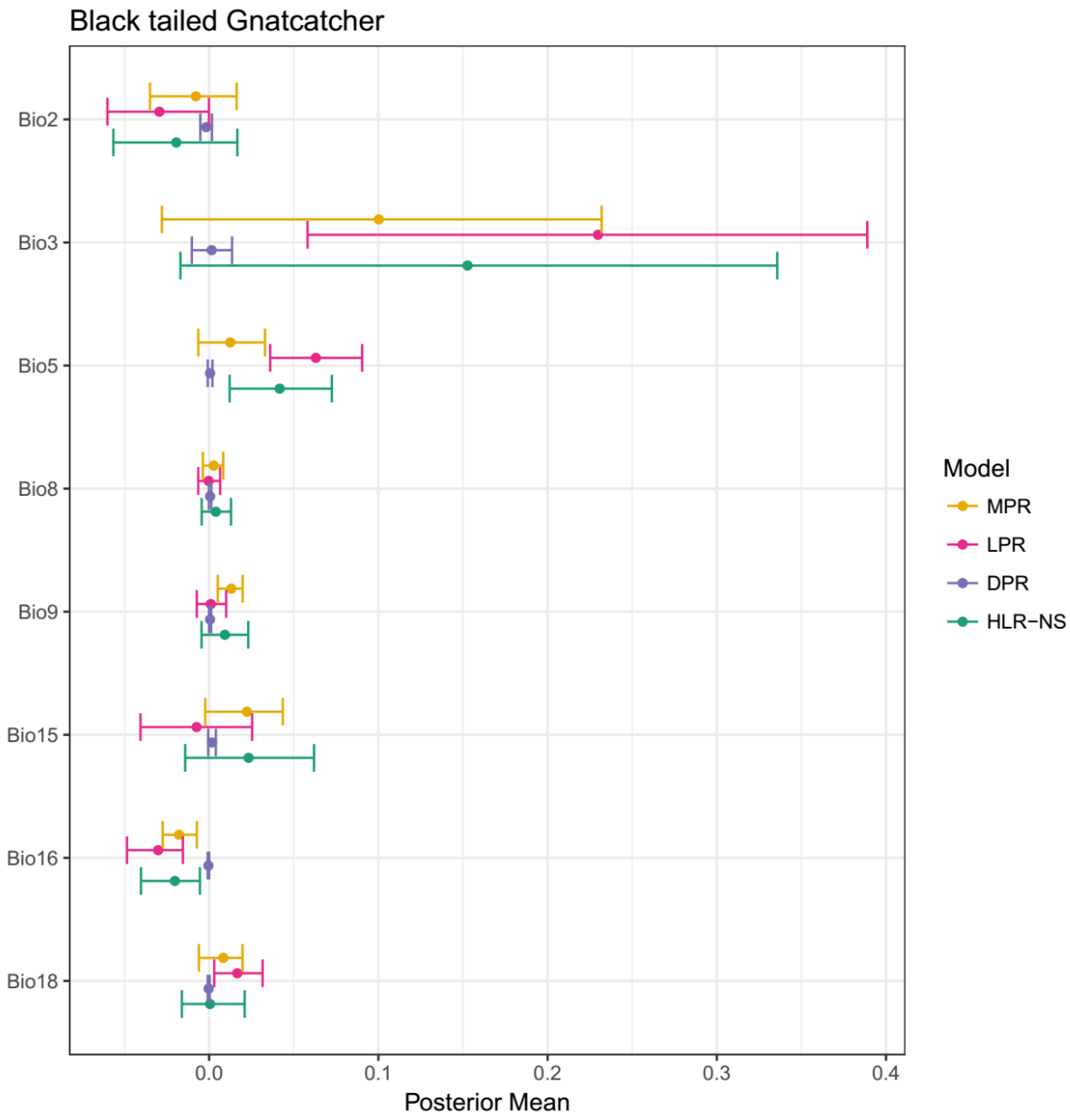


Figure 53: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-tailed gnatcatcher, *Polioptila melanura*.

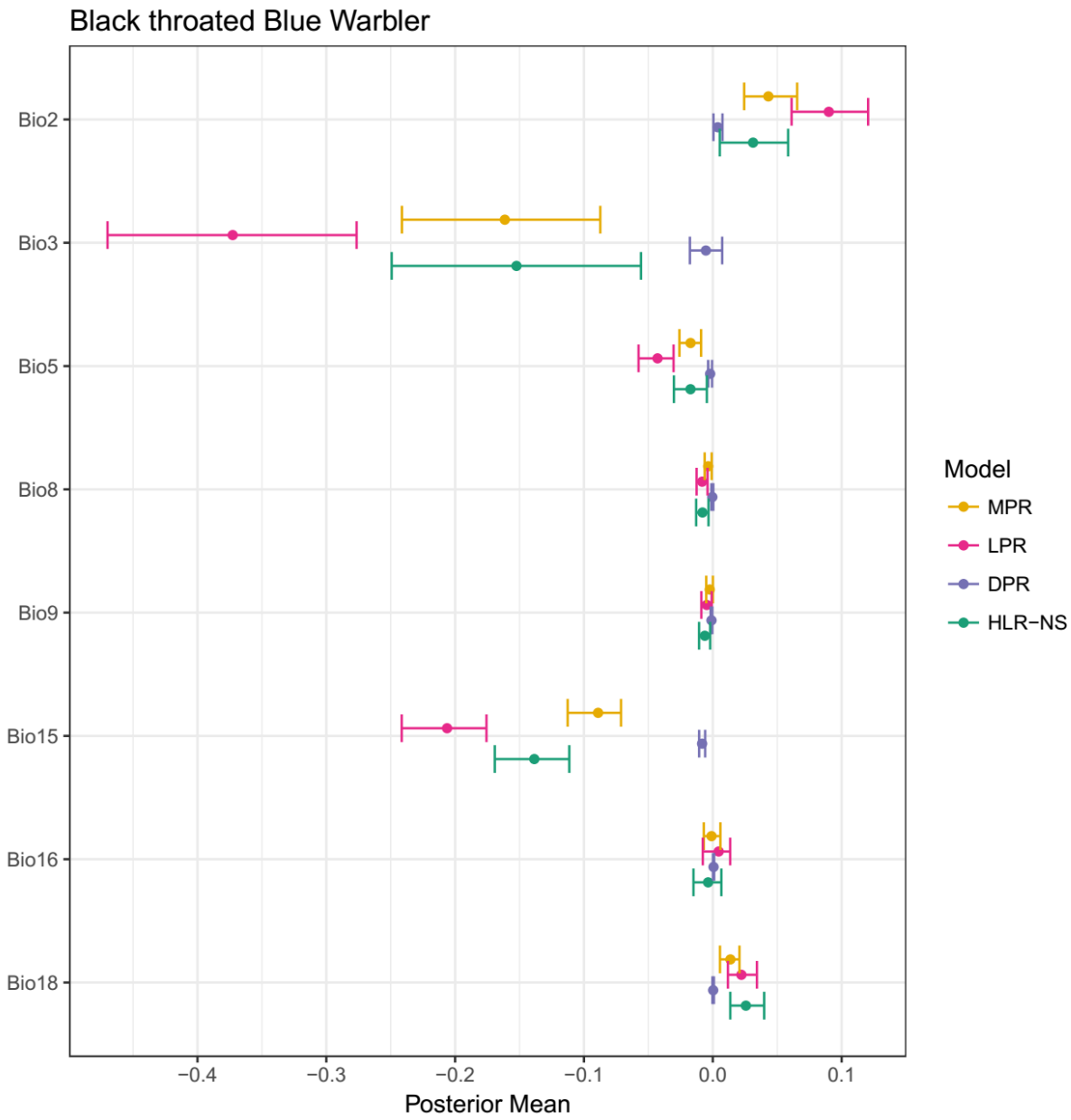


Figure 54: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-throated blue warbler, *Setophaga caerulescens*.

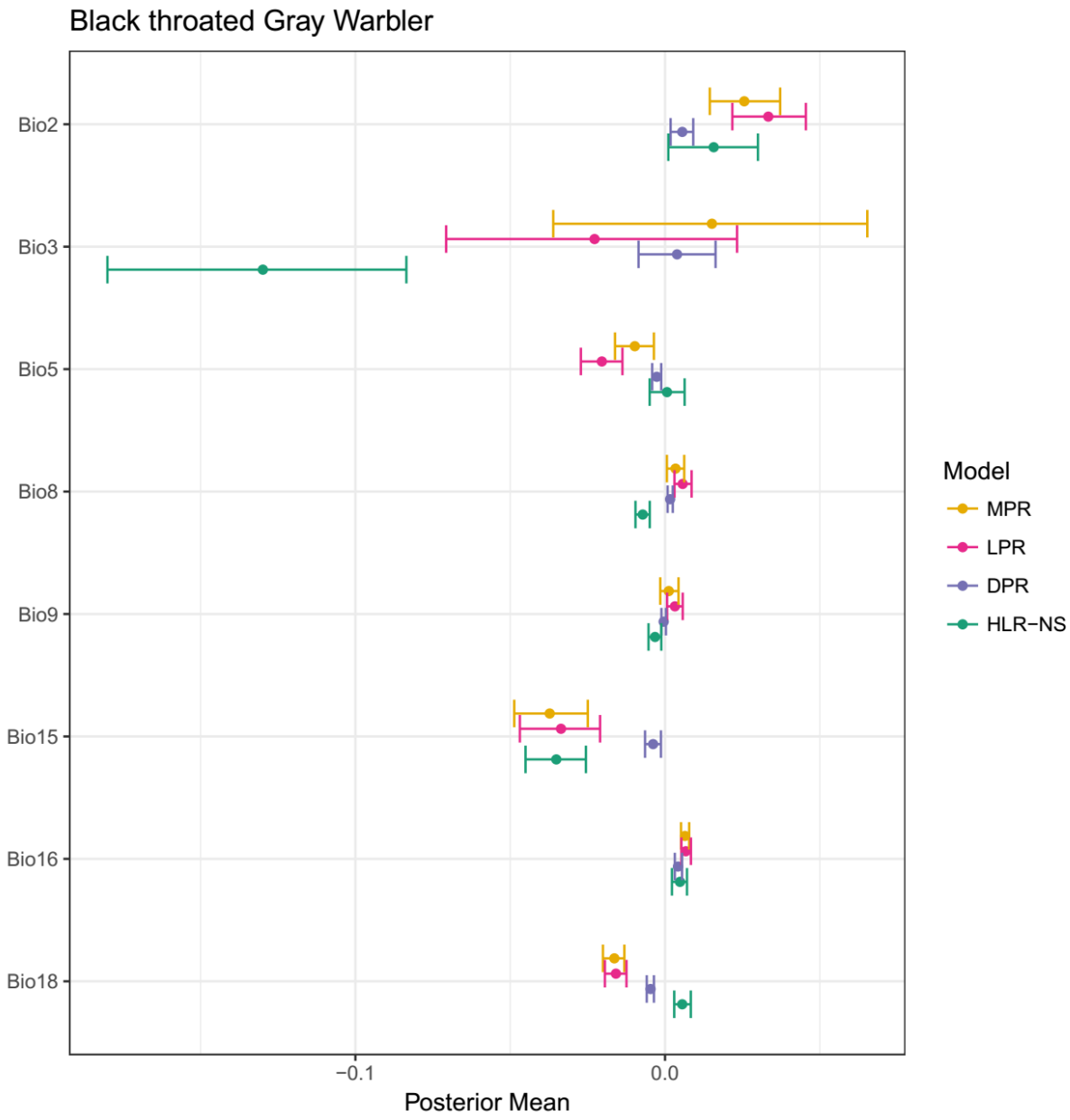


Figure 55: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-throated gray warbler, *Setophaga nigrescens*.

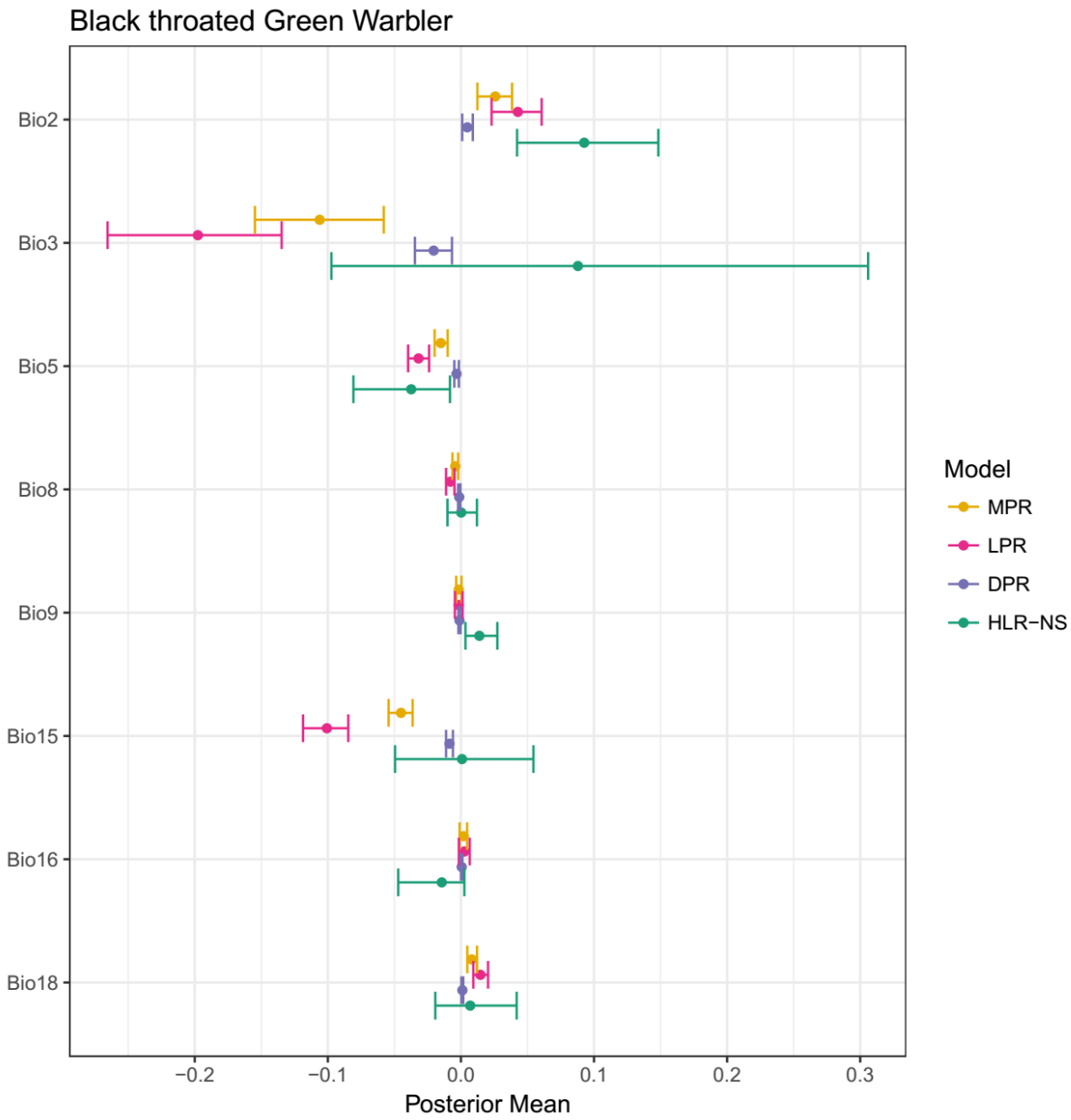


Figure 56: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-throated green warbler, *Setophaga virens*.

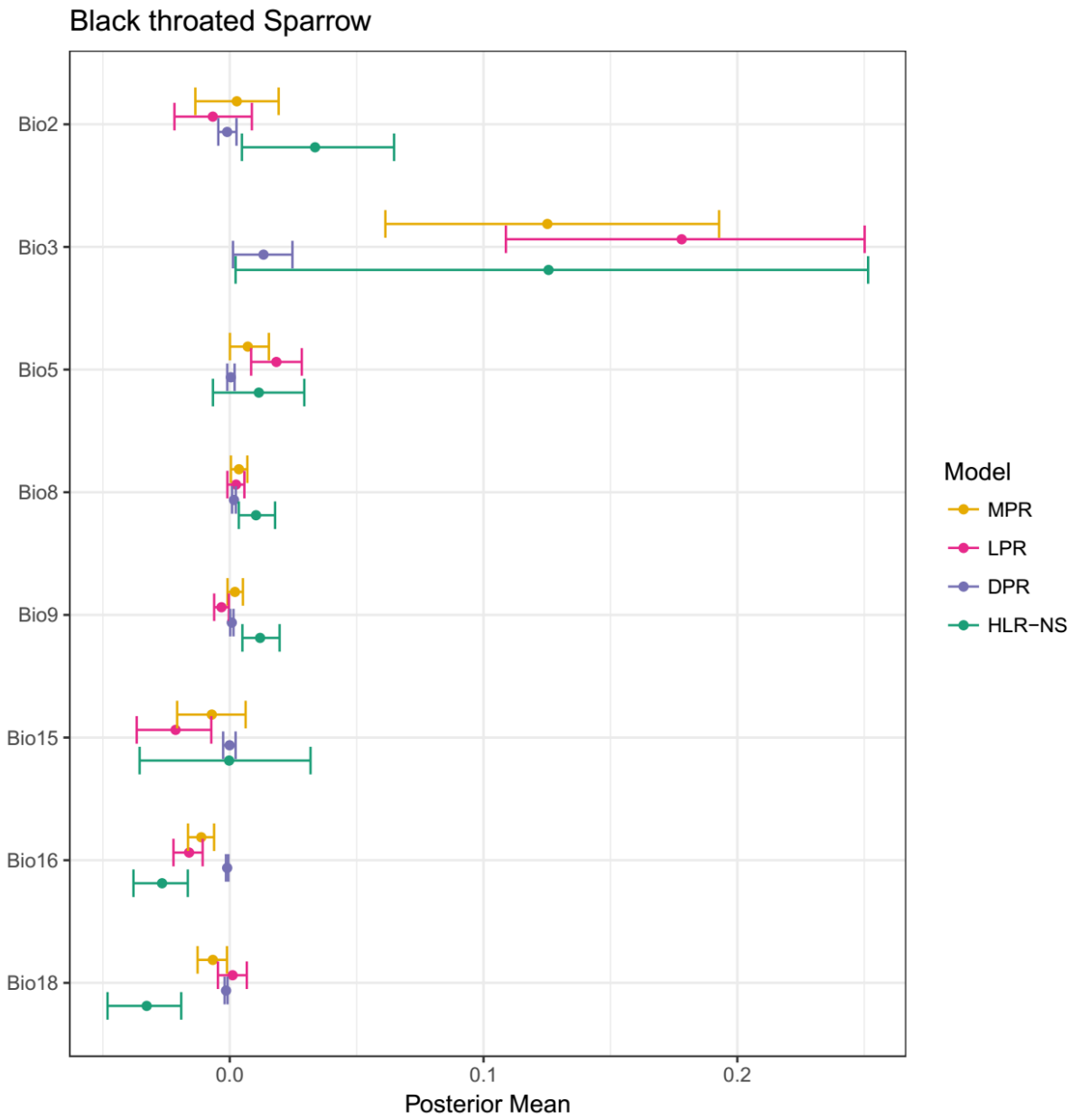


Figure 57: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black-throated sparrow, *Amphispiza bilineata*.

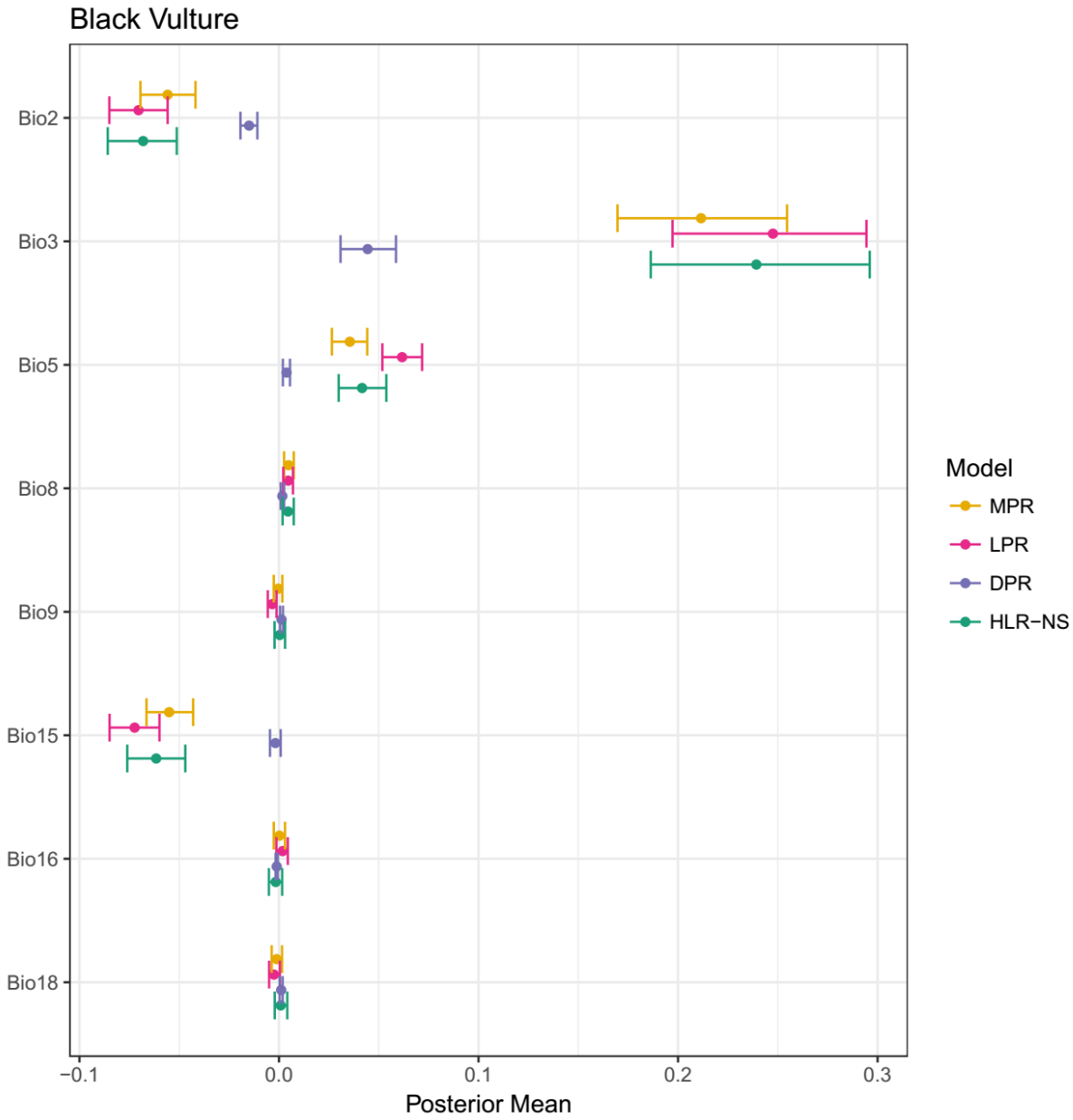


Figure 58: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black vulture, *Coragyps atratus*.

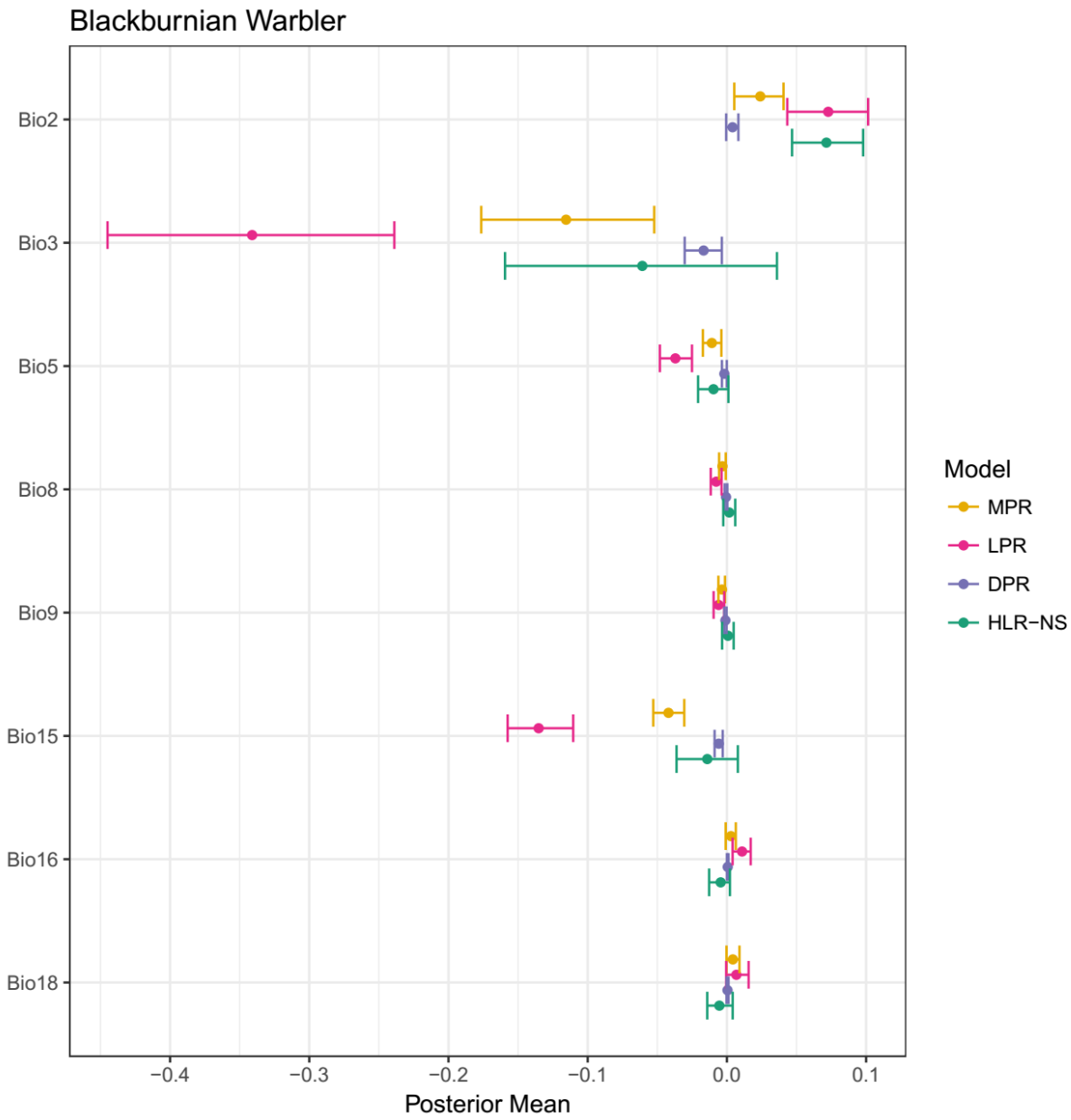


Figure 59: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Blackburnian warbler, *Setophaga fusca*.

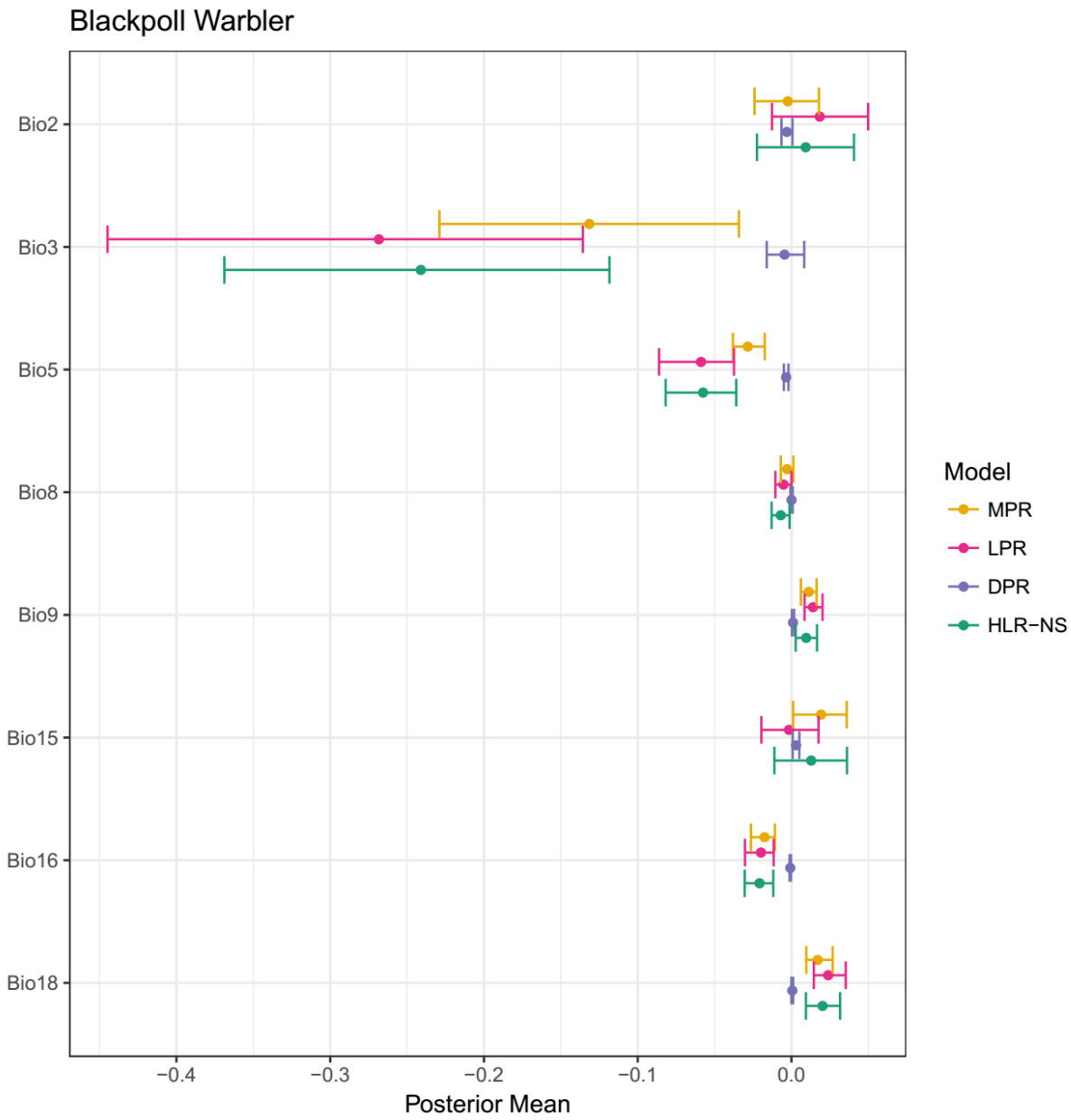


Figure 60: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blackpoll warbler, *Setophaga striata*.

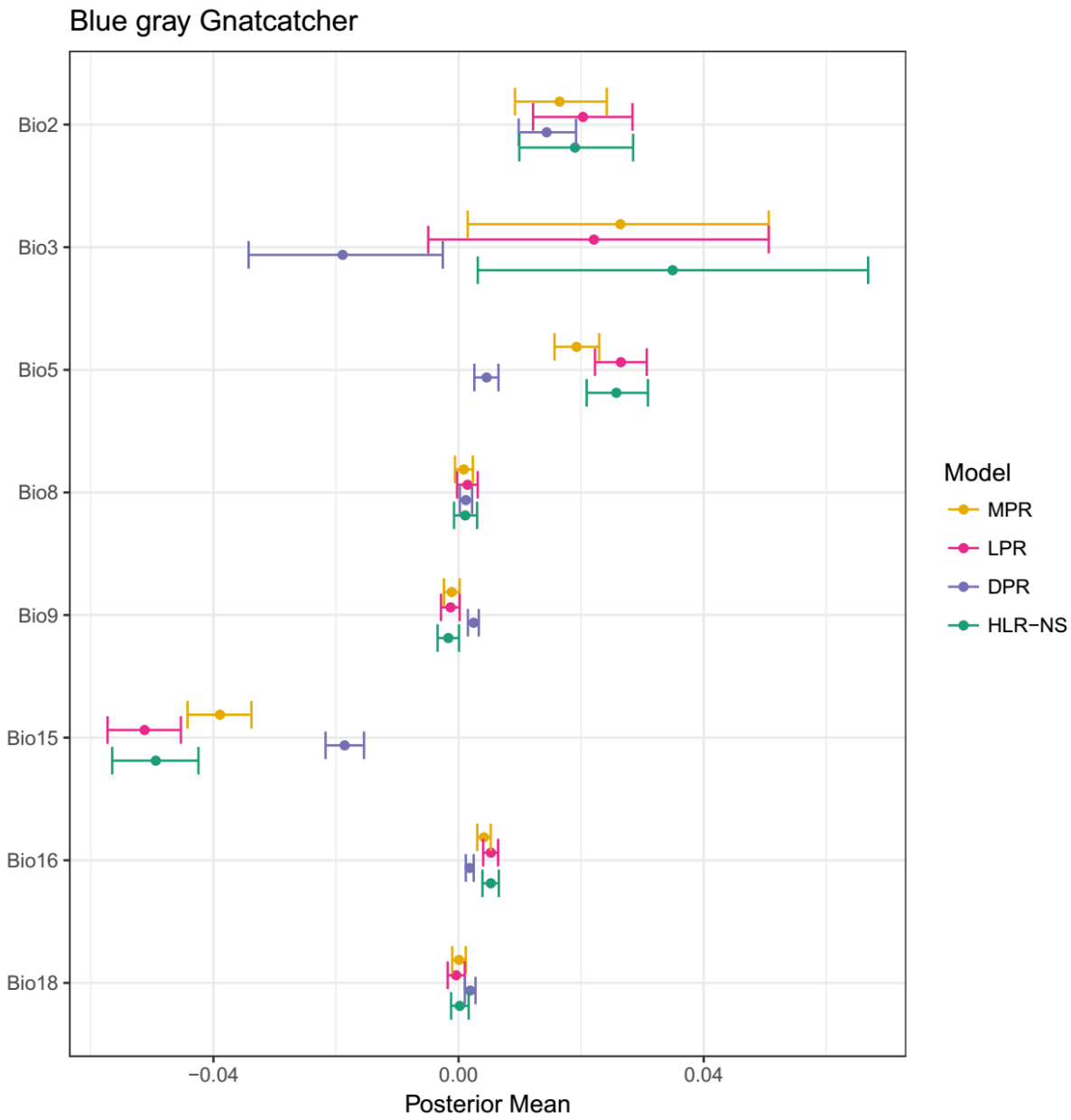


Figure 61: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blue-gray gnatcatcher, *Polioptila caerulea*.

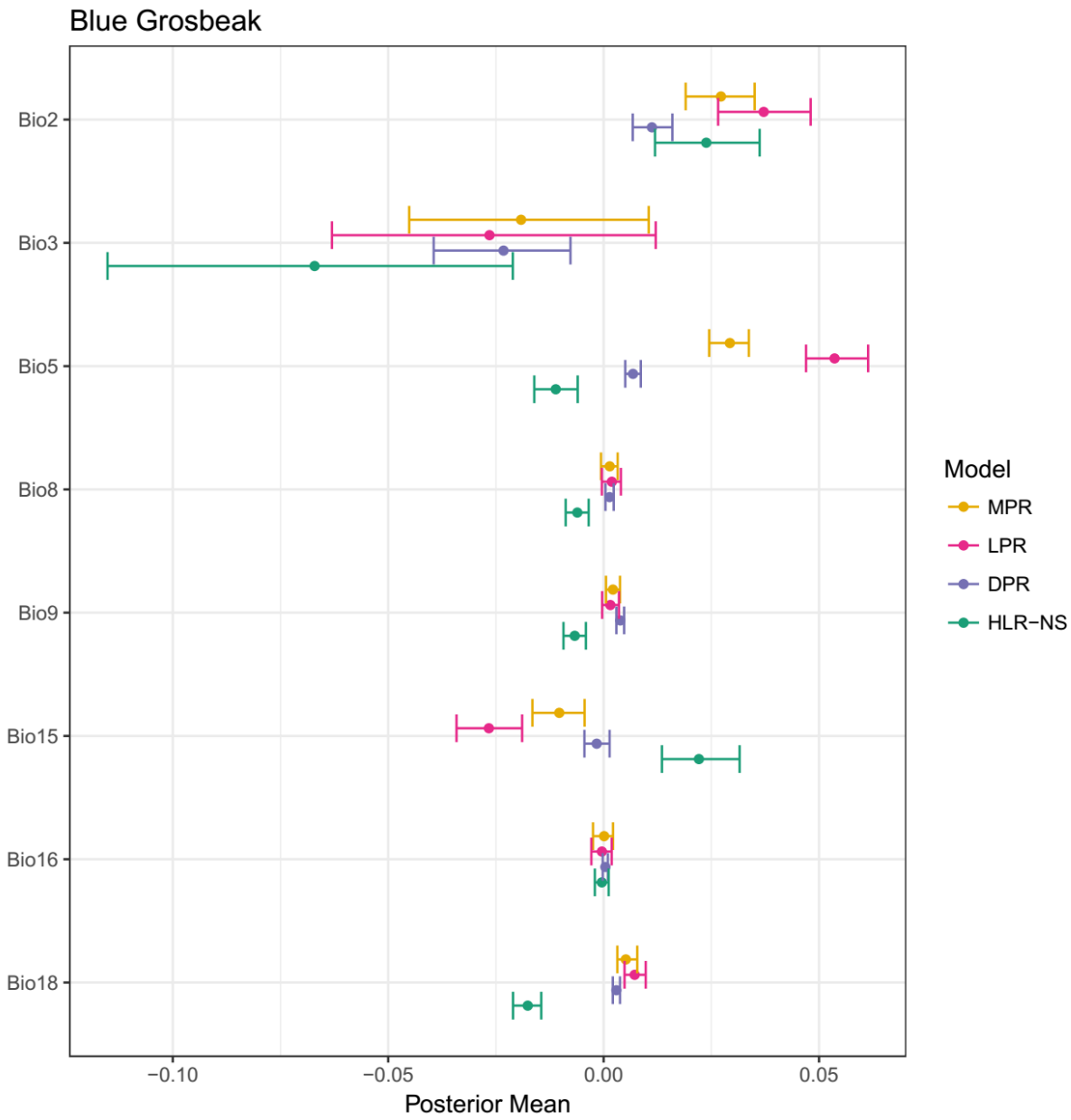


Figure 62: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blue grosbeak, *Passerina caerulea*.

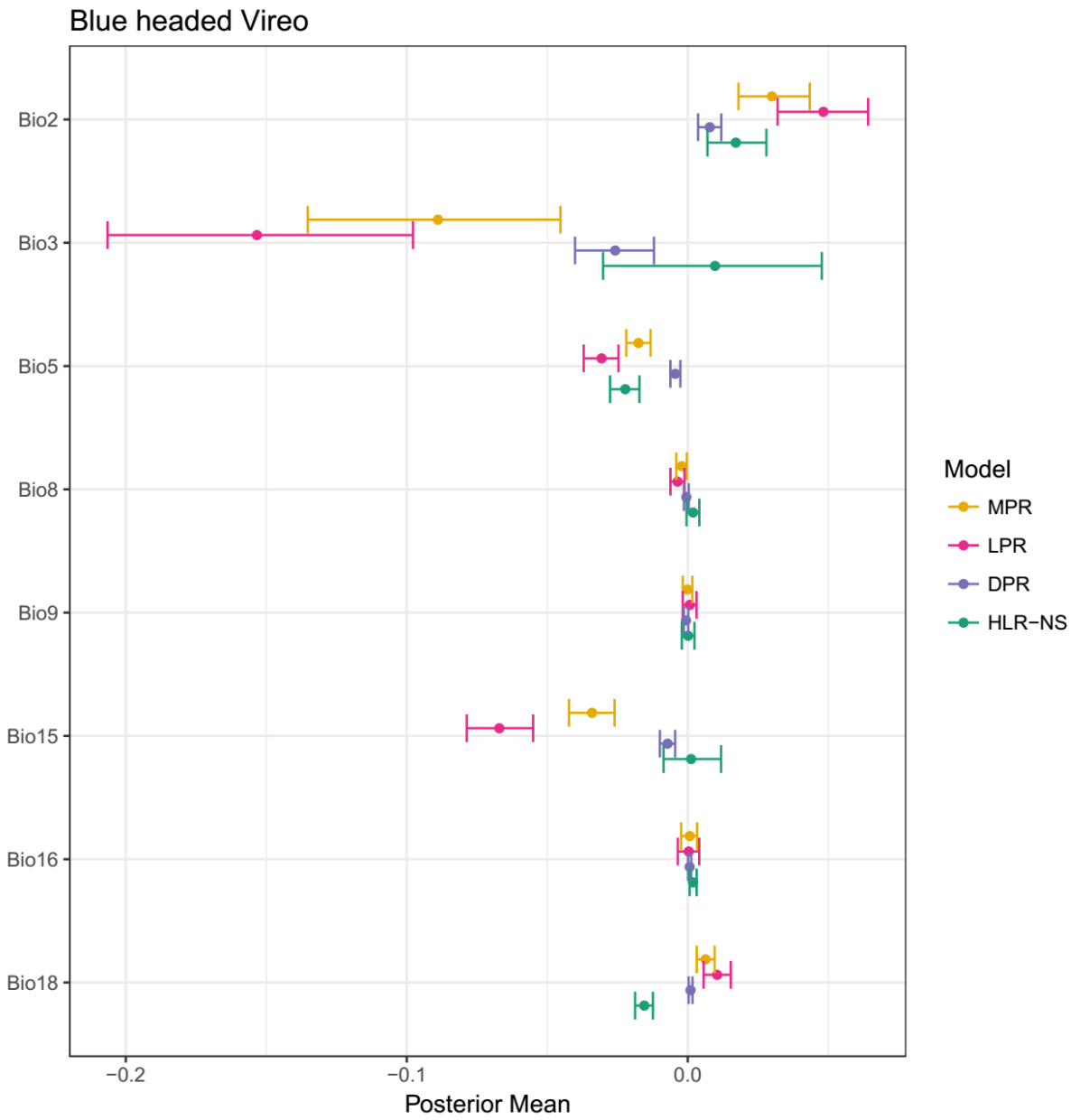


Figure 63: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blue-headed vireo, *Vireo solitarius*.

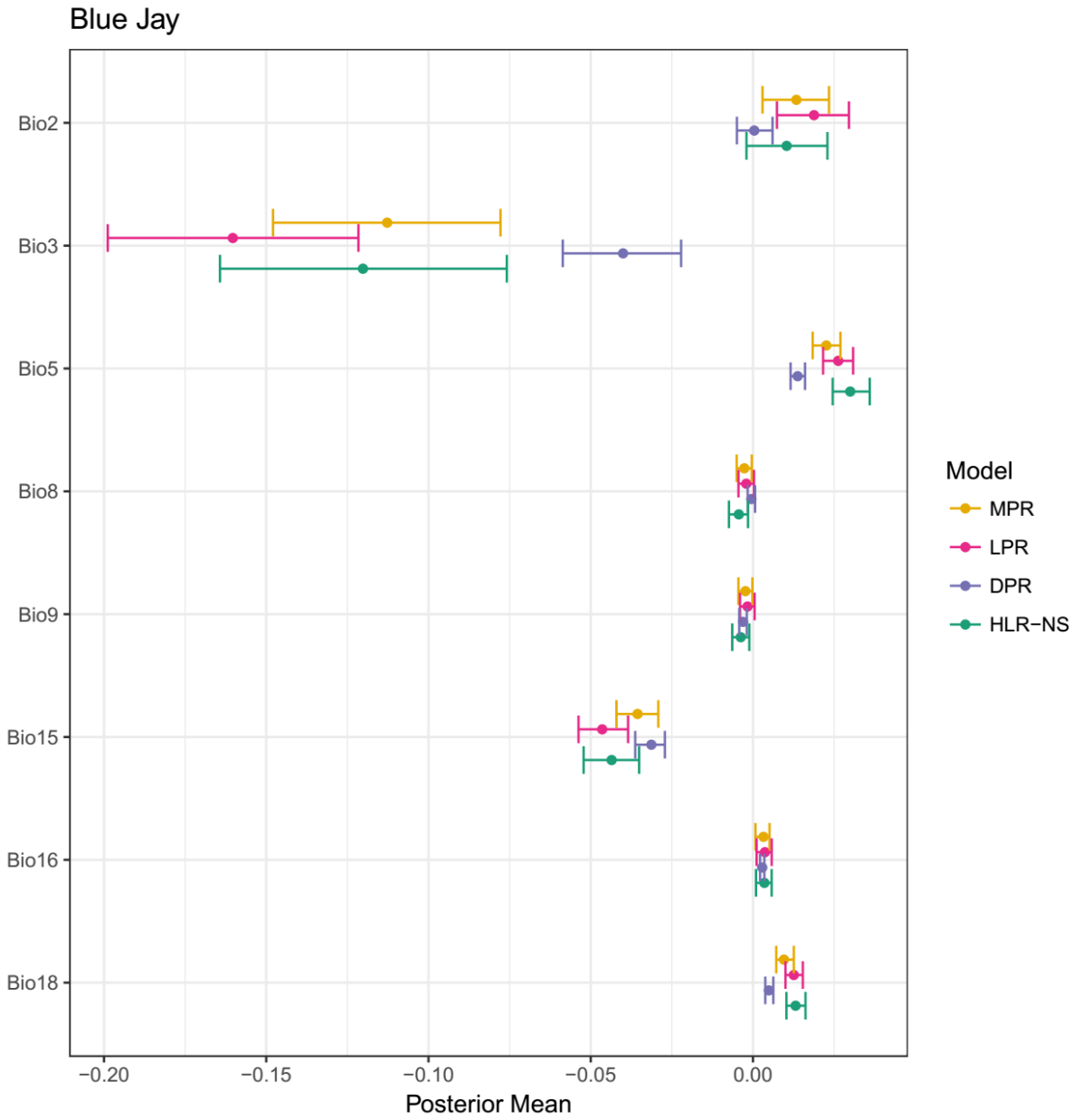


Figure 64: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blue jay, *Cyanocitta cristata*.

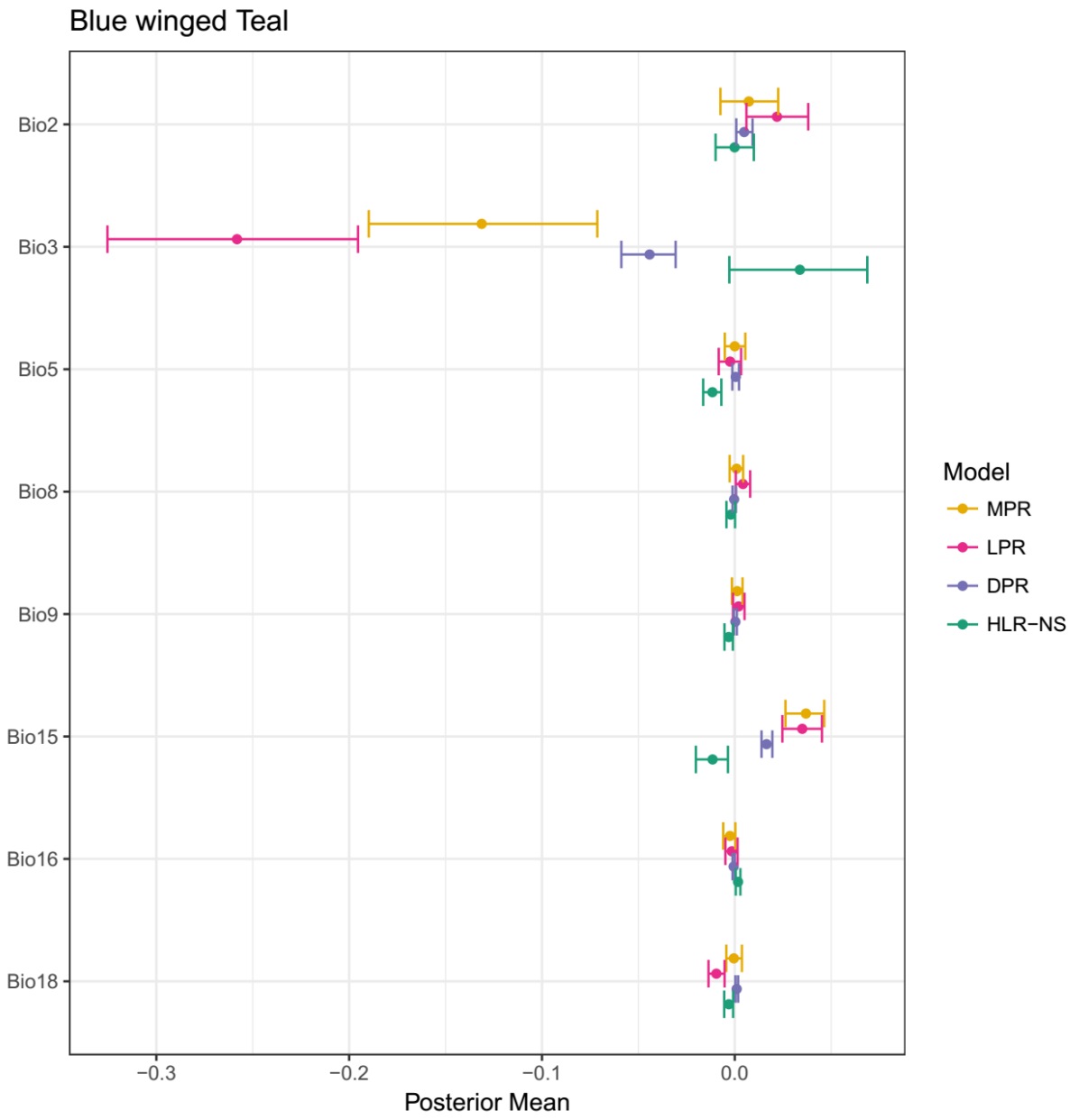


Figure 65: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blue-winged teal, *Anas discors*.

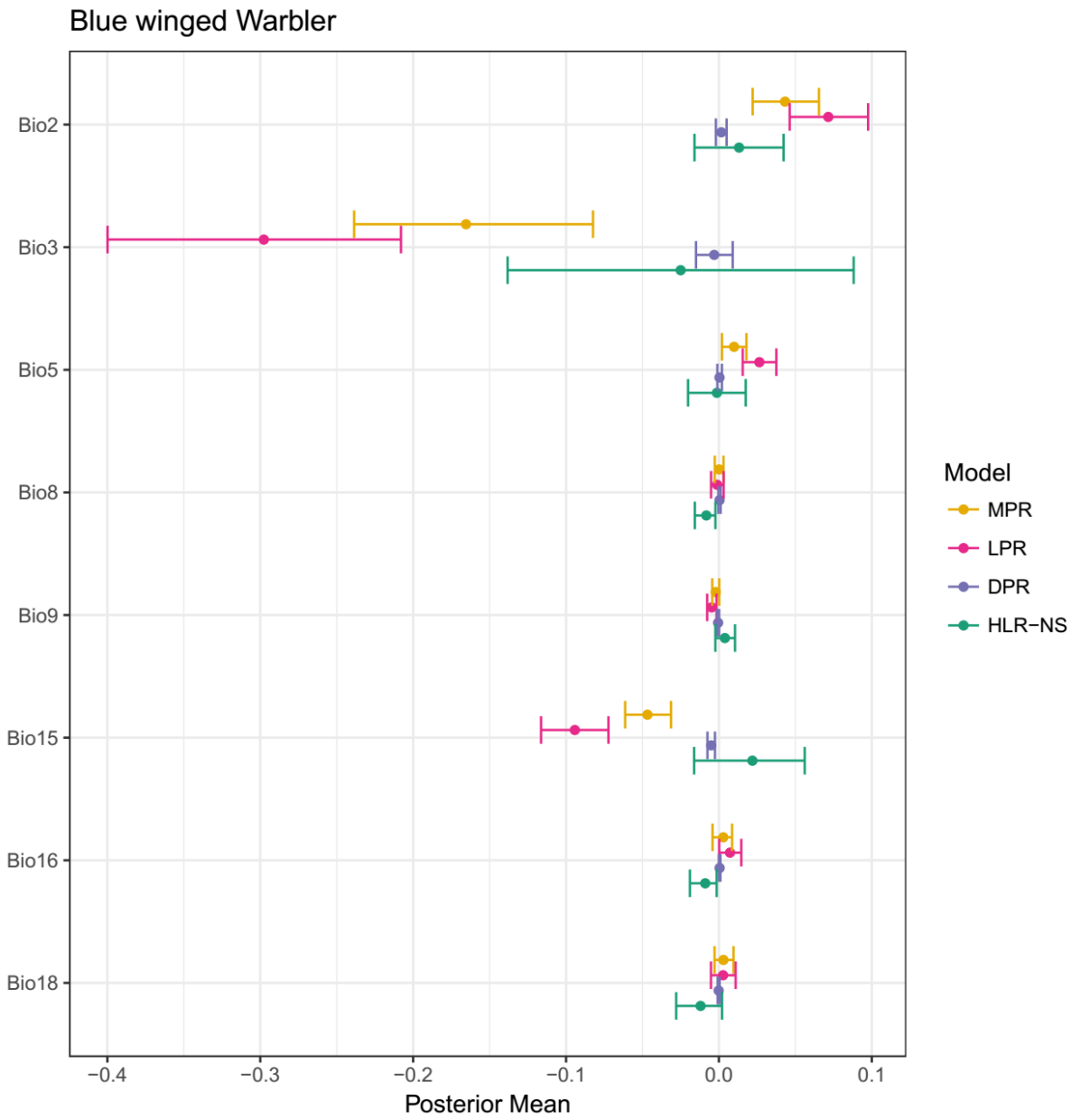


Figure 66: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the blue-winged warbler, *Vermivora cyanoptera*.

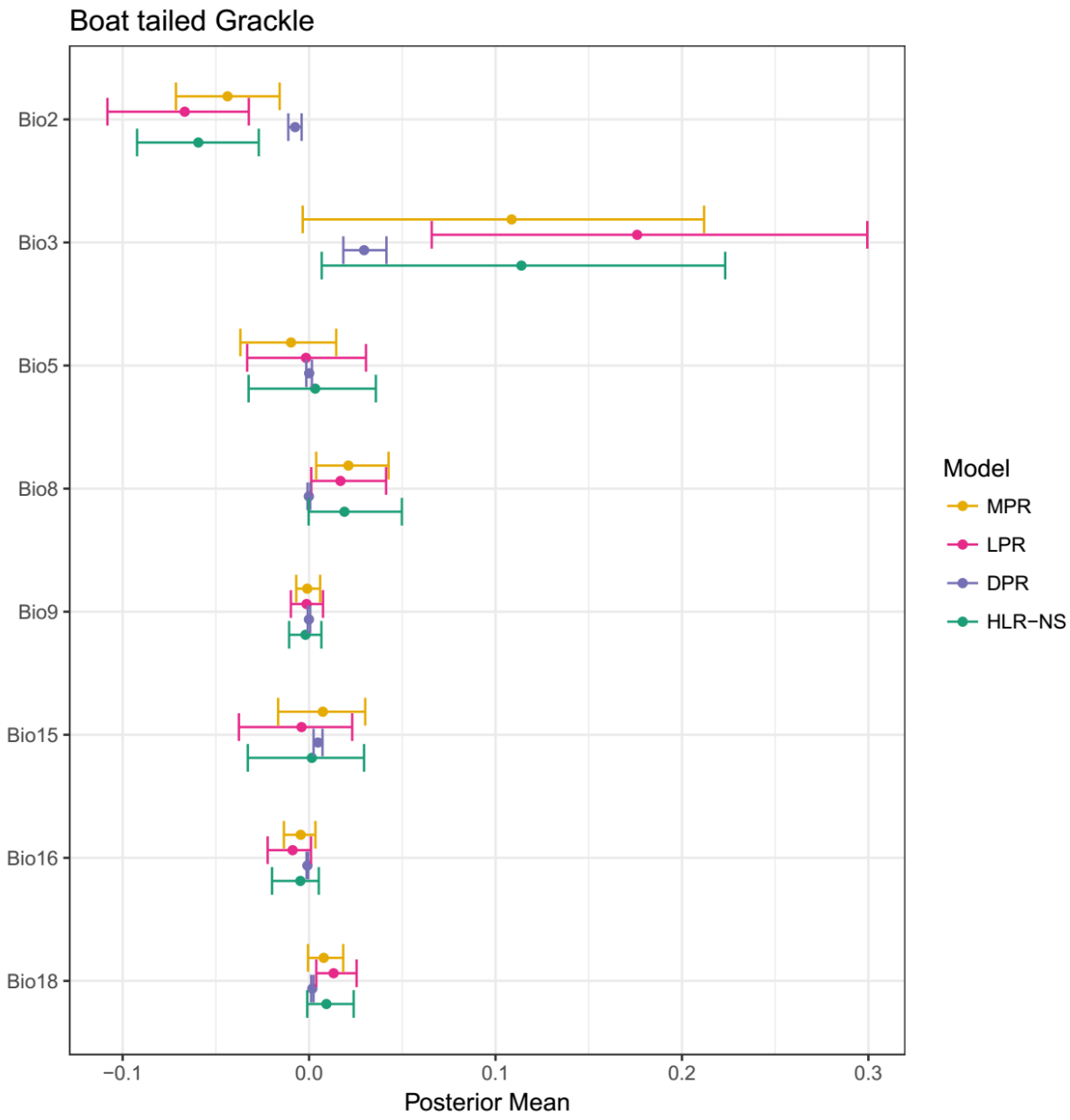


Figure 67: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the boat-tailed grackle, *Quiscalus major*.

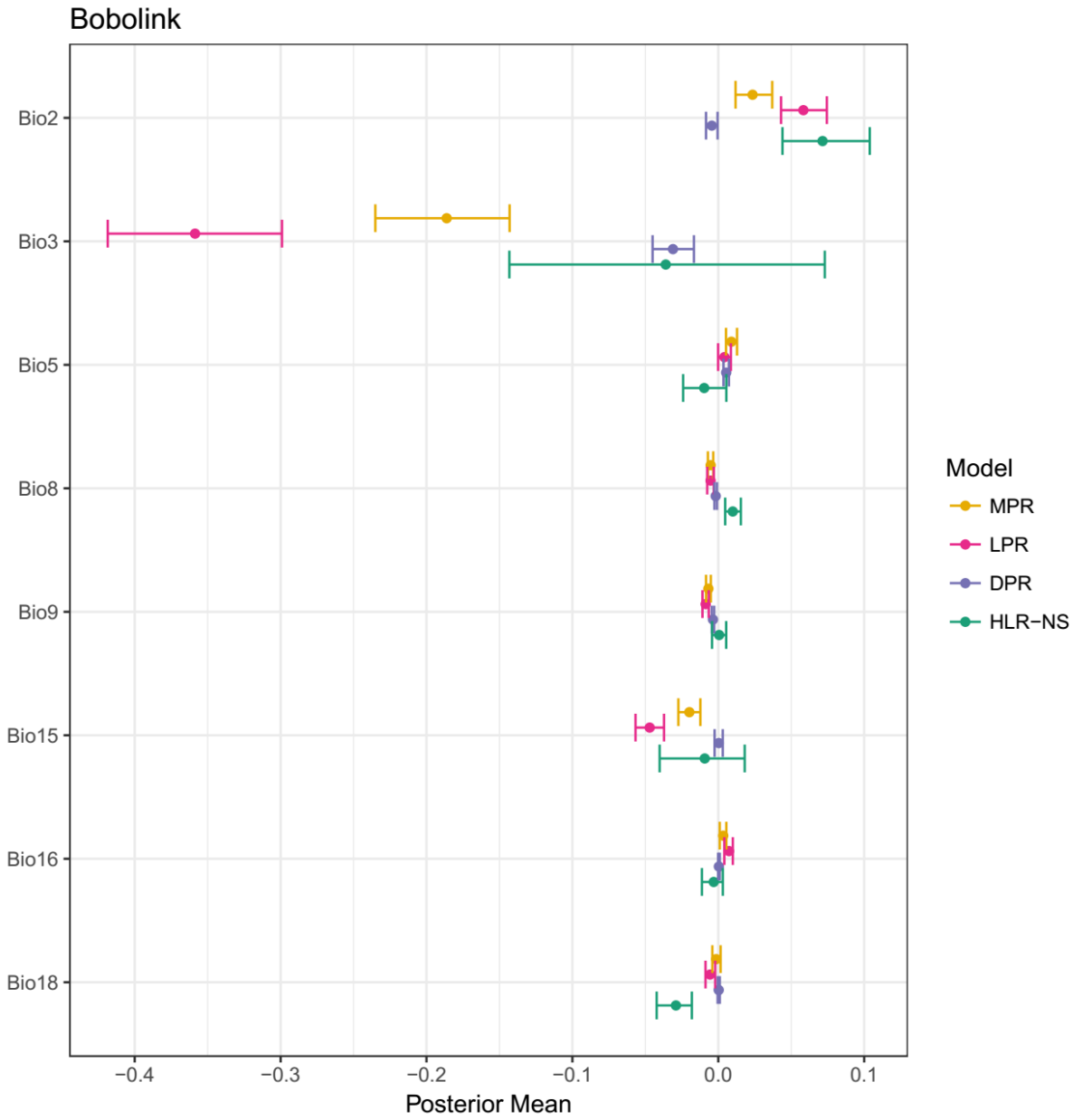


Figure 68: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bobolink, *Dolichonyx oryzivorus*.

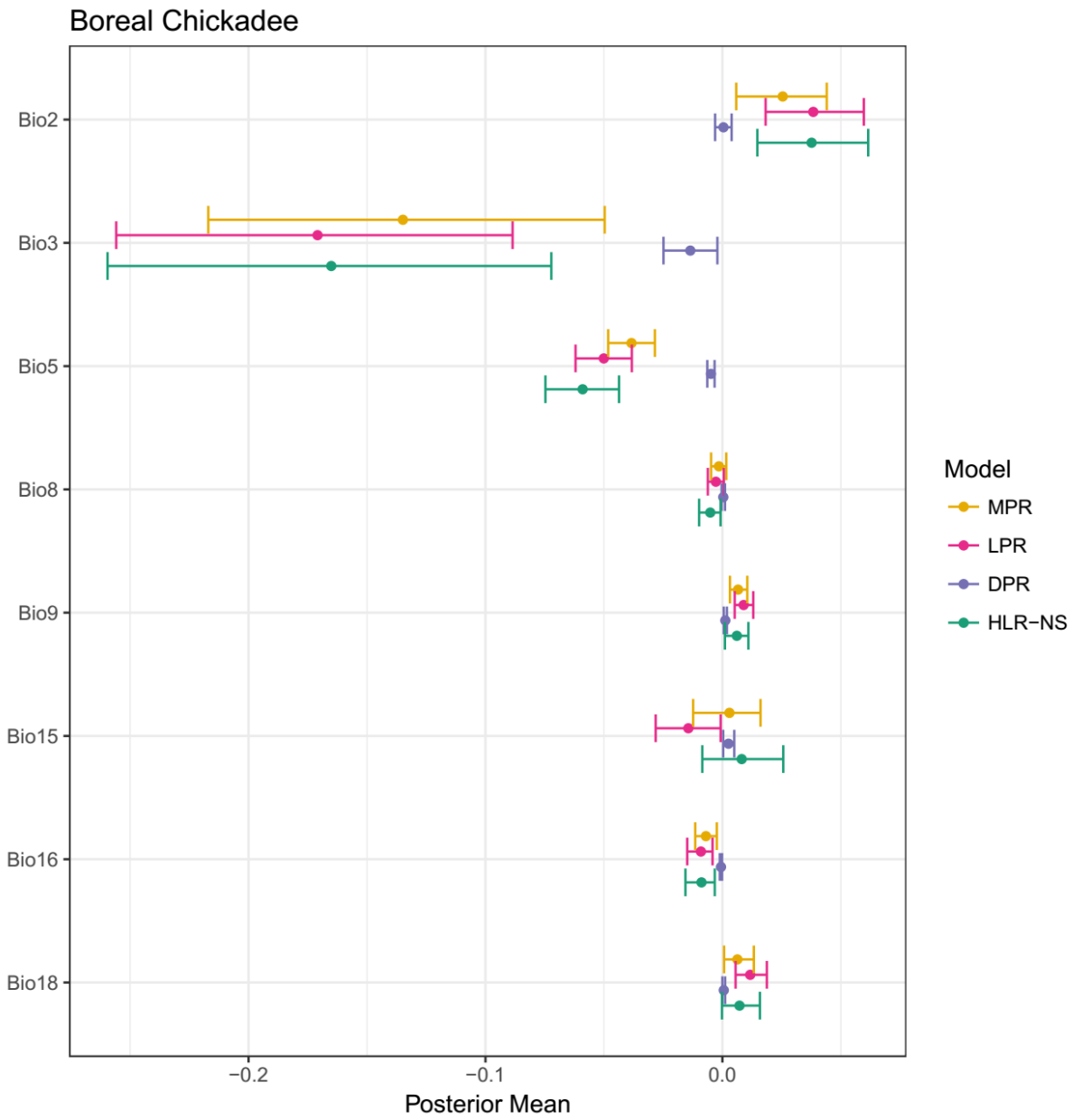


Figure 69: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the boreal chickadee, *Poecile hudsonicus*.

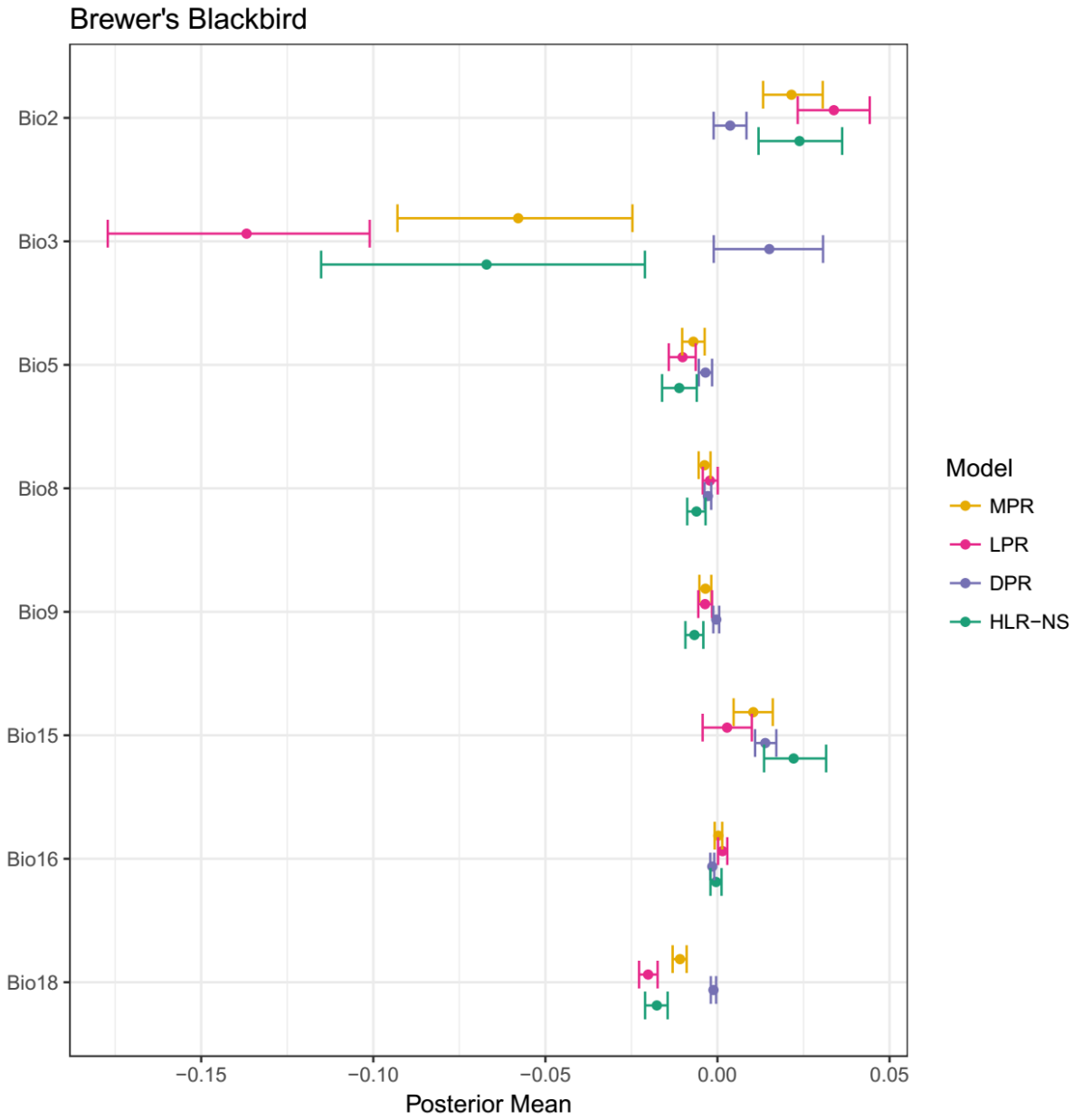


Figure 70: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Brewer's blackbird, *Euphagus cyanocephalus*.

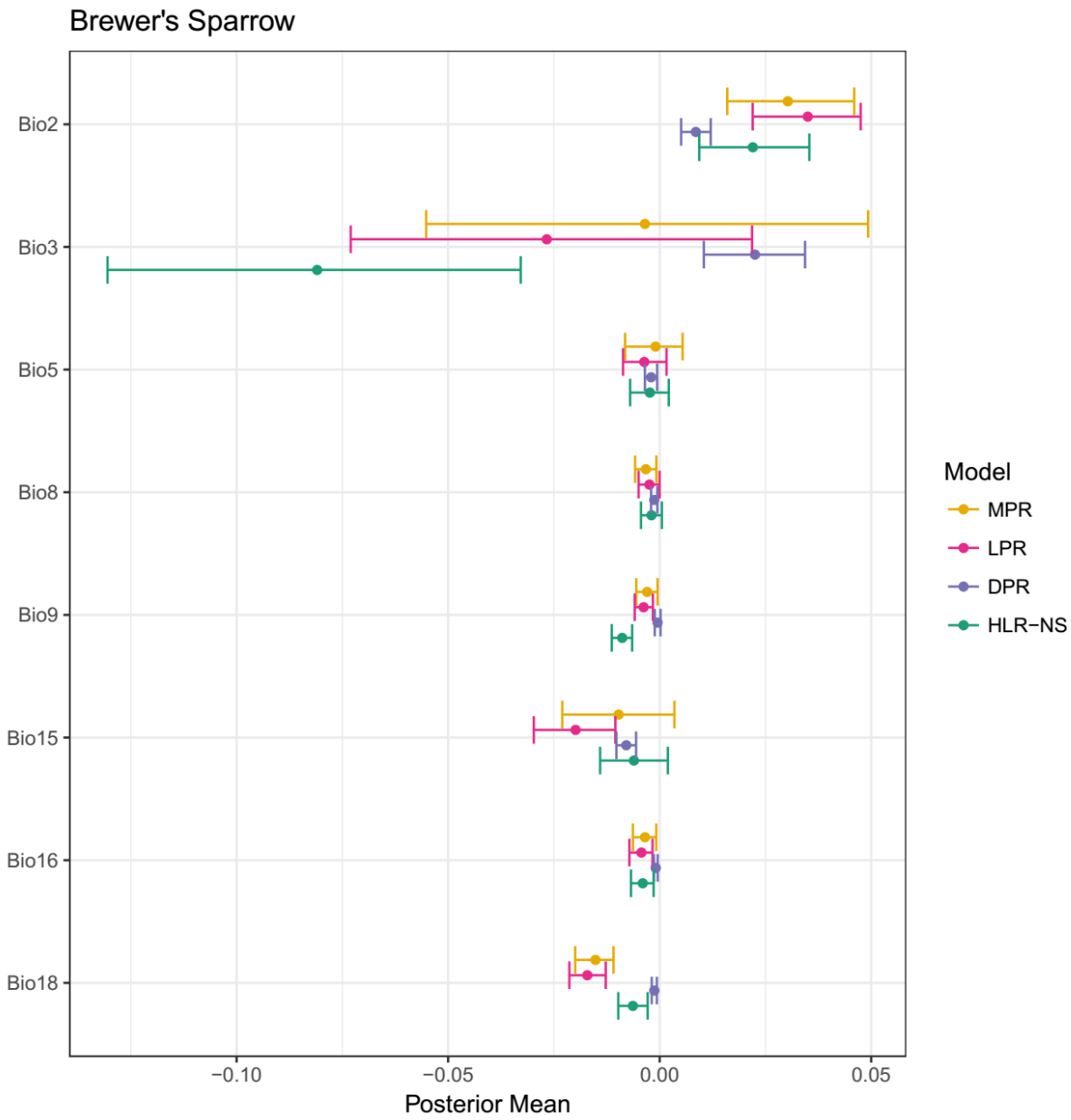


Figure 71: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Brewer's sparrow, *Spizella breweri*.

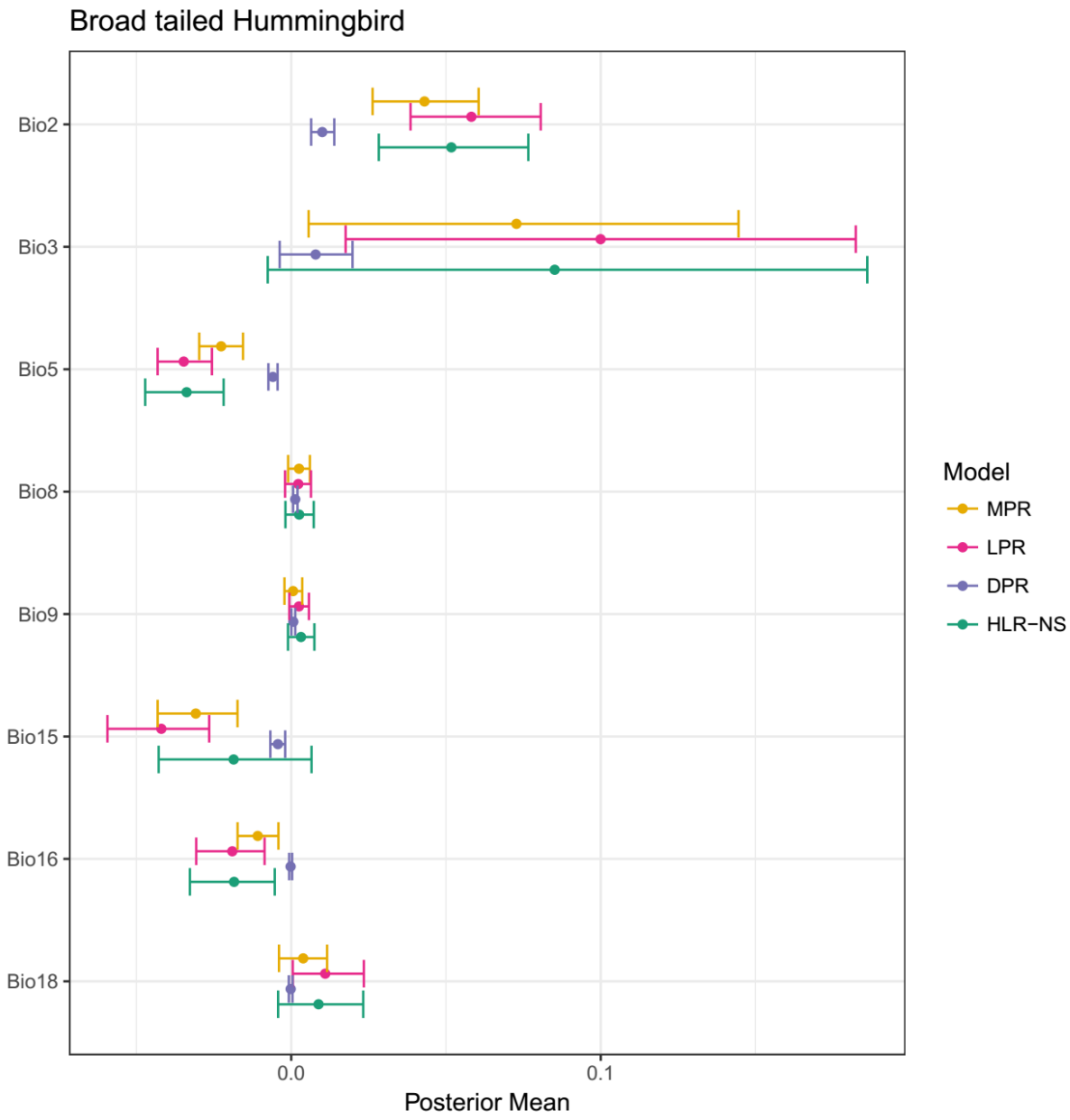


Figure 72: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the broad-tailed hummingbird, *Selasphorus platycercus*.

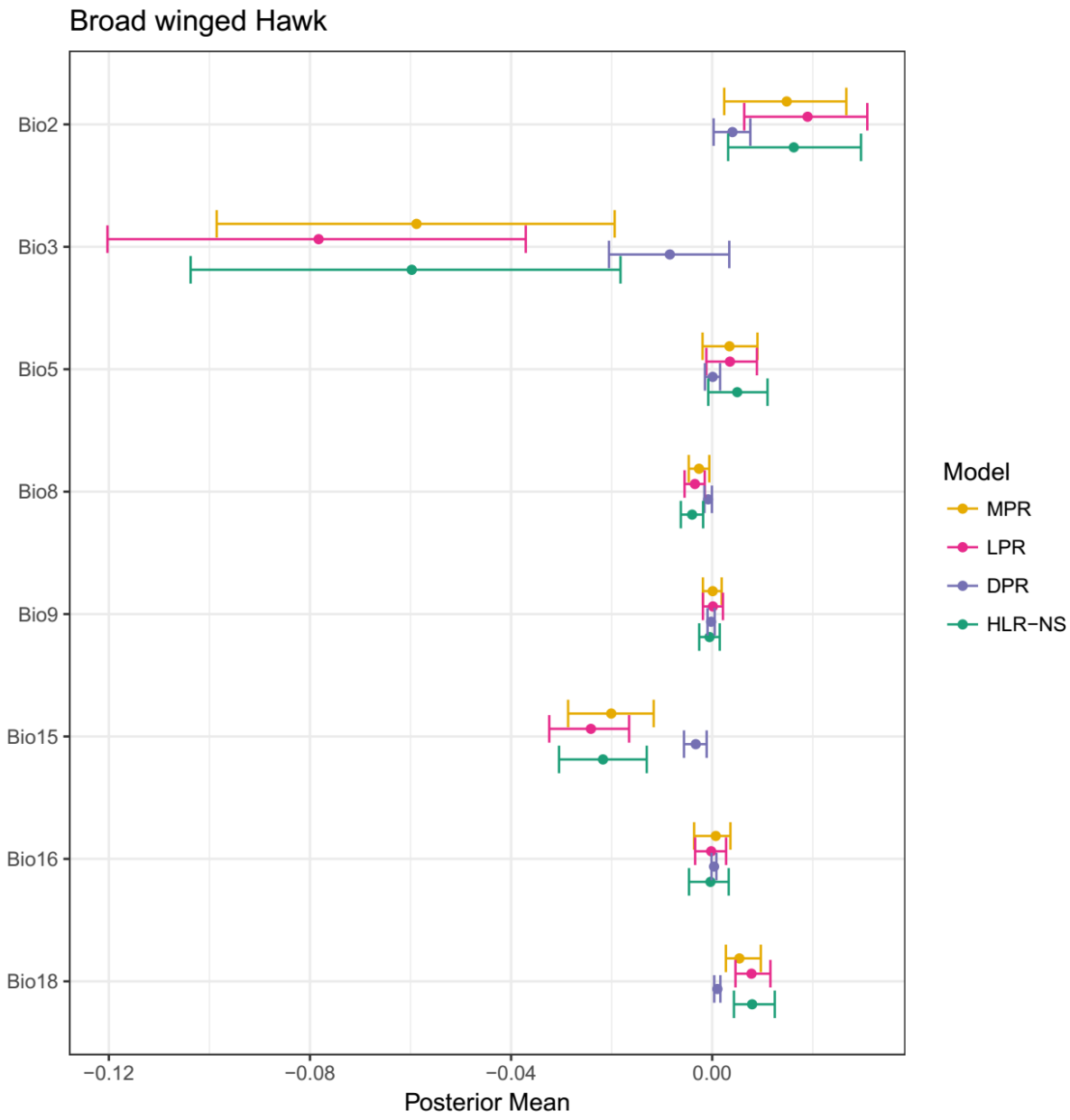


Figure 73: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the broad-winged hawk, *Buteo platypterus*.

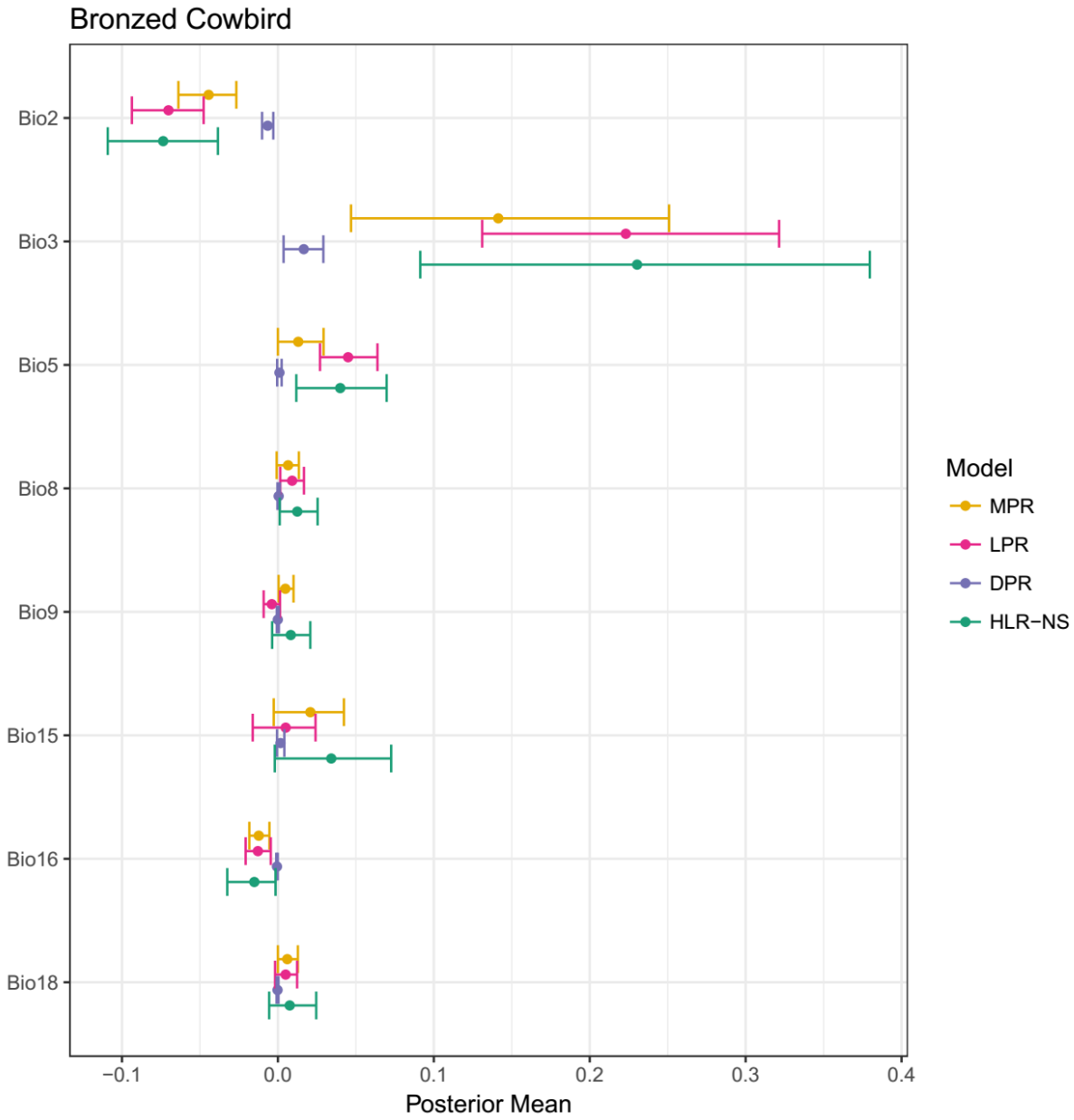


Figure 74: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bronzed cowbird, *Molothrus aeneus*.

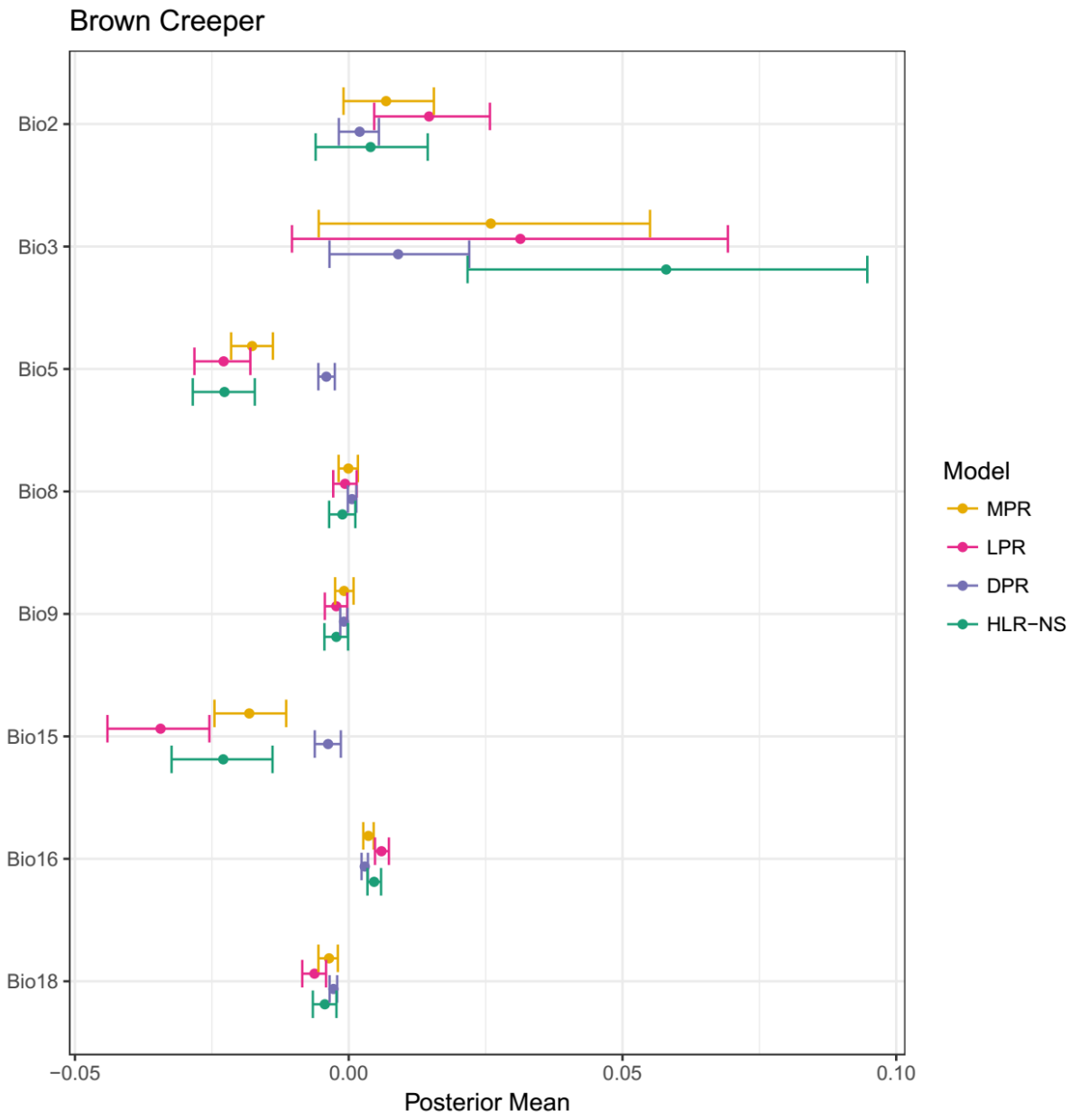


Figure 75: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown creeper, *Certhia americana*.

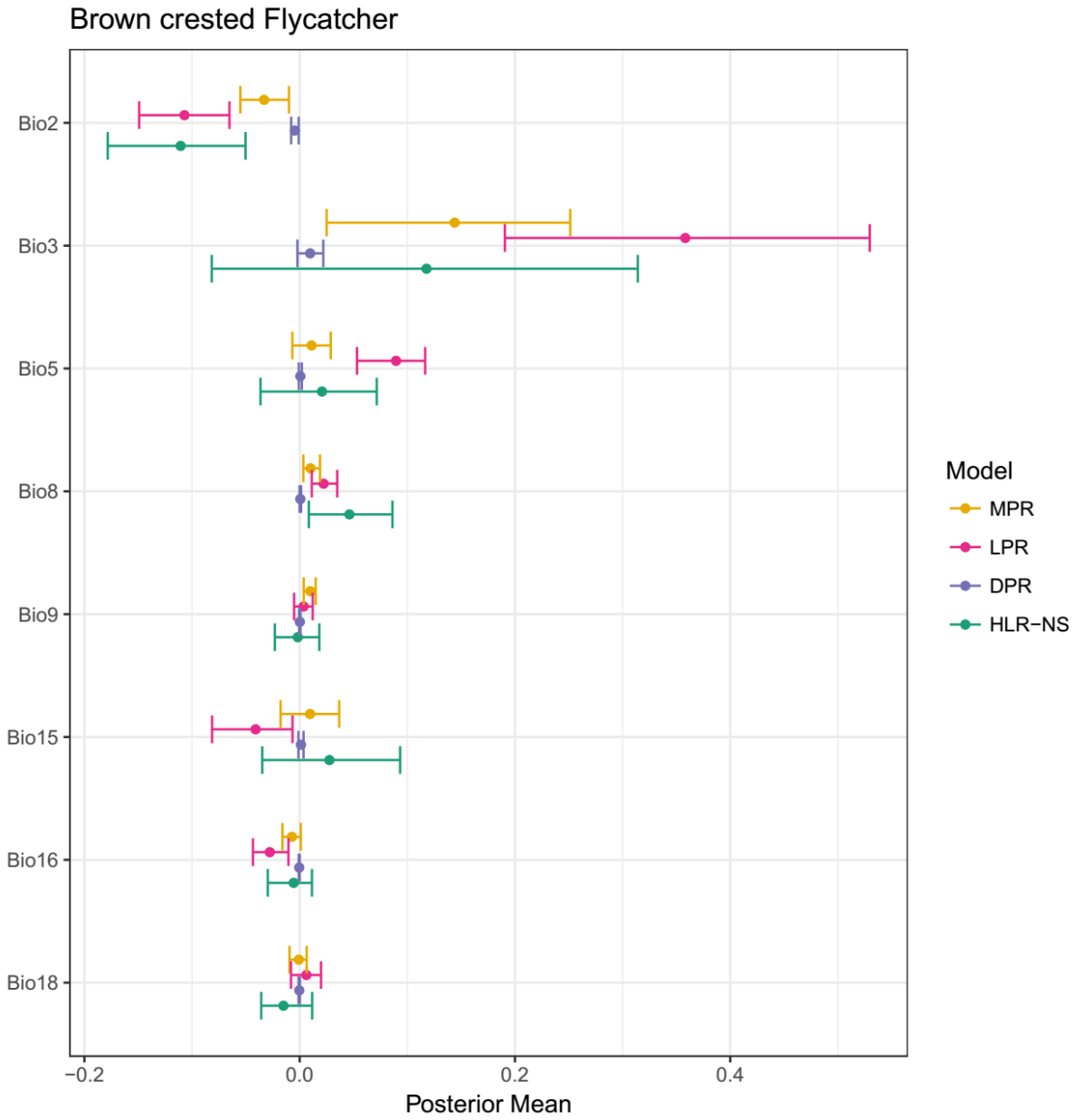


Figure 76: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown-crested flycatcher, *Myiarchus tyrannulus*.

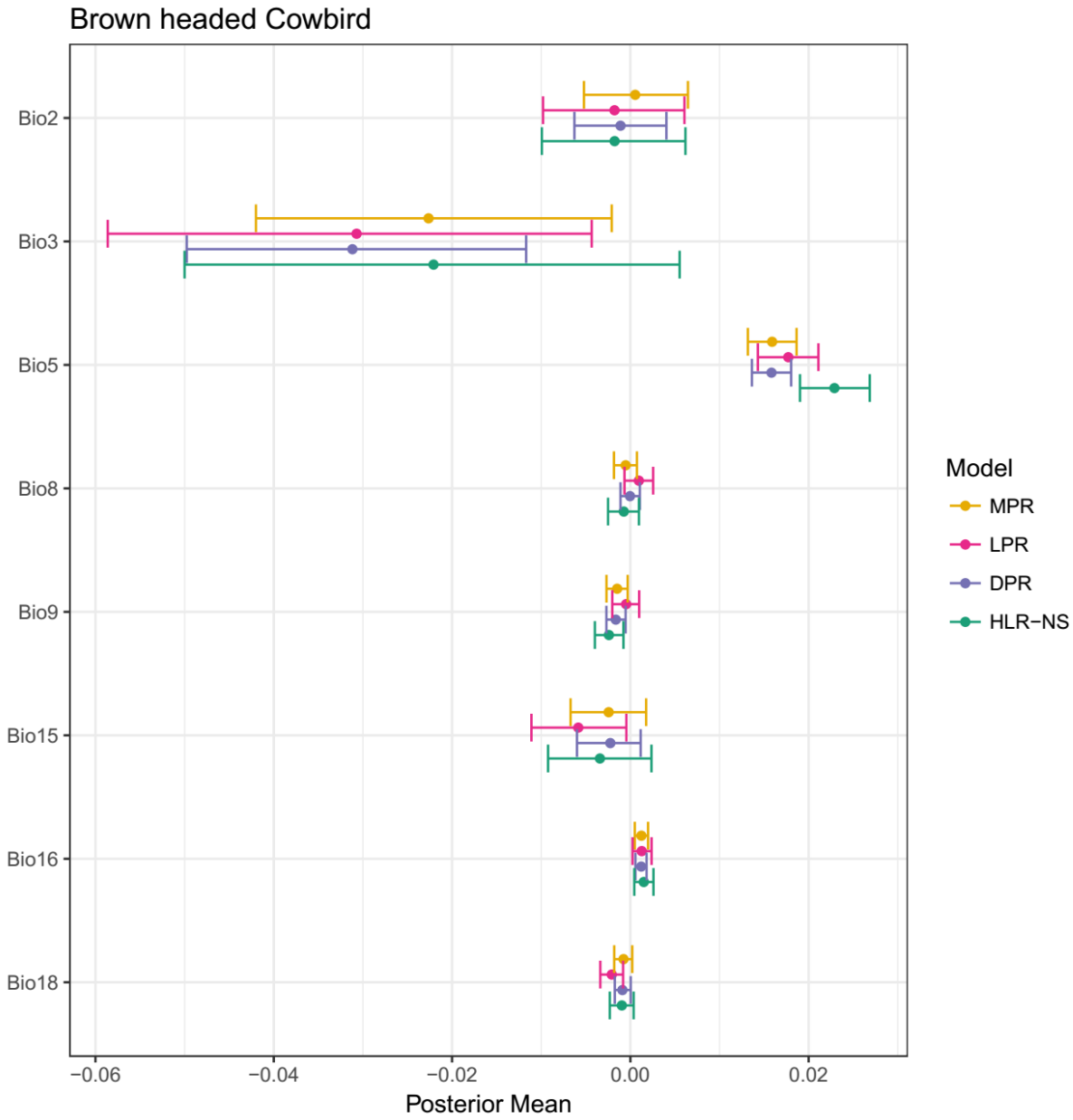


Figure 77: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown-headed cowbird, *Molothrus ater*.

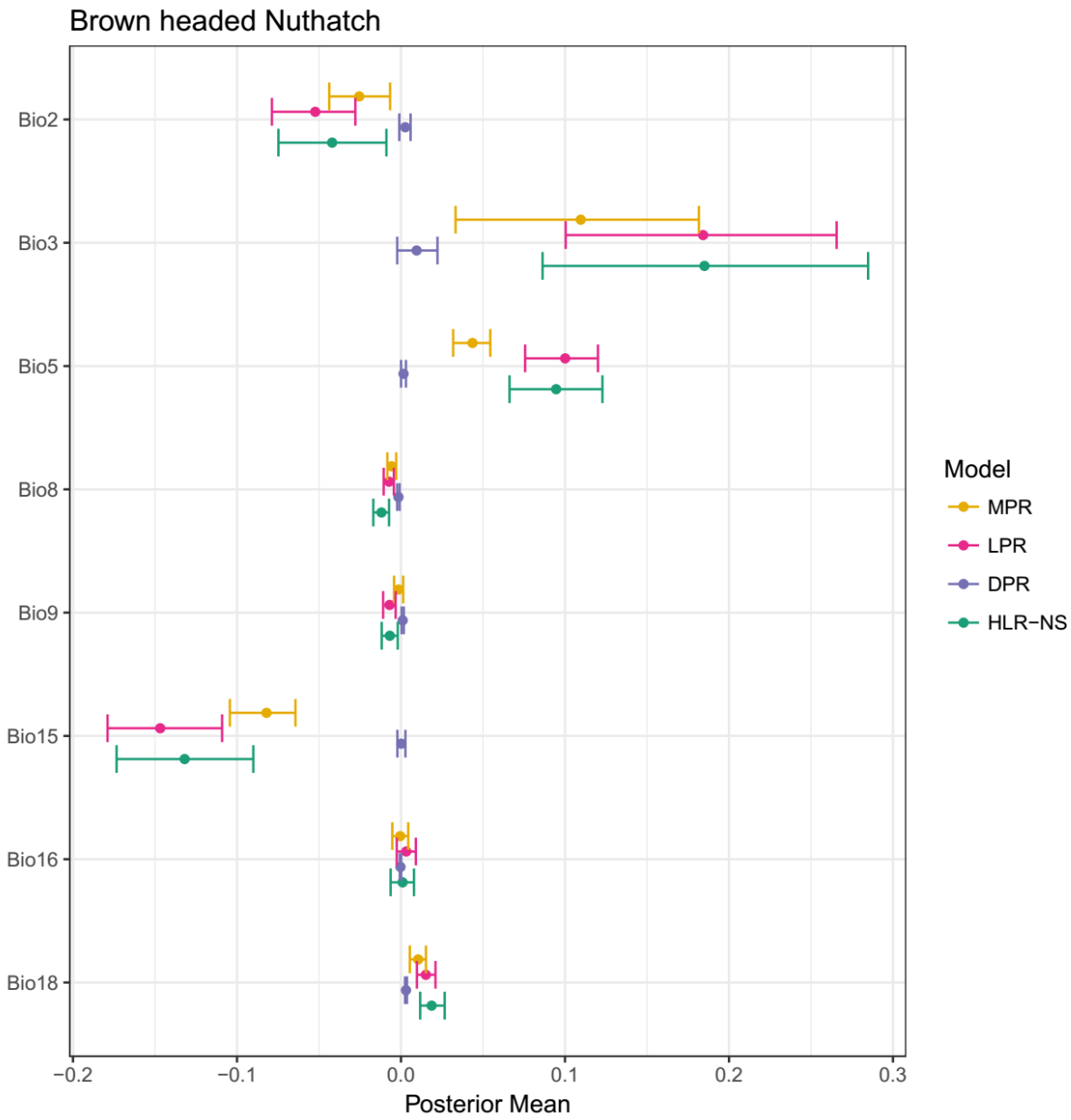


Figure 78: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown-headed nuthatch, *Sitta pusilla*.

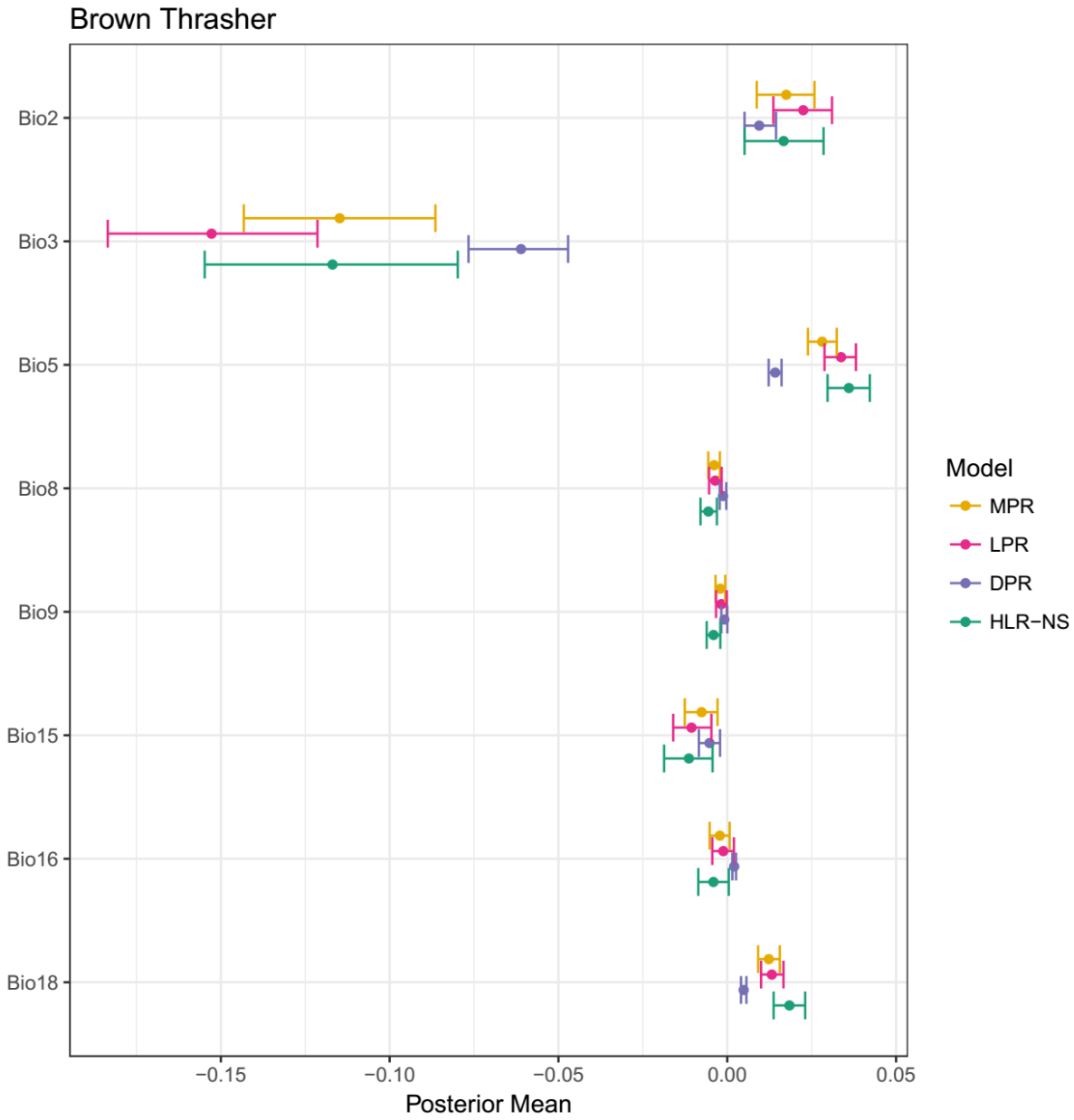


Figure 79: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown thrasher, *Toxostoma rufum*.

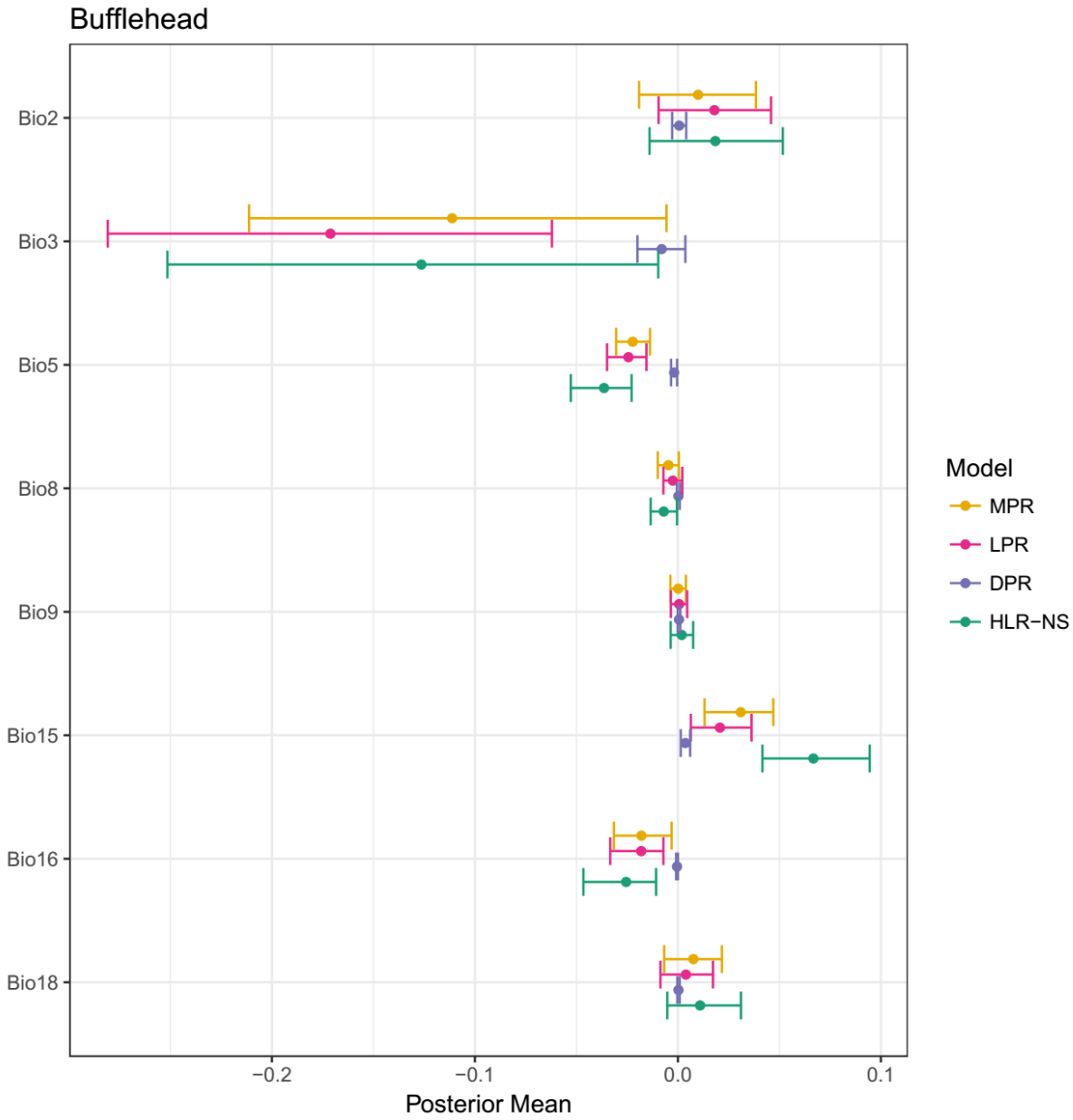


Figure 80: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bufflehead, *Bucephala albeola*.

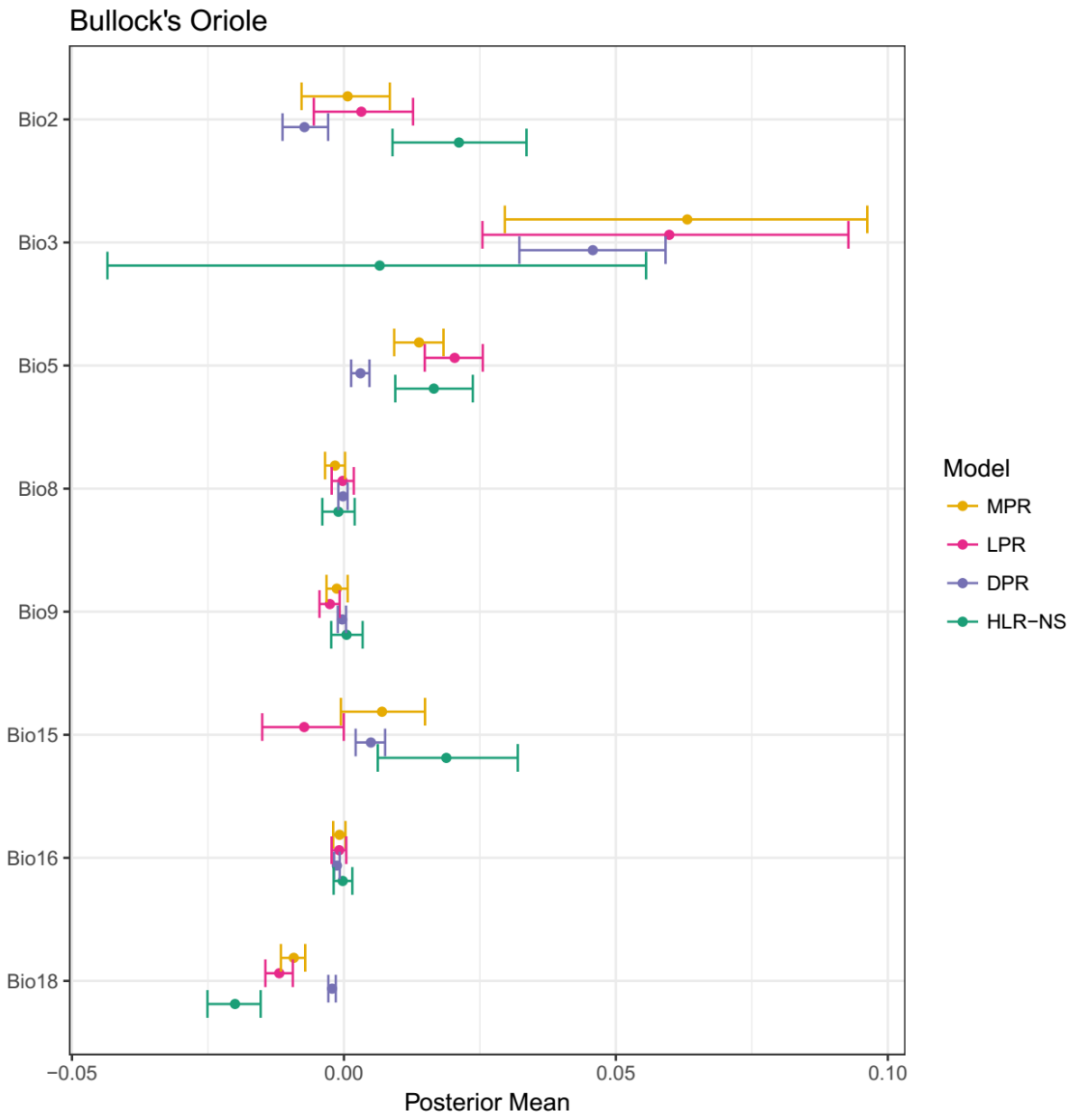


Figure 81: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Bullock's oriole, *Icterus bullockii*.

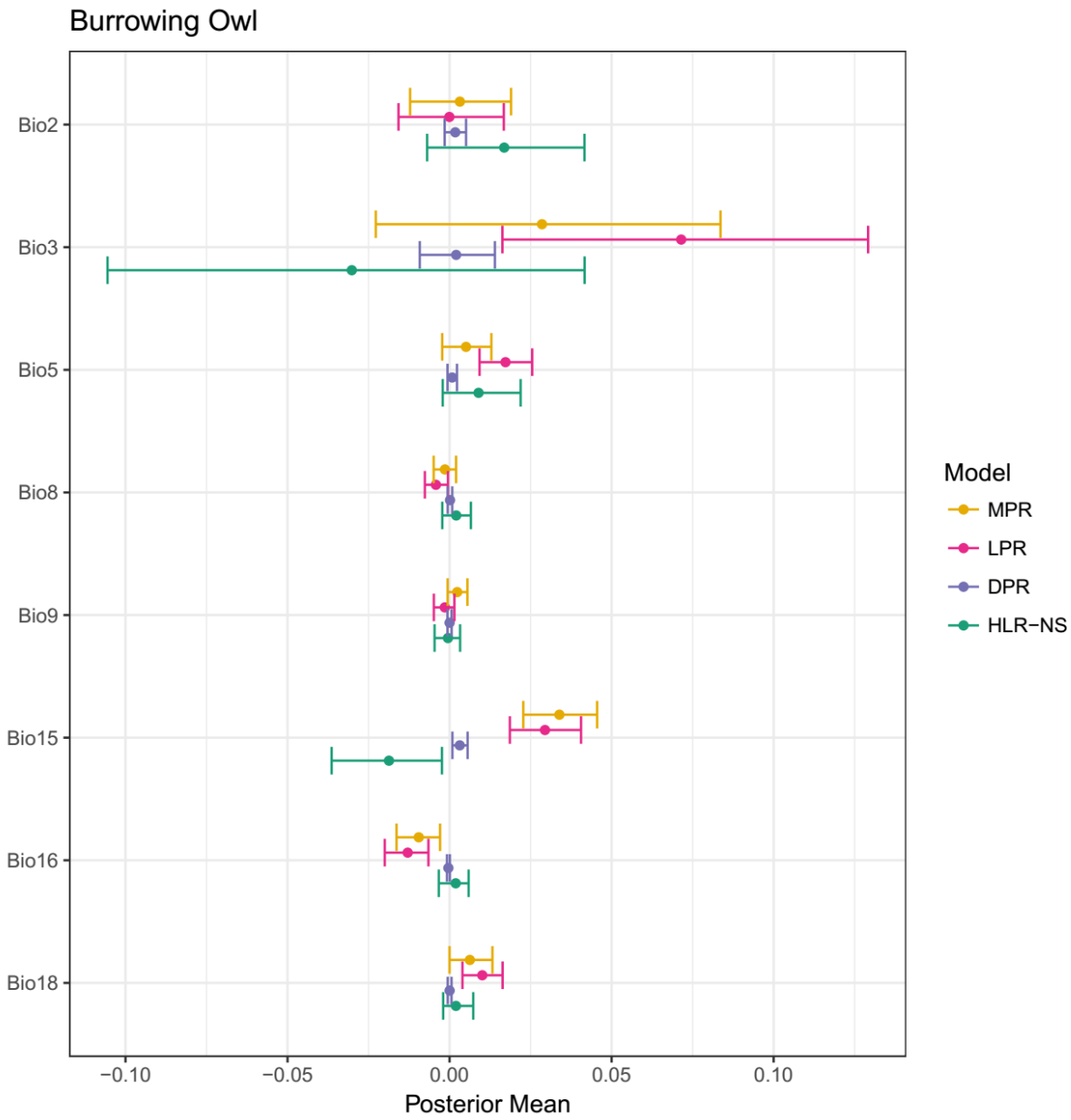


Figure 82: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the burrowing owl, *Athene cunicularia*.

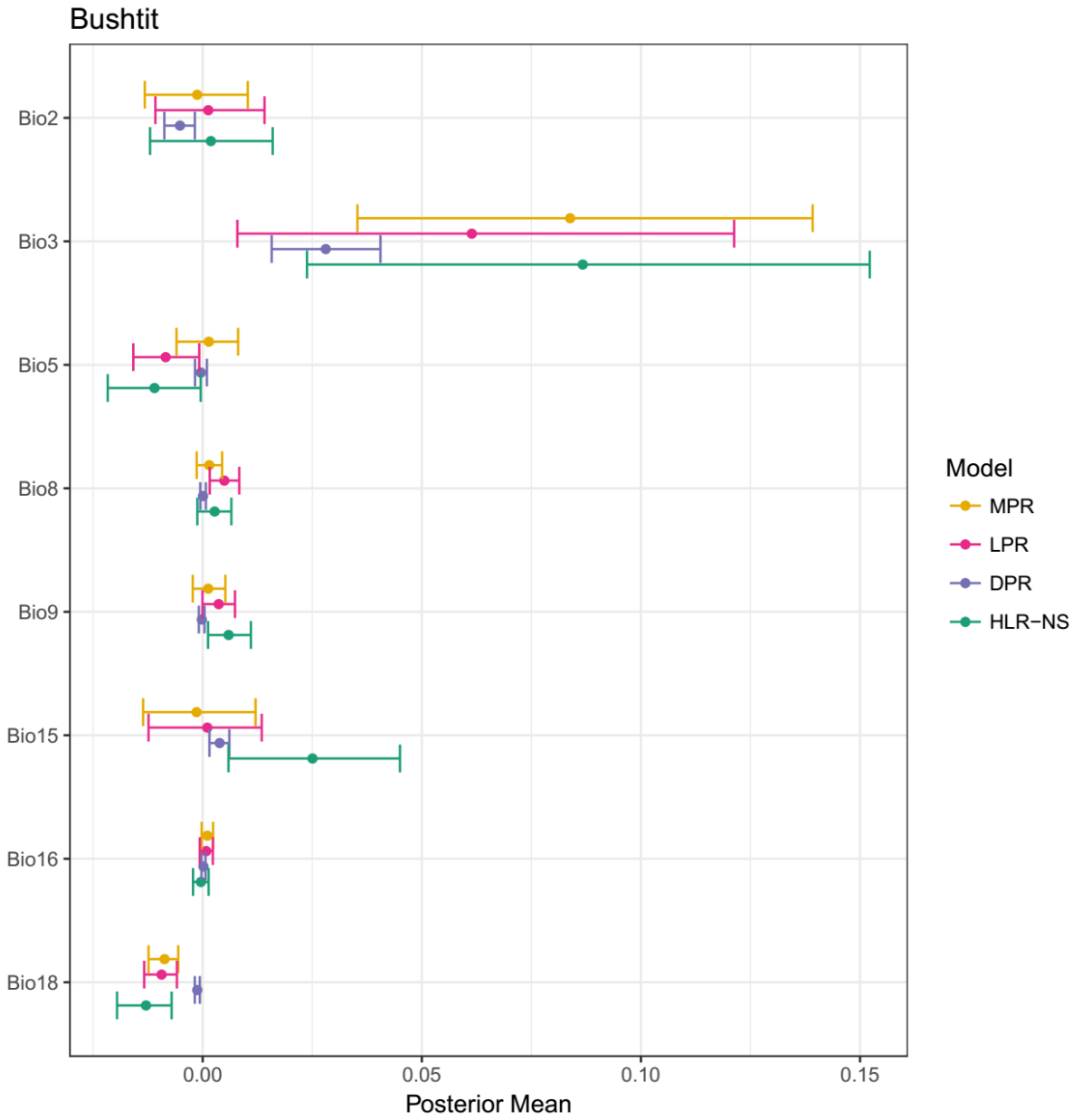


Figure 83: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the bushtit, *Psaltriparus minimus*.

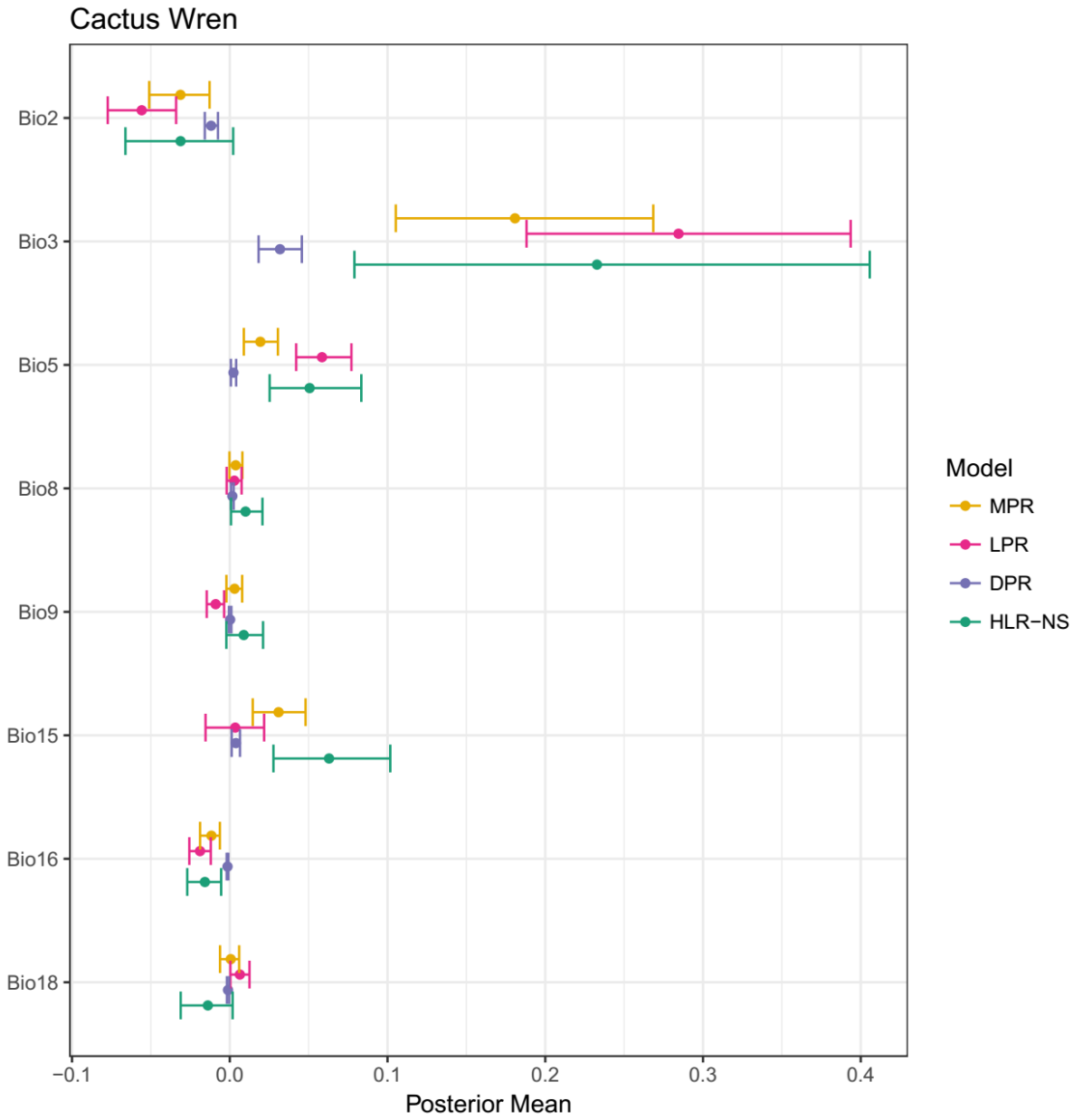


Figure 84: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cactus wren, *Campylorhynchus brunneicapillus*.

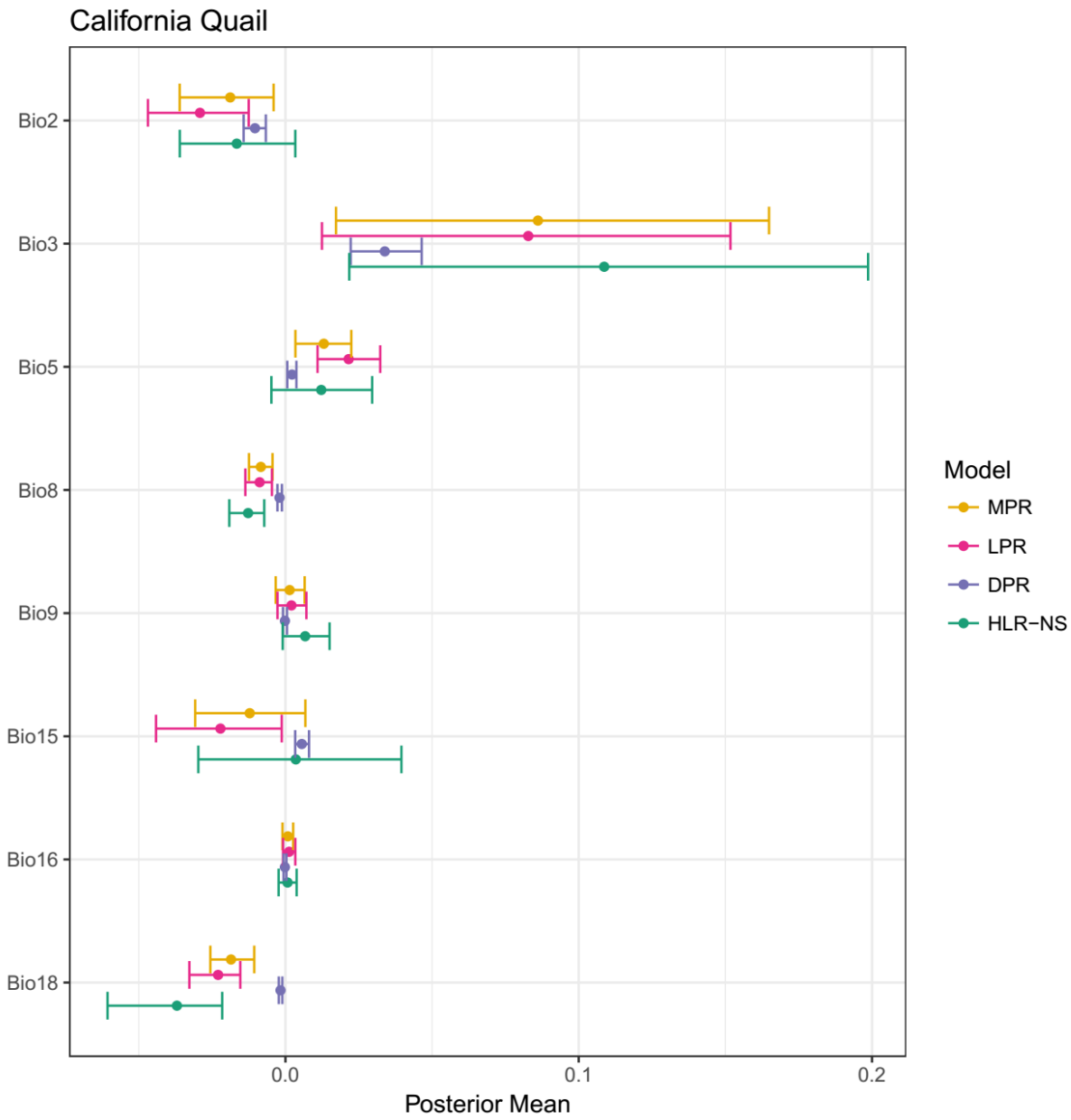


Figure 85: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the California quail, *Callipepla californica*.

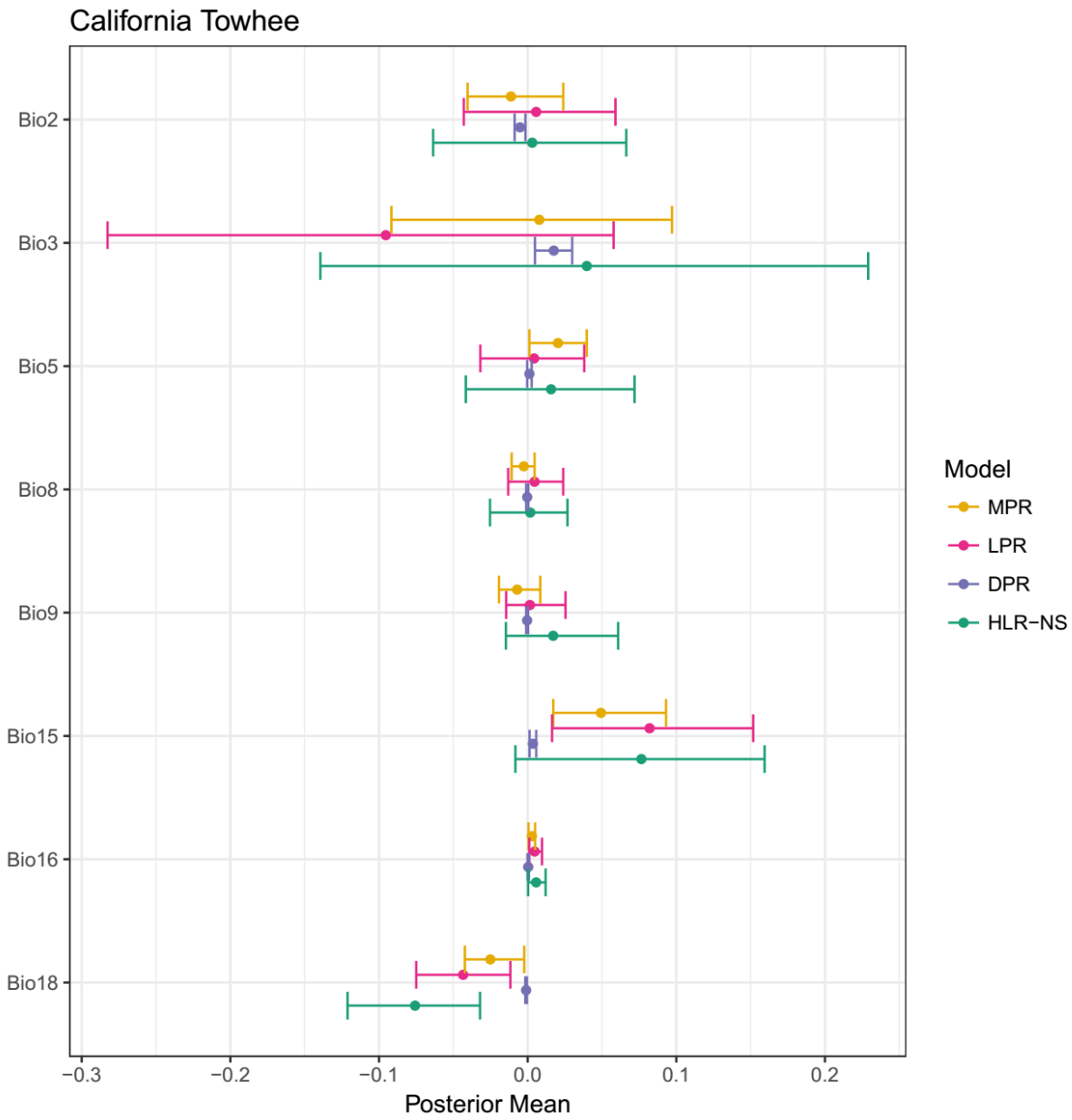


Figure 86: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the California towhee, *Melozone crissalis*.

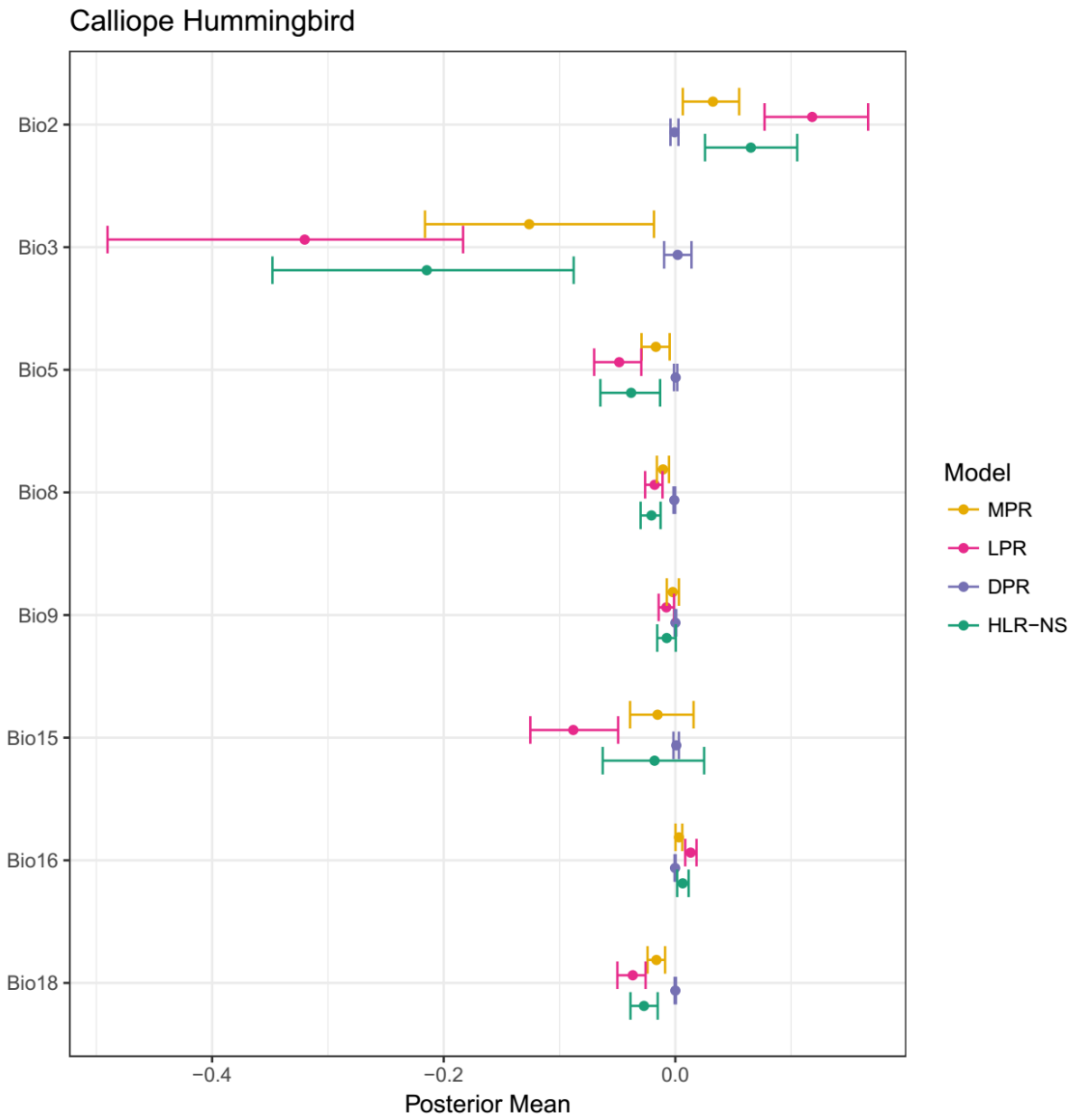


Figure 87: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the calliope hummingbird, *Selasphorus calliope*.

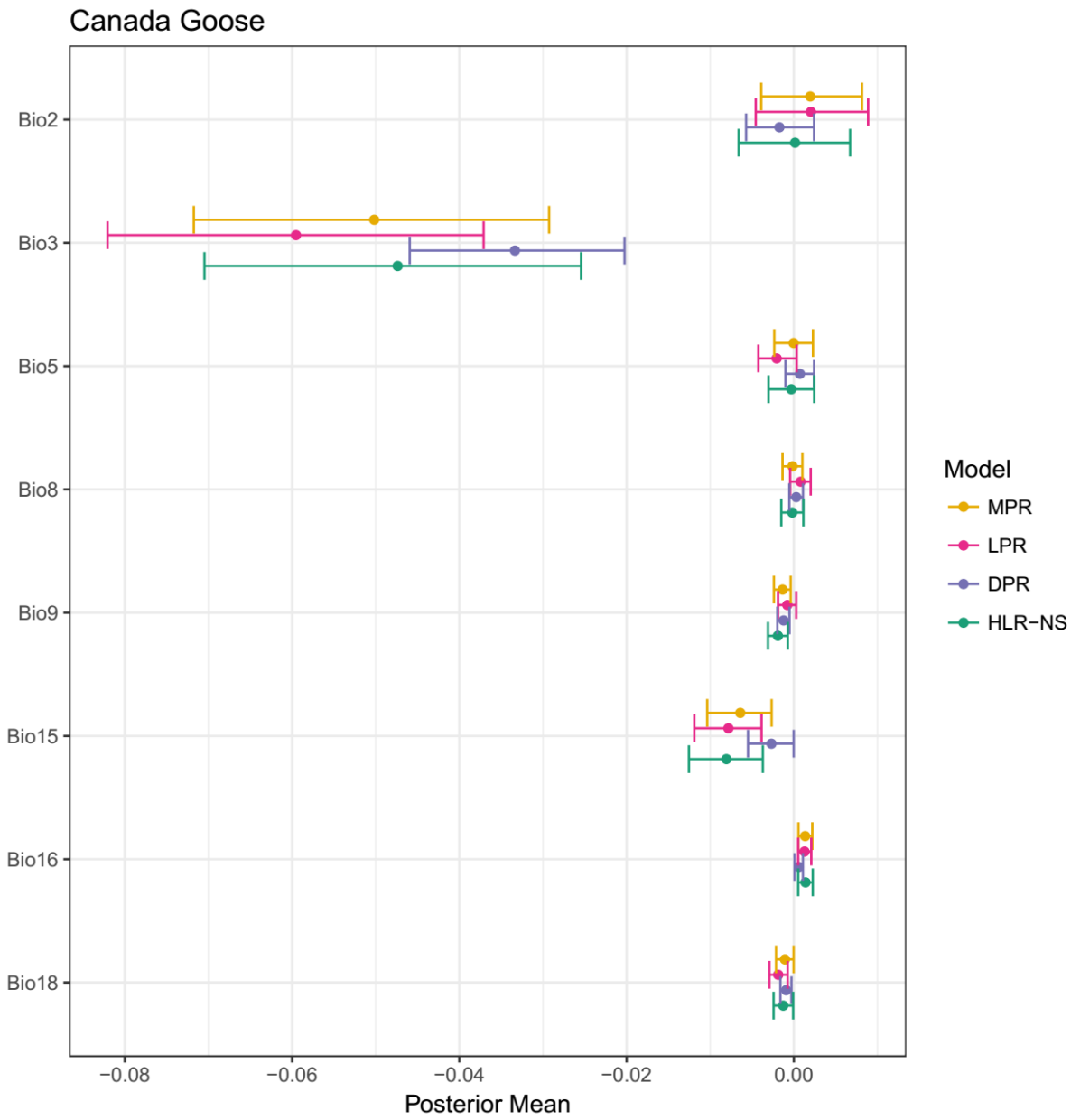


Figure 88: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Canada goose, *Branta canadensis*.

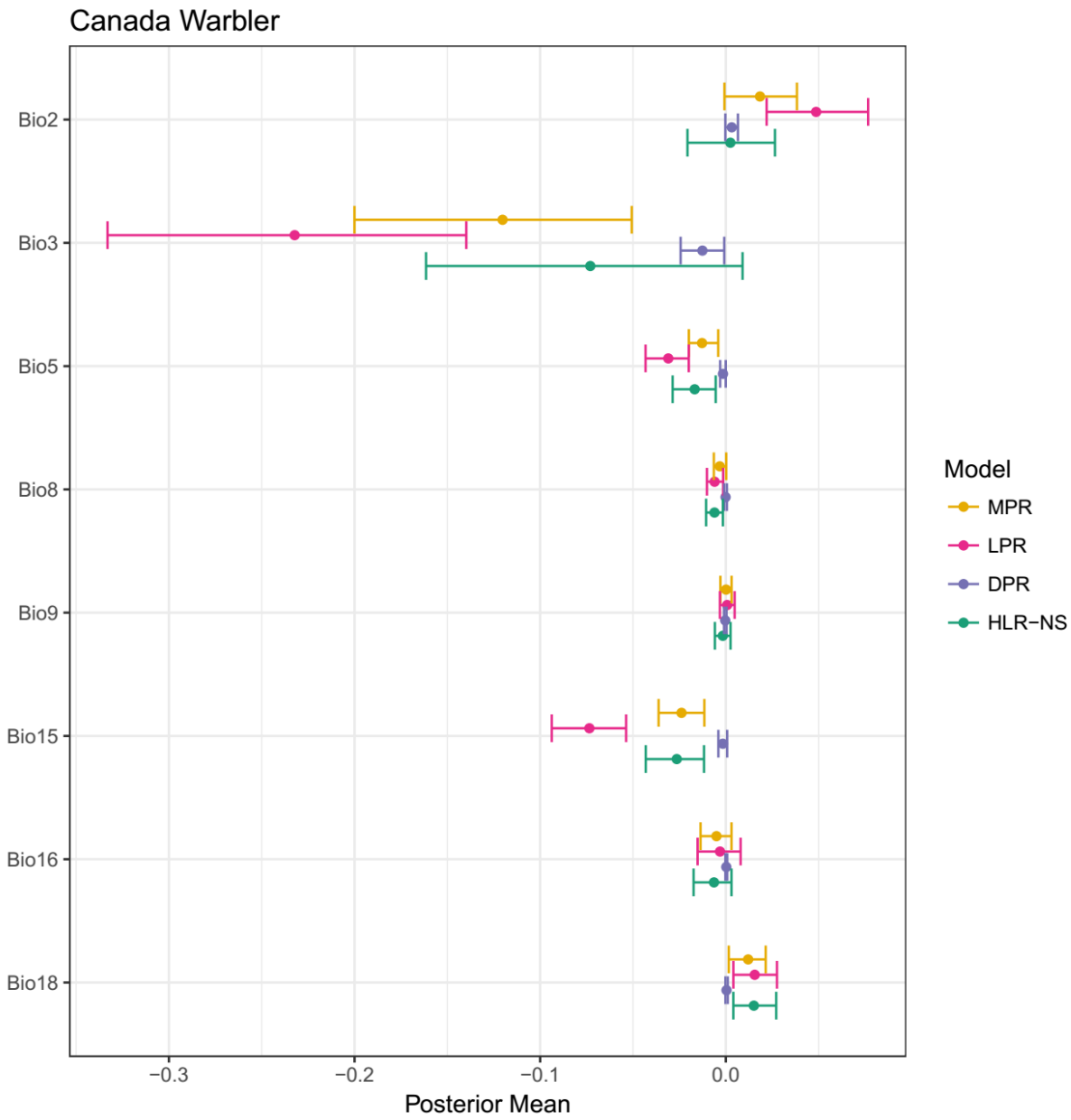


Figure 89: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Canada warbler, *Cardellina canadensis*.

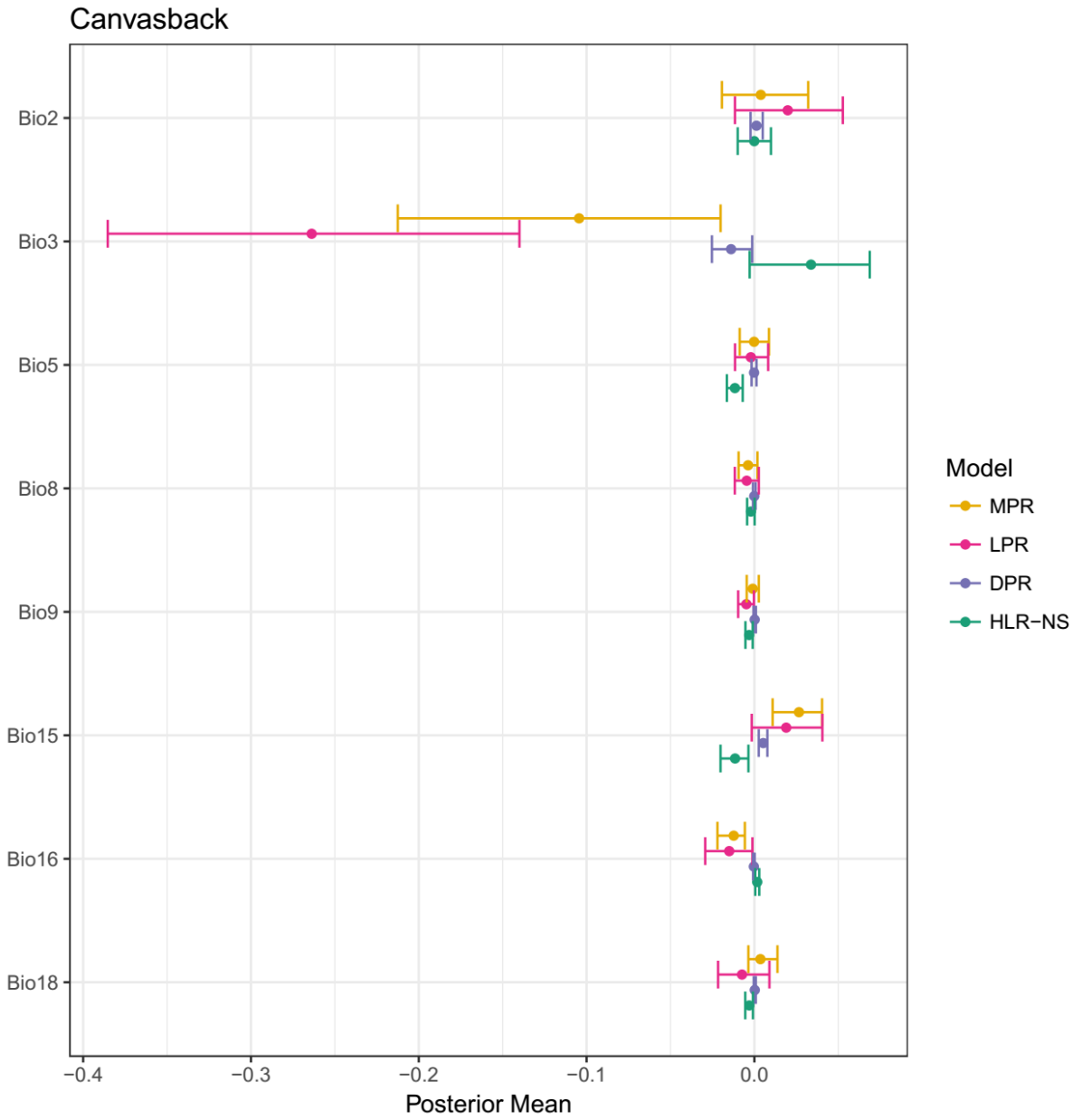


Figure 90: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the canvasback, *Aythya valisineria*.

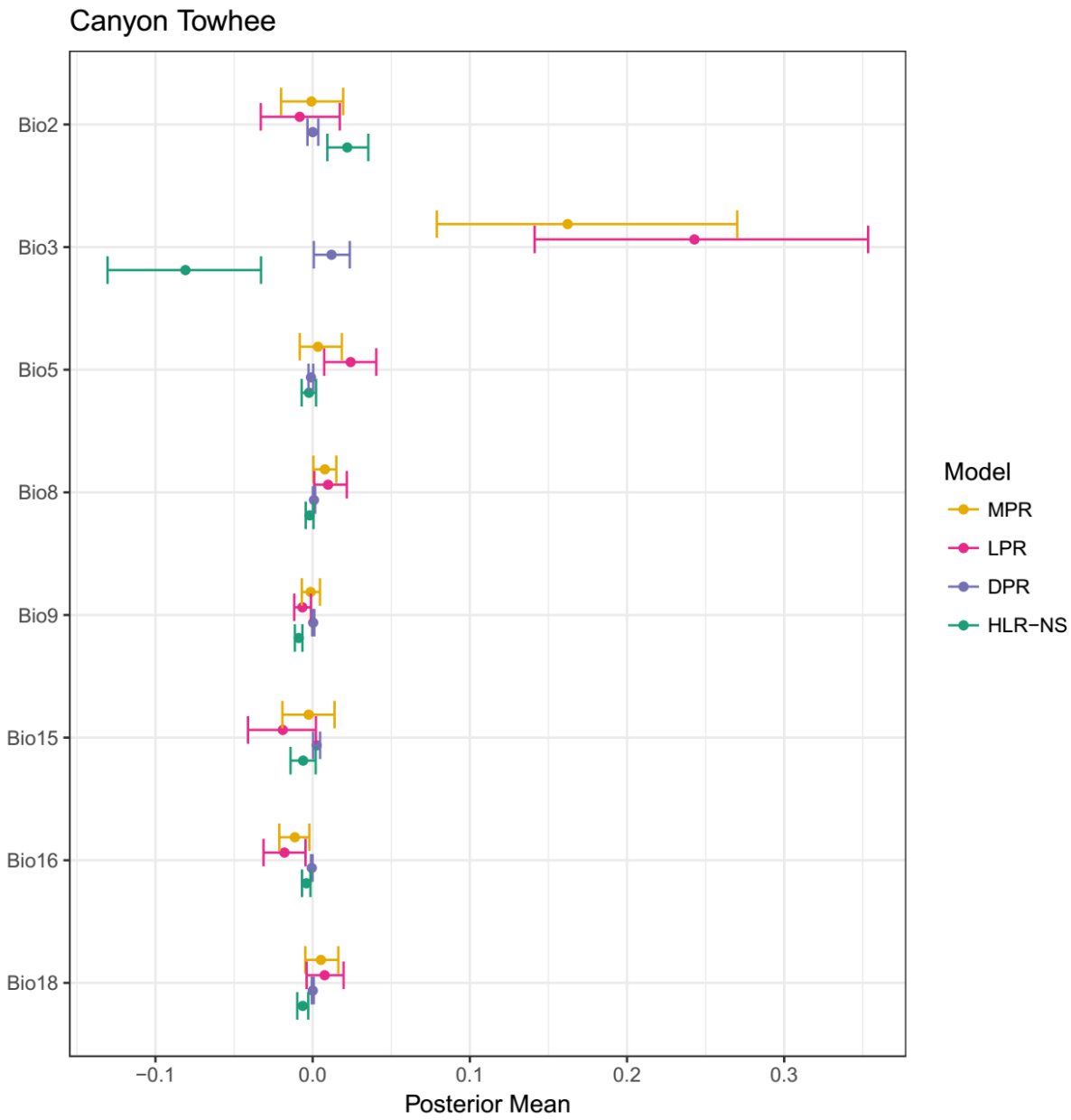


Figure 91: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the canyon towhee, *Melozone fusca*.

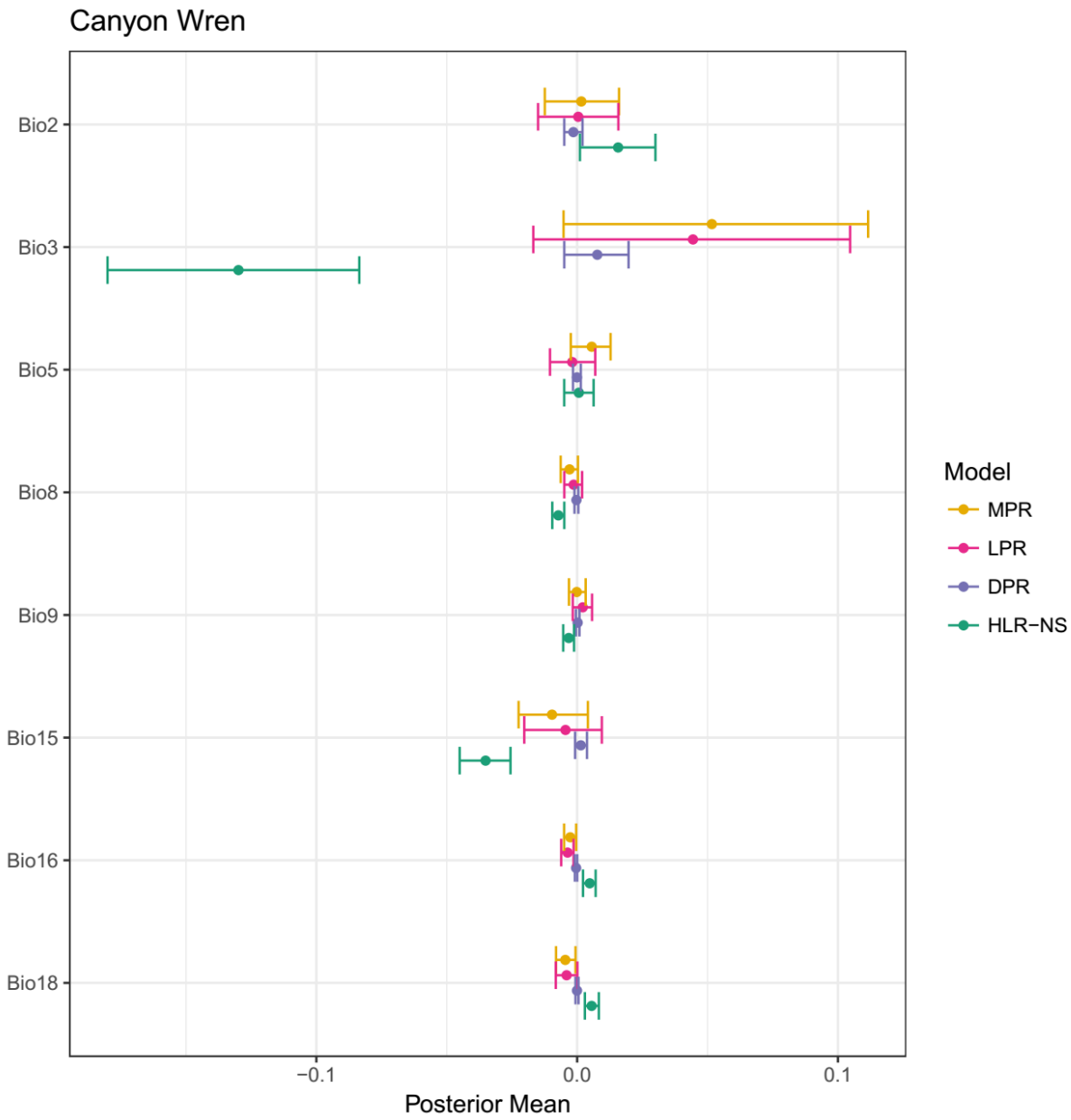


Figure 92: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Canyon wren, *Catherpes mexicanus*.

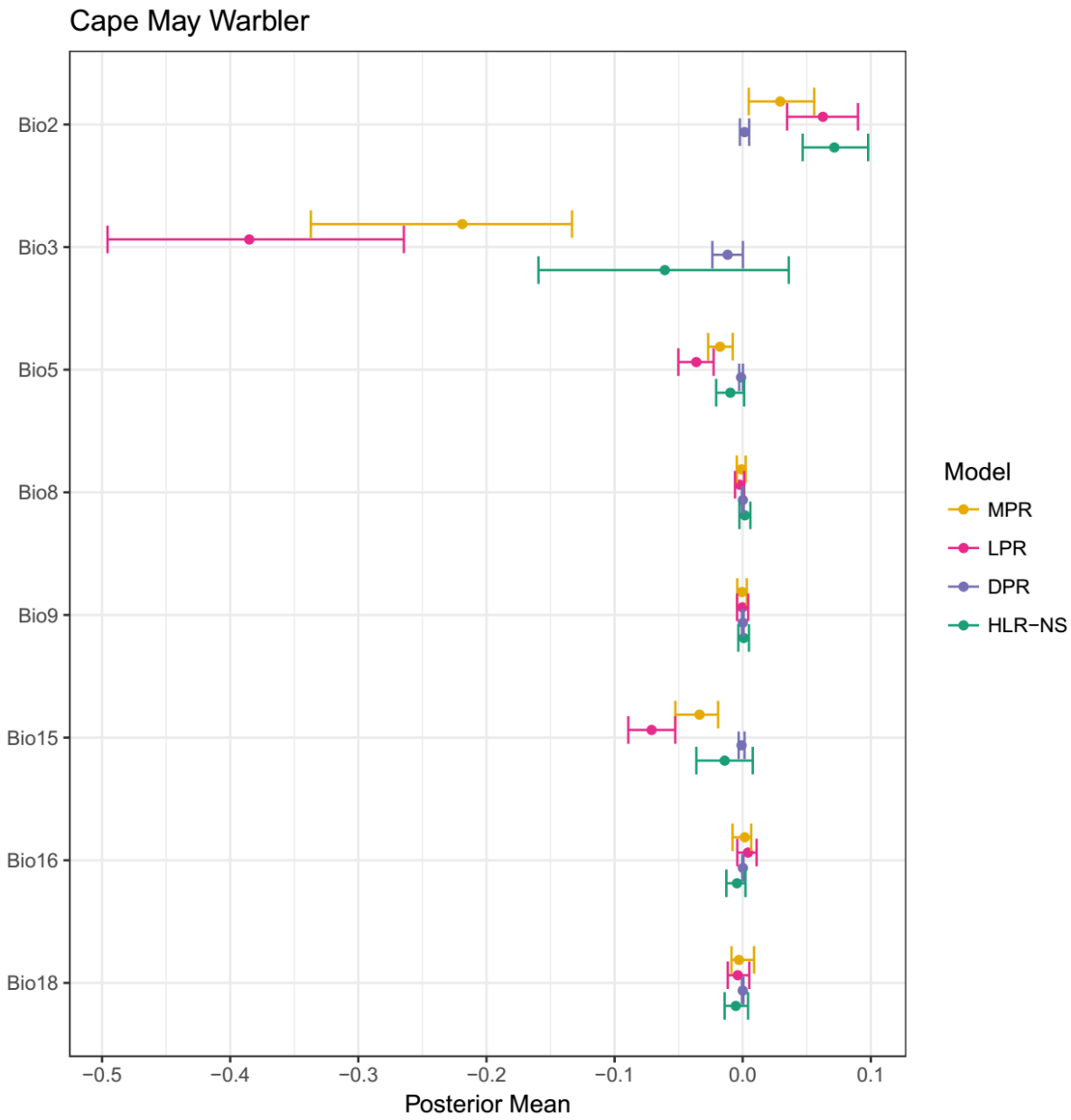


Figure 93: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cape May warbler, *Setophaga tigrina*.

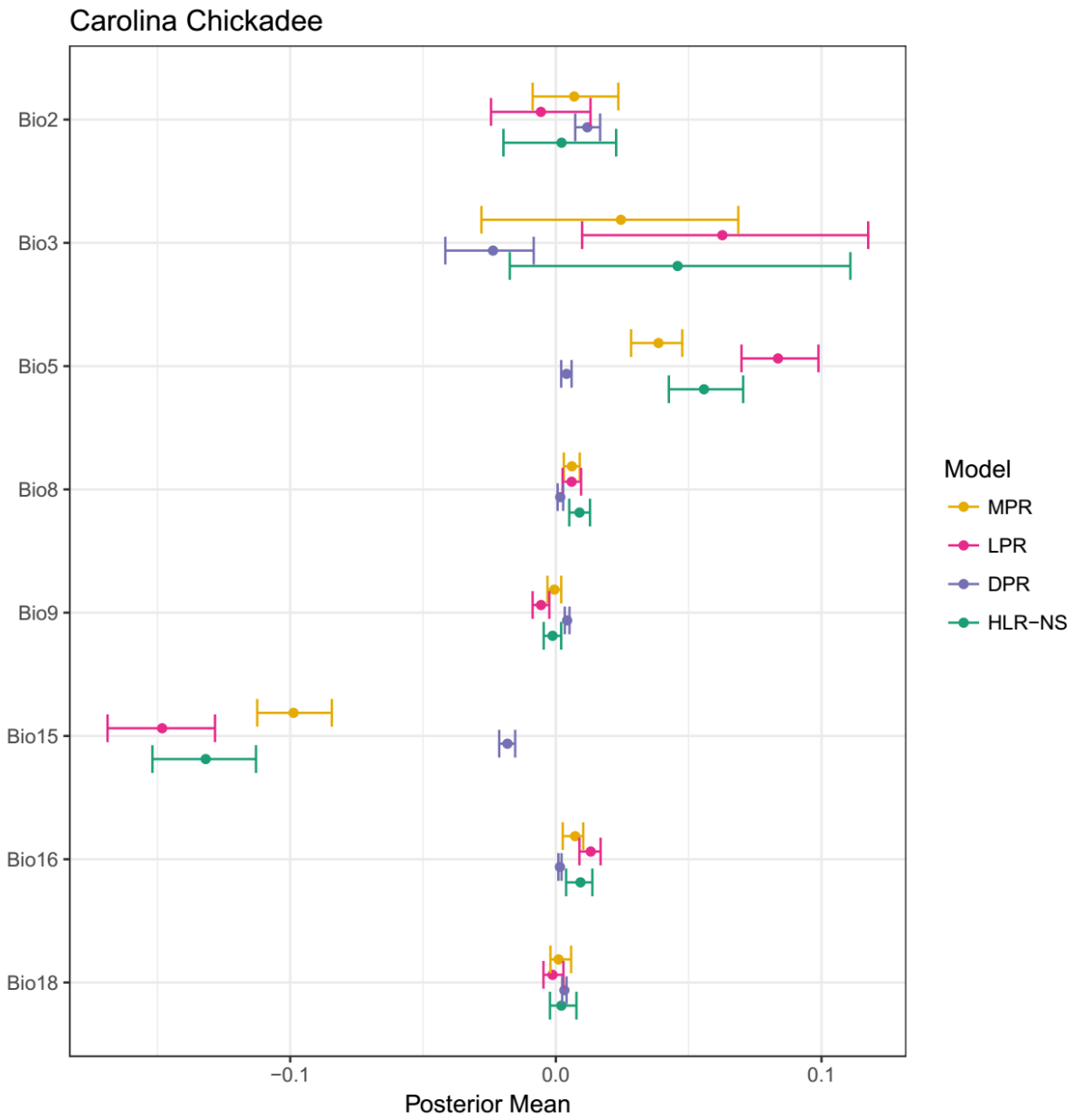


Figure 94: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Carolina chickadee, *Poecile carolinensis*.

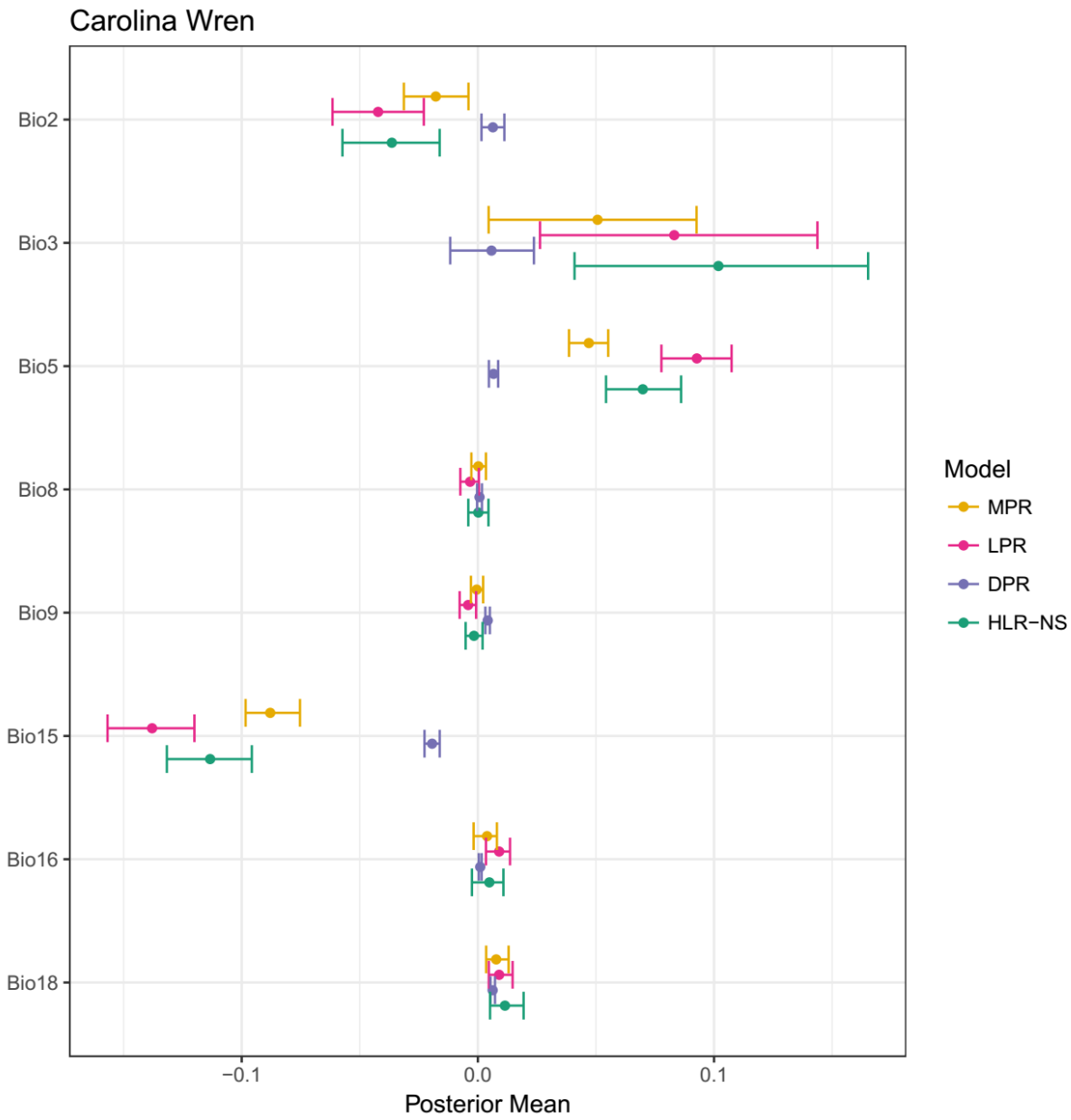


Figure 95: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Carolina wren, *Thryothorus ludovicianus*.

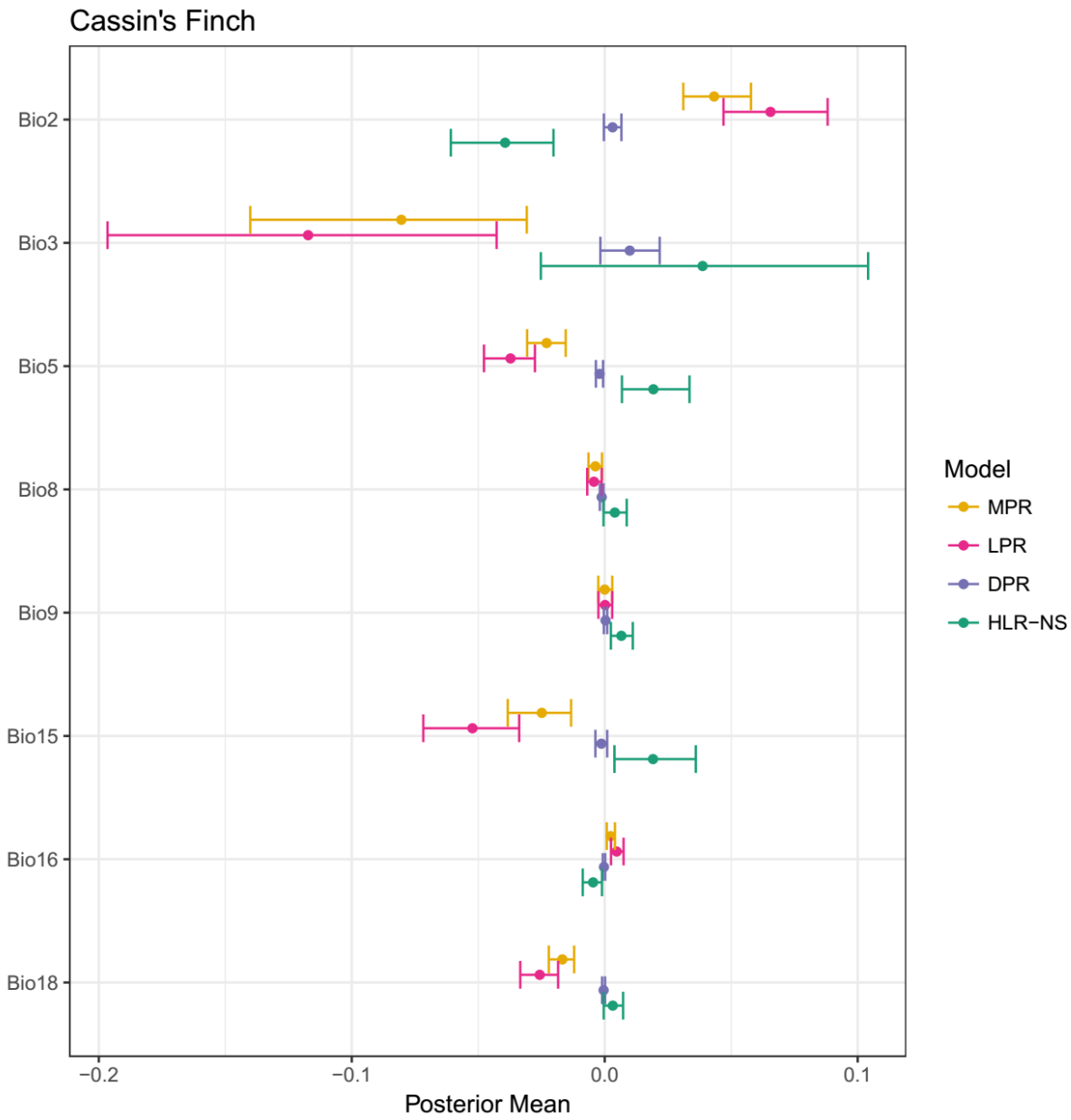


Figure 96: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cassin's finch, *Haemorhous cassinii*.

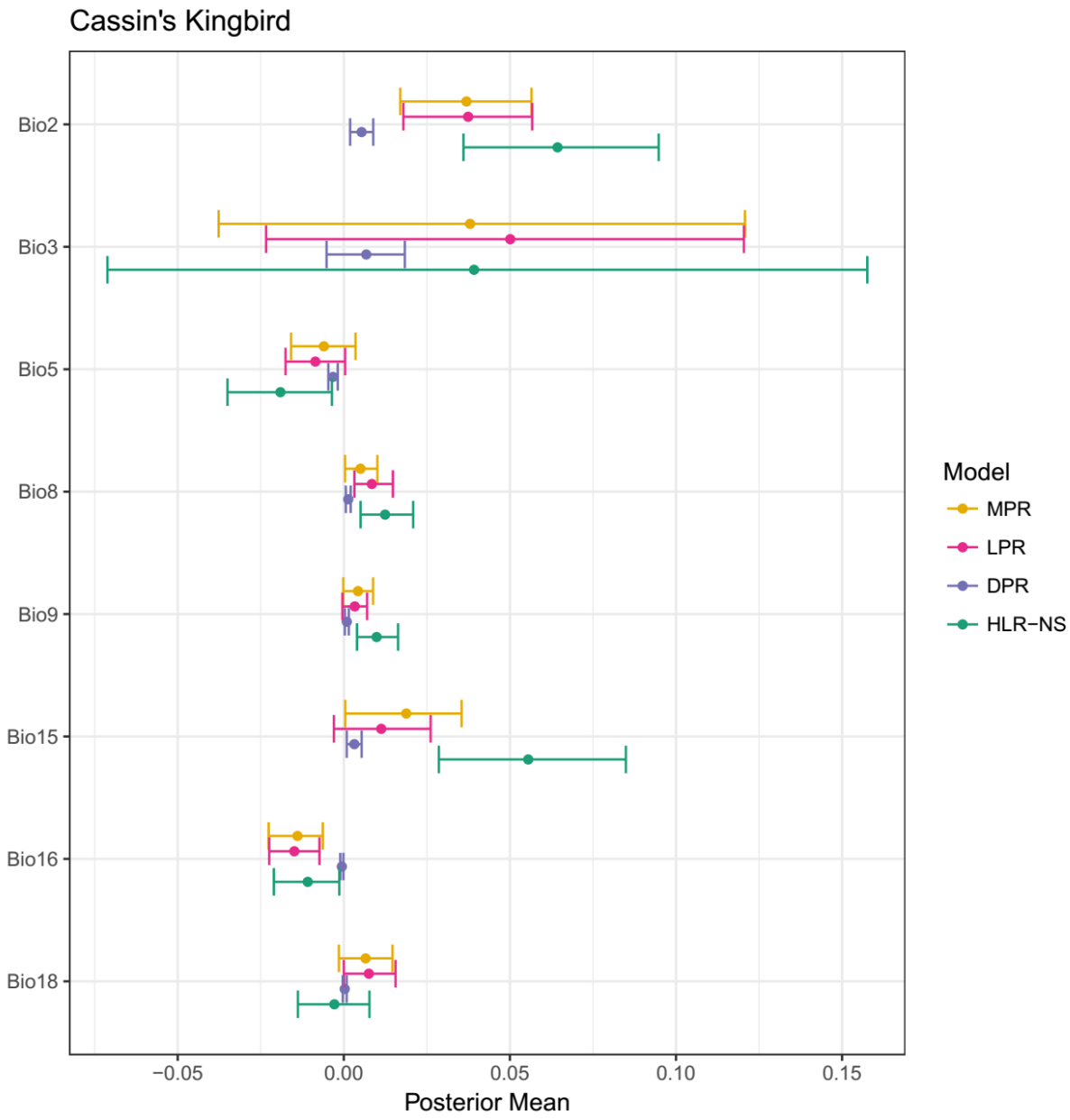


Figure 97: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cassin's kingbird, *Tyrannus vociferans*.

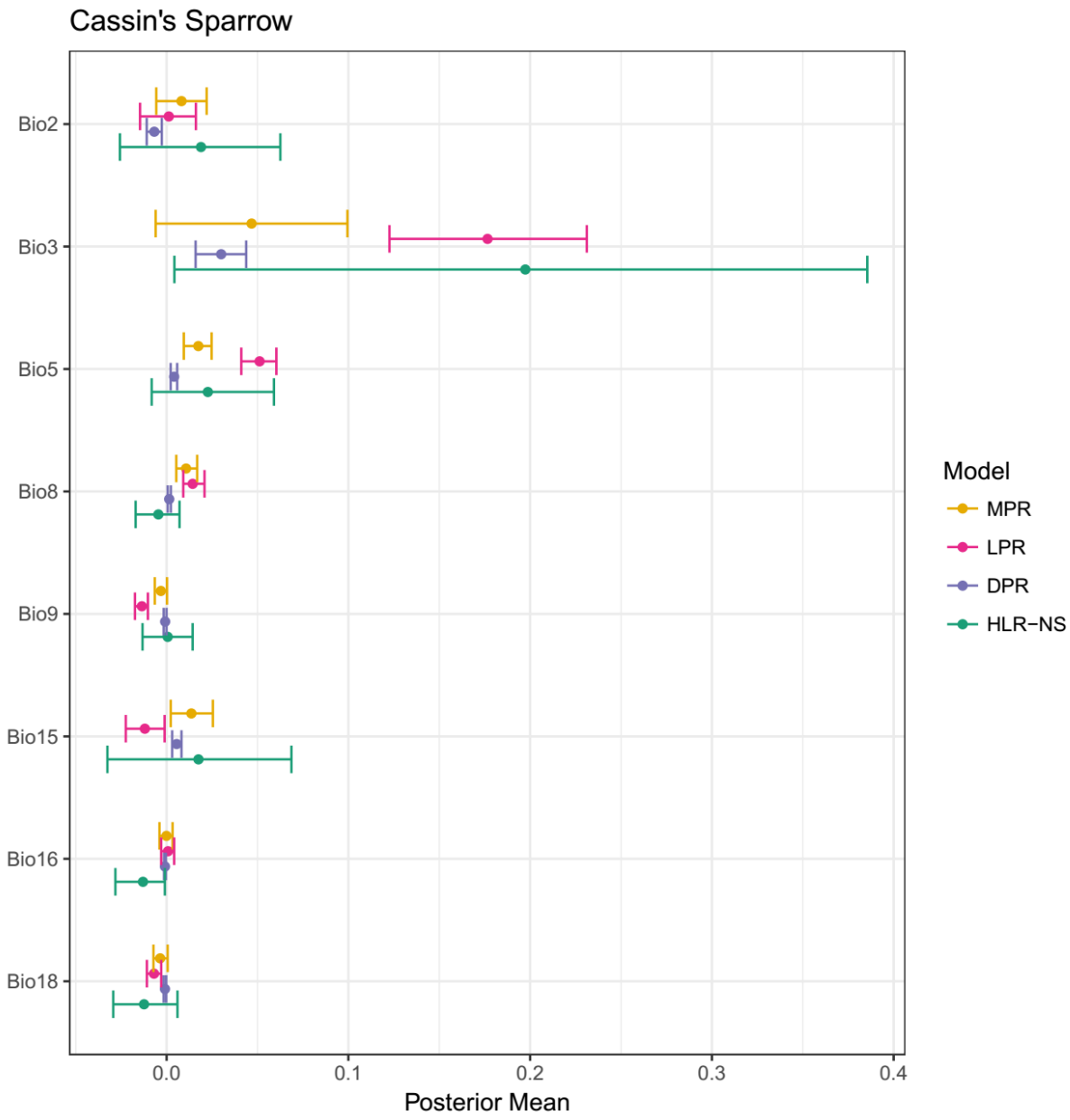


Figure 98: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cassin's sparrow, *Peucaea cassinii*.

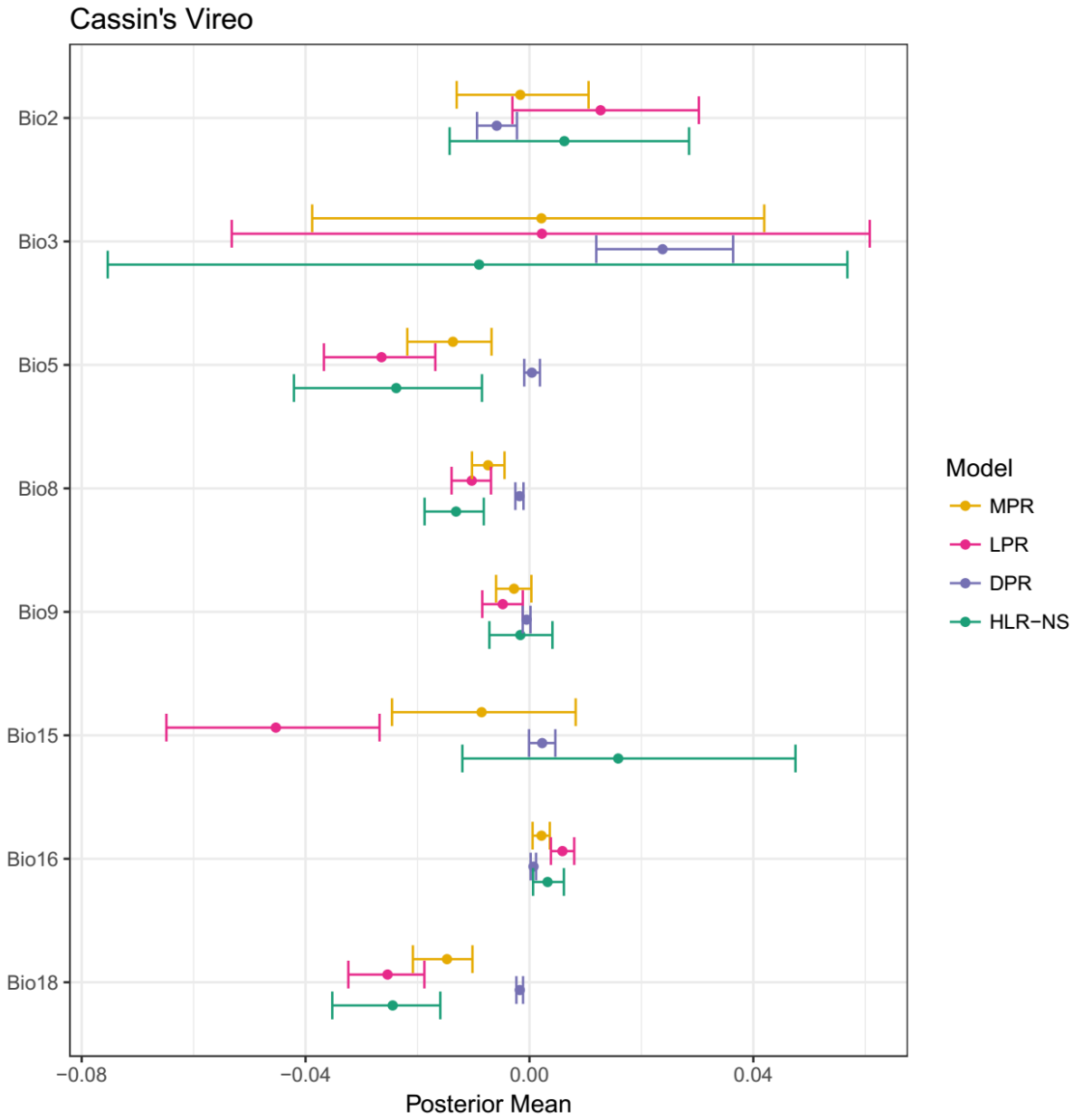


Figure 99: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cassin's vireo, *Vireo cassinii*.

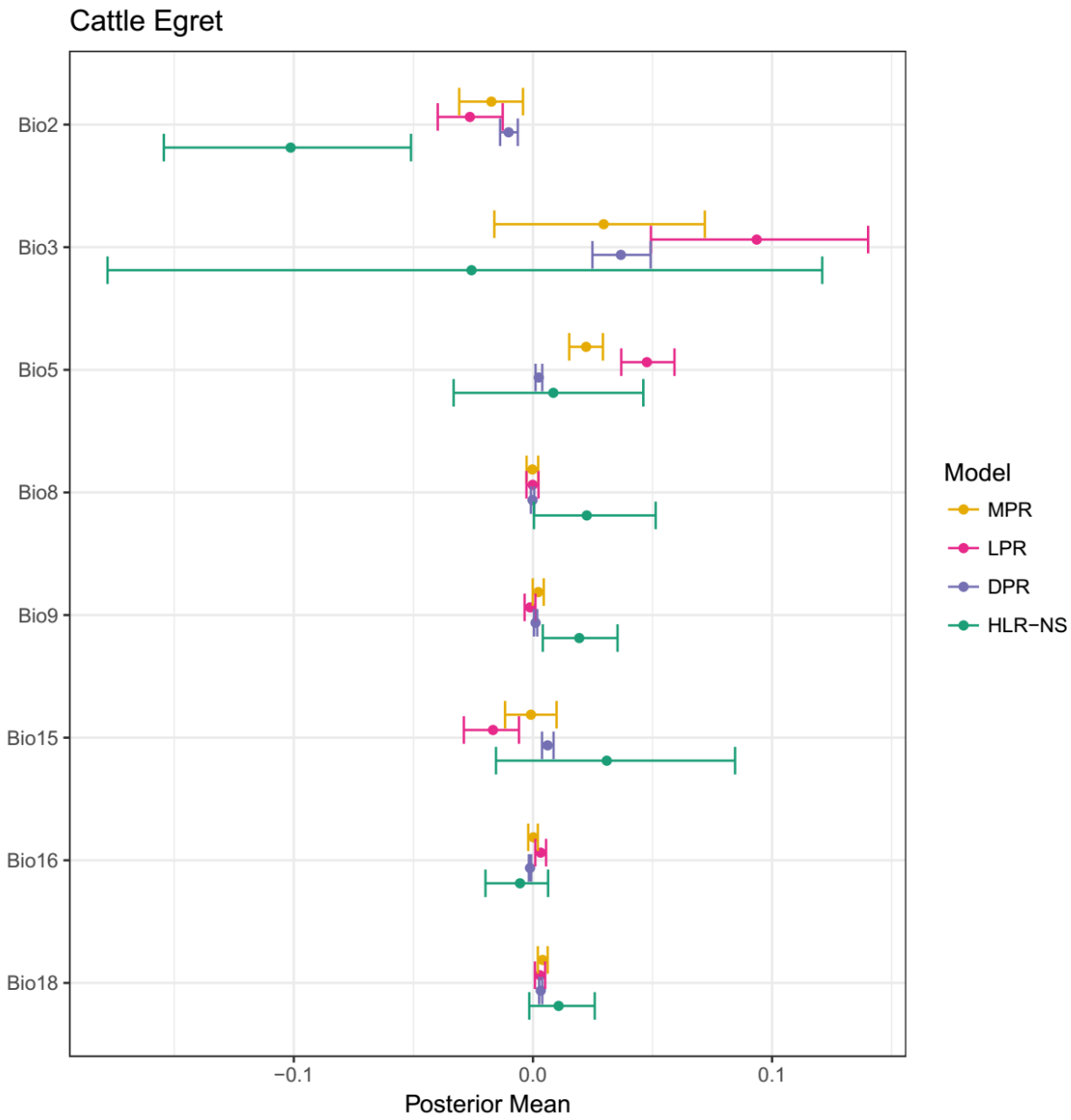


Figure 100: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cattle egret, *Bubulcus ibis*.

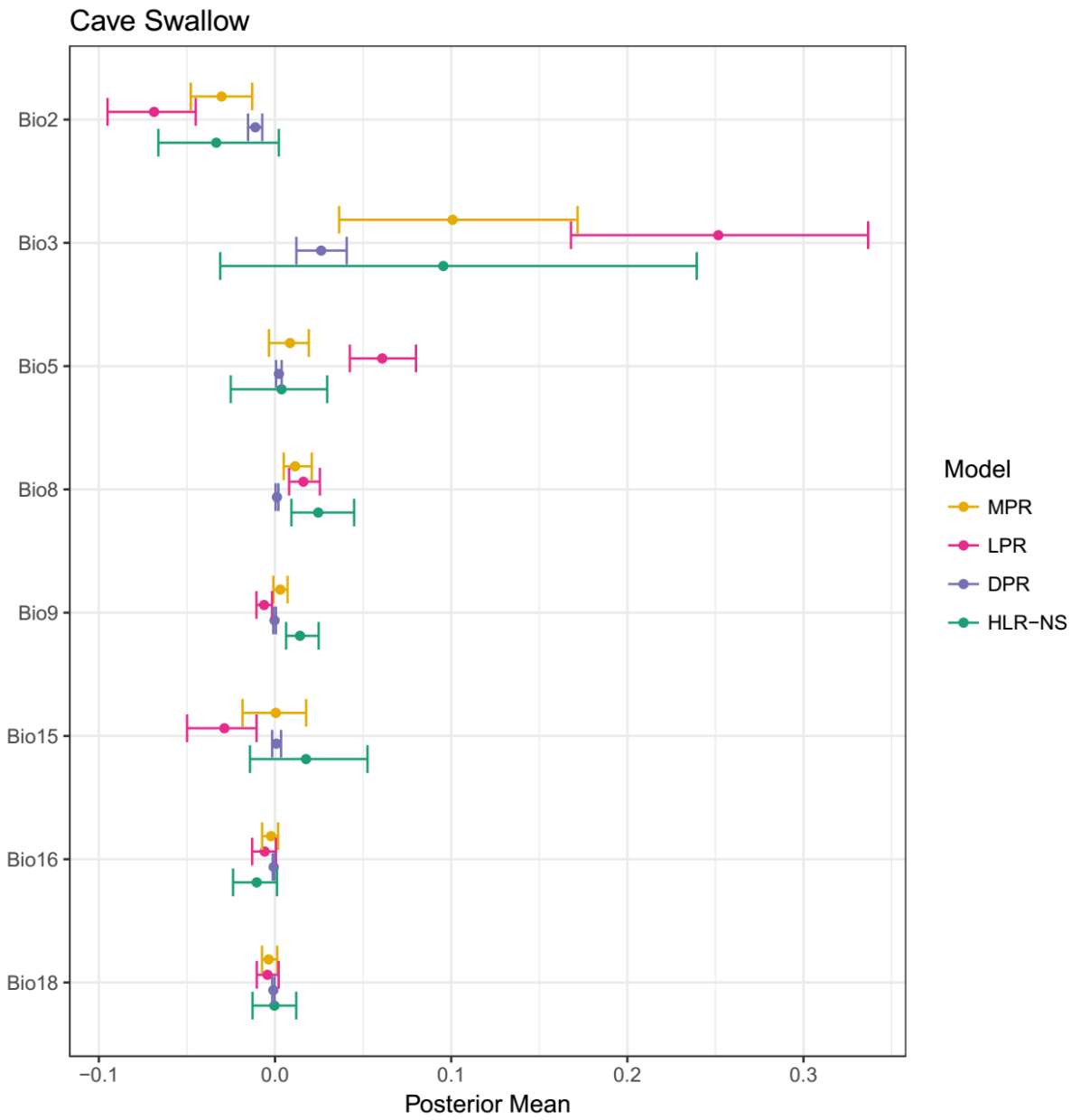


Figure 101: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cave swallow, *Petrochelidon fulva*.

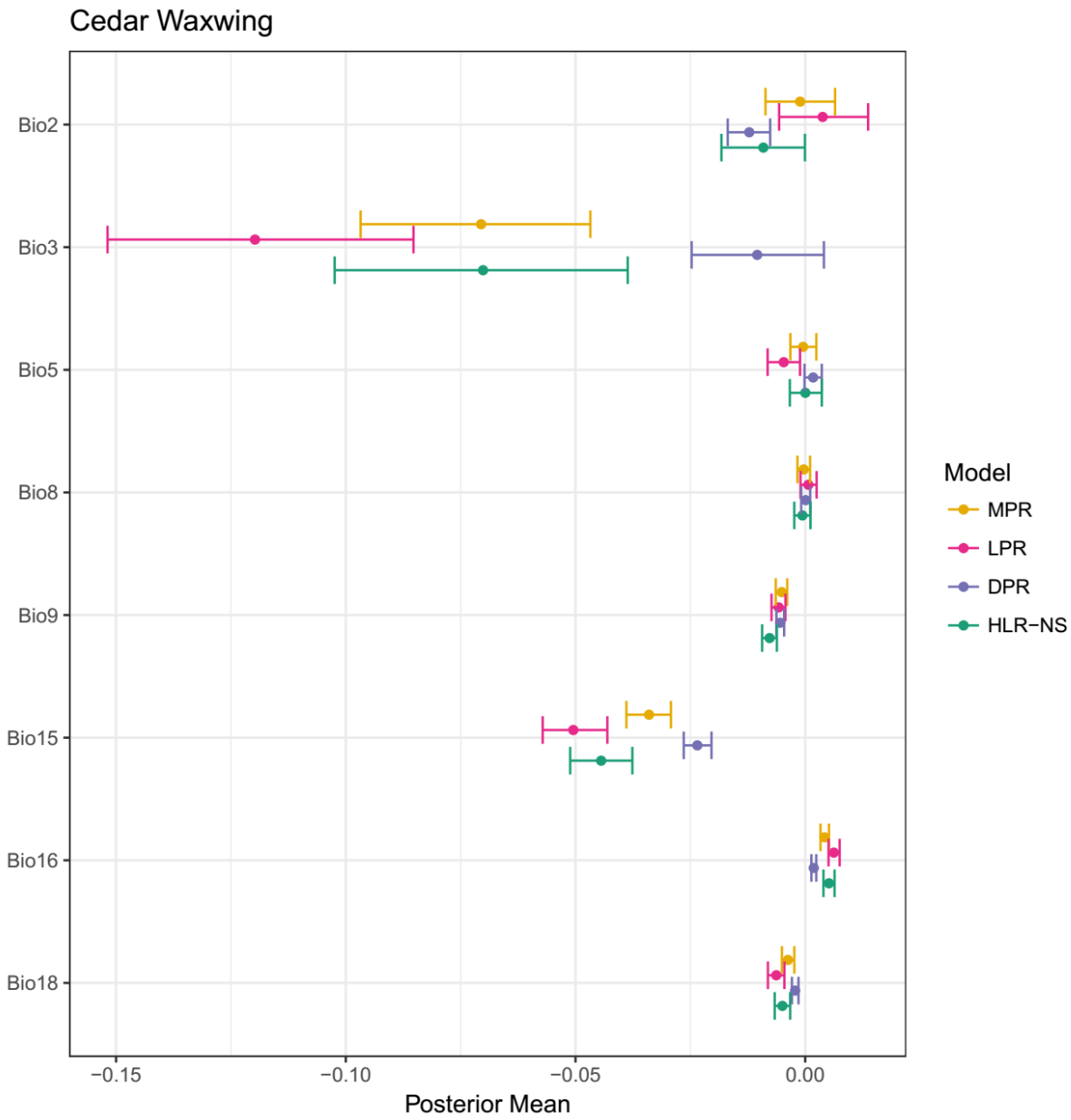


Figure 102: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cedar waxwing, *Bombycilla cedrorum*.

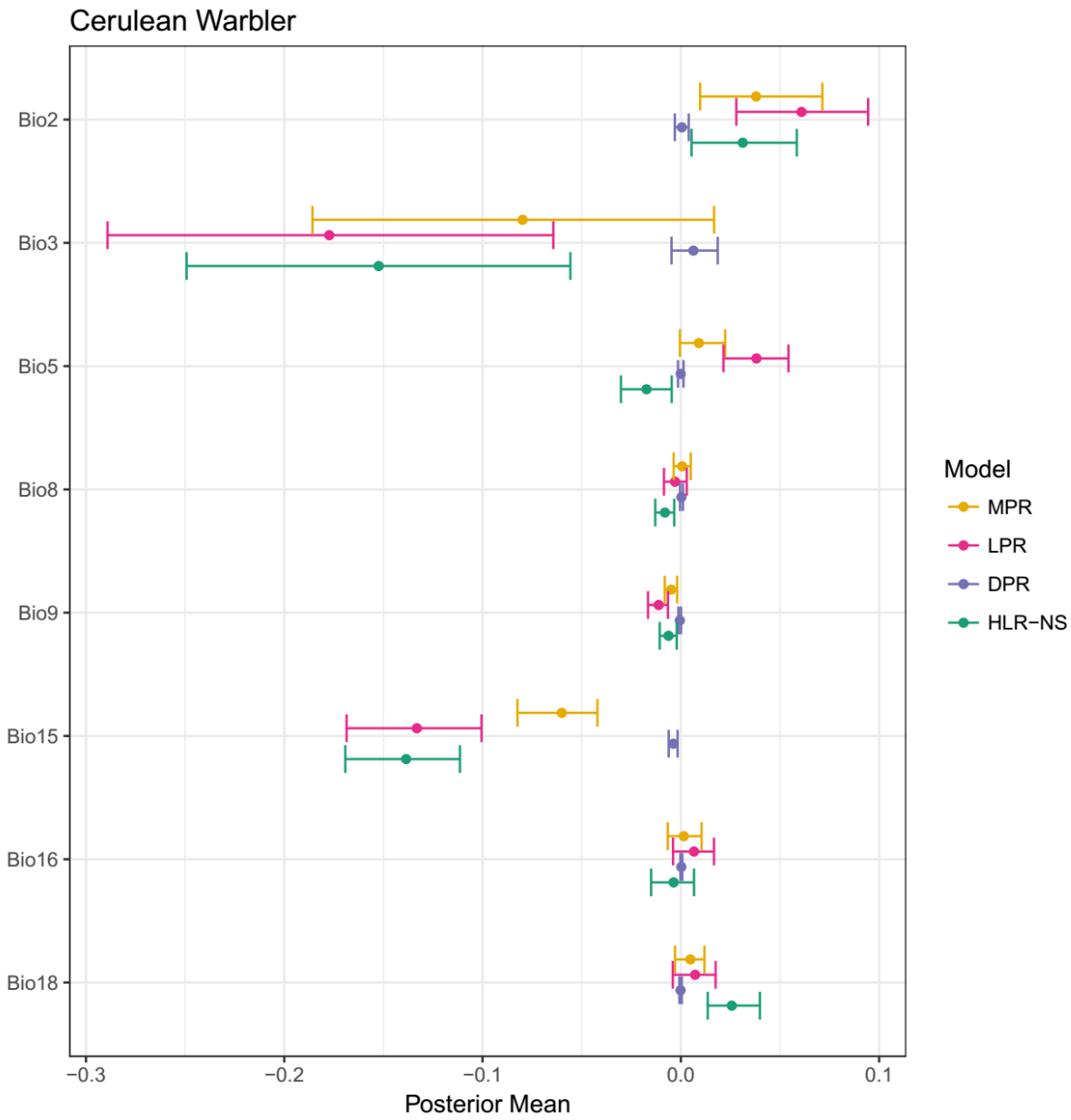


Figure 103: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cerulean warbler, *Setophaga cerulea*.

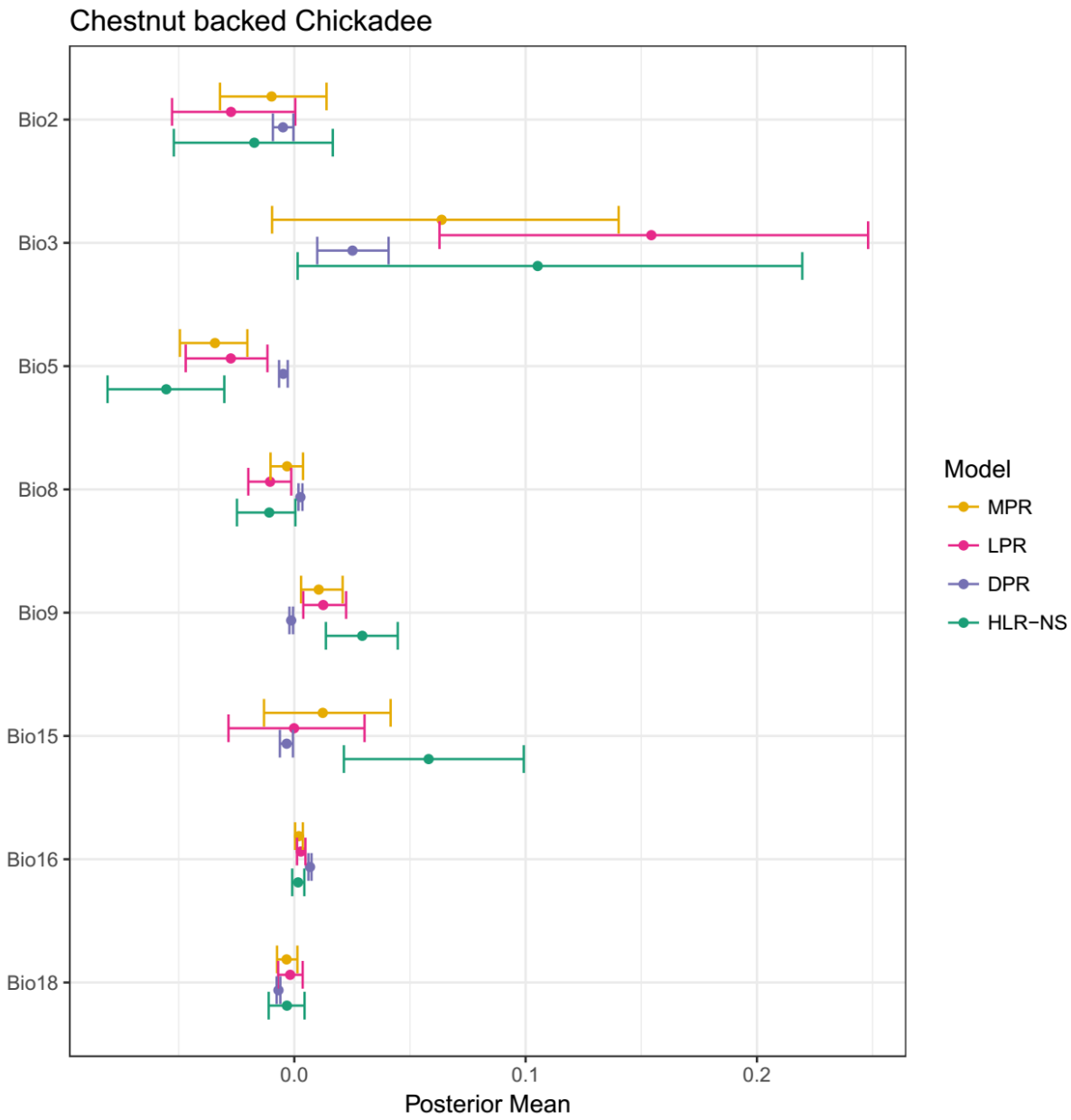


Figure 104: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chestnut-backed chickadee, *Poecile rufescens*.

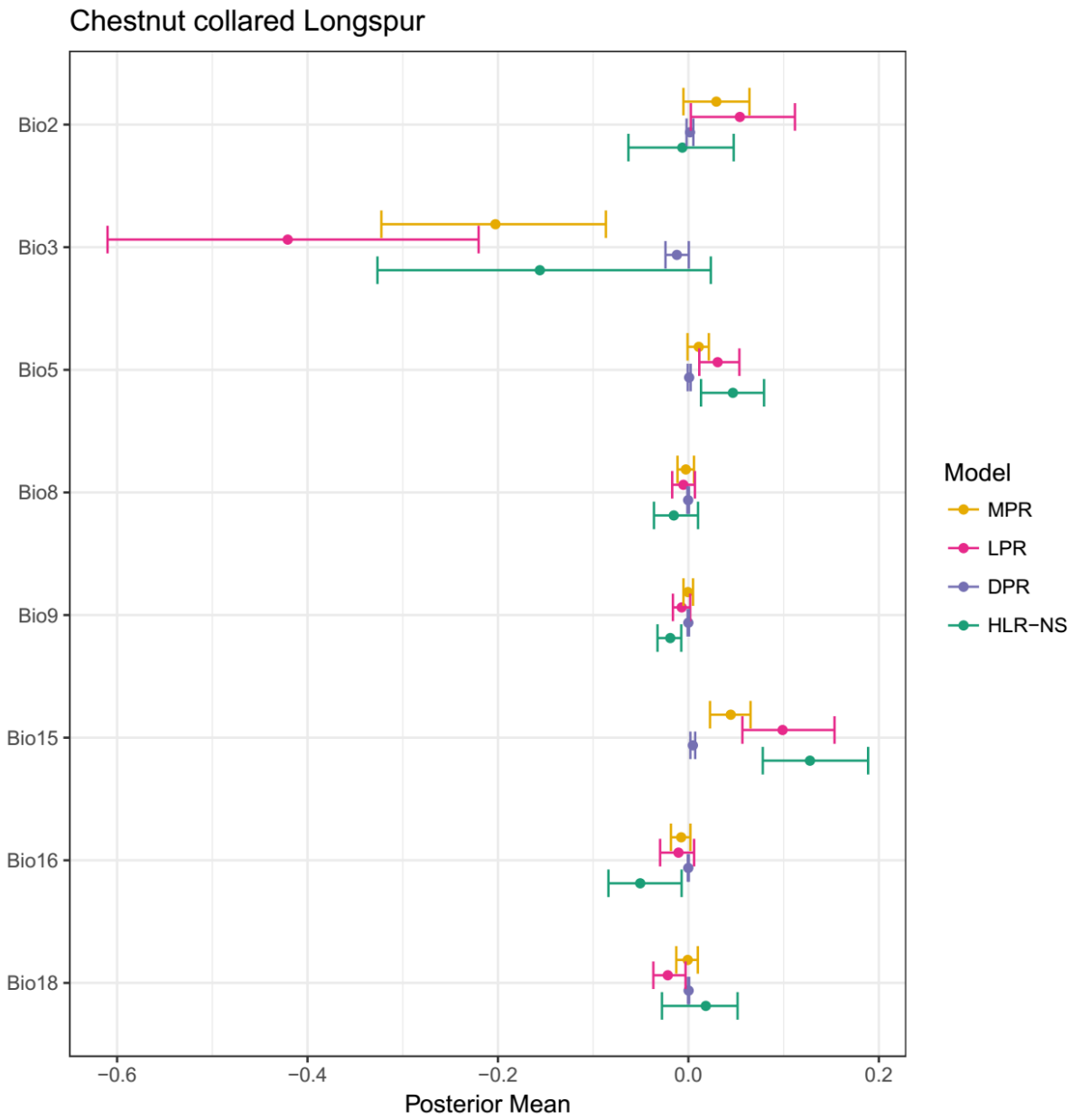


Figure 105: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chestnut-collared longspur, *Calcarius ornatus*.

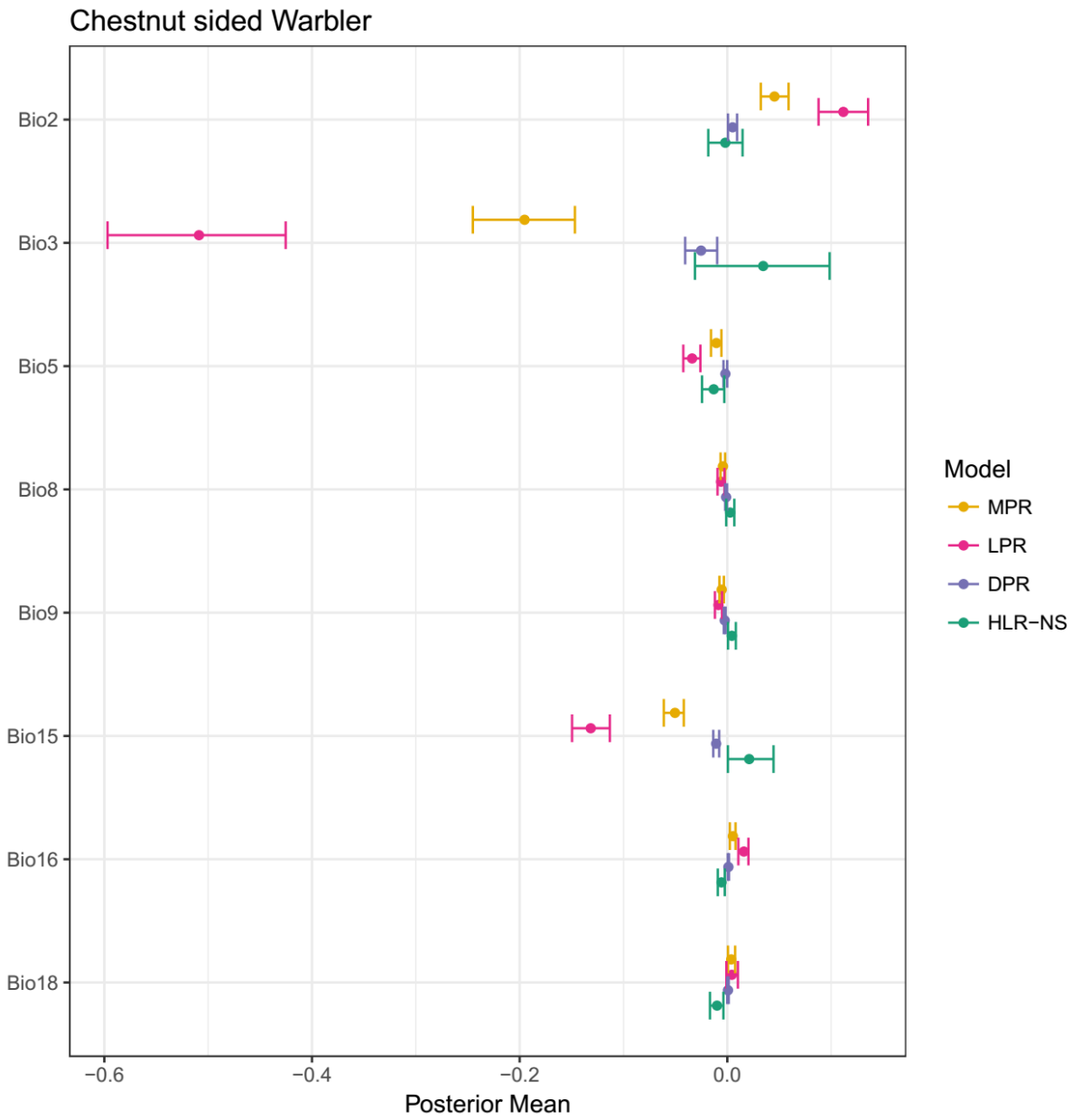


Figure 106: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chestnut-sided warbler, *Setophaga pensylvanica*.

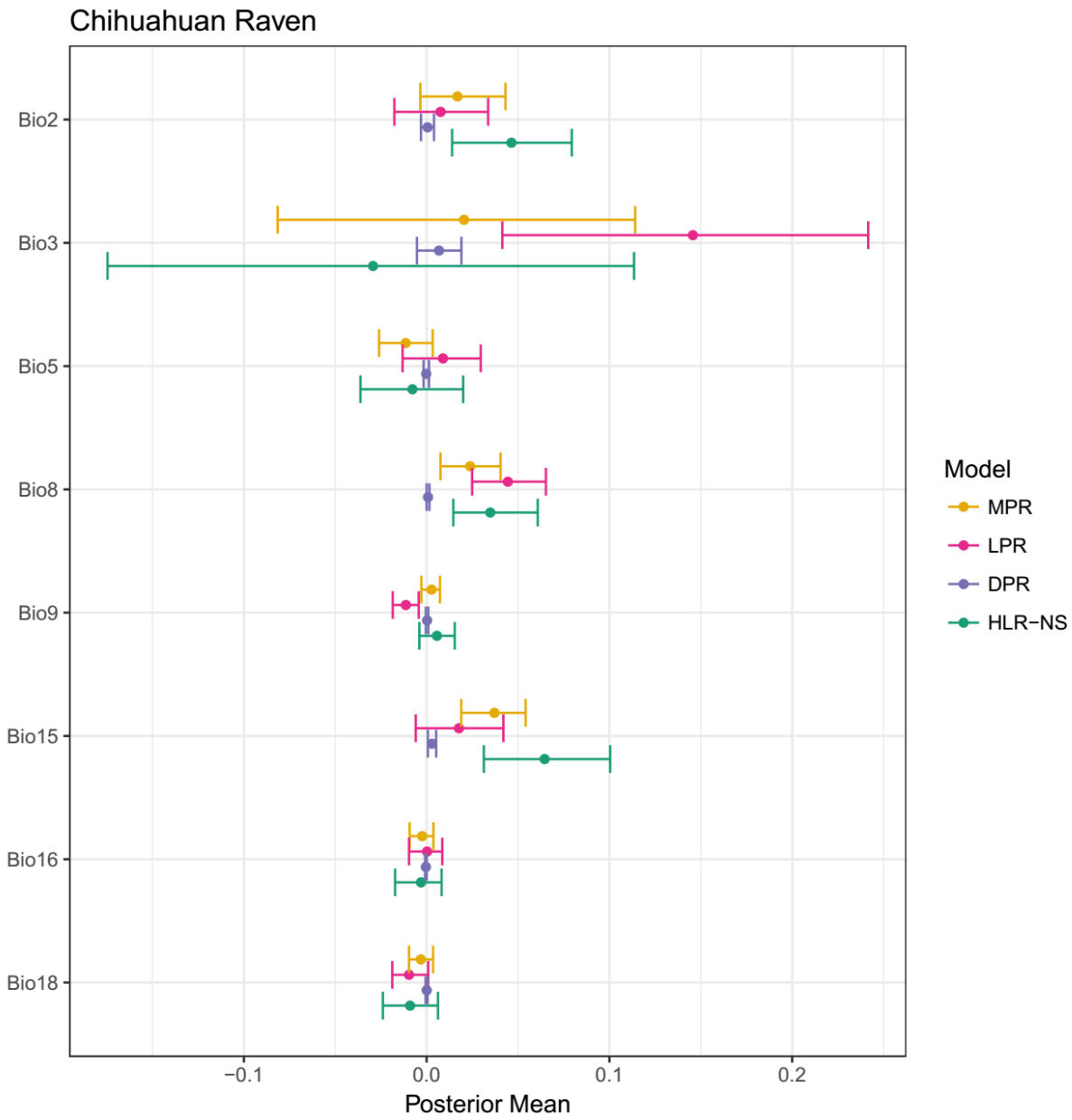


Figure 107: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Chihuahuan raven, *Corvus cryptoleucus*.

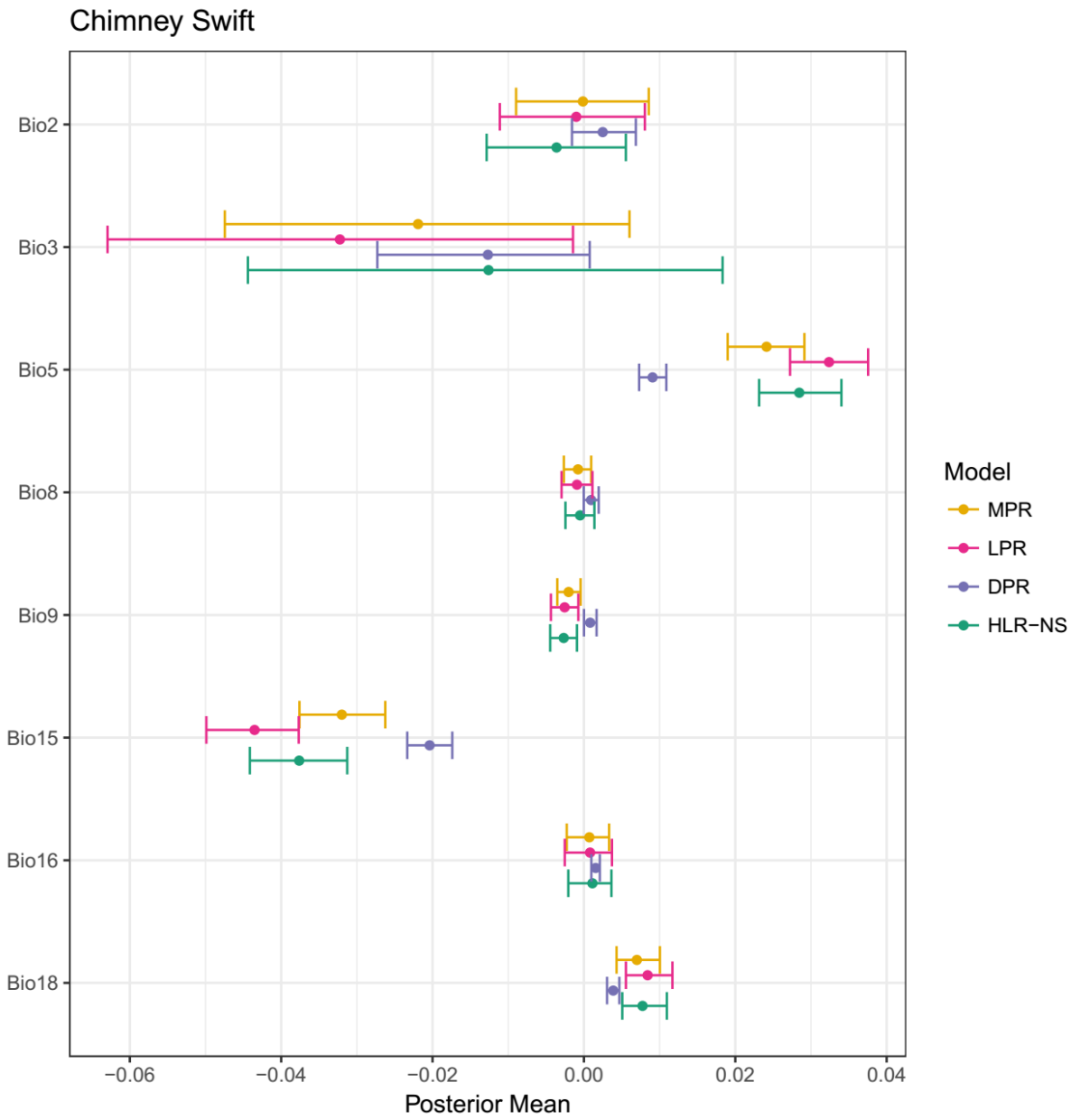


Figure 108: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chimney swift, *Chaetura pelagica*.

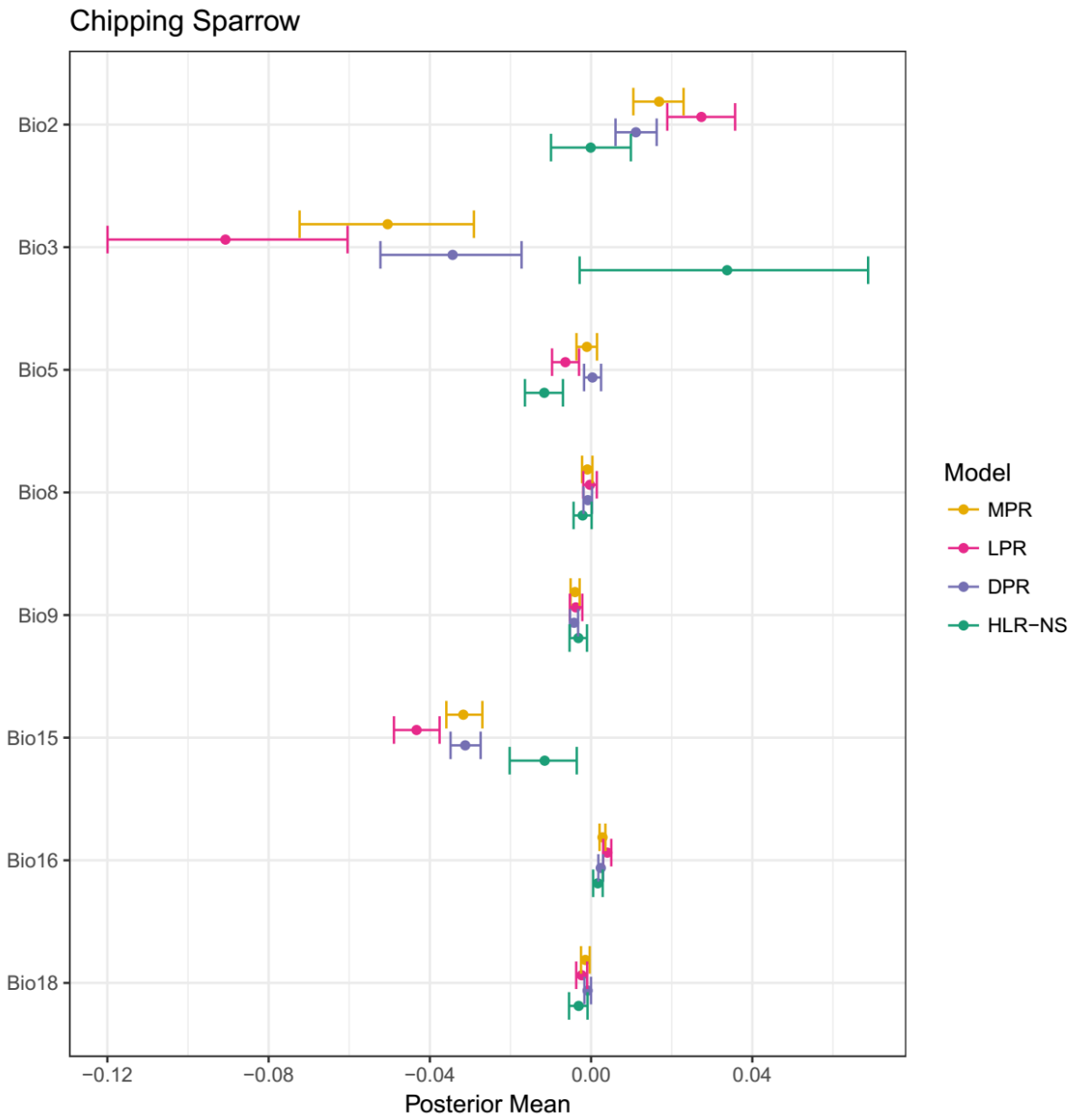


Figure 109: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chipping sparrow, *Spizella passerina*.

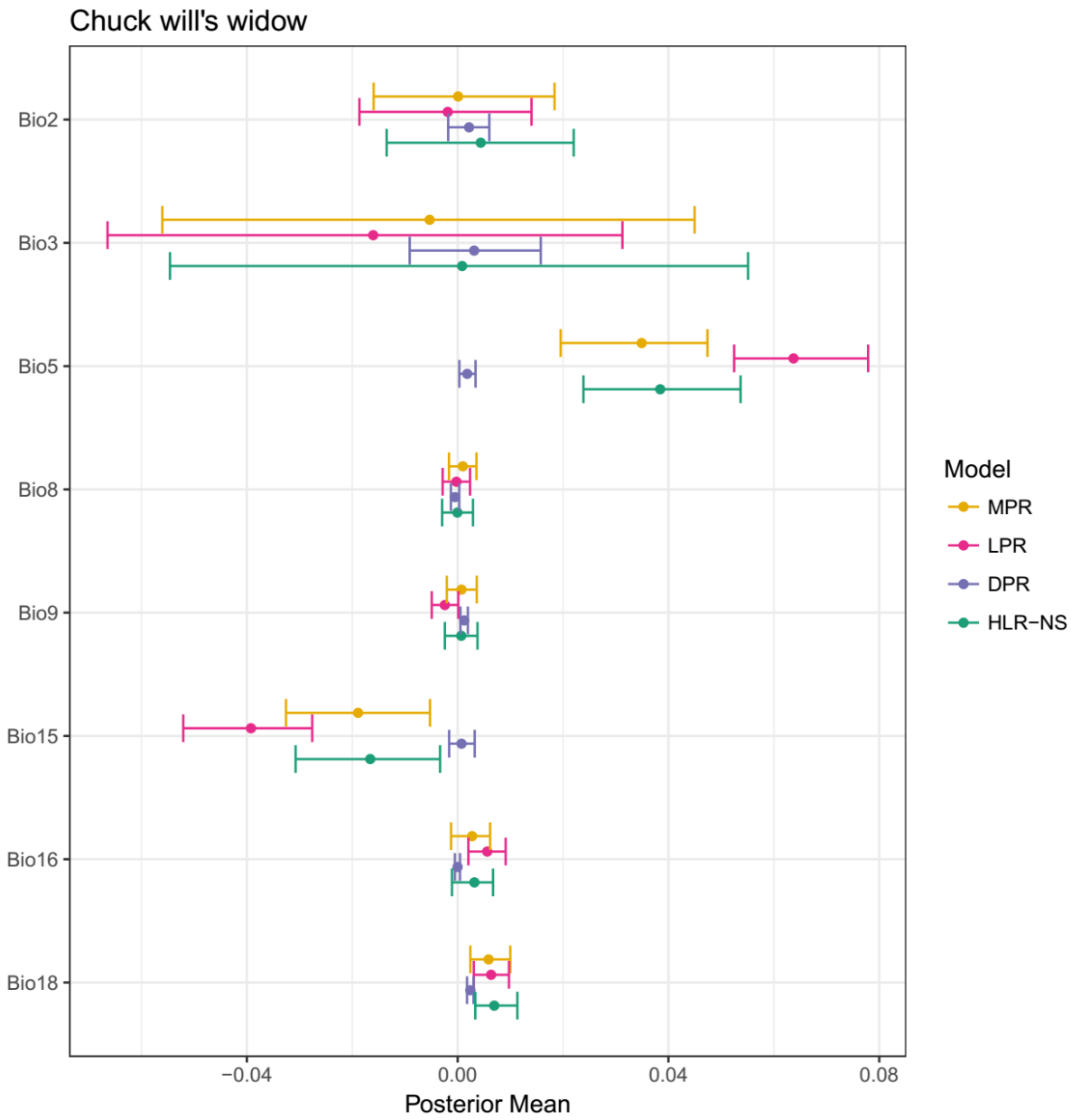


Figure 110: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the huck-will's-widow, *Antrostomus carolinensis*.

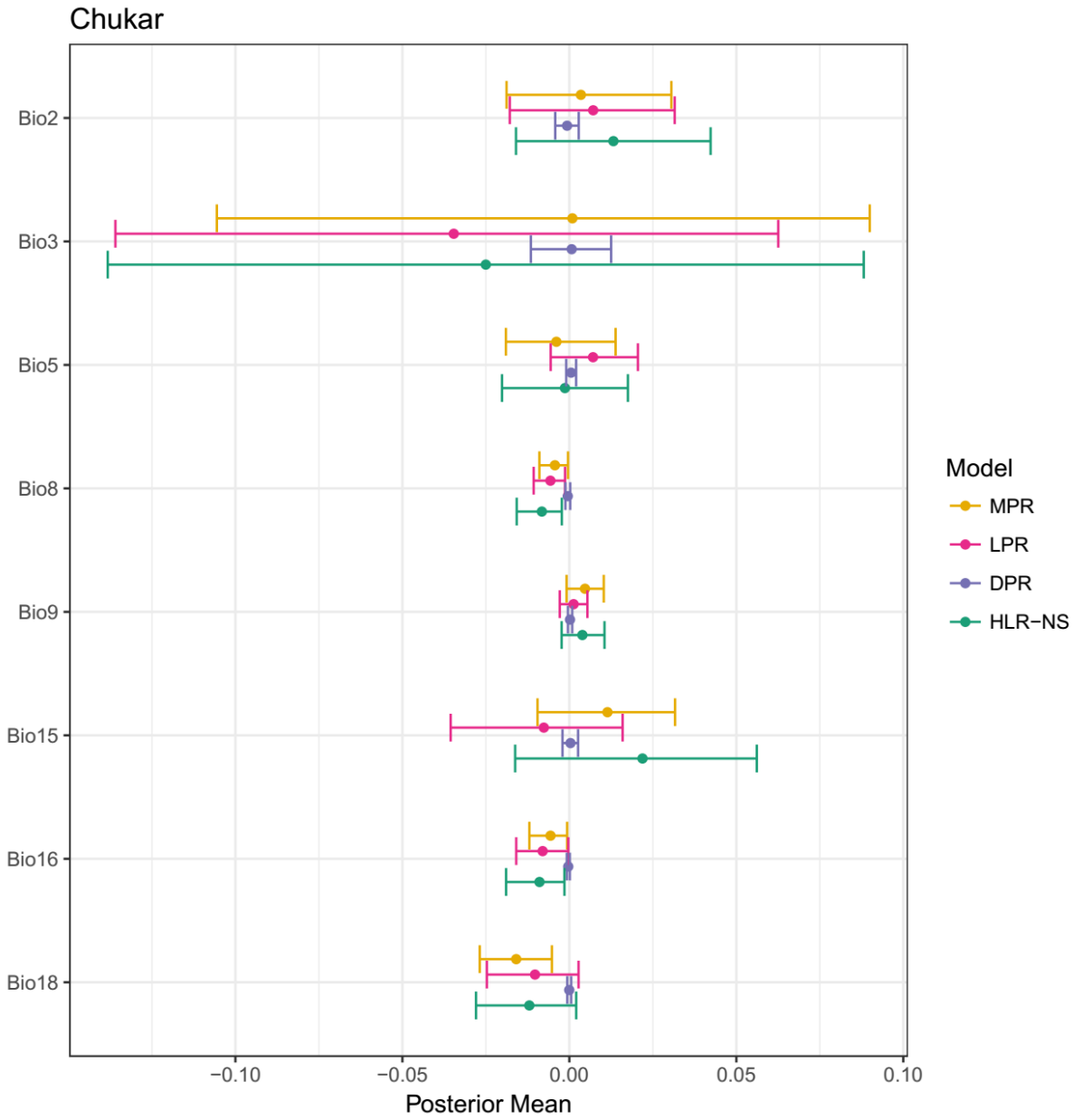


Figure 111: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chukar, *Alectoris chukar*.

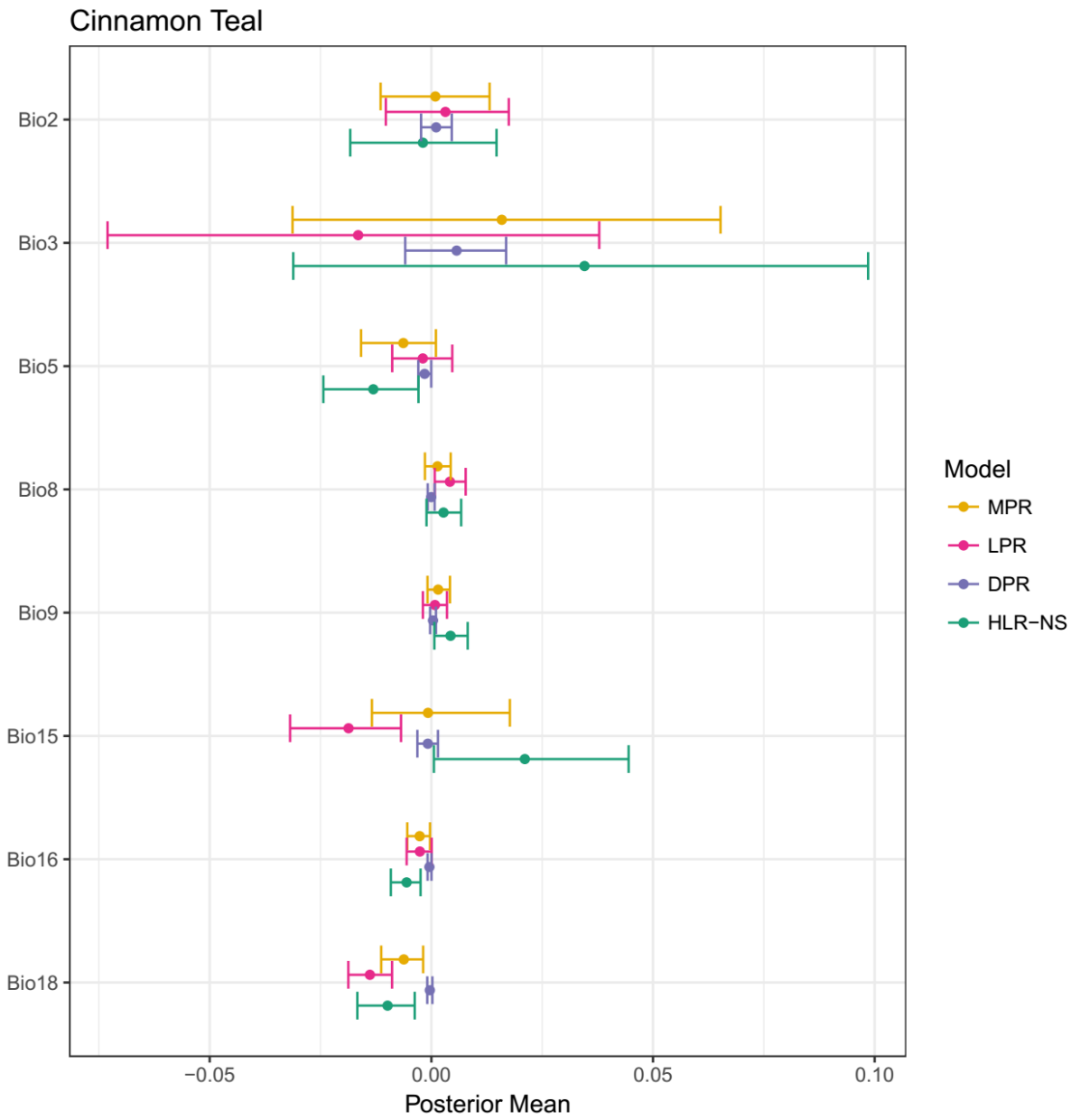


Figure 112: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cinnamon teal, *Anas cyanoptera*.

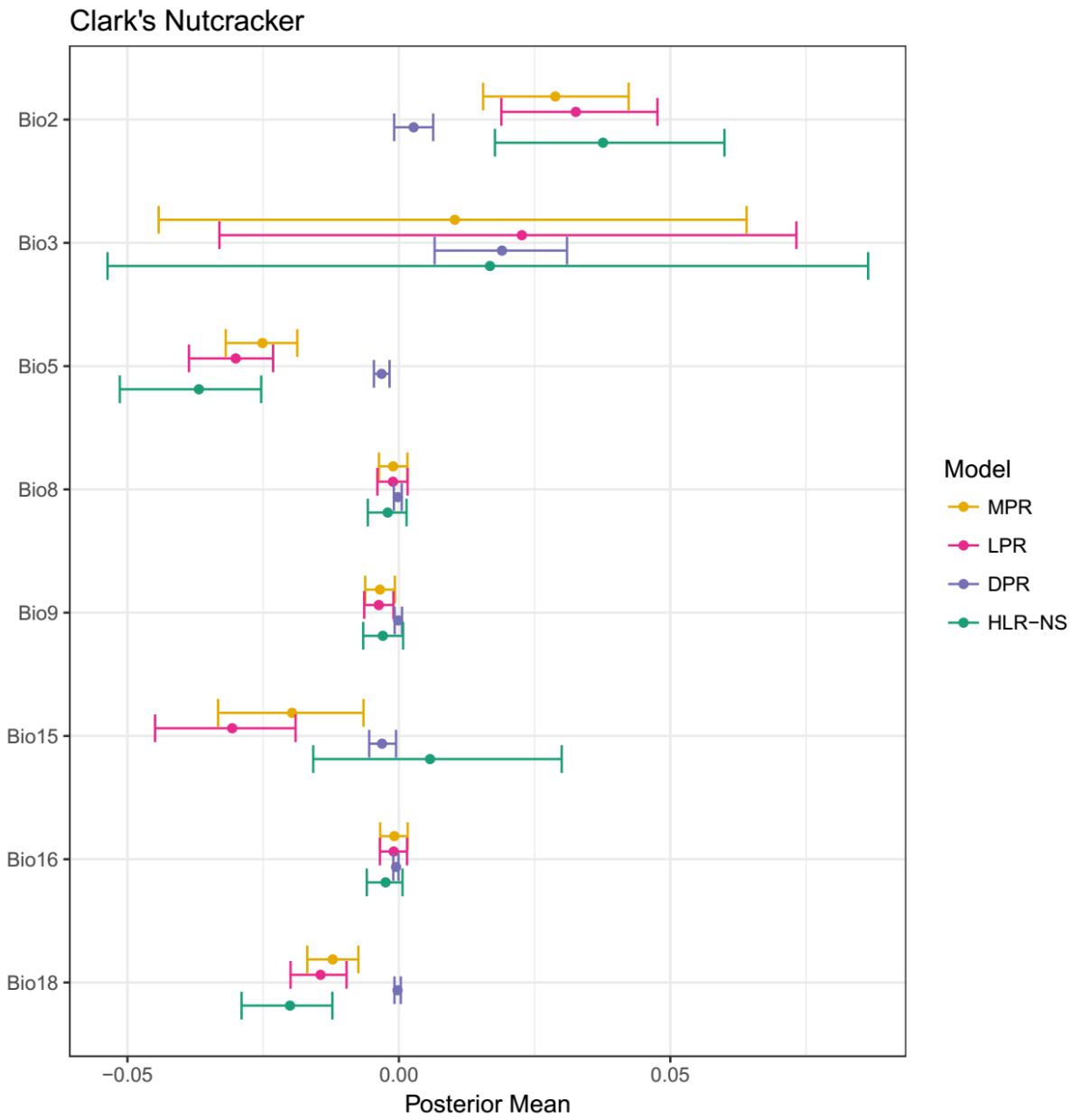


Figure 113: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Clark's nutcracker, *Nucifraga columbiana*.

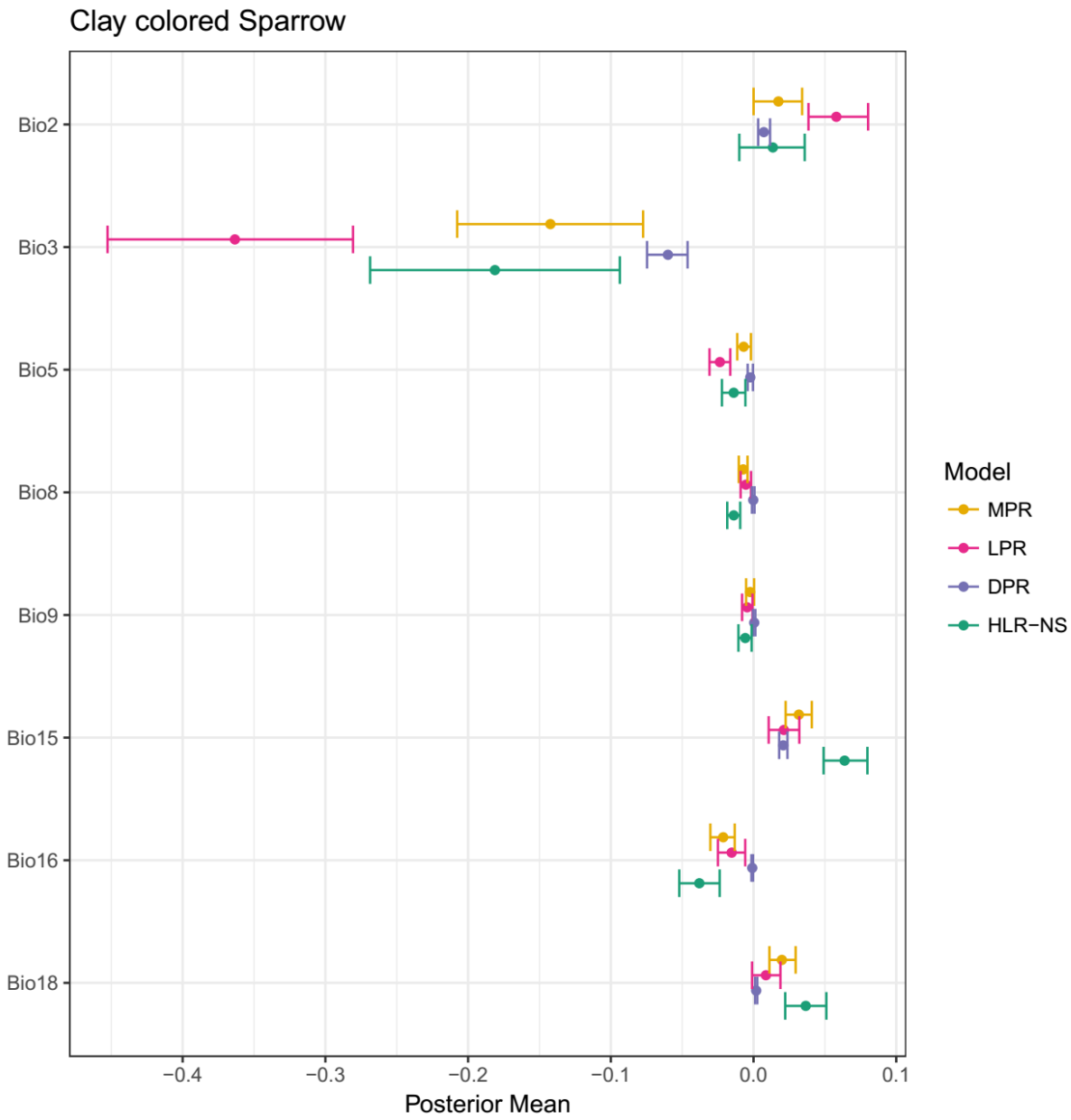


Figure 114: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the clay-colored sparrow, *Spizella pallida*.

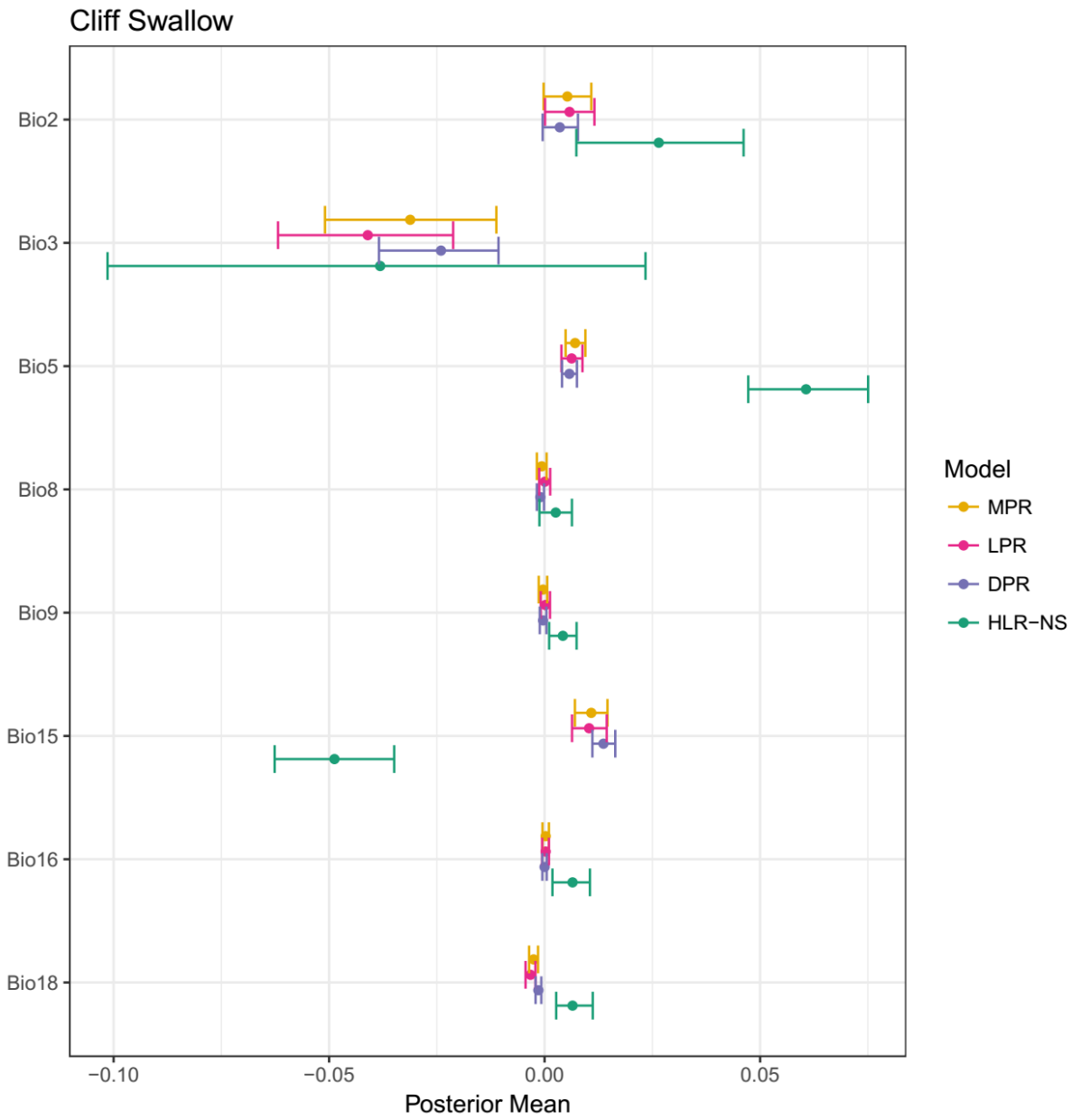


Figure 115: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the cliff swallow, *Petrochelidon pyrrhonota*.

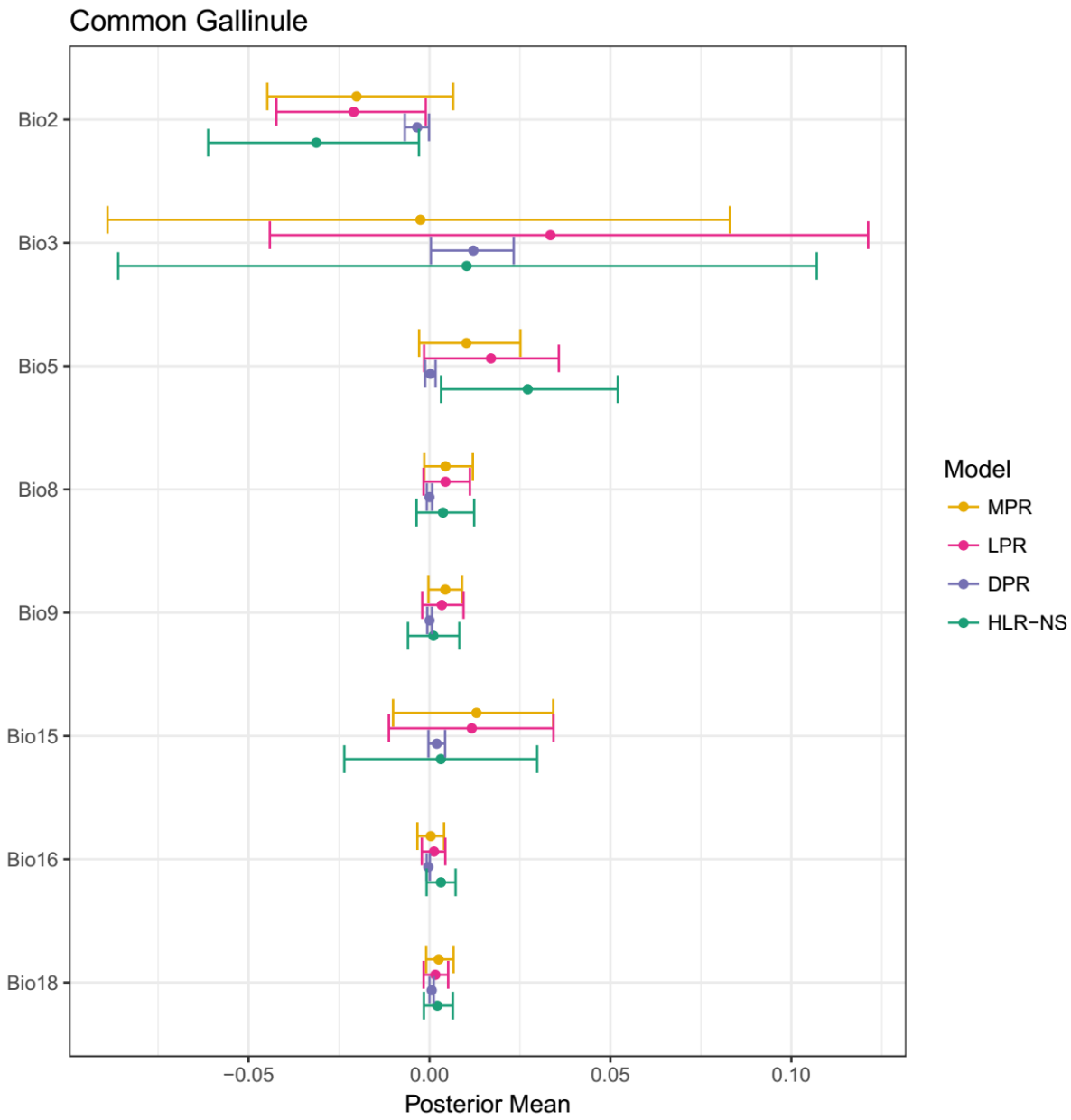


Figure 116: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common gallinule, *Gallinula galeata*.

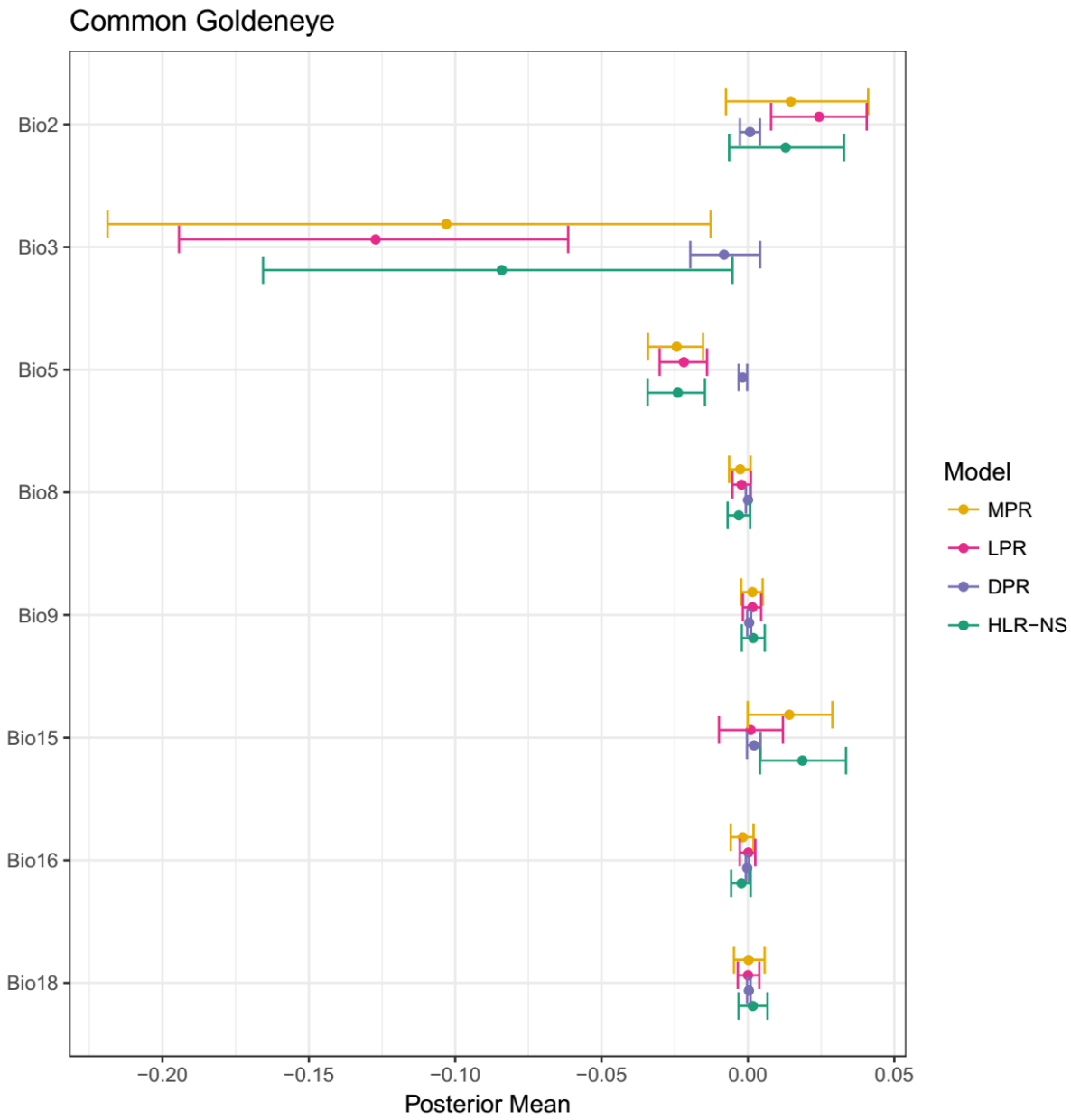


Figure 117: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common goldeneye, *Bucephala clangula*.

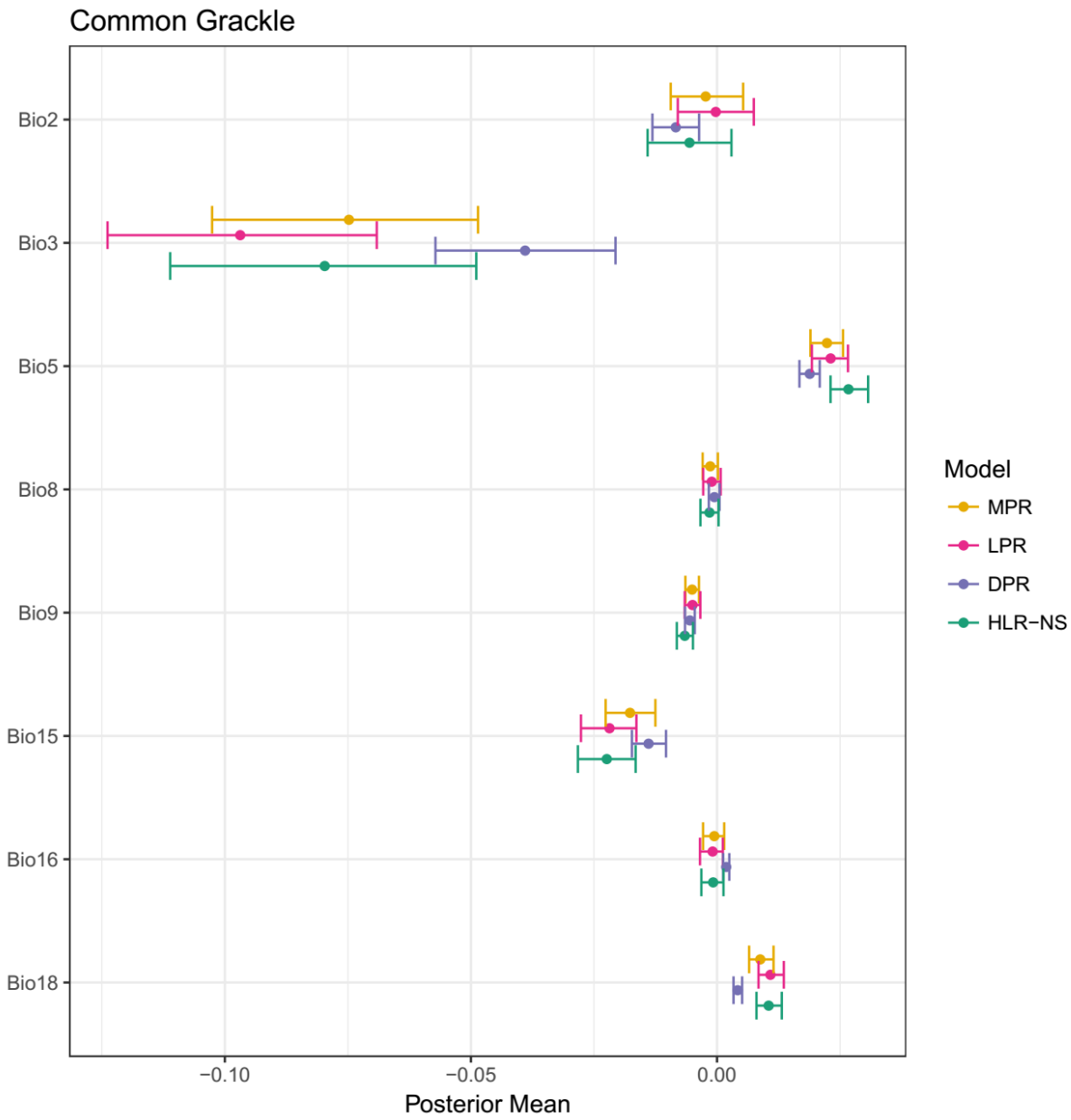


Figure 118: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common grackle, *Quiscalus quiscula*.

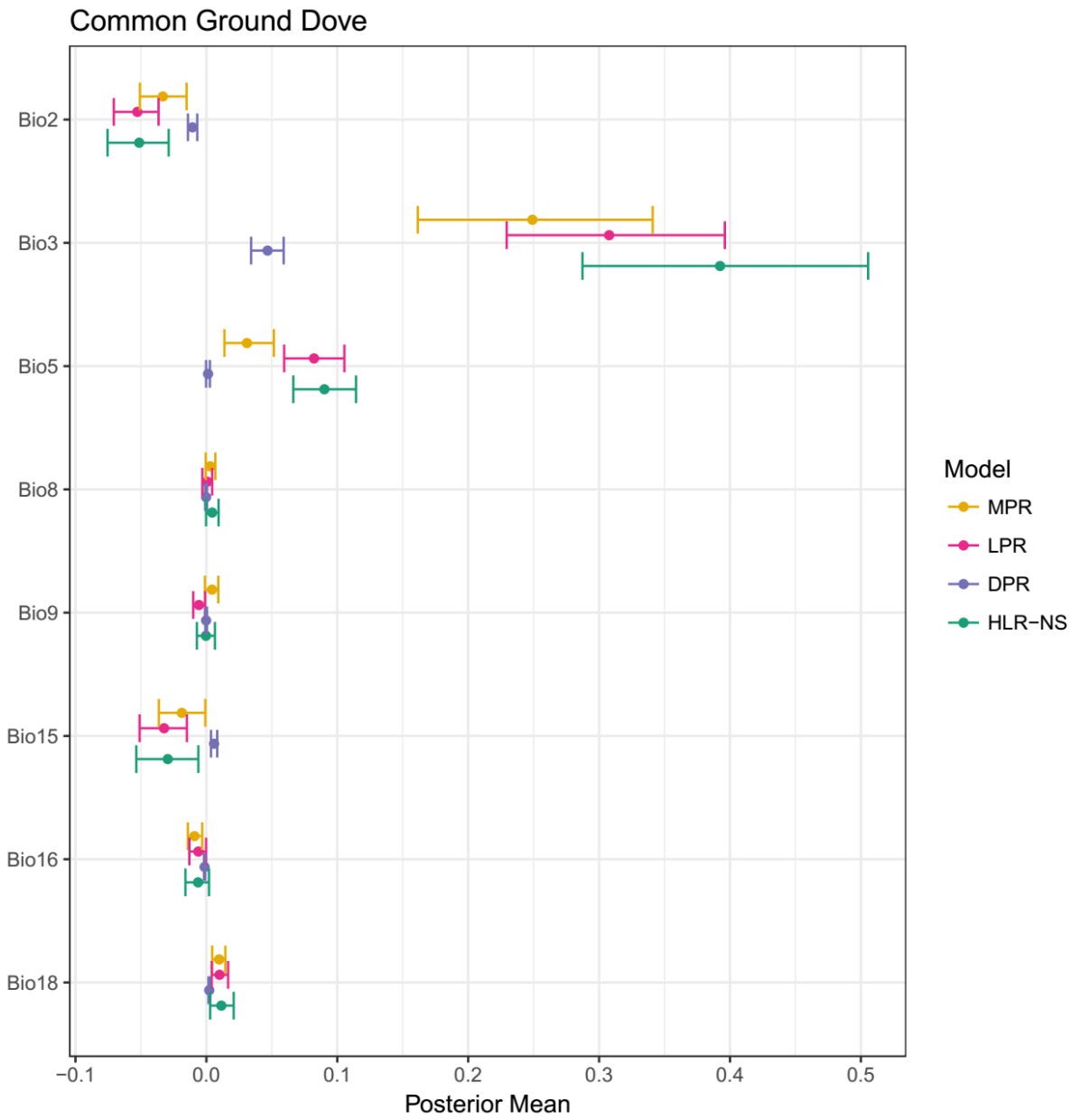


Figure 119: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common ground dove, *Columbina passerina*.

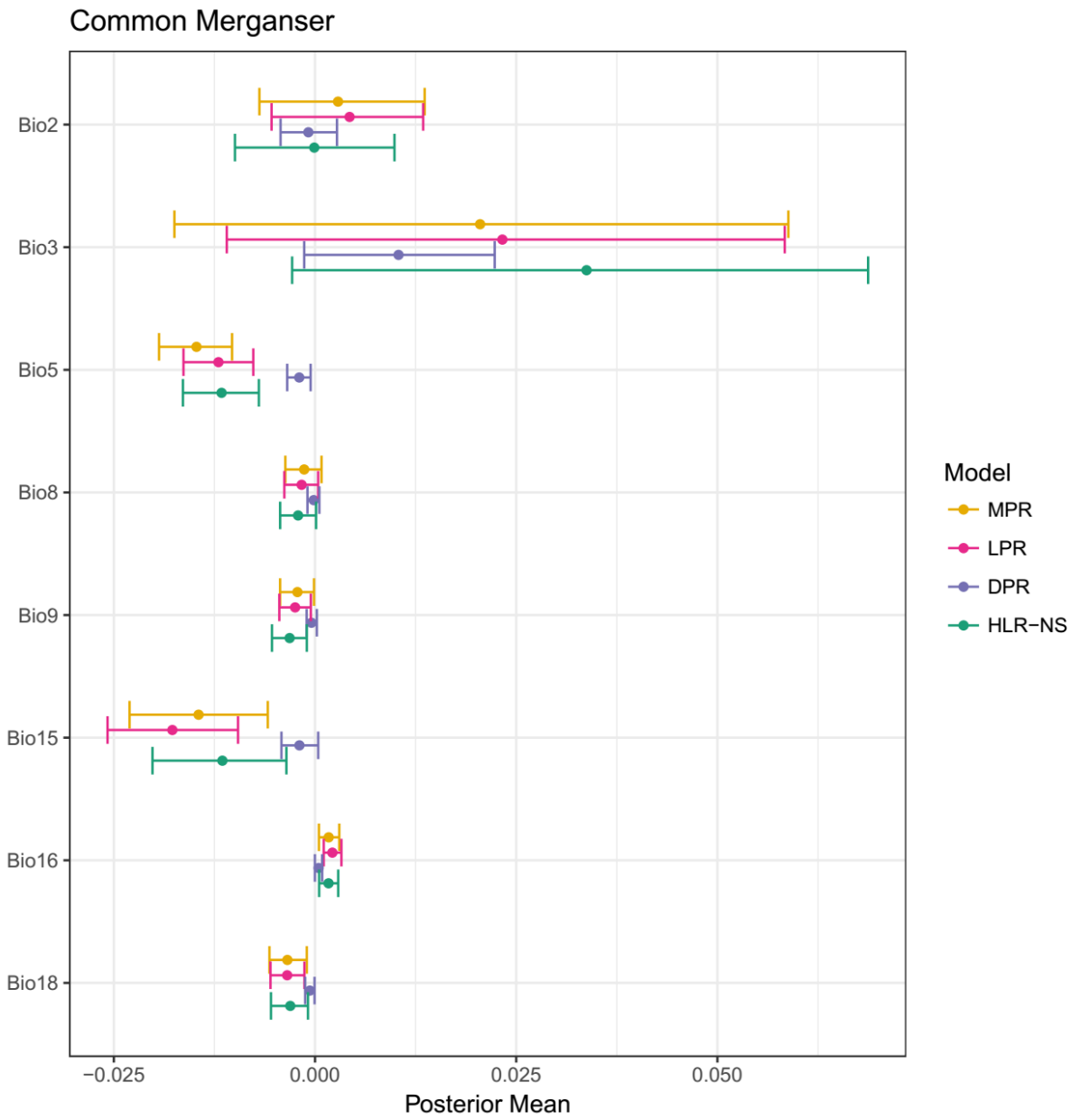


Figure 120: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common merganser, *Mergus merganser*.

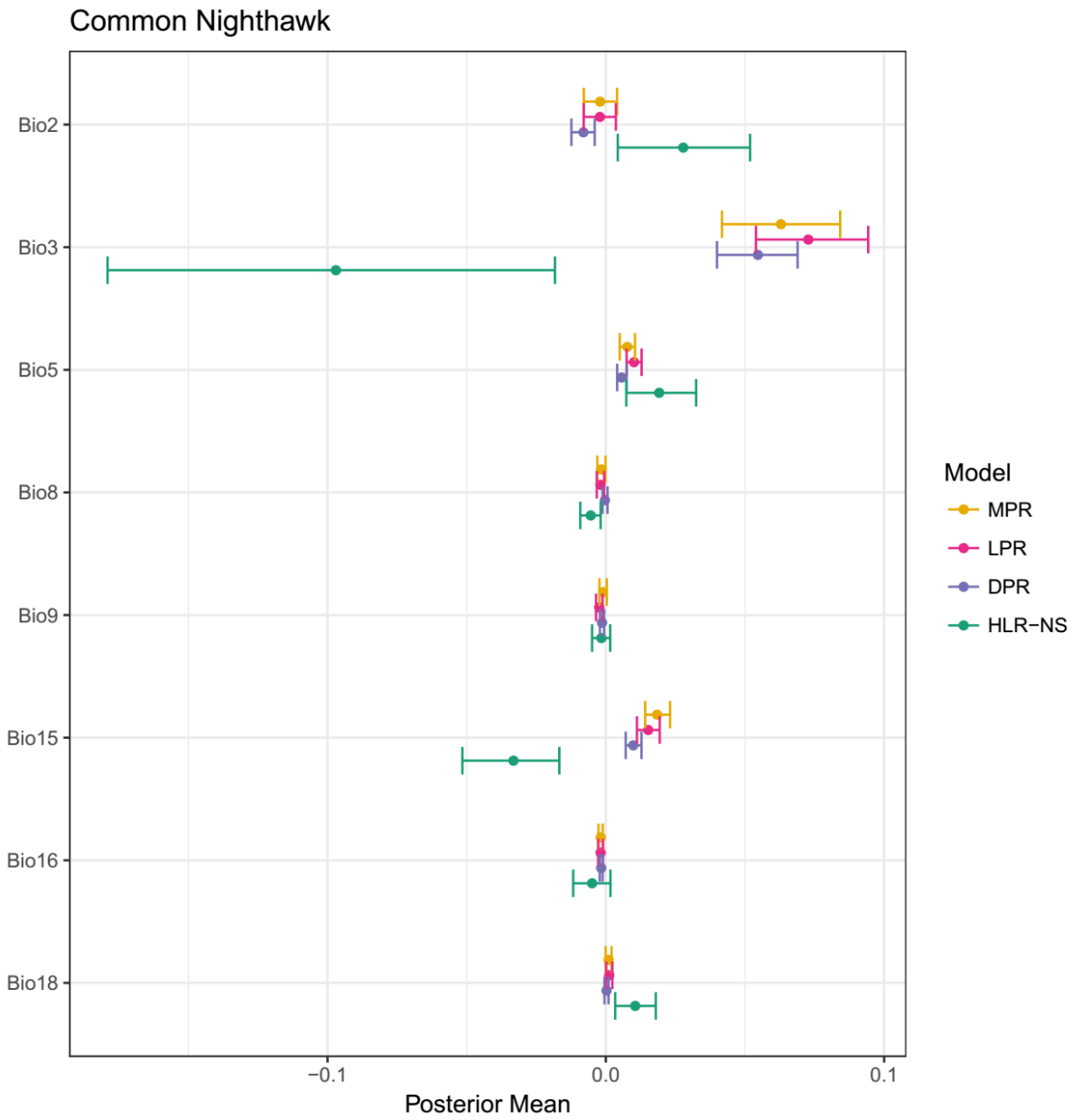


Figure 121: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common nighthawk, *Chordeiles minor*.

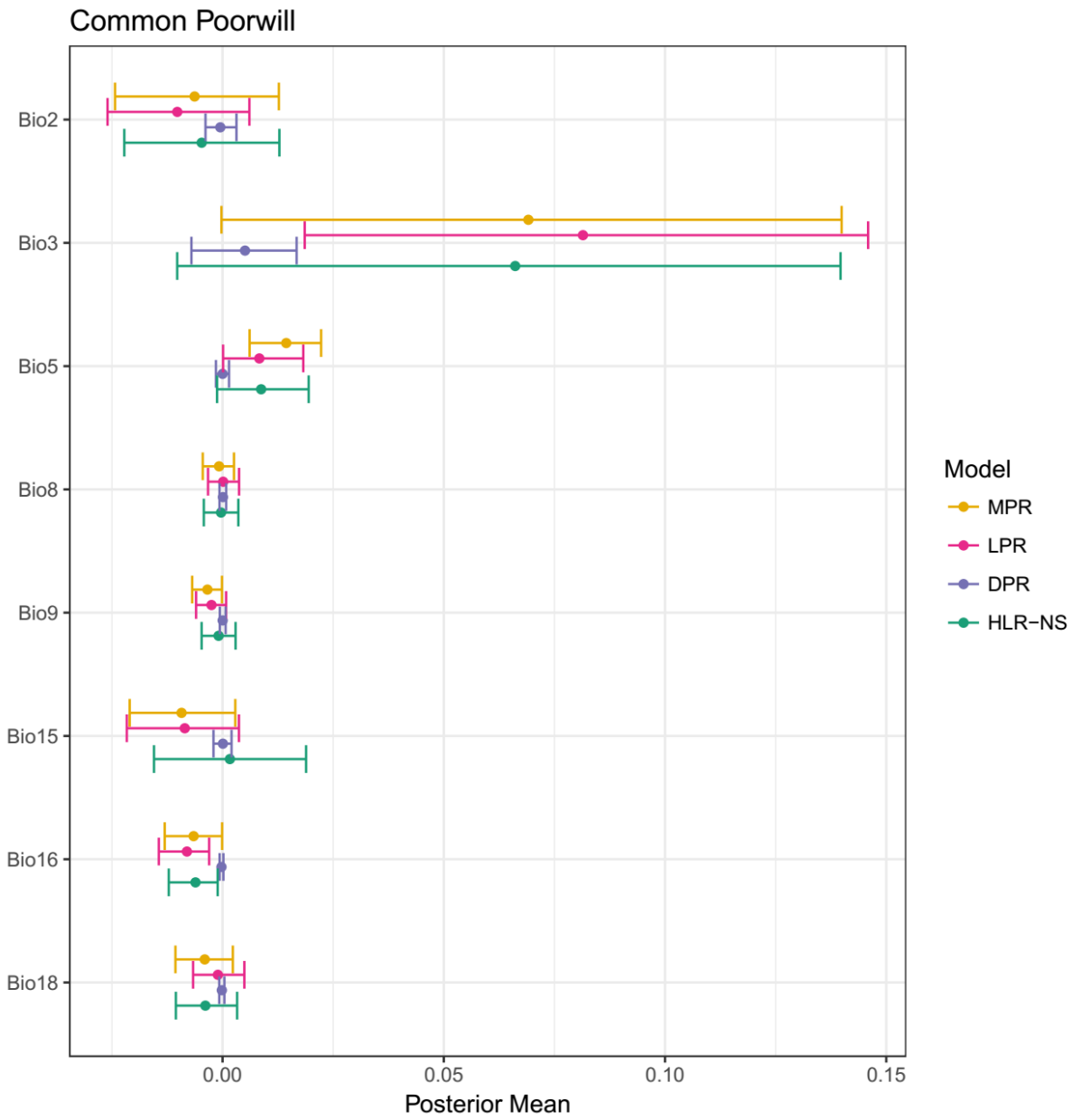


Figure 122: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common poorwill, *Phalaenoptilus nuttallii*.

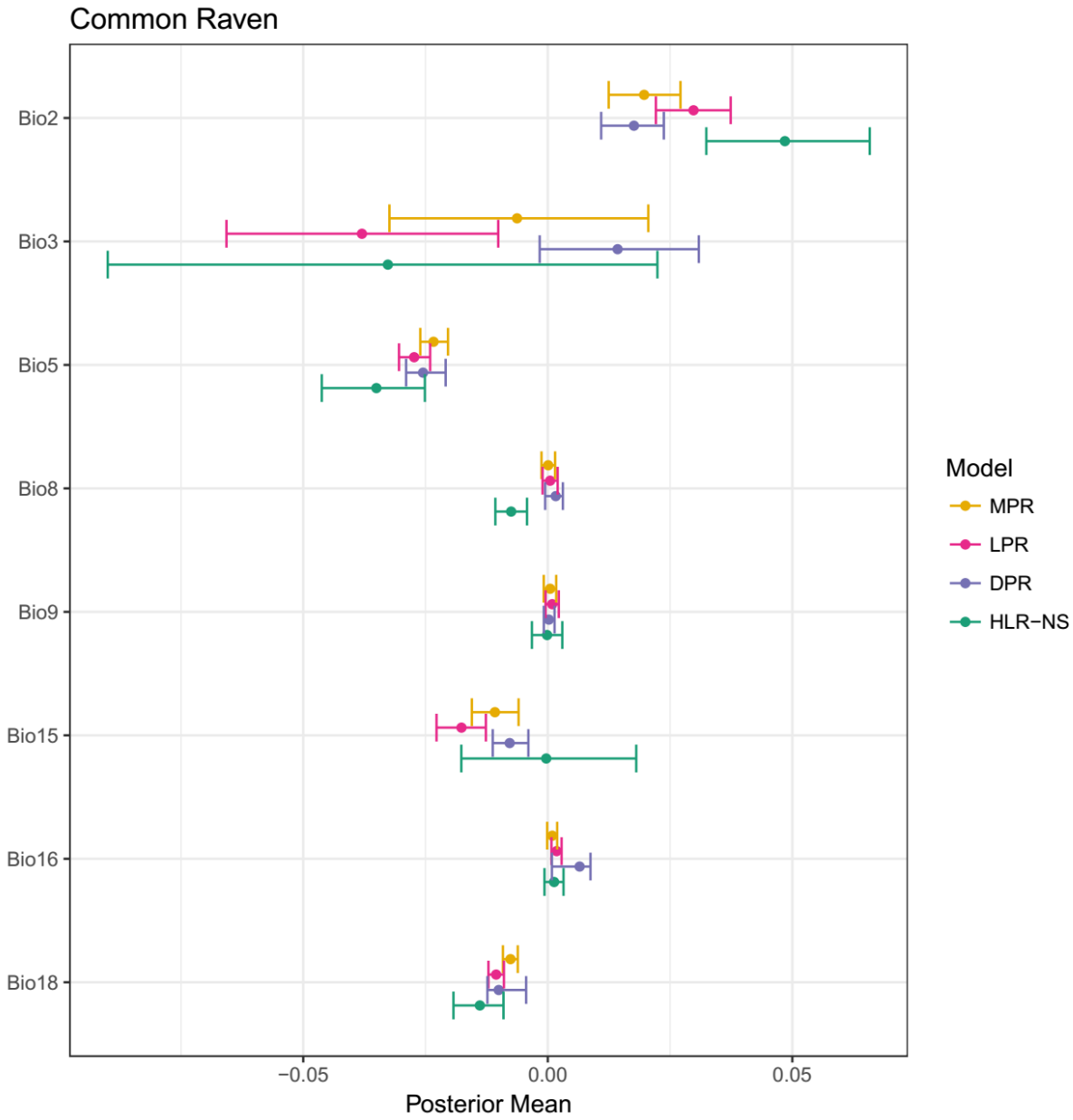


Figure 123: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common raven, *Corvus corax*.

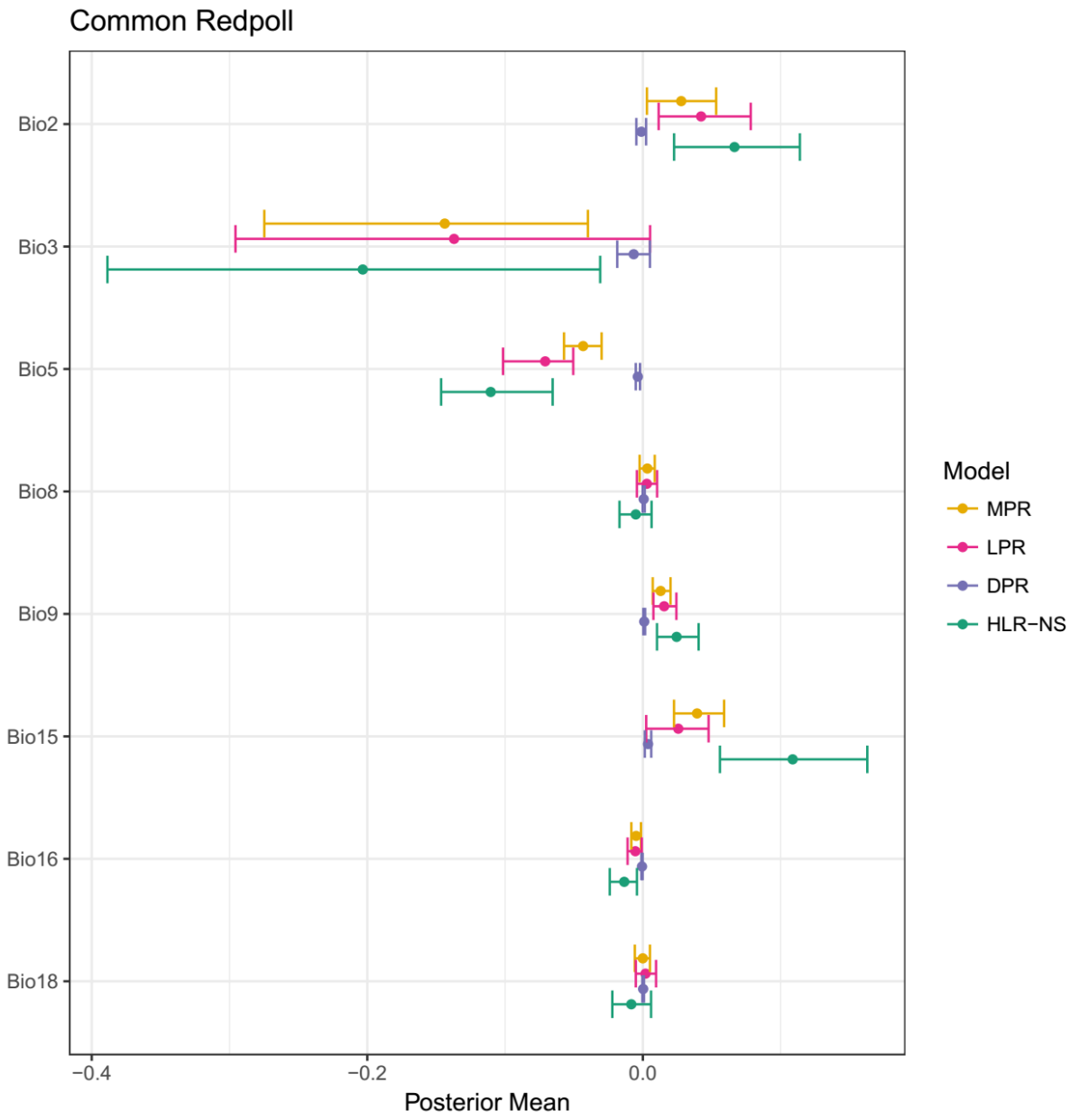


Figure 124: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common redpoll, *Acanthis flammea*.

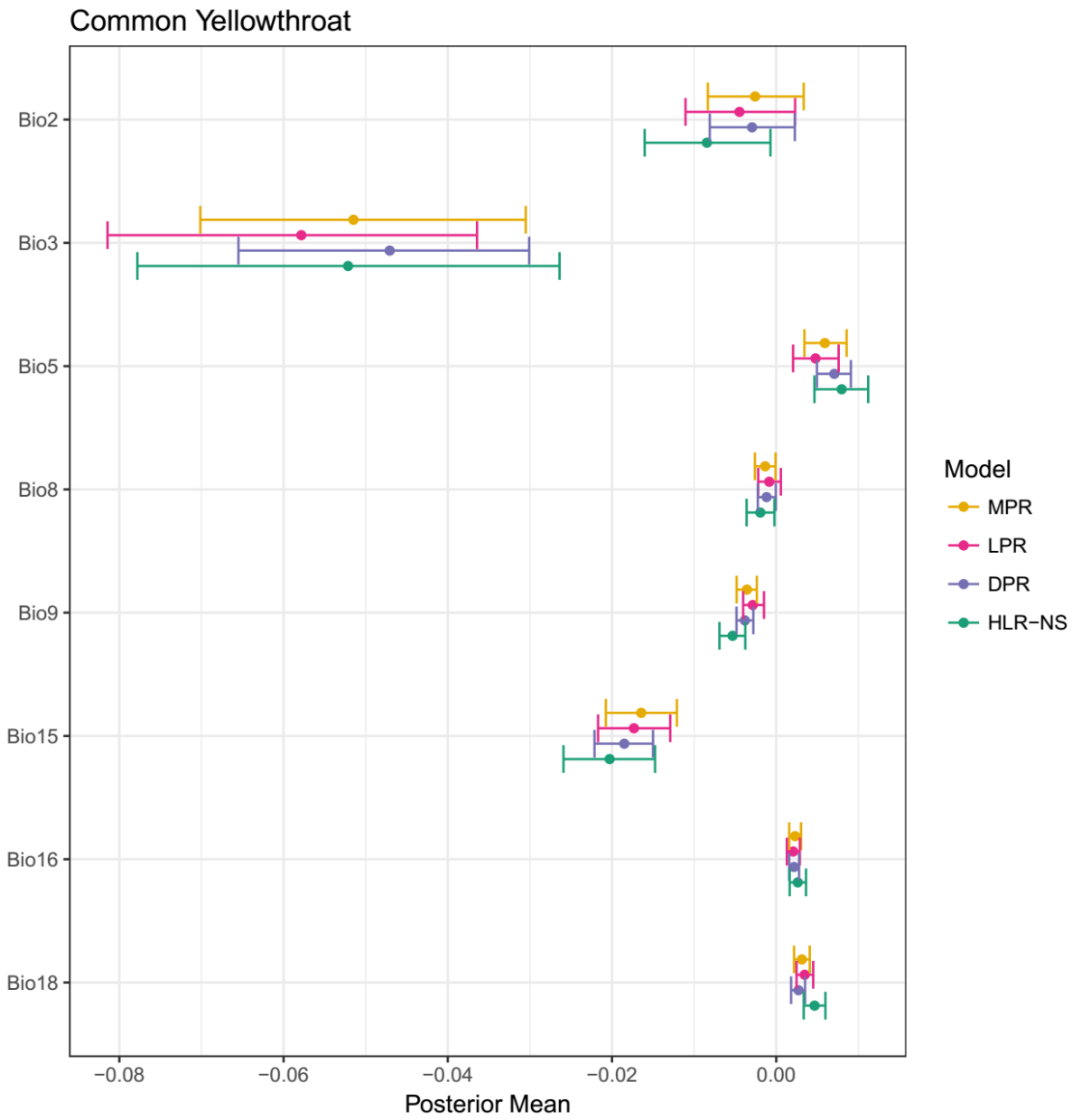


Figure 125: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common yellowthroat, *Geothlypis trichas*.

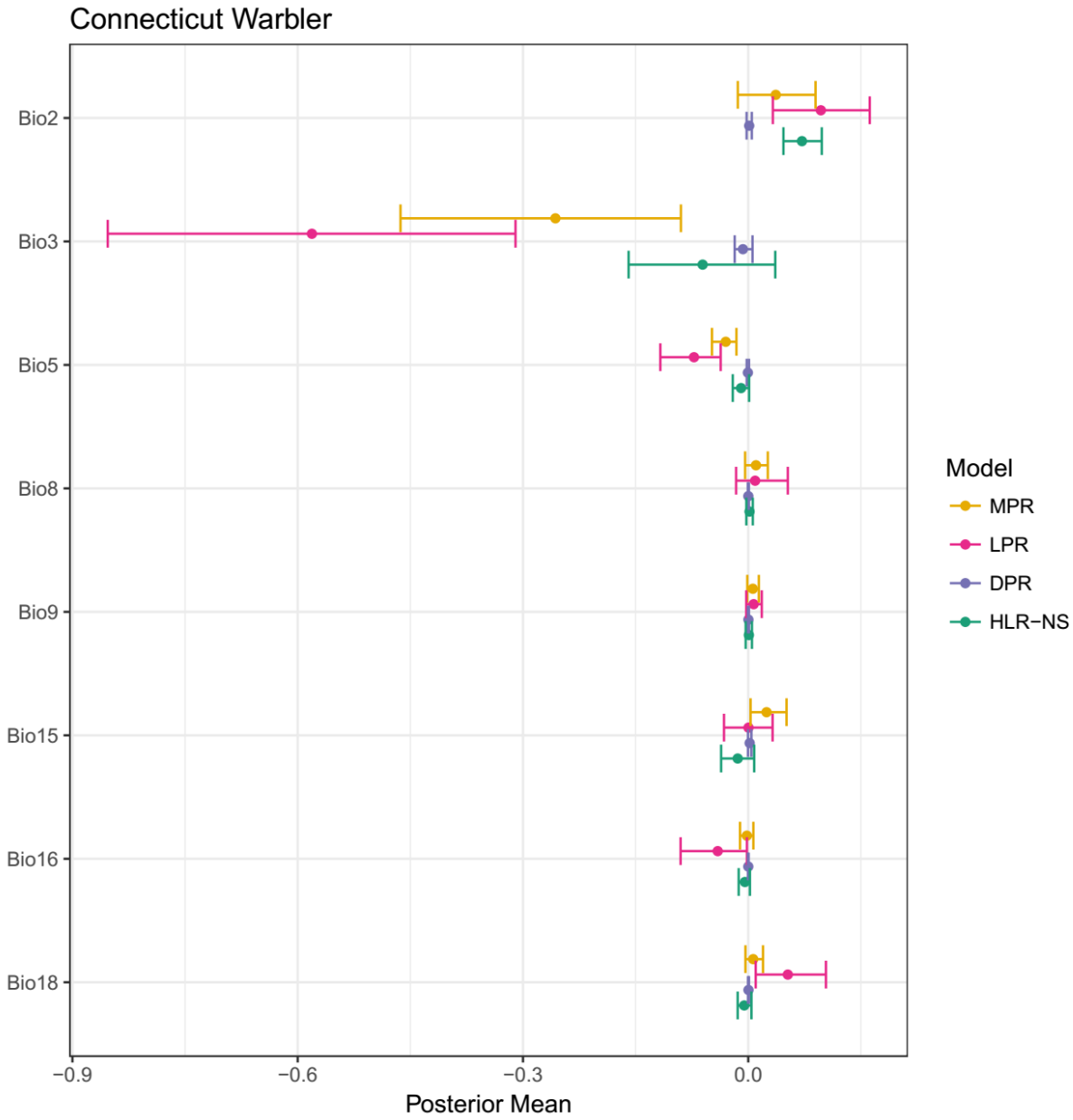


Figure 126: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Connecticet warbler, *Oporornis agilis*.

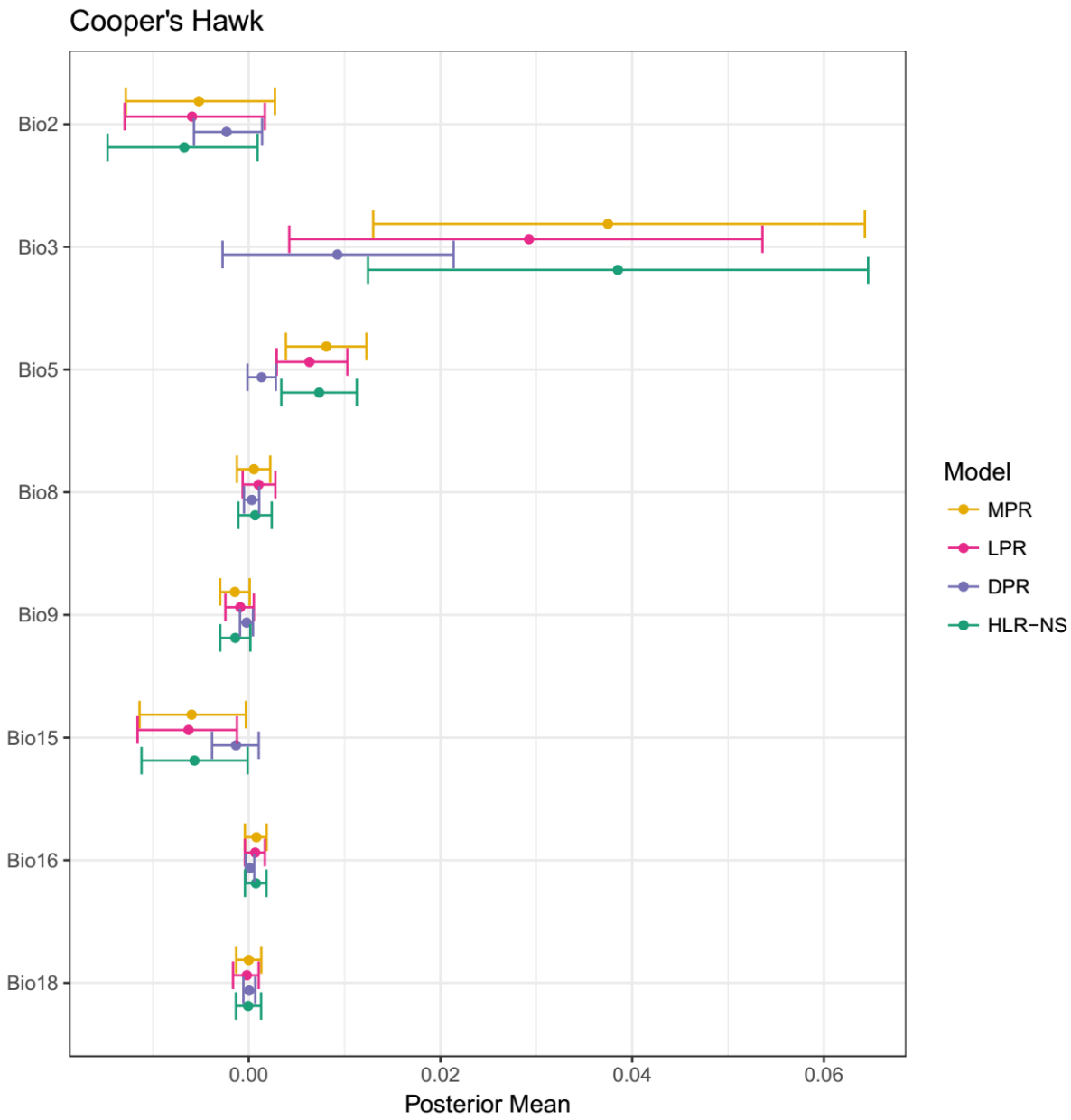


Figure 127: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cooper's hawk, *Accipiter cooperii*.

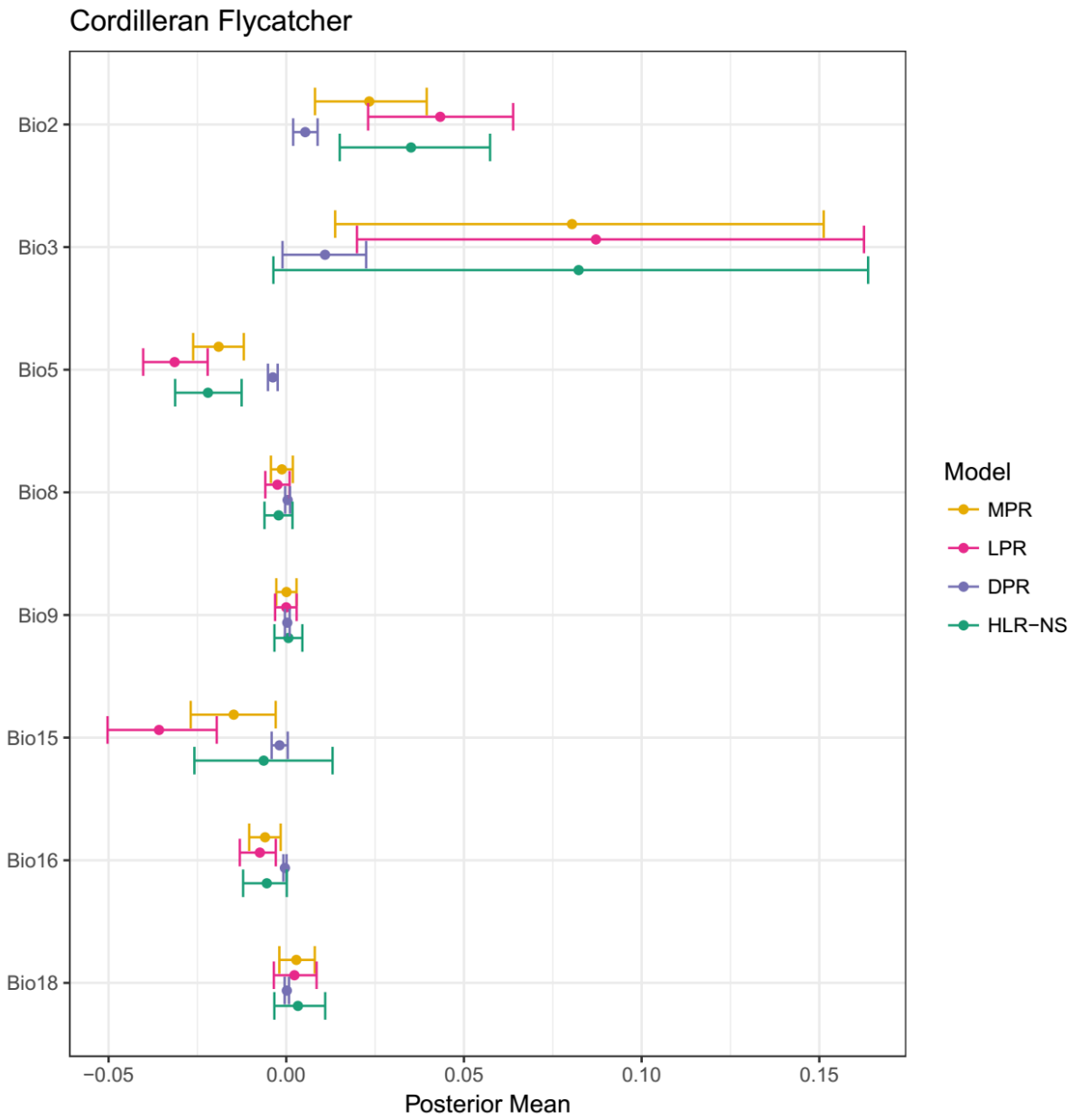


Figure 128: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Cordilleran flycatcher, *Empidonax occidentalis*.

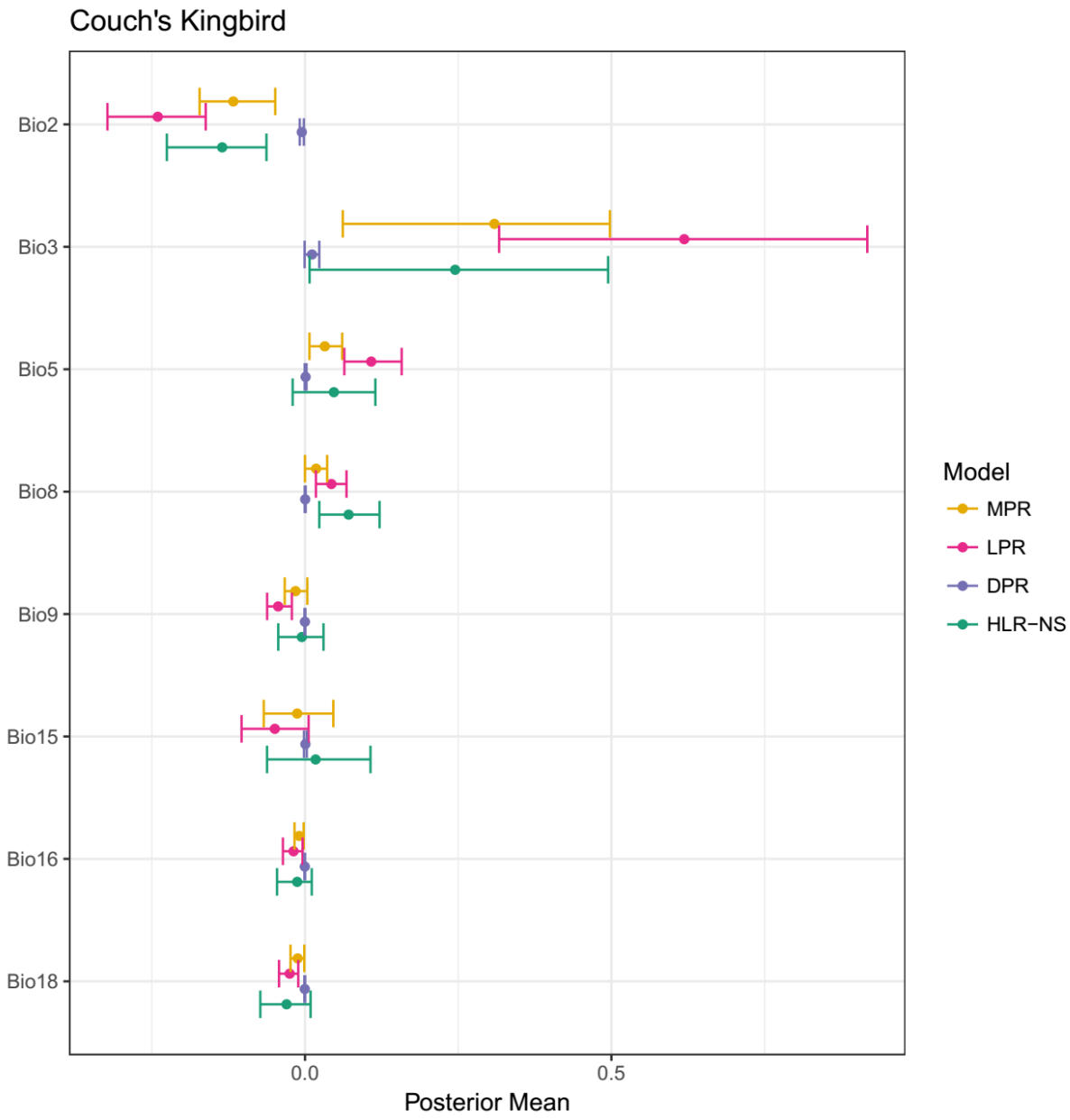


Figure 129: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Couch's kingbird, *Tyrannus couchii*.

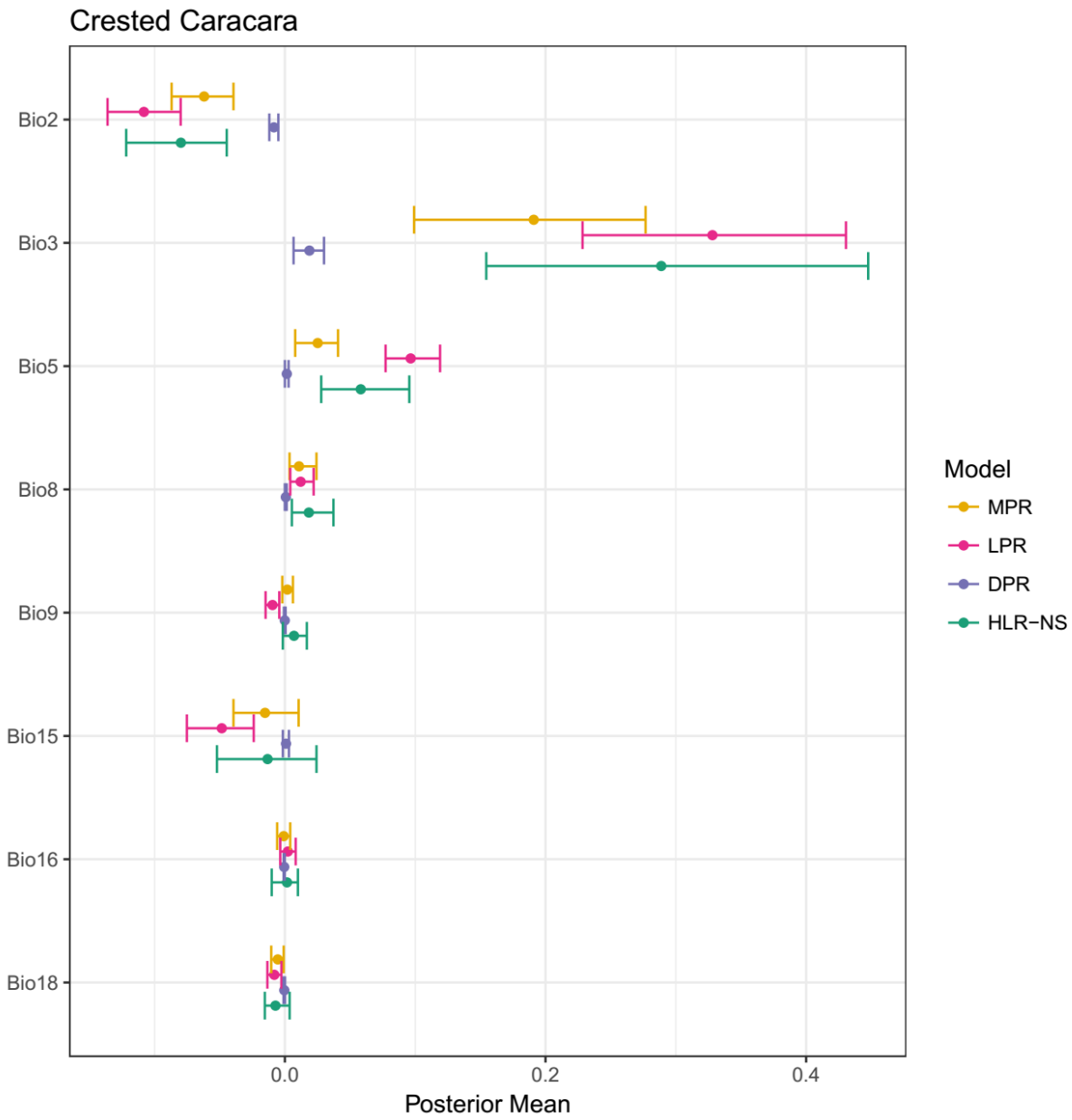


Figure 130: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the crested caracara, *Caracara cheriway*.

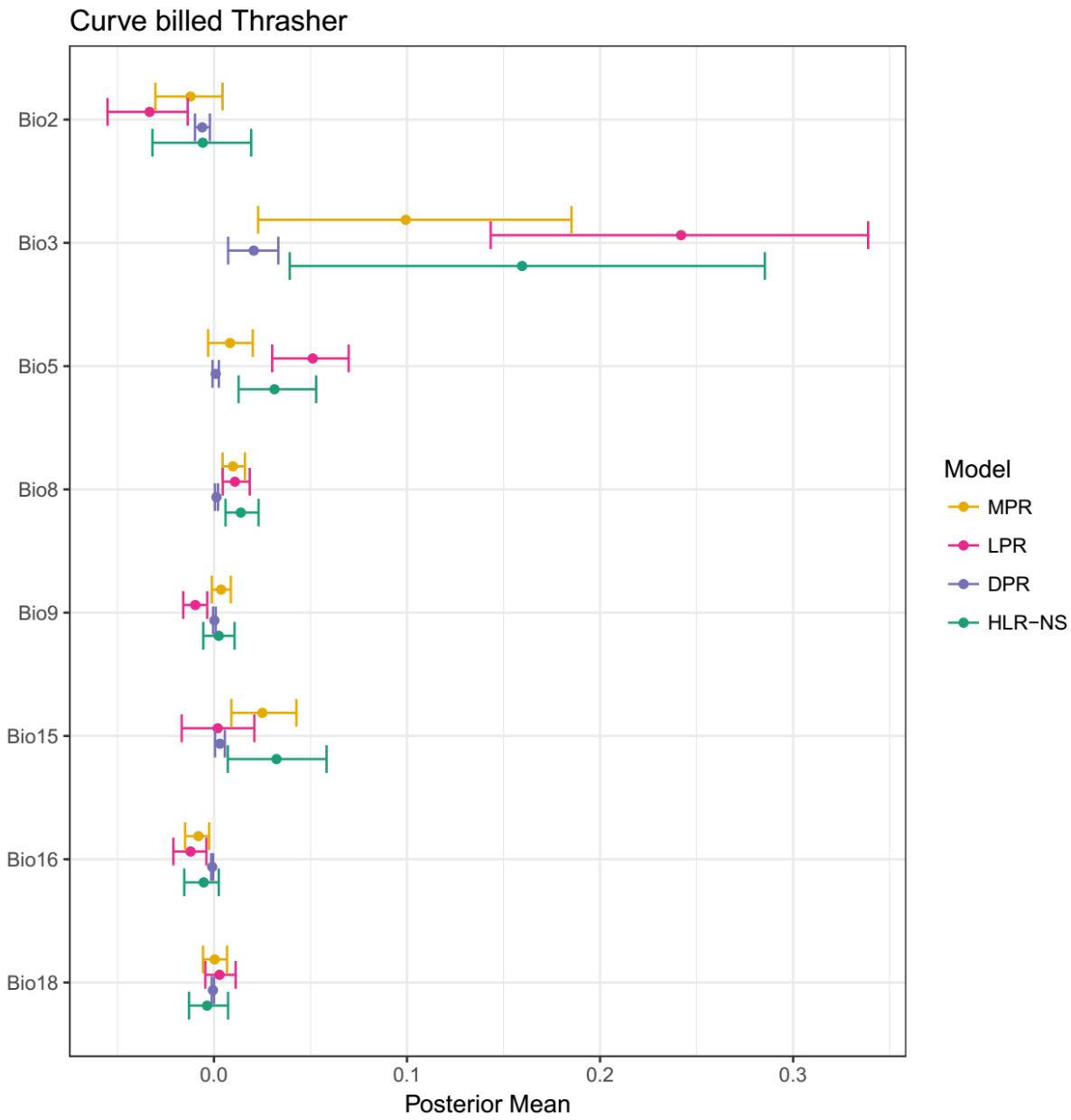


Figure 131: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the curve-billed thrasher, *Toxostoma curvirostre*.

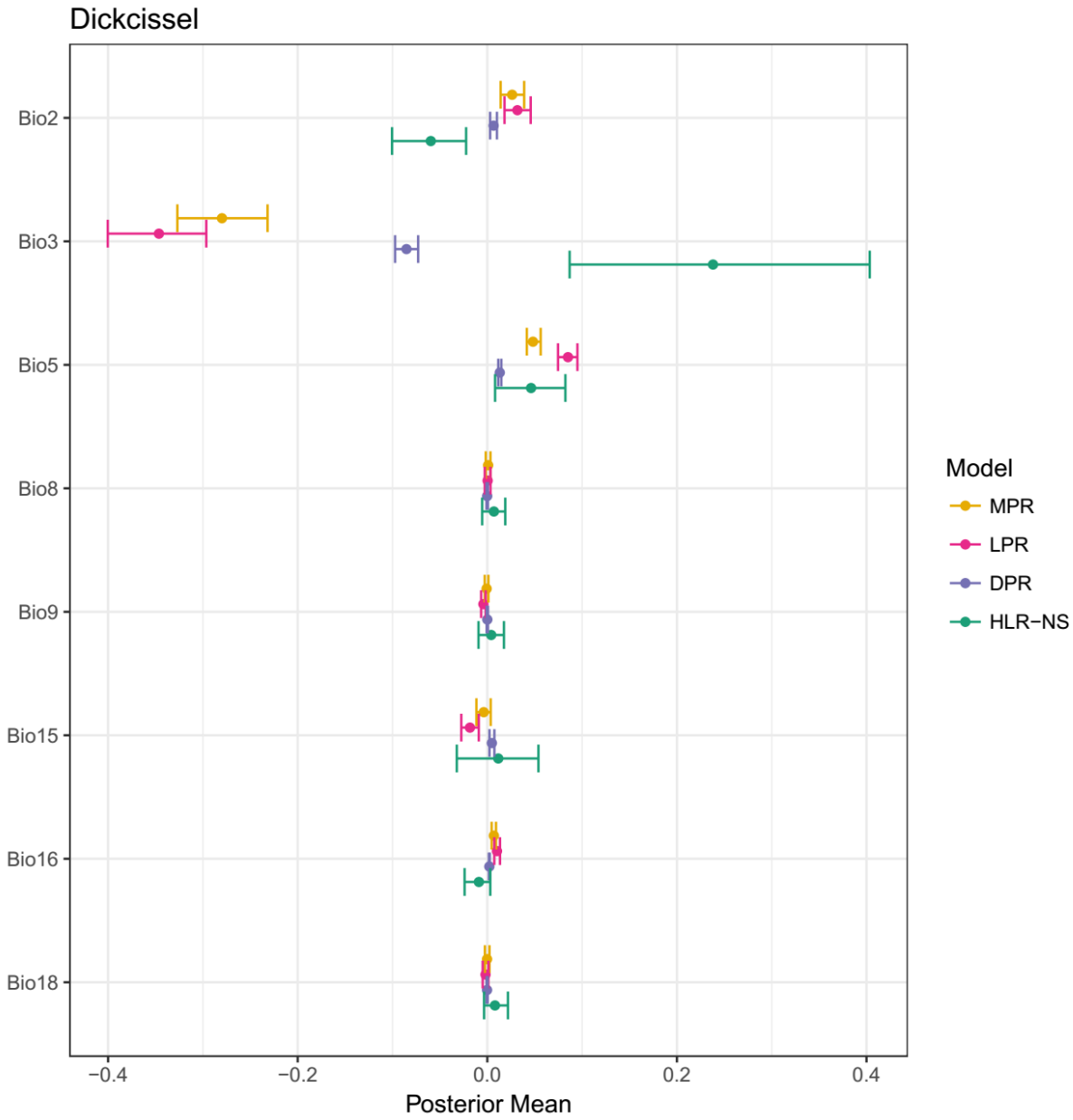


Figure 132: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the dickcissel, *Spiza americana*.

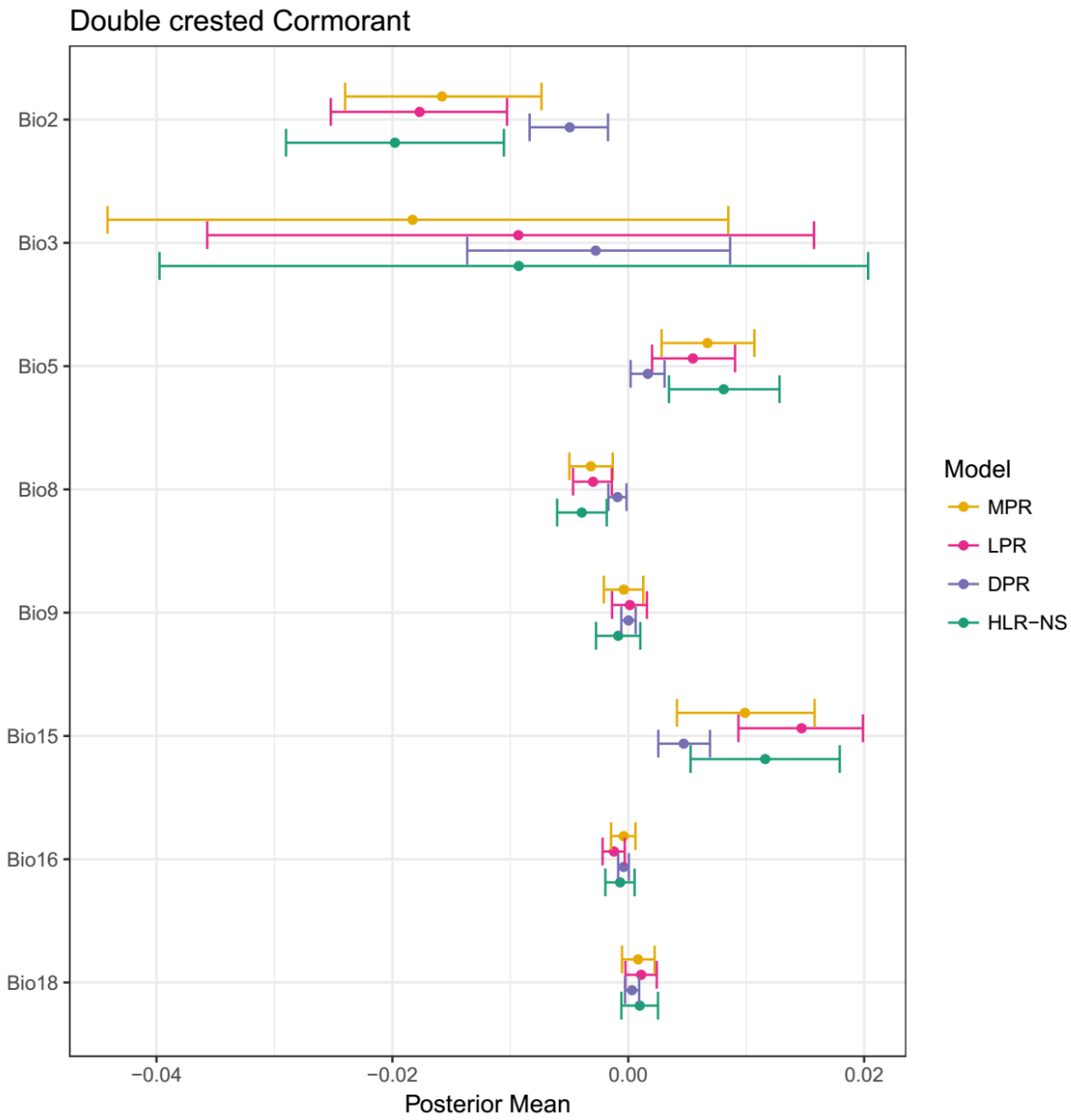


Figure 133: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the double-crested cormorant, *Phalacrocorax auritus*.

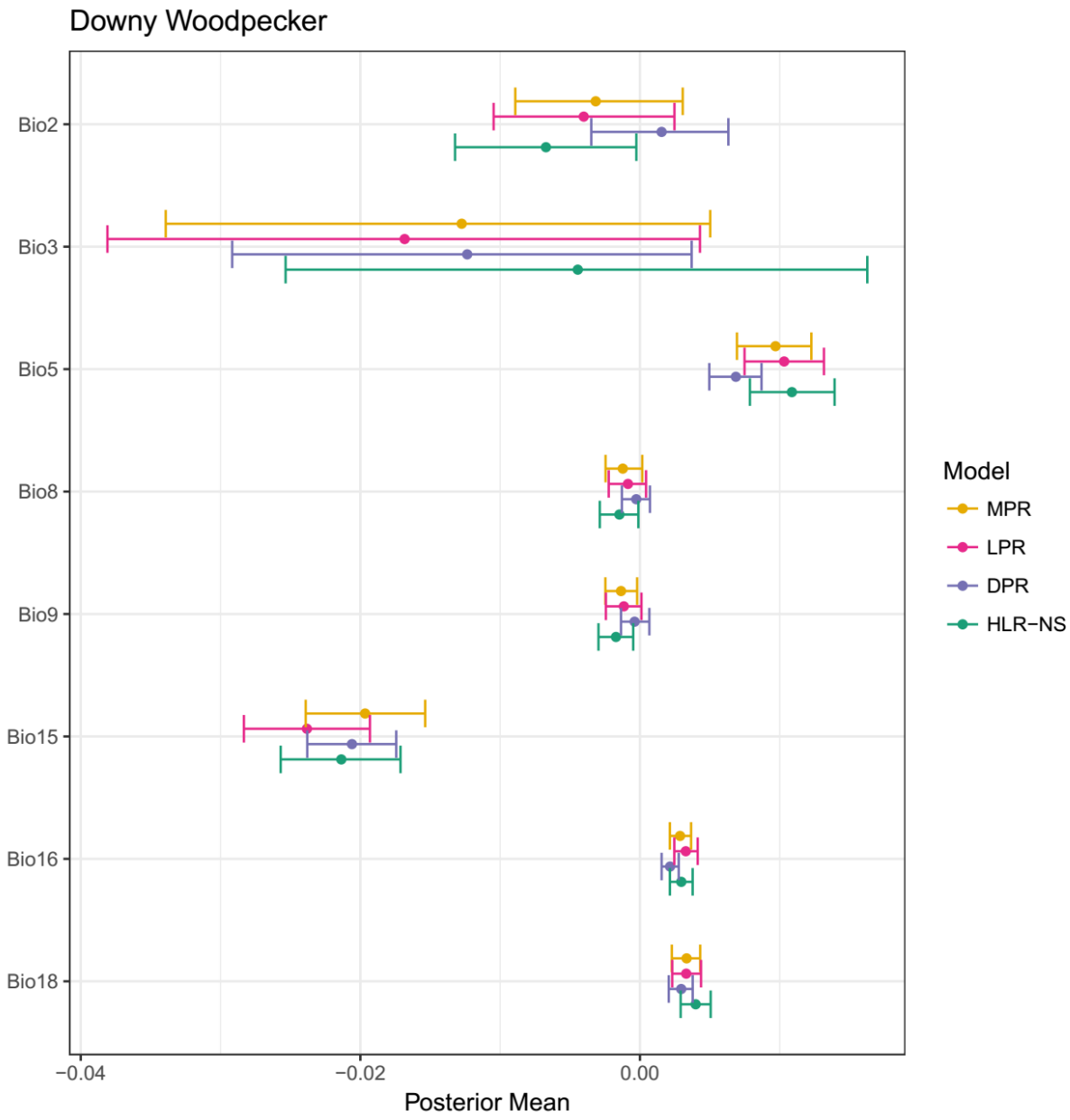


Figure 134: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the downy woodpecker, *Picoides pubescens*.

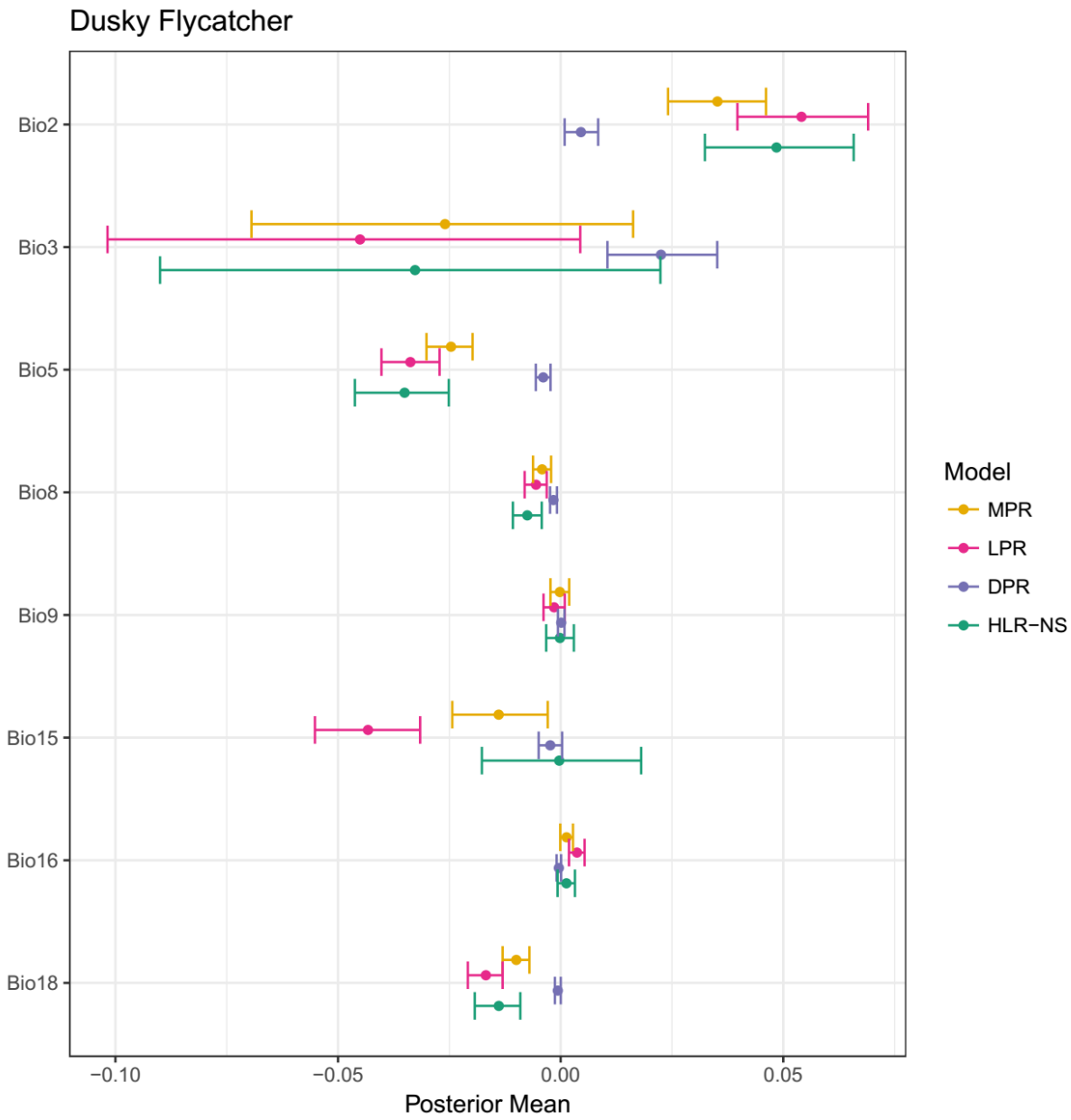


Figure 135: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the dusky flycatcher, *Empidonax oberholseri*.

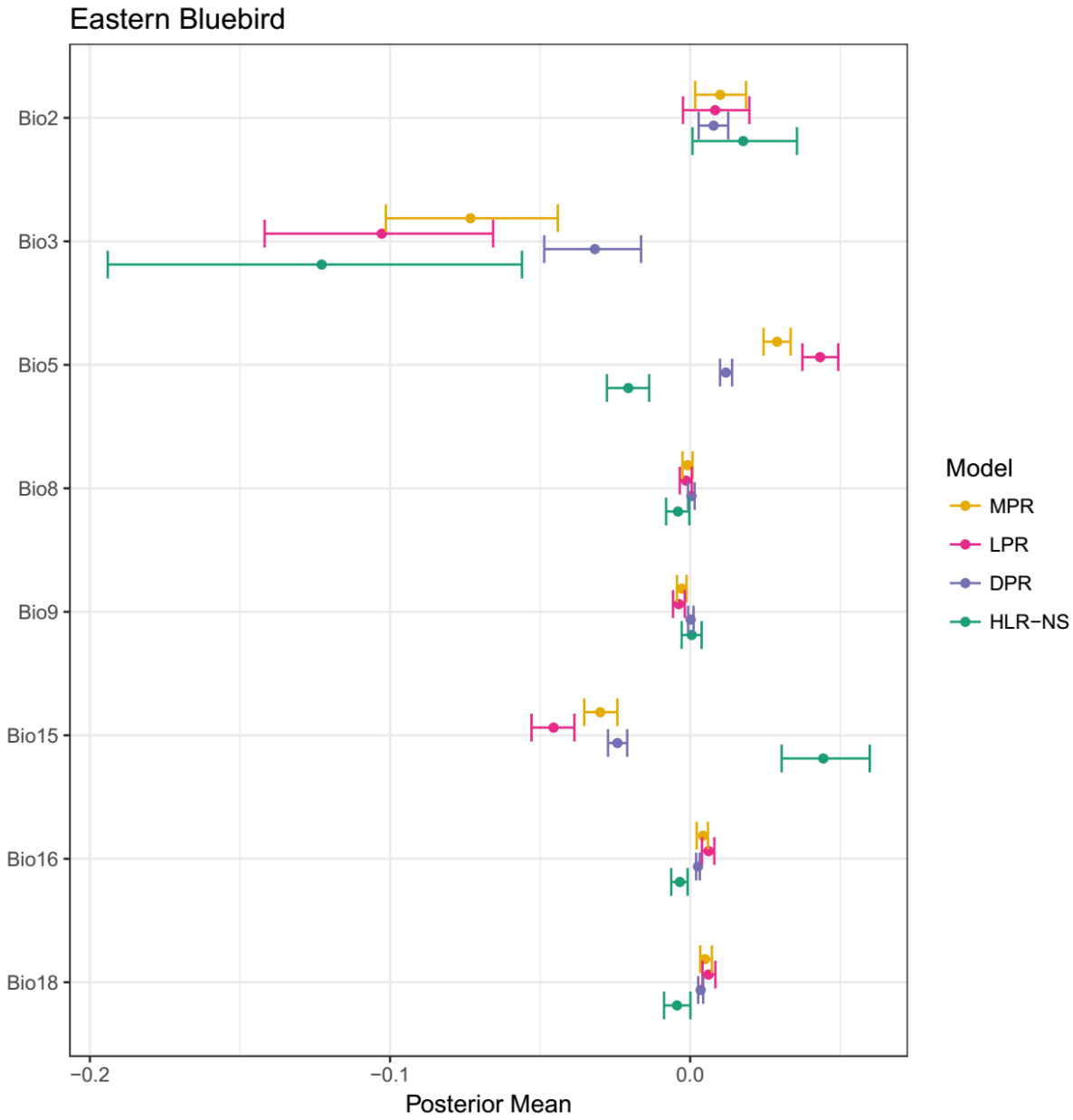


Figure 136: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern bluebird, *Sialia sialis*.

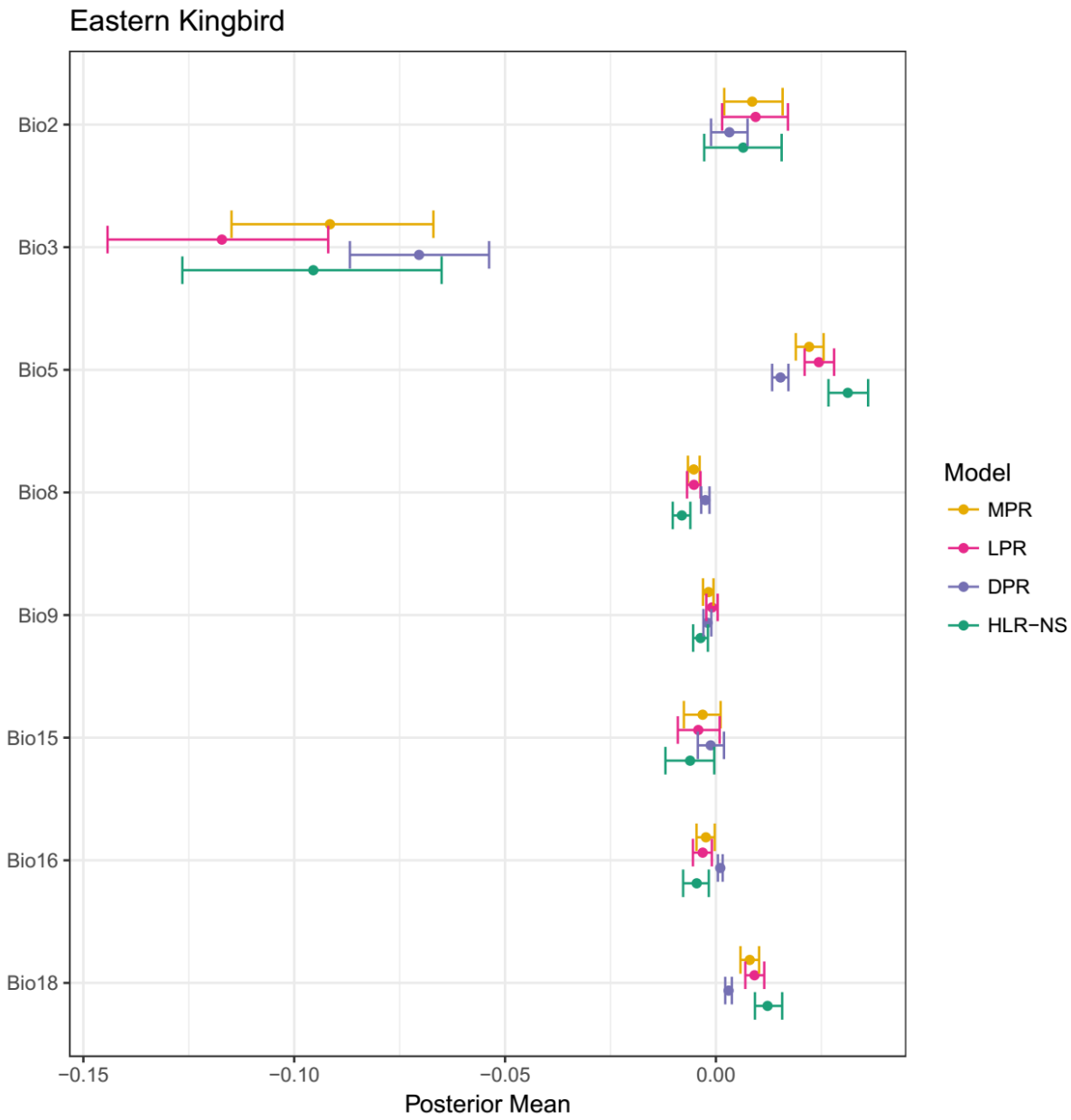


Figure 137: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern kingbird, *Tyrannus tyrannus*.

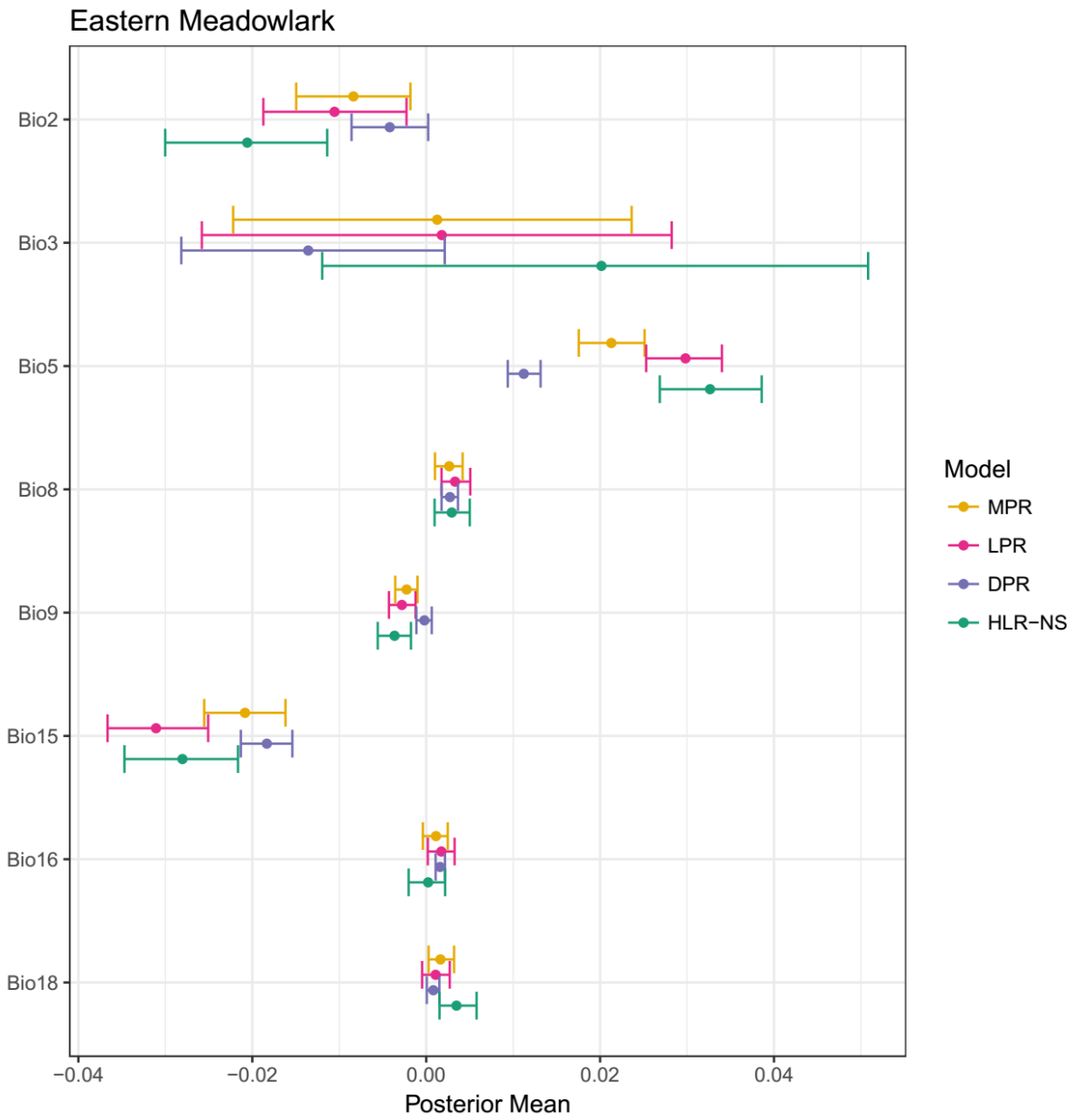


Figure 138: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern meadowlark, *Sturnella magna*.

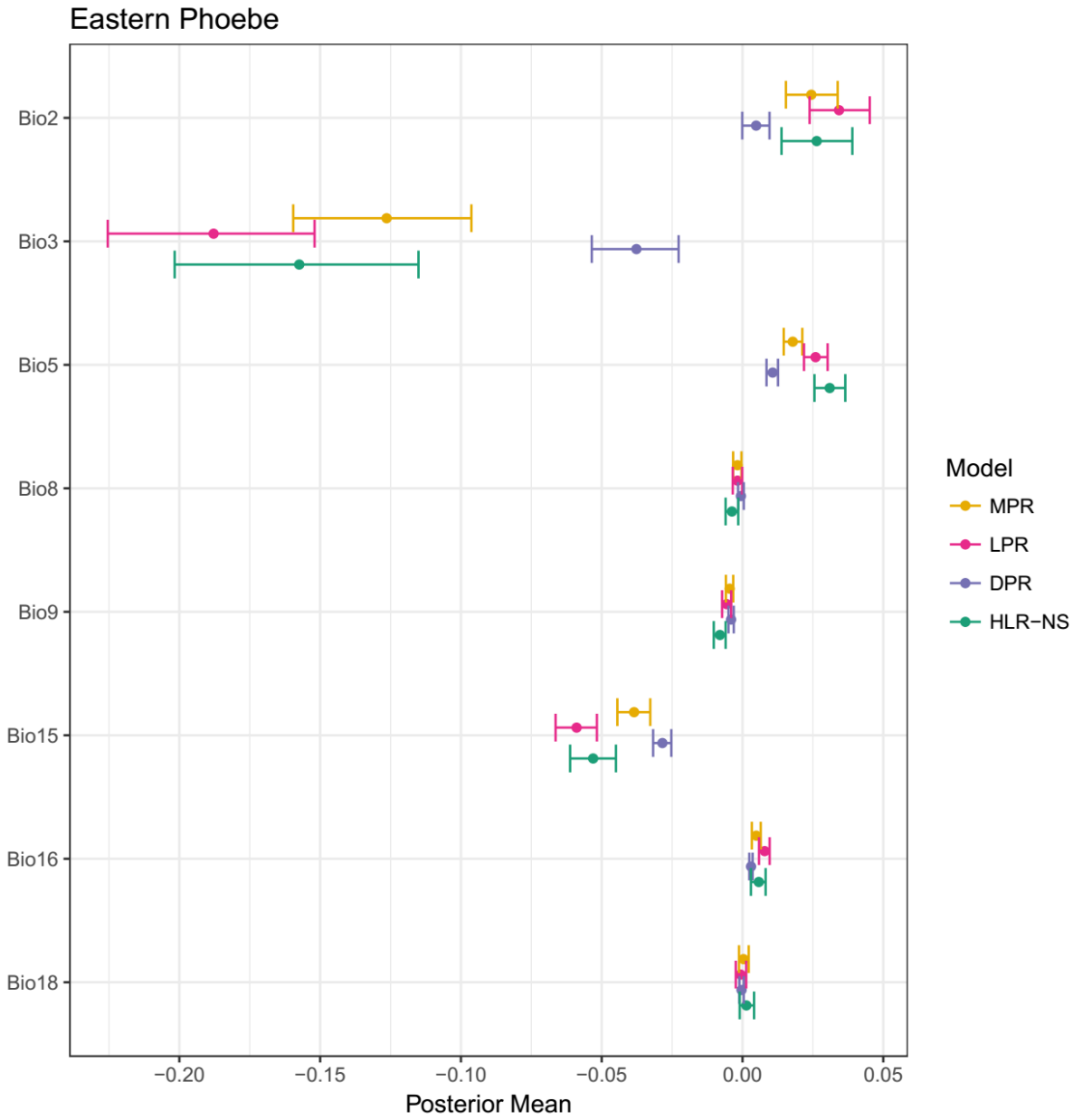


Figure 139: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern phoebe, *Sayornis phoebe*.

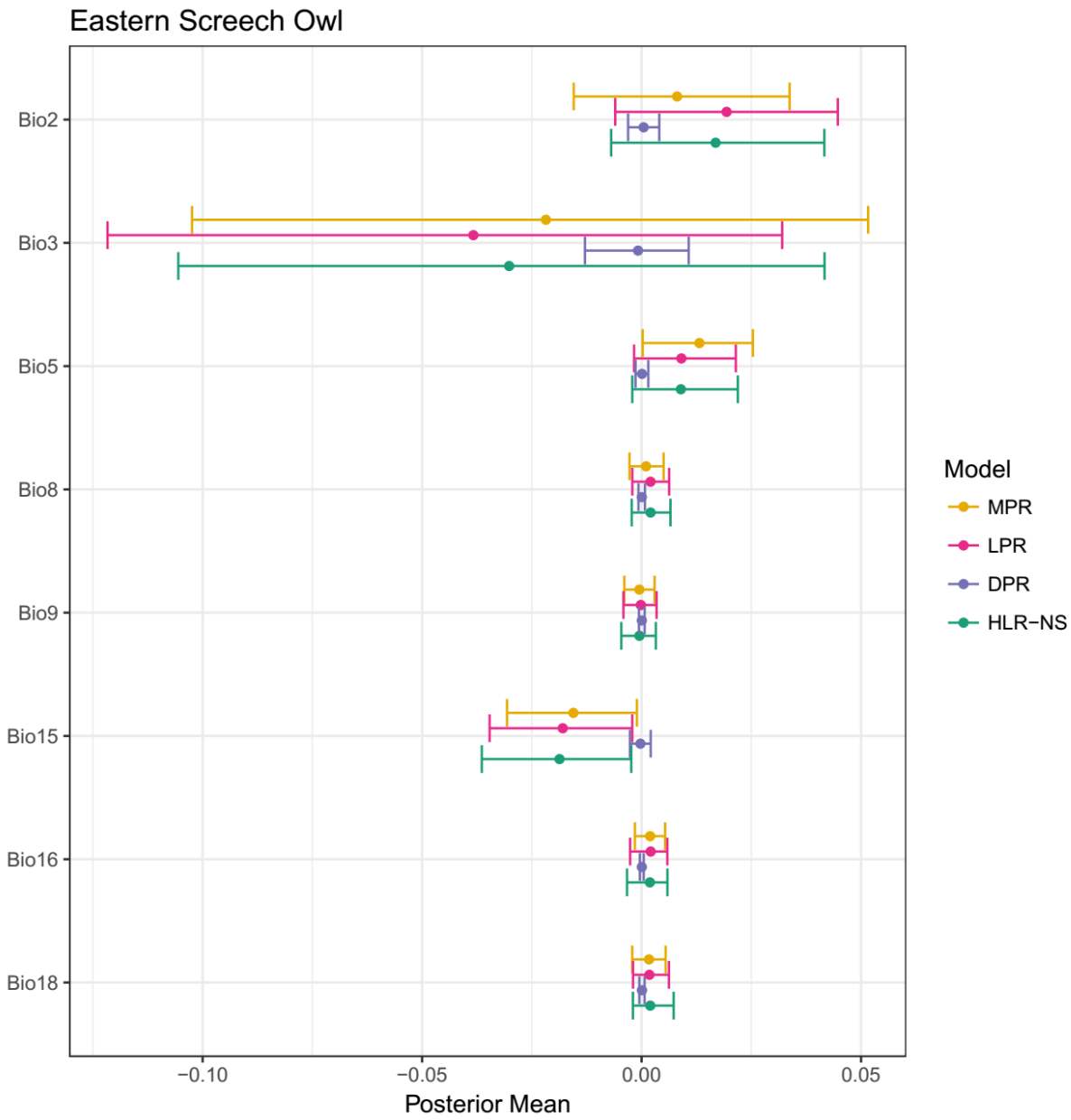


Figure 140: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern screech owl, *Megascops asio*.

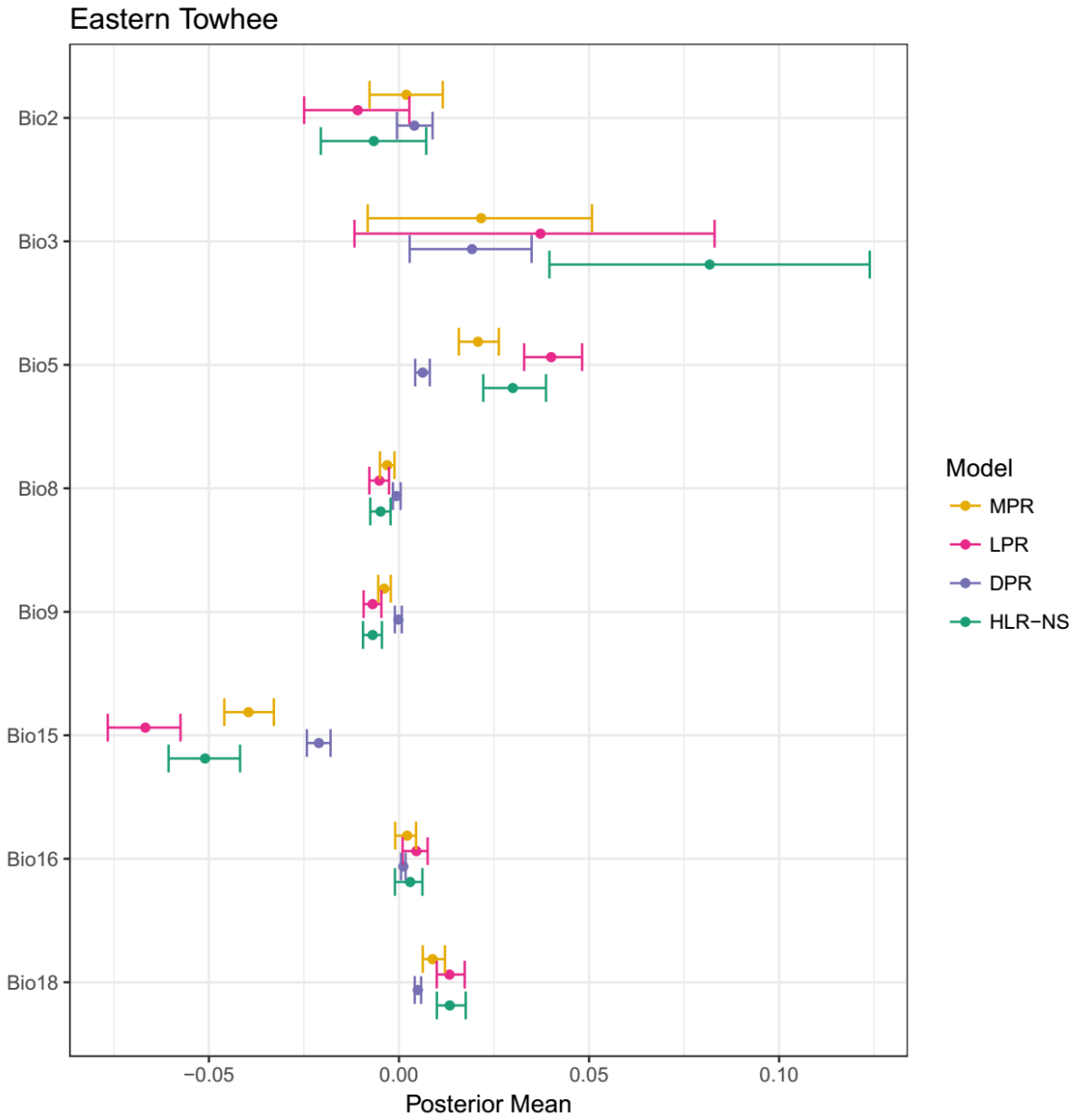


Figure 141: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern towhee, *Pipilo erythrophthalmus*.

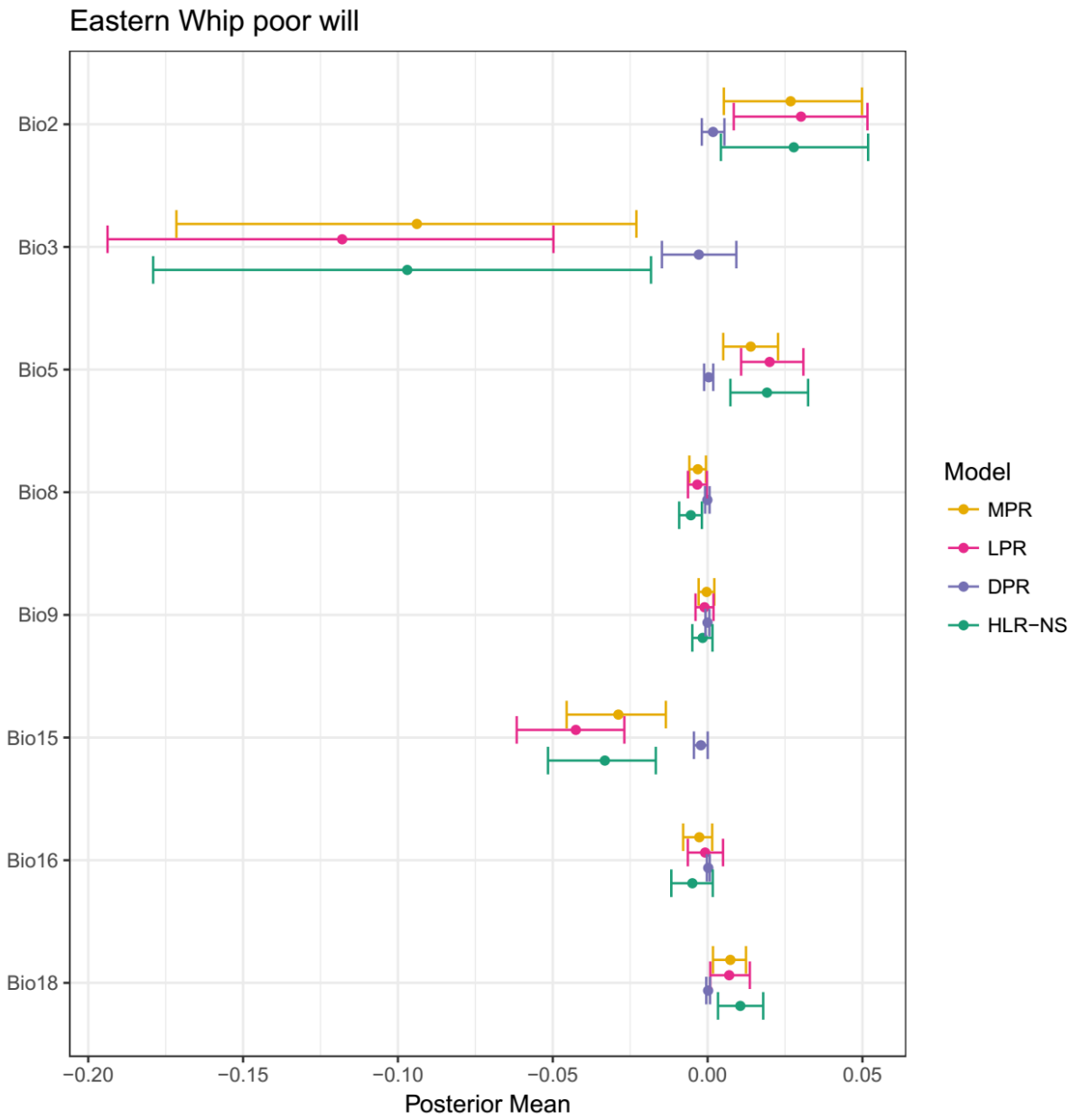


Figure 142: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern whip-poor-will, *Caprimulgus vociferus*.

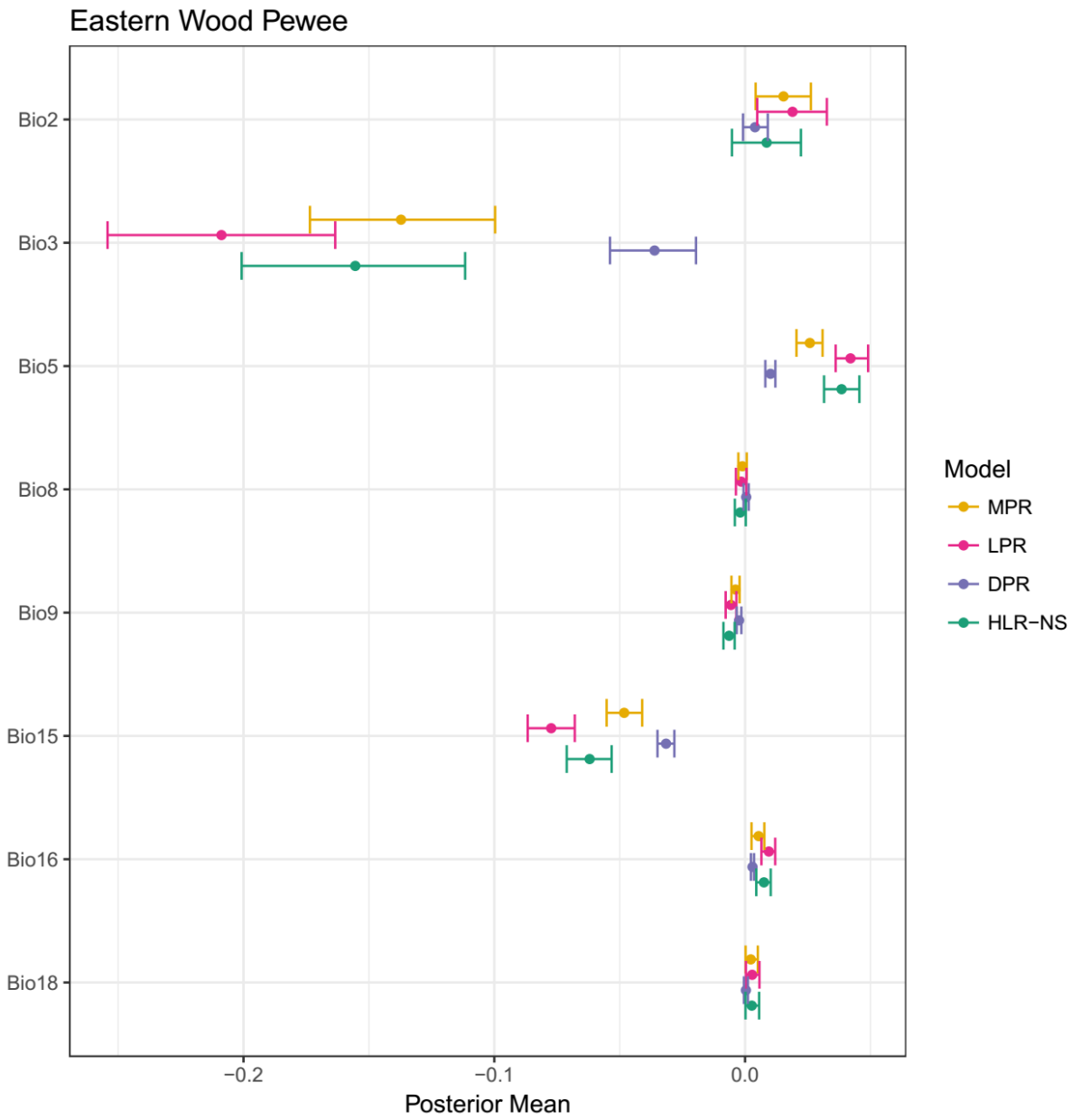


Figure 143: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the eastern wood peewee, *Contopus virens*.

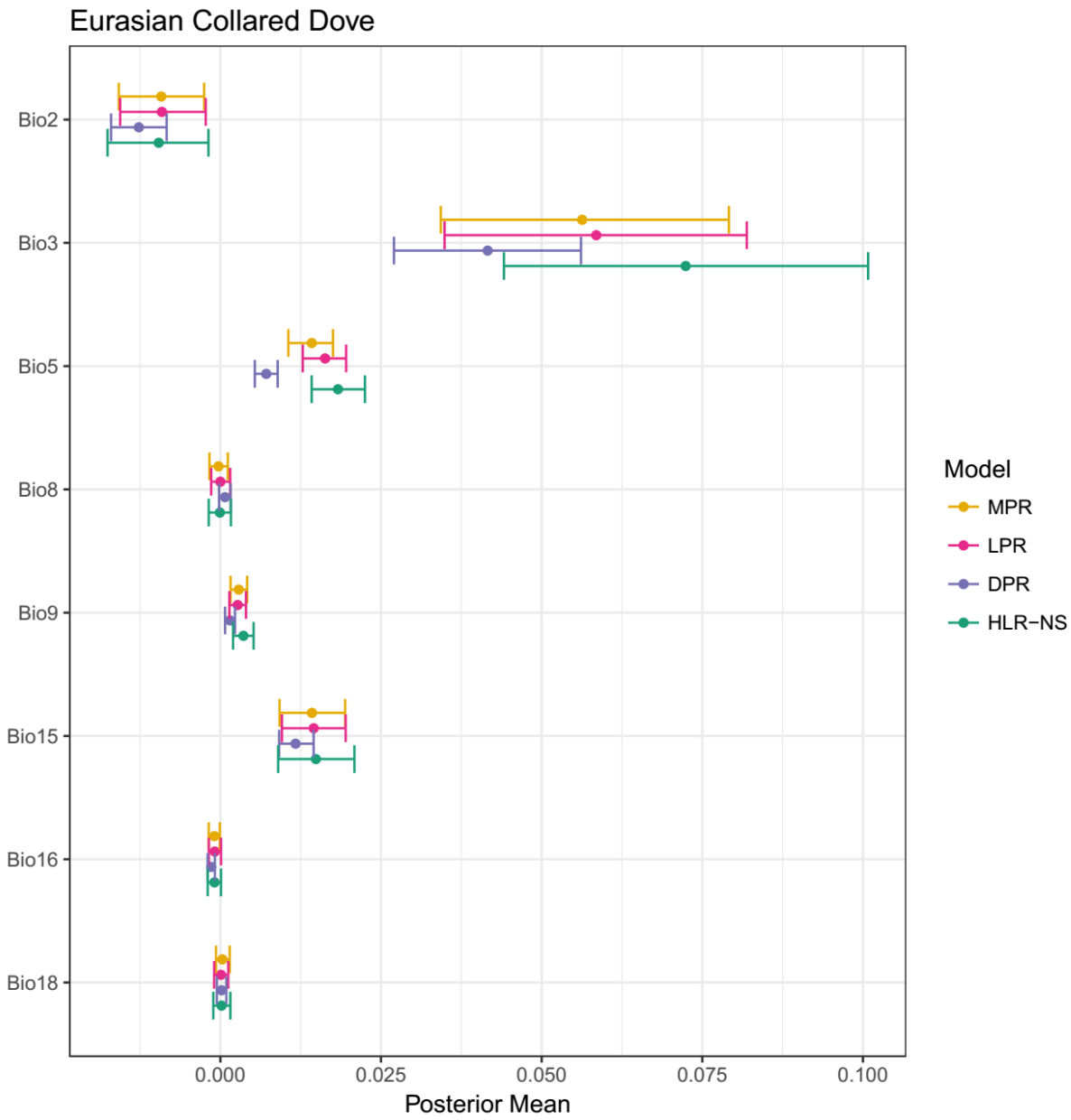


Figure 144: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Eurasian collared dove, *Streptopelia decaocto*.

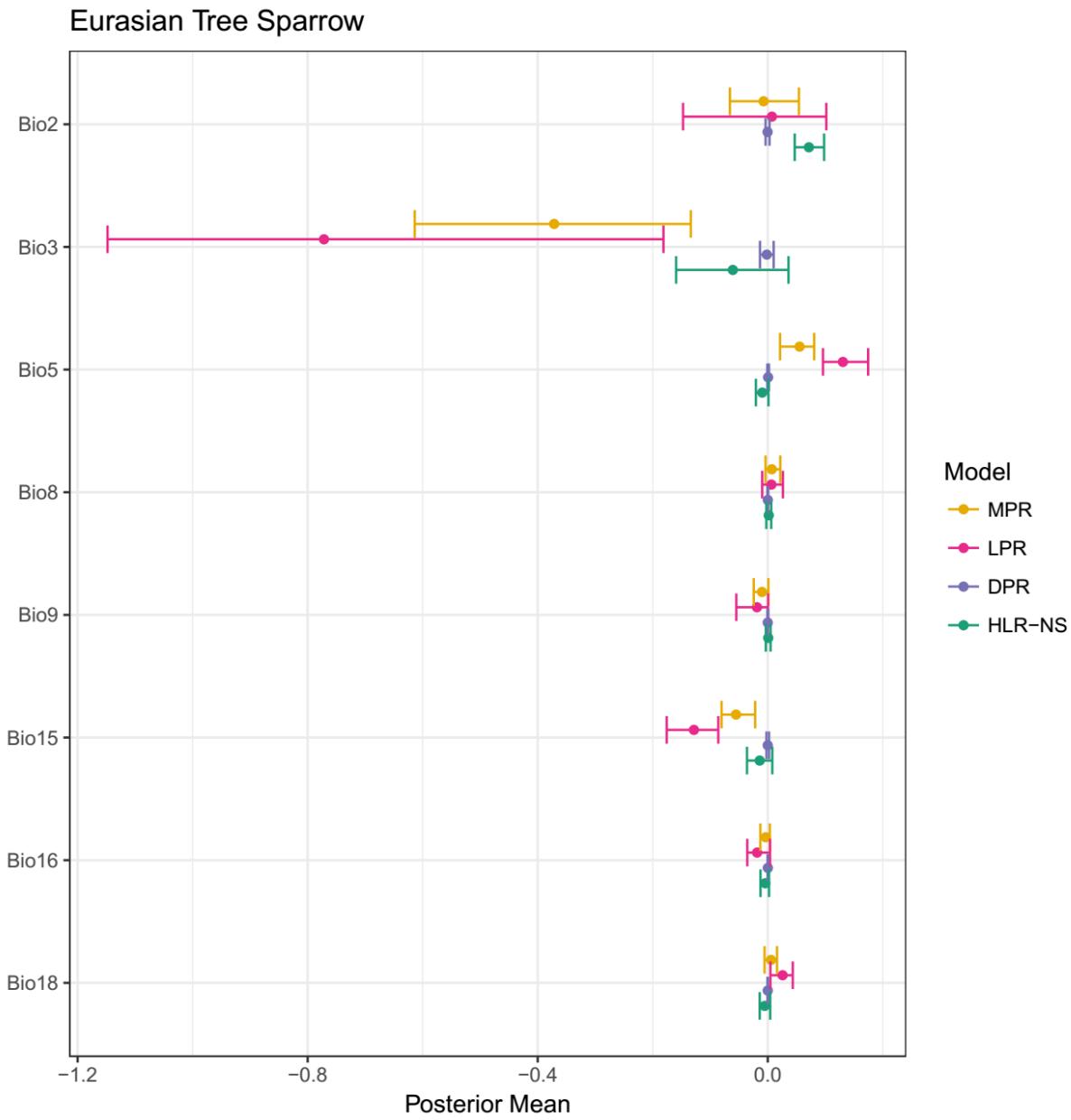


Figure 145: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Eurasian tree sparrow, *Passer montanus*.

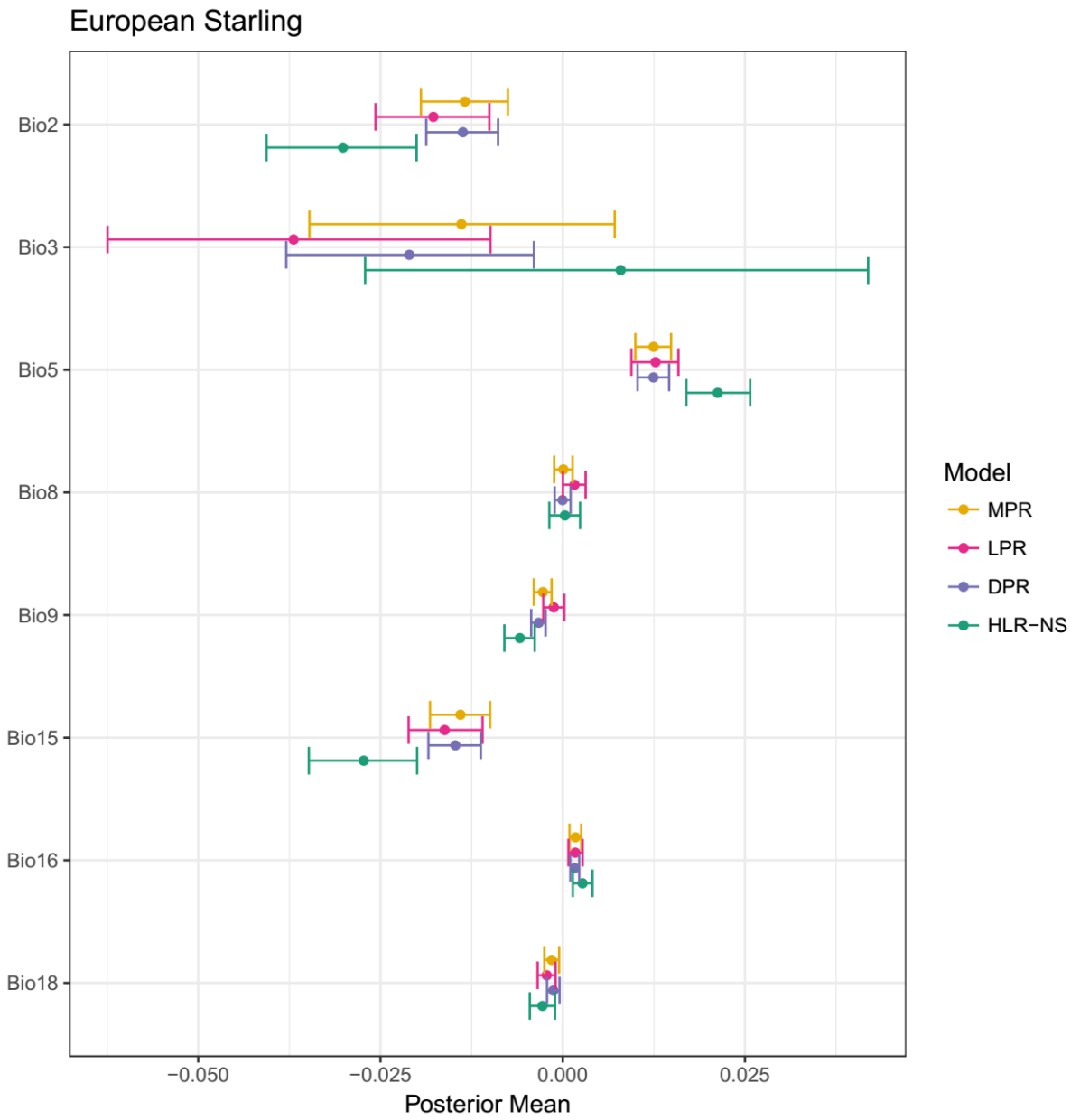


Figure 146: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the European starling, *Sturnus vulgaris*.

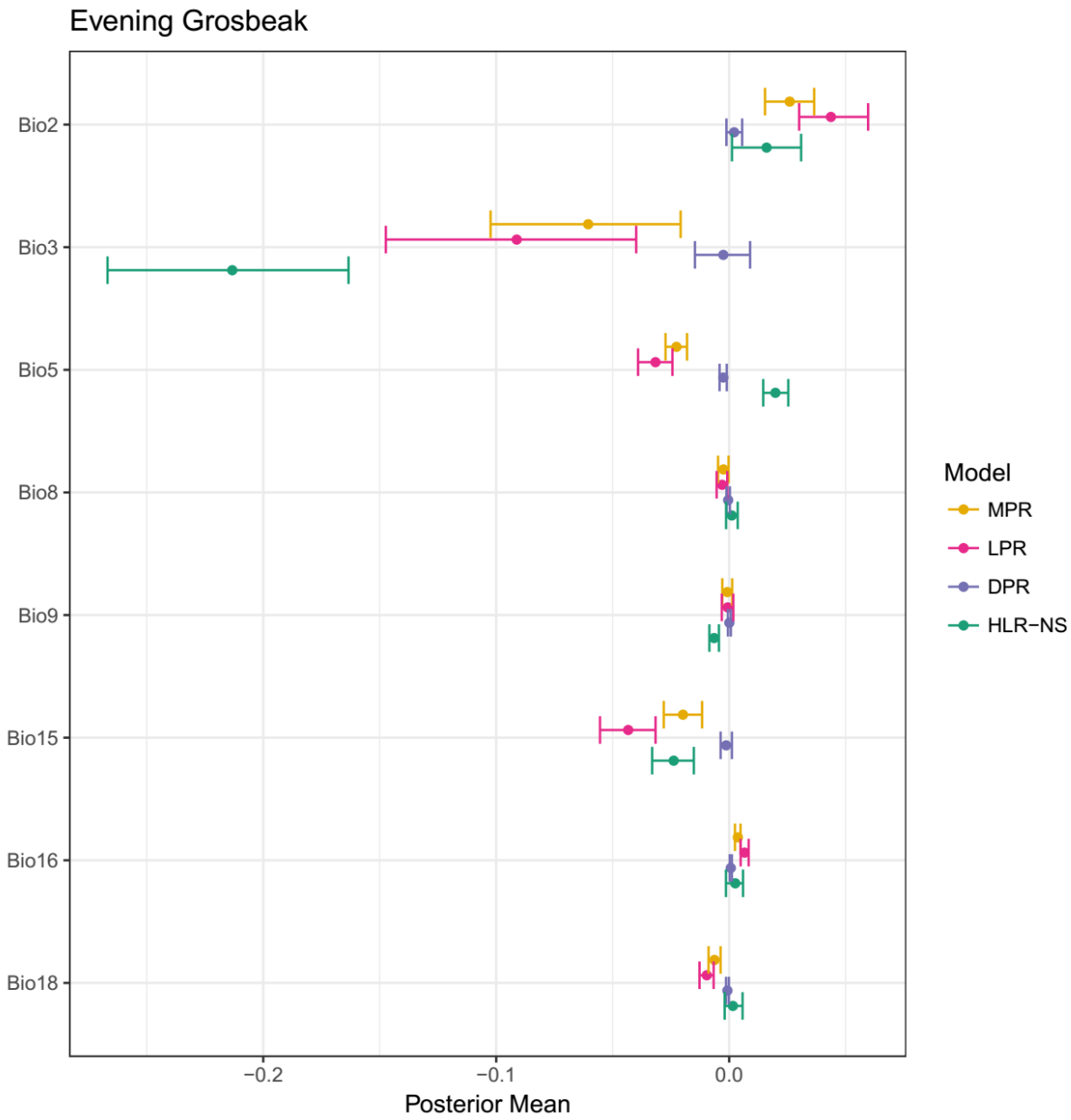


Figure 147: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the evening grosbeak, *Coccothraustes vespertinus.*

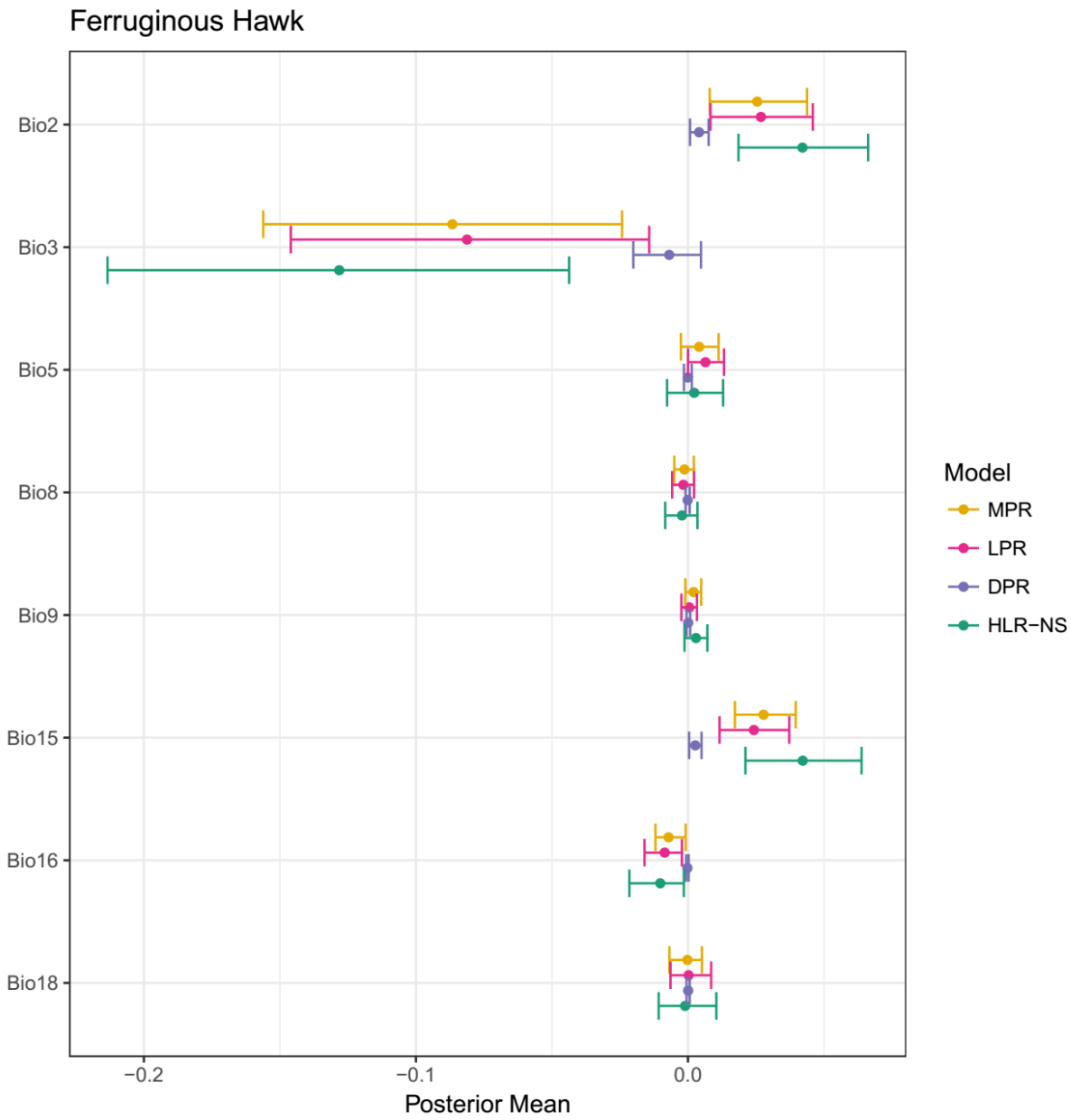


Figure 148: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ferruginous hawk, *Buteo regalis*.

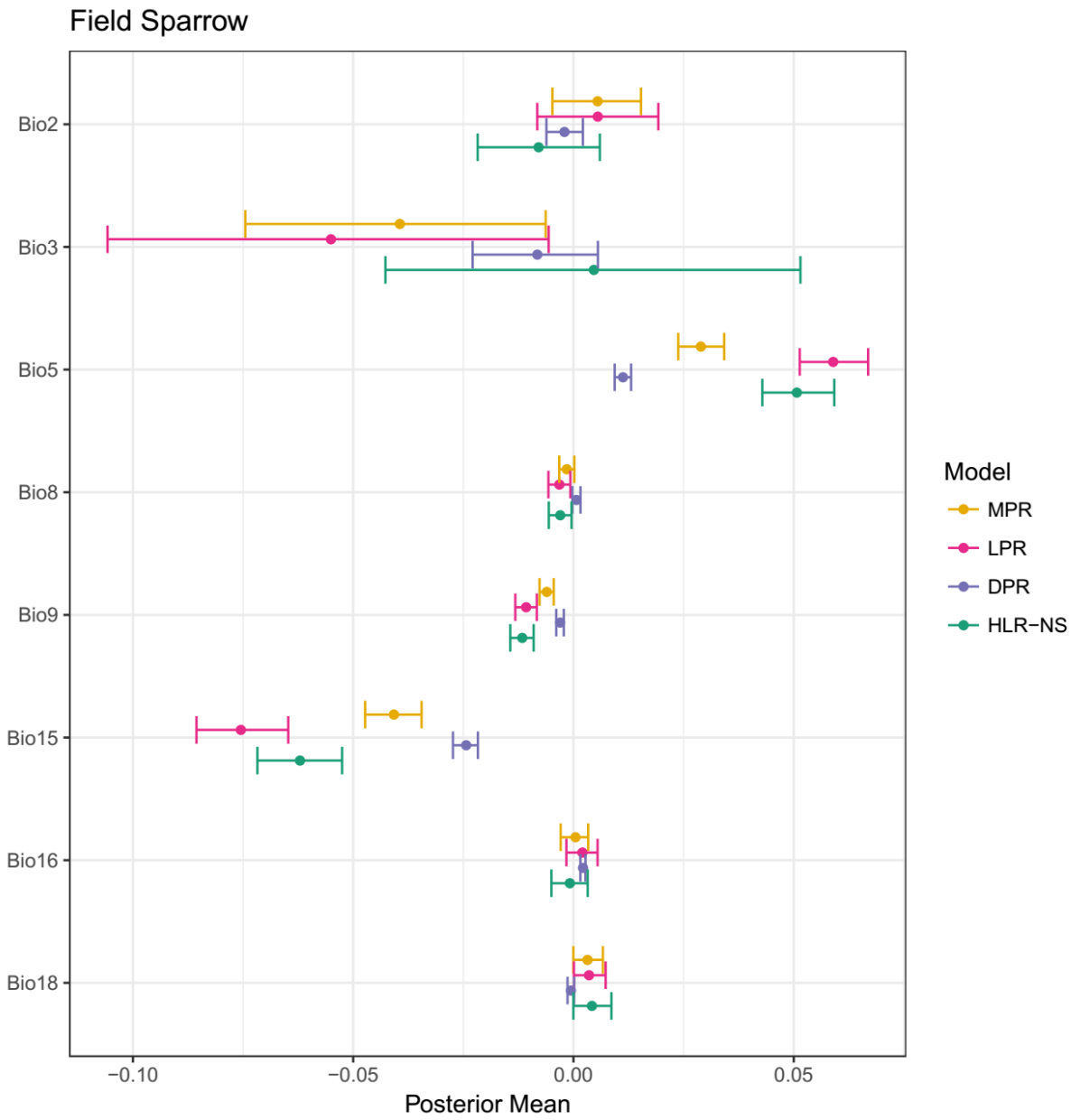


Figure 149: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the field sparrow, *Spizella pusilla*.

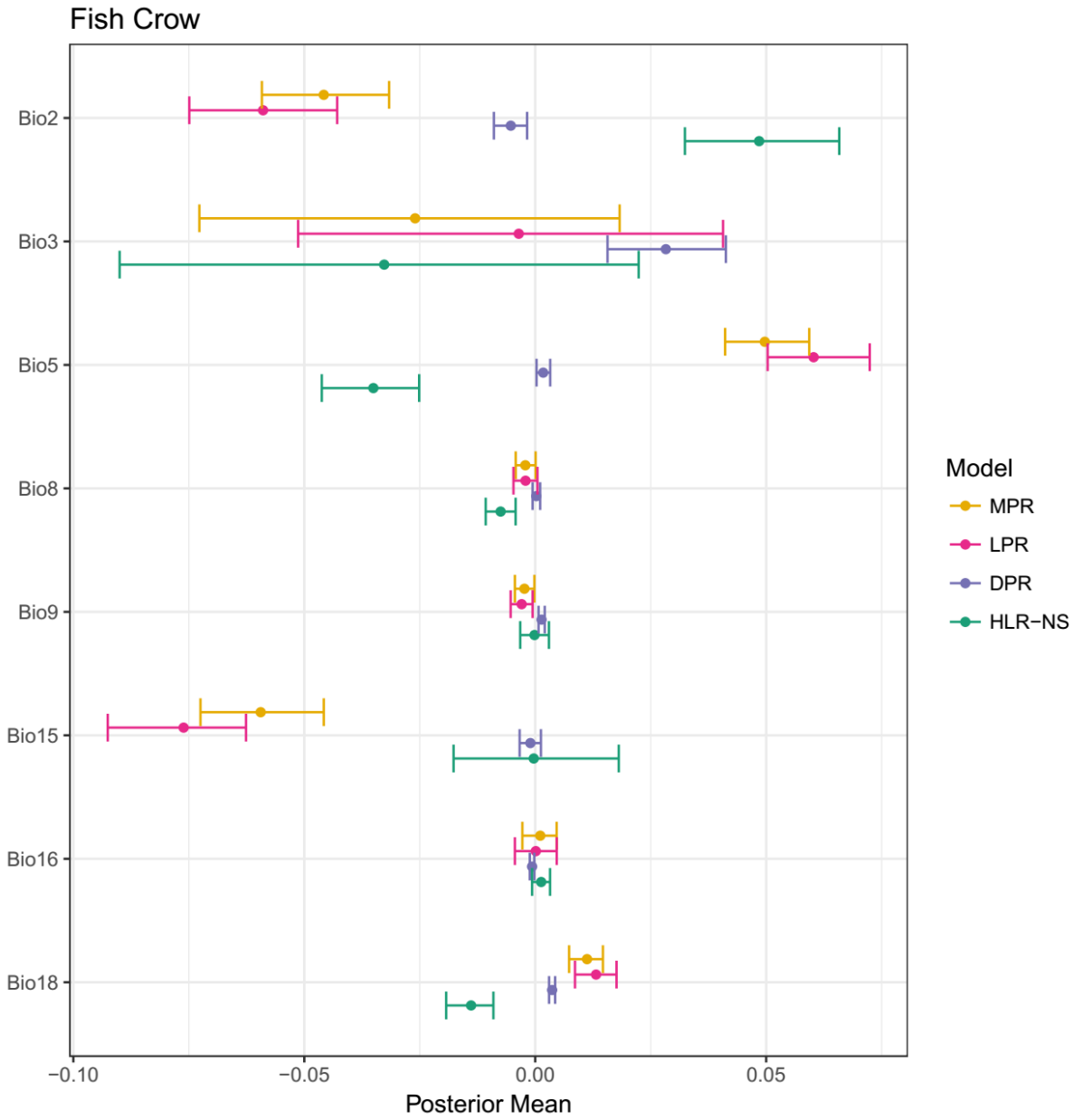


Figure 150: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the fish crow, *Corvus ossifragus*.

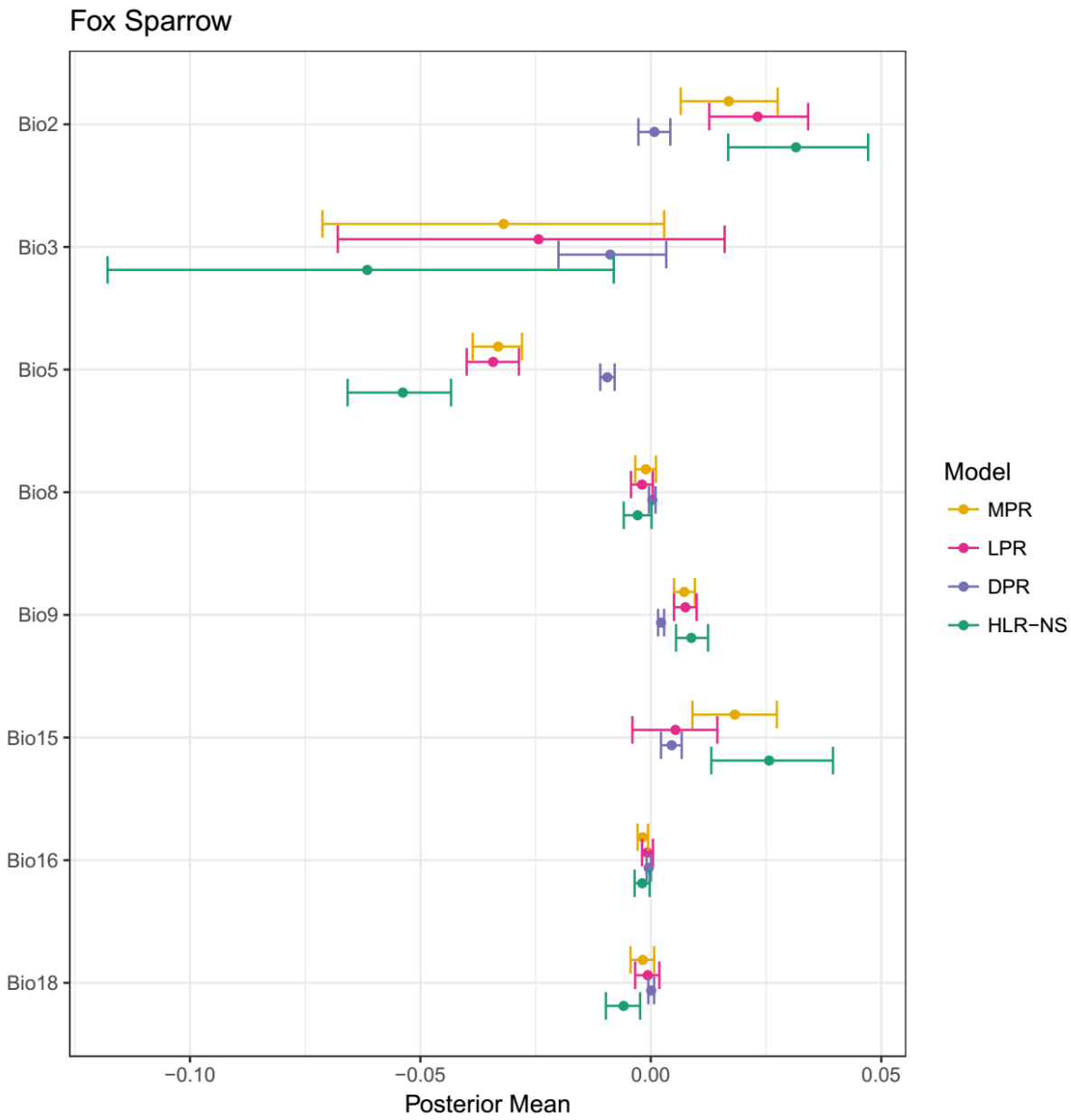


Figure 151: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the fox sparrow, *Passerella iliaca*.

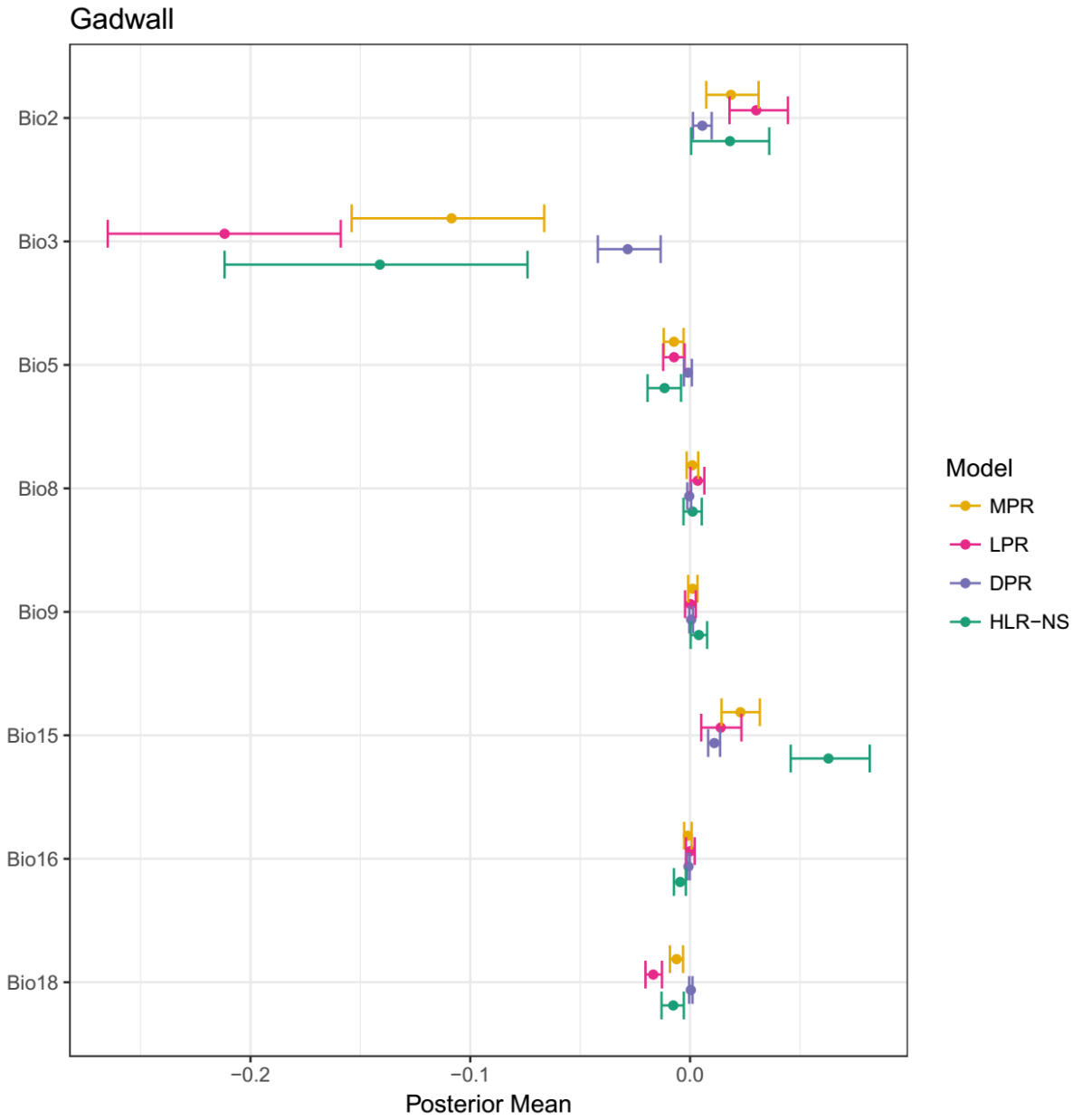


Figure 152: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gadwall, *Anas strepera*.

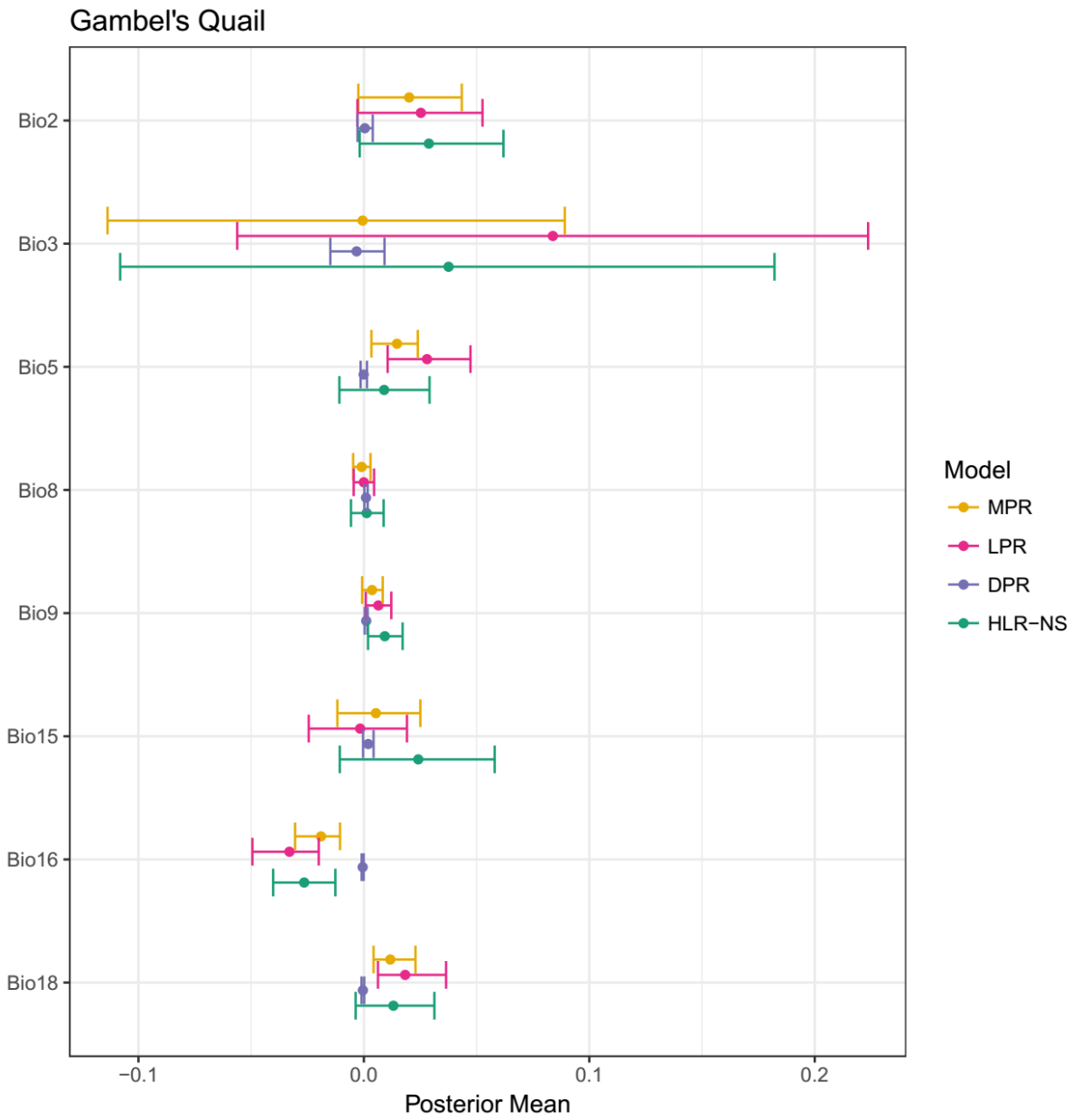


Figure 153: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Gambel's quail, *Callipepla gambelii*.

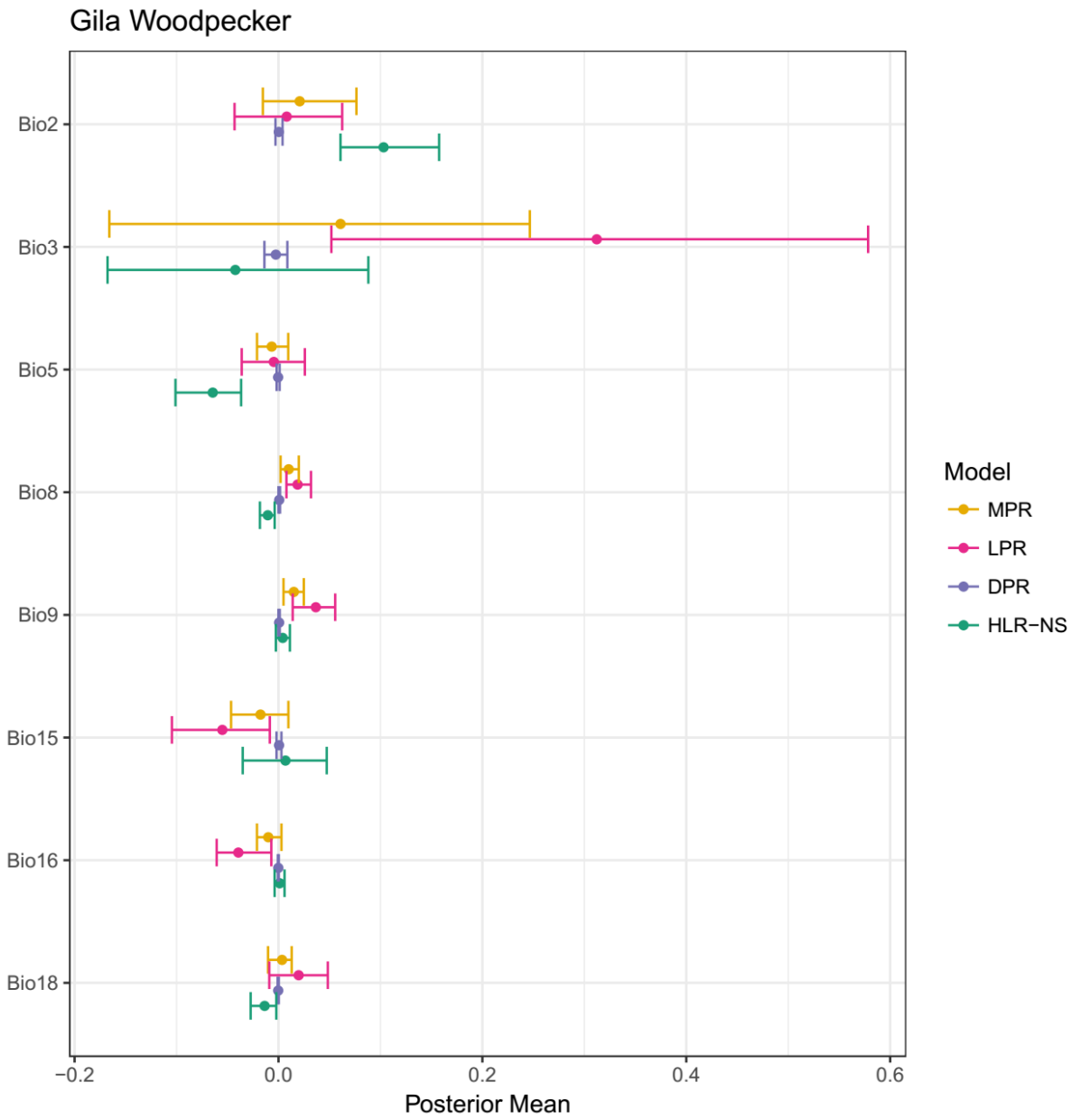


Figure 154: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Gila woodpecker, *Melanerpes uropygialis*.

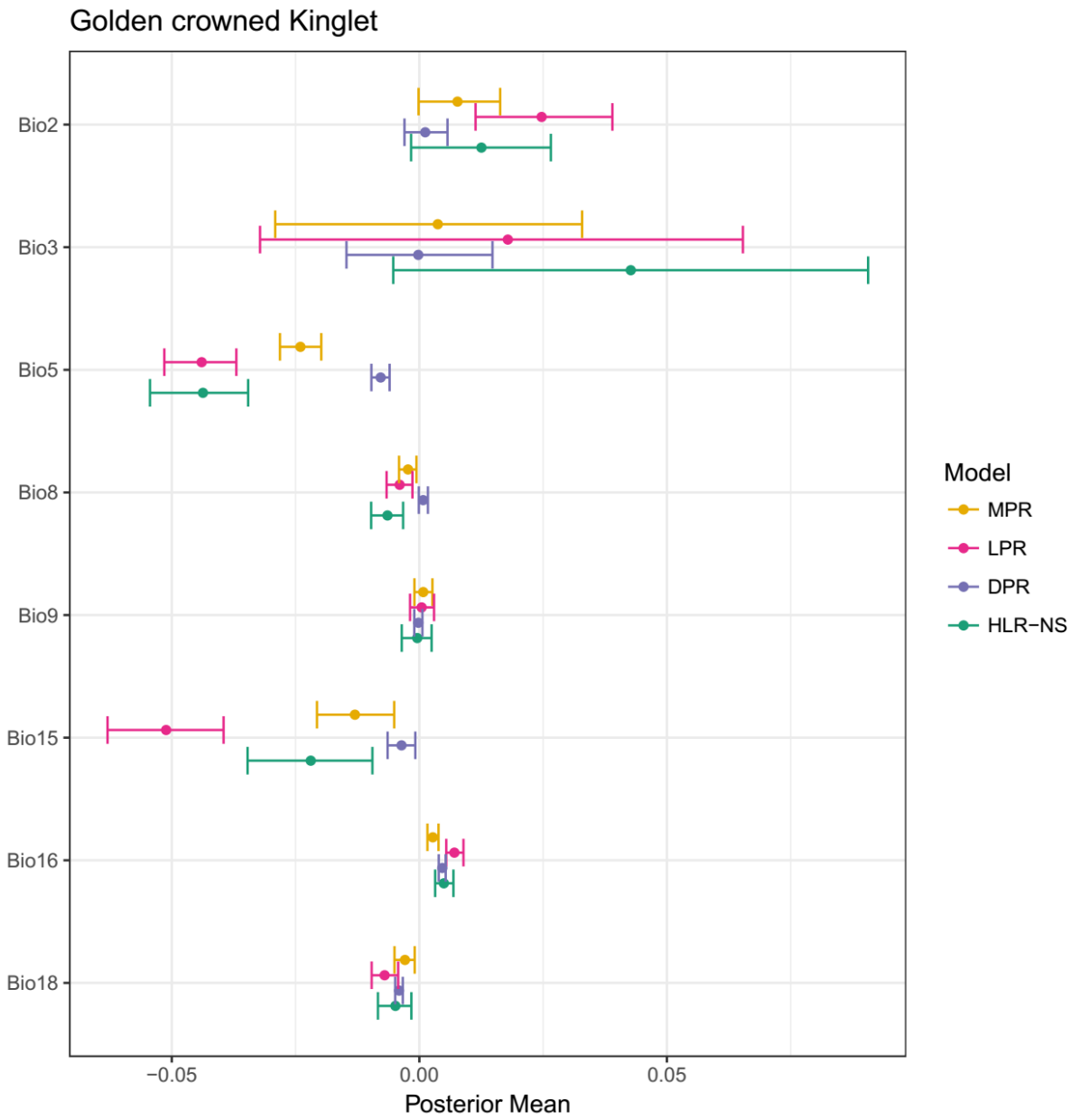


Figure 155: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the golden-crowned kinglet, *Regulus satrapa*.

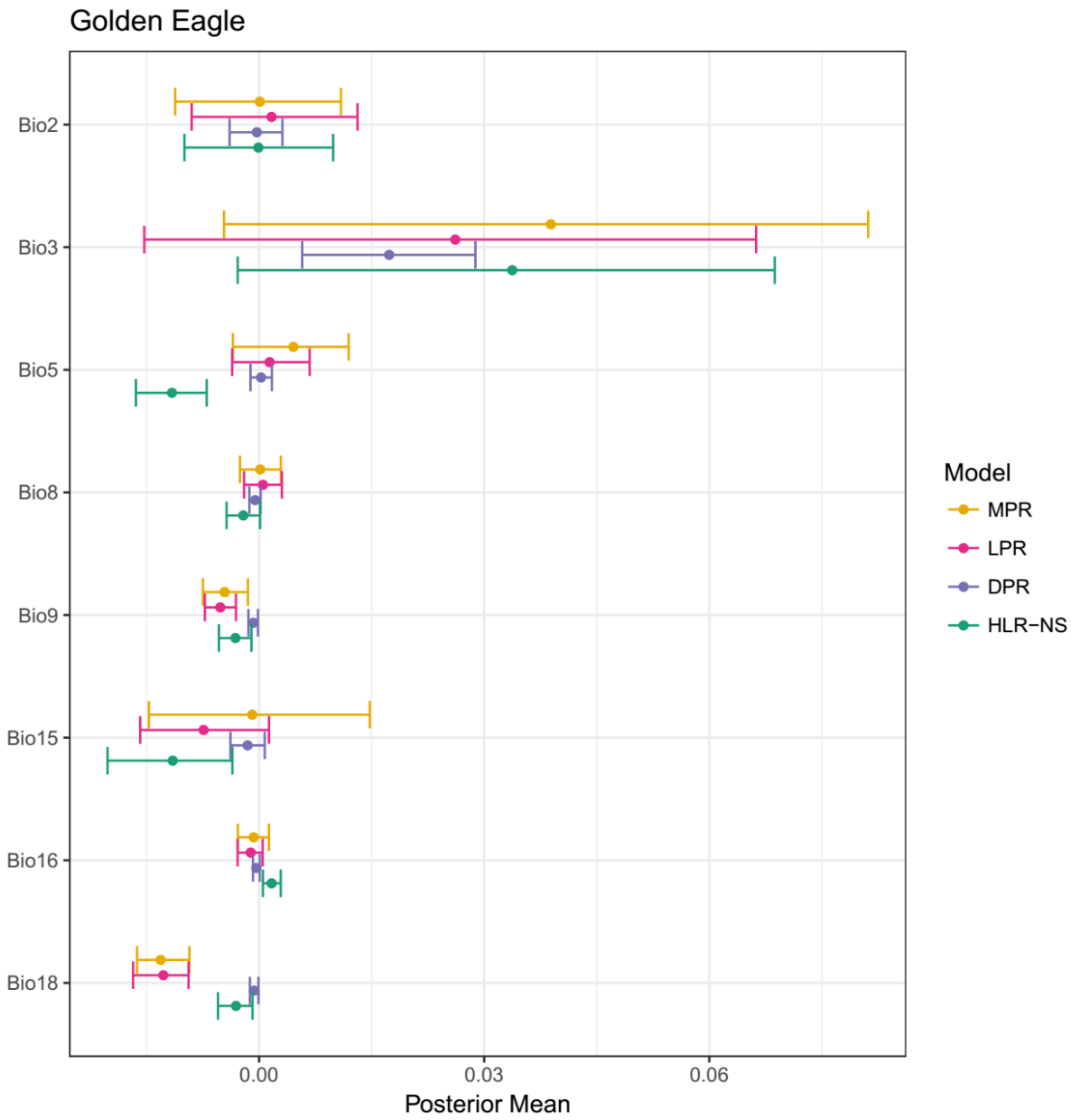


Figure 156: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the golden eagle, *Aquila chrysaetos*.

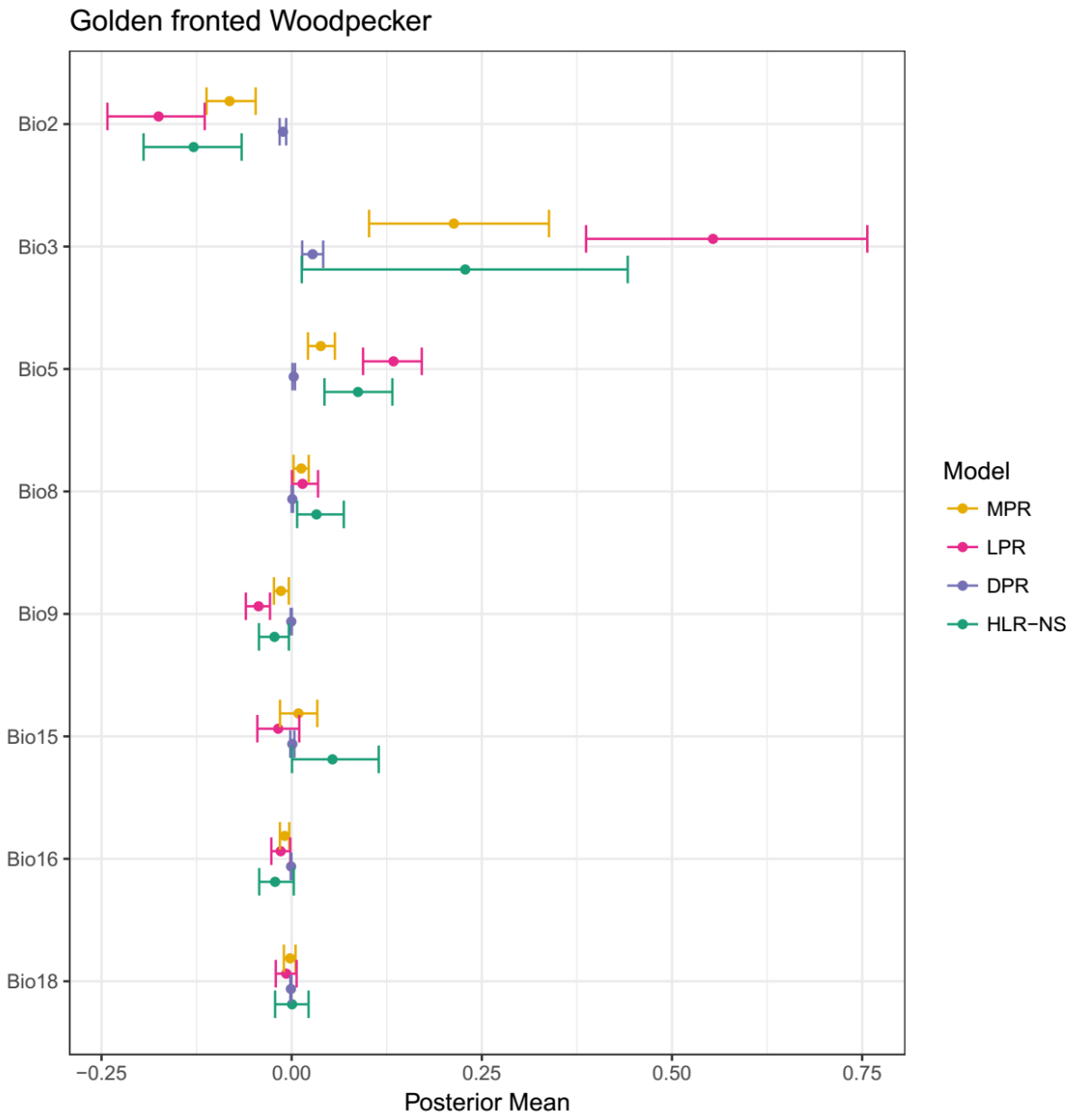


Figure 157: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the golden-fronted woodpecker, *Melanerpes aurifrons*.

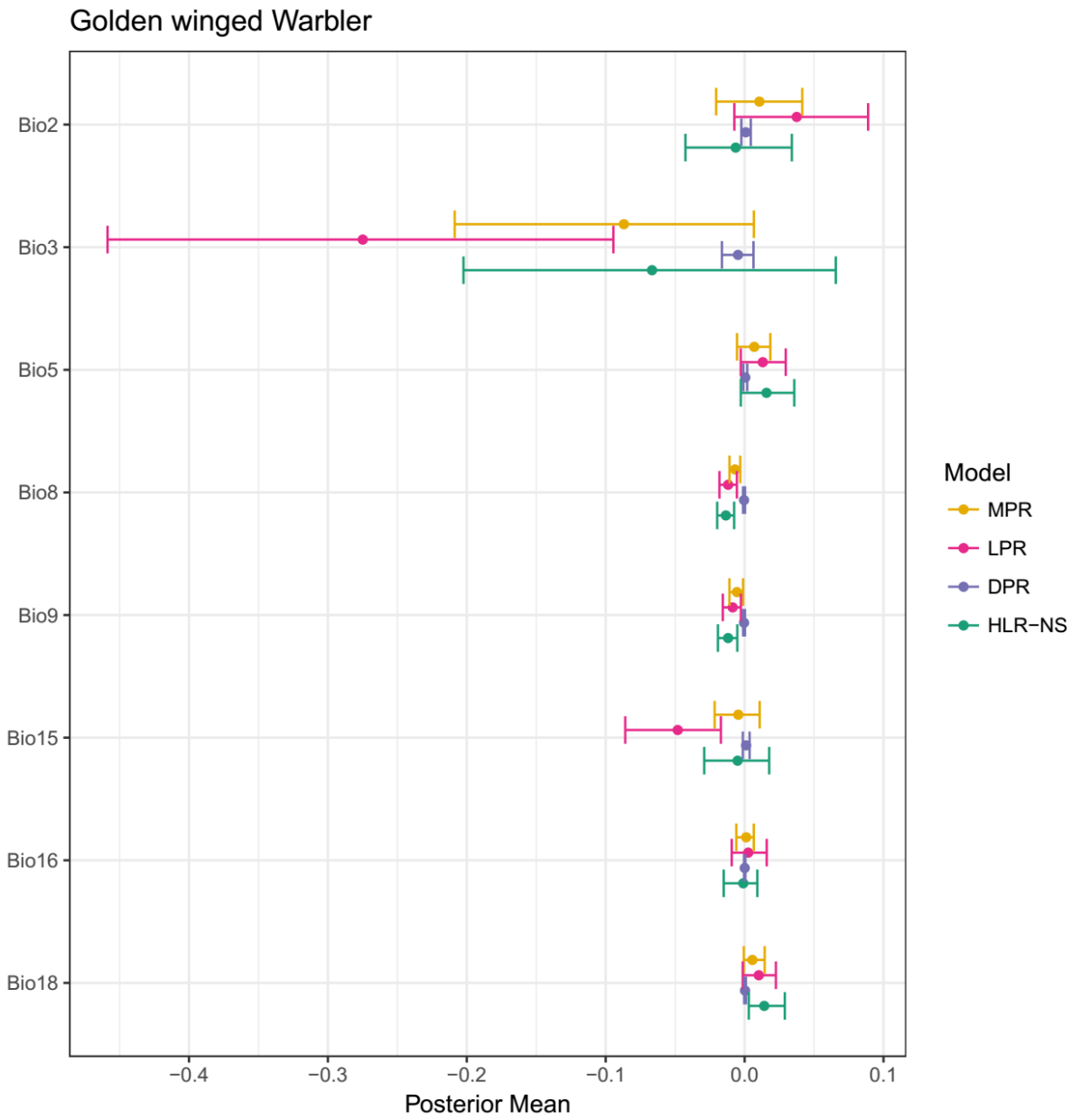


Figure 158: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the golden-winged warbler, *Vermivora chrysoptera*.

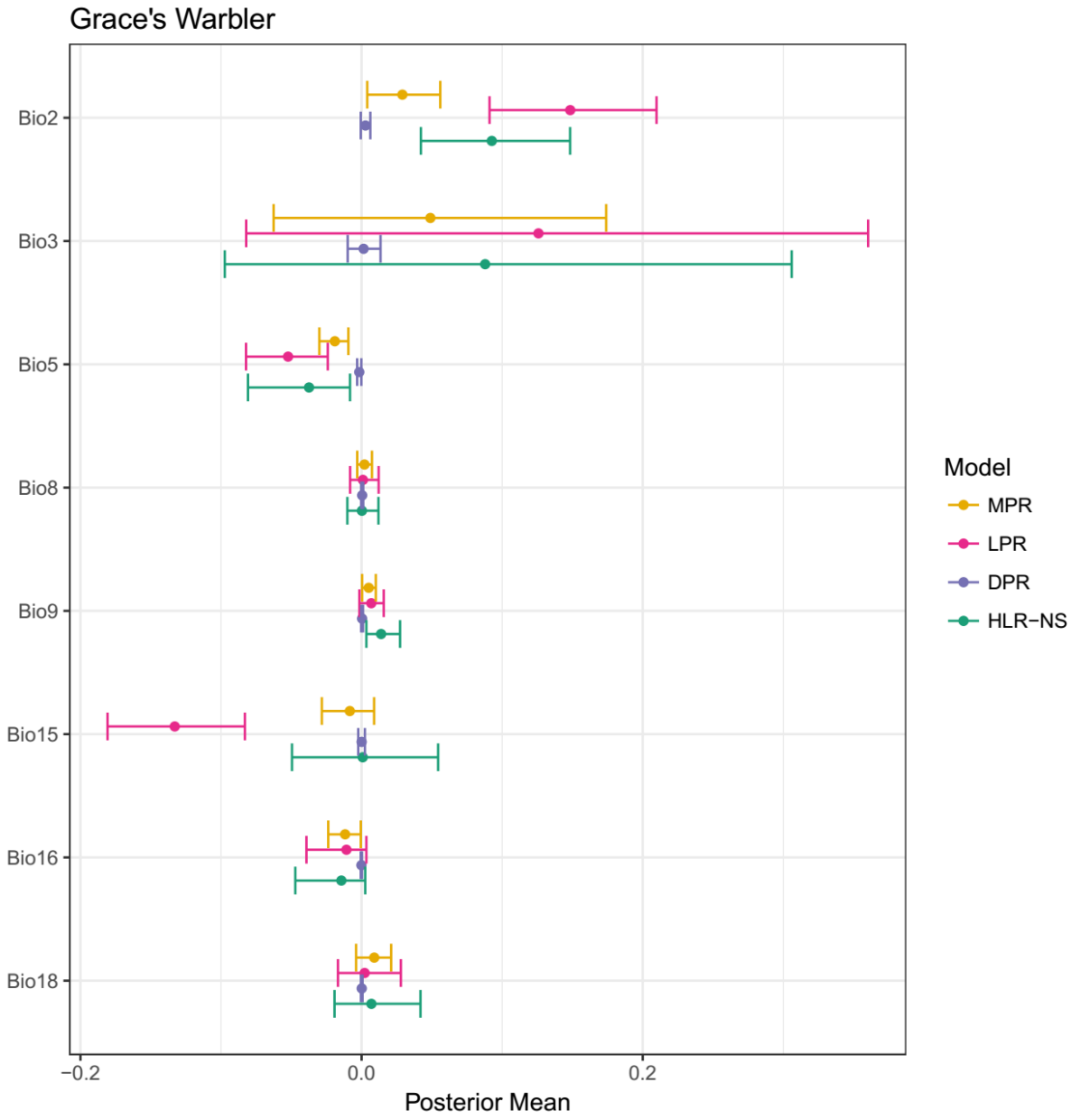


Figure 159: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Grace's warbler, *Setophaga graciae*.

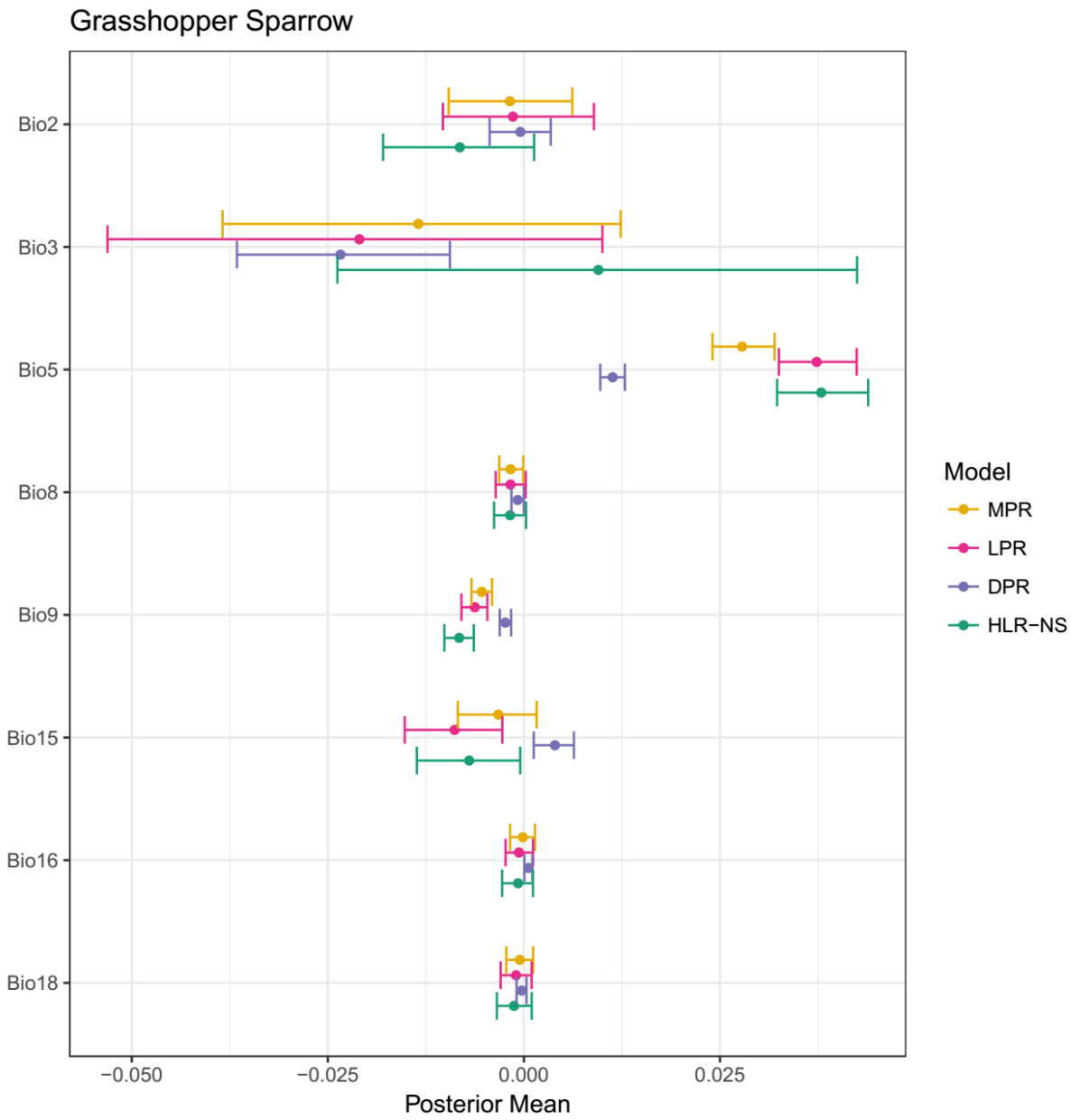


Figure 160: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the grasshopper sparrow, *Ammodramus savannarum*.

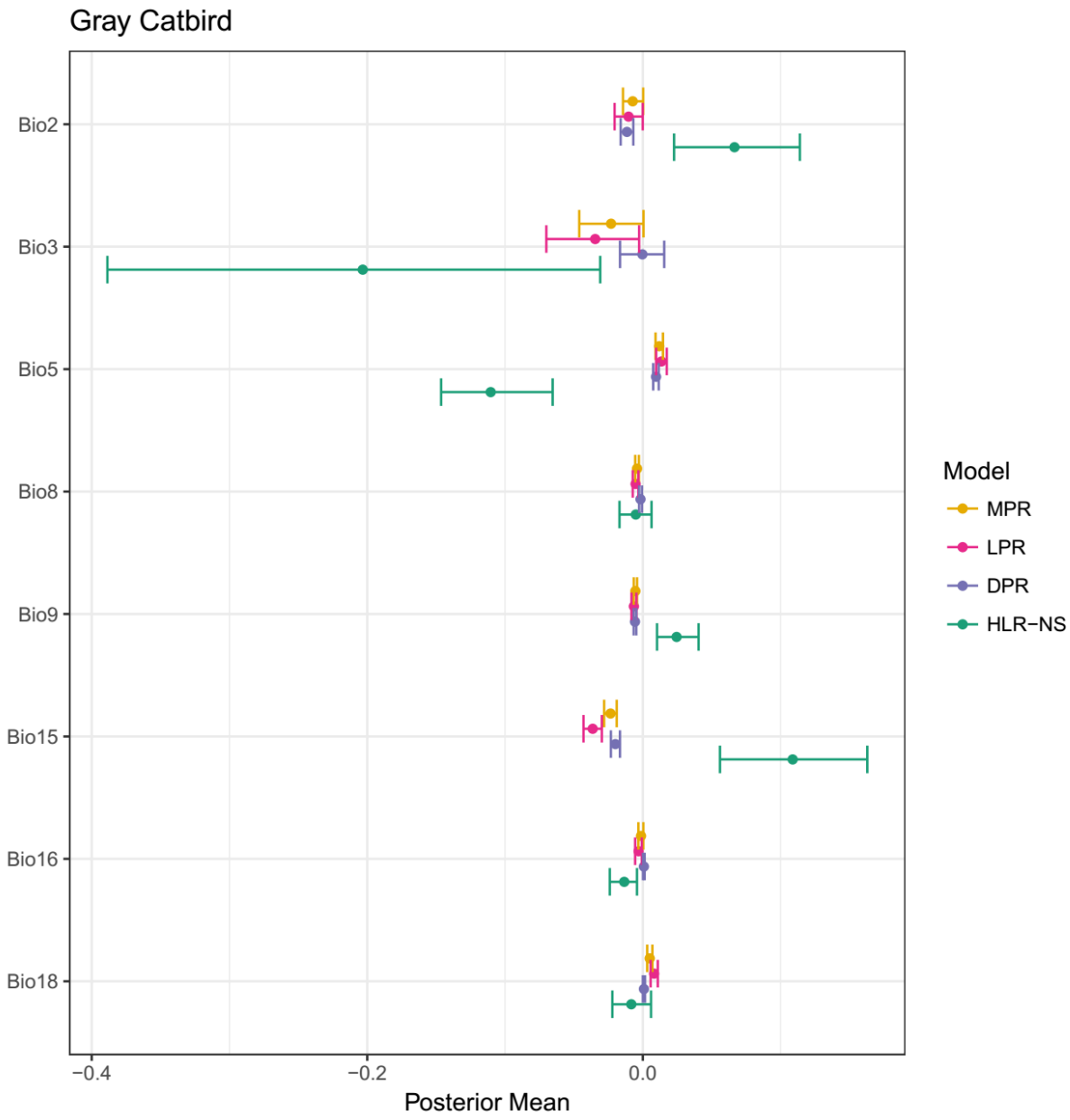


Figure 161: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gray catbird, *Dumetella carolinensis*.

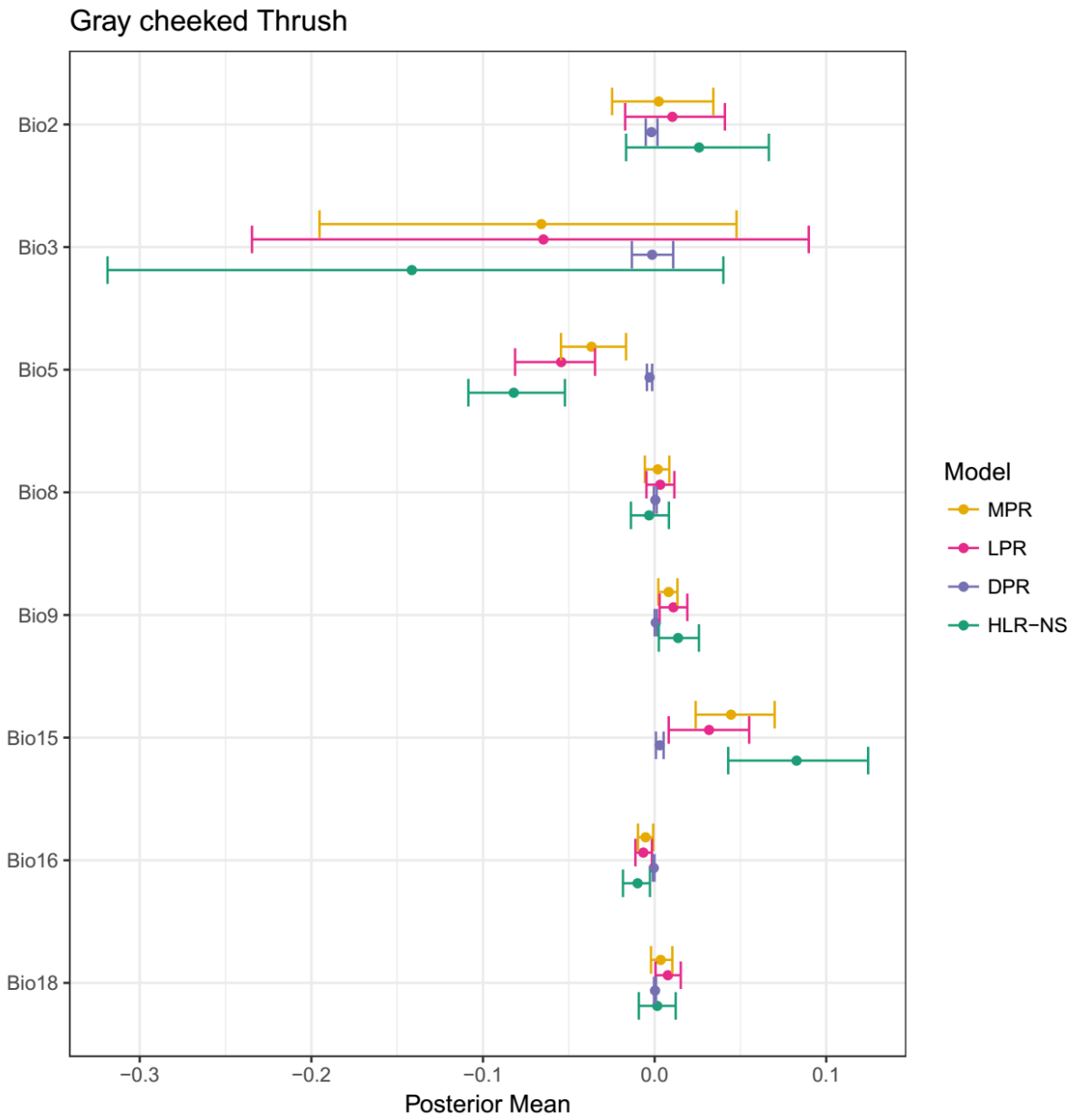


Figure 162: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gray-cheeked thrush, *Catharus minimus*.

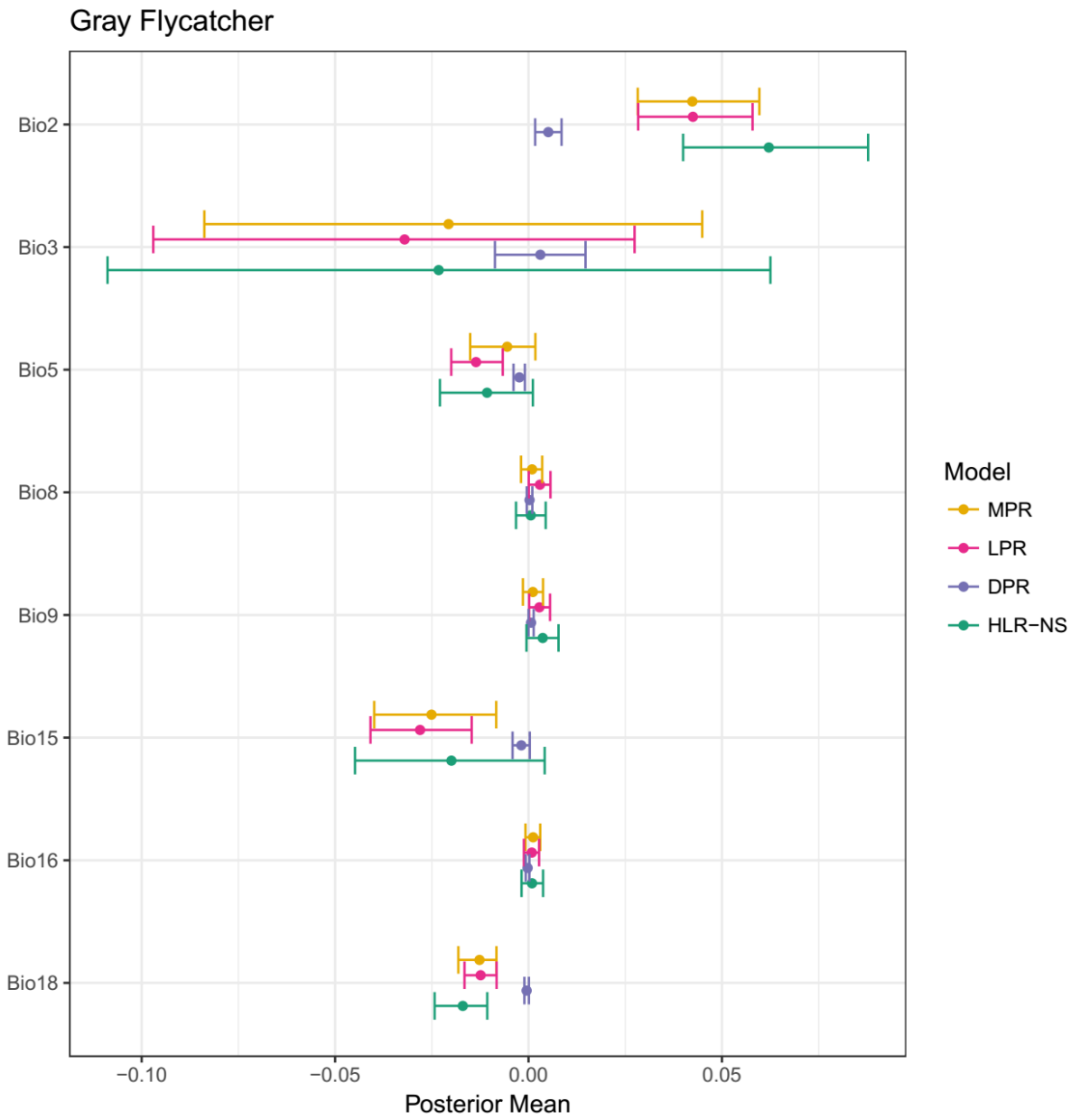


Figure 163: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gray flycatcher, *Empidonax wrightii*.

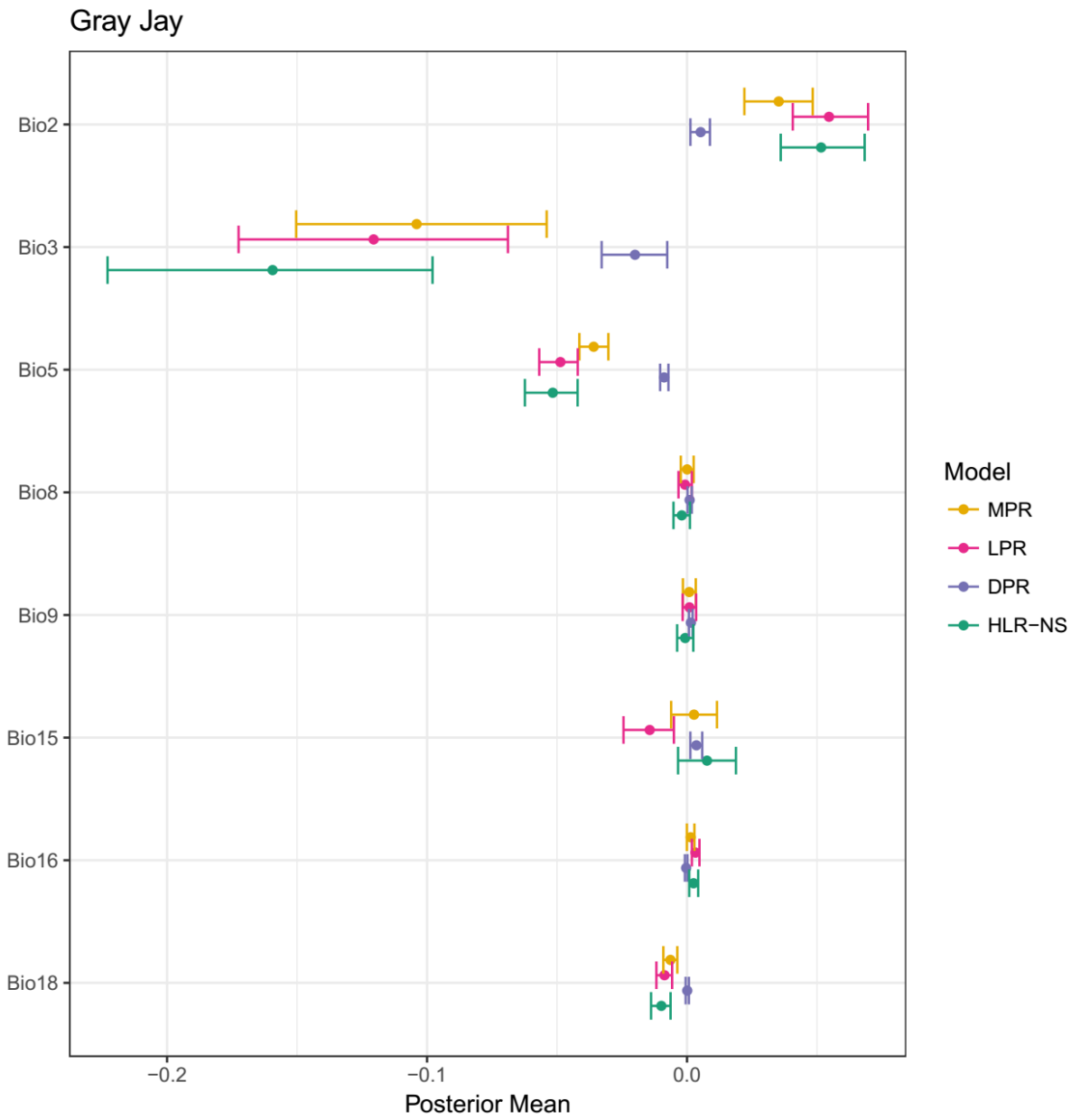


Figure 164: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gray jay, *Perisoreus canadensis*.

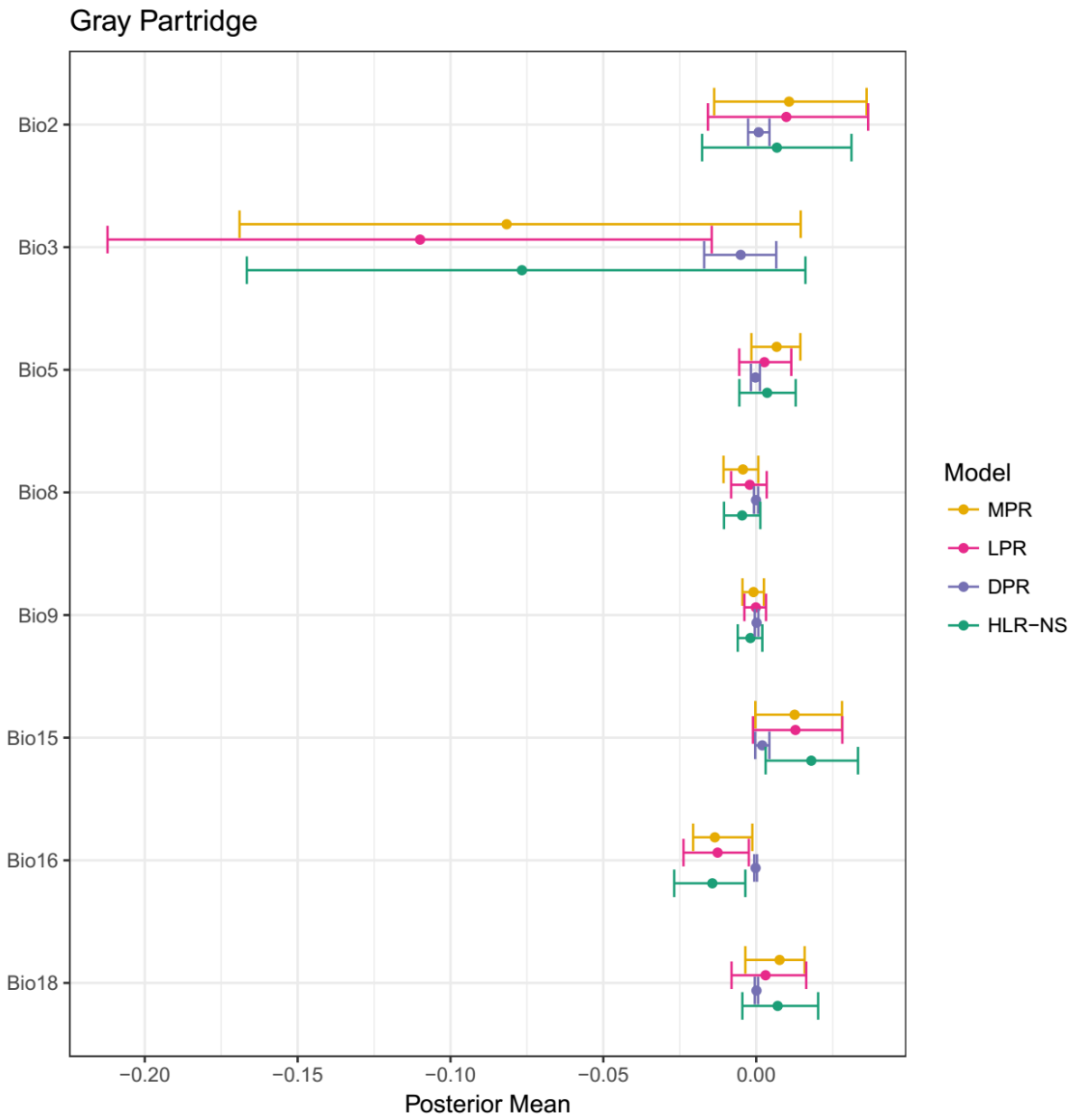


Figure 165: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gray partridge, *Perdix perdix*.

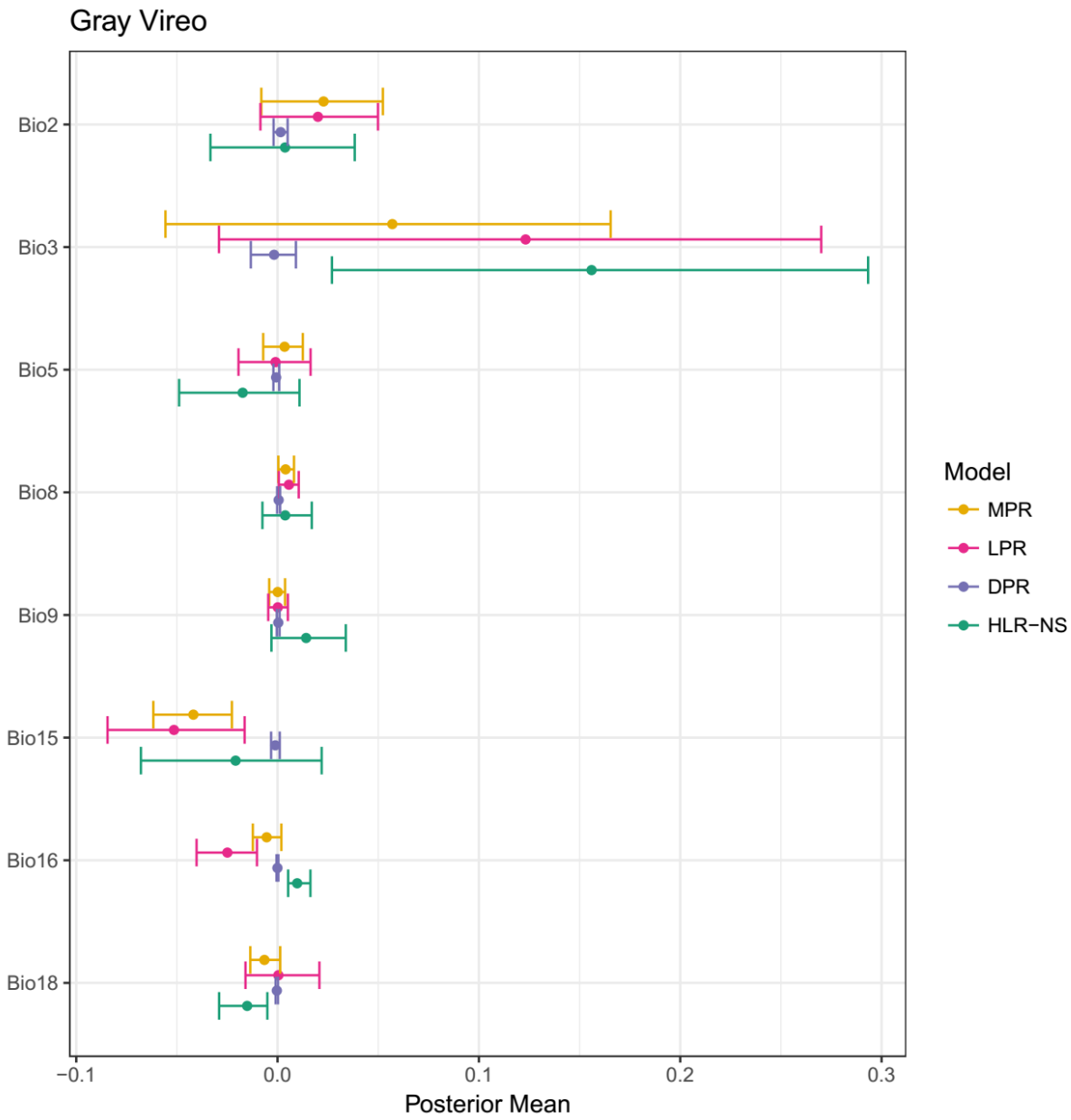


Figure 166: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gray vireo, *Vireo vicinior*.

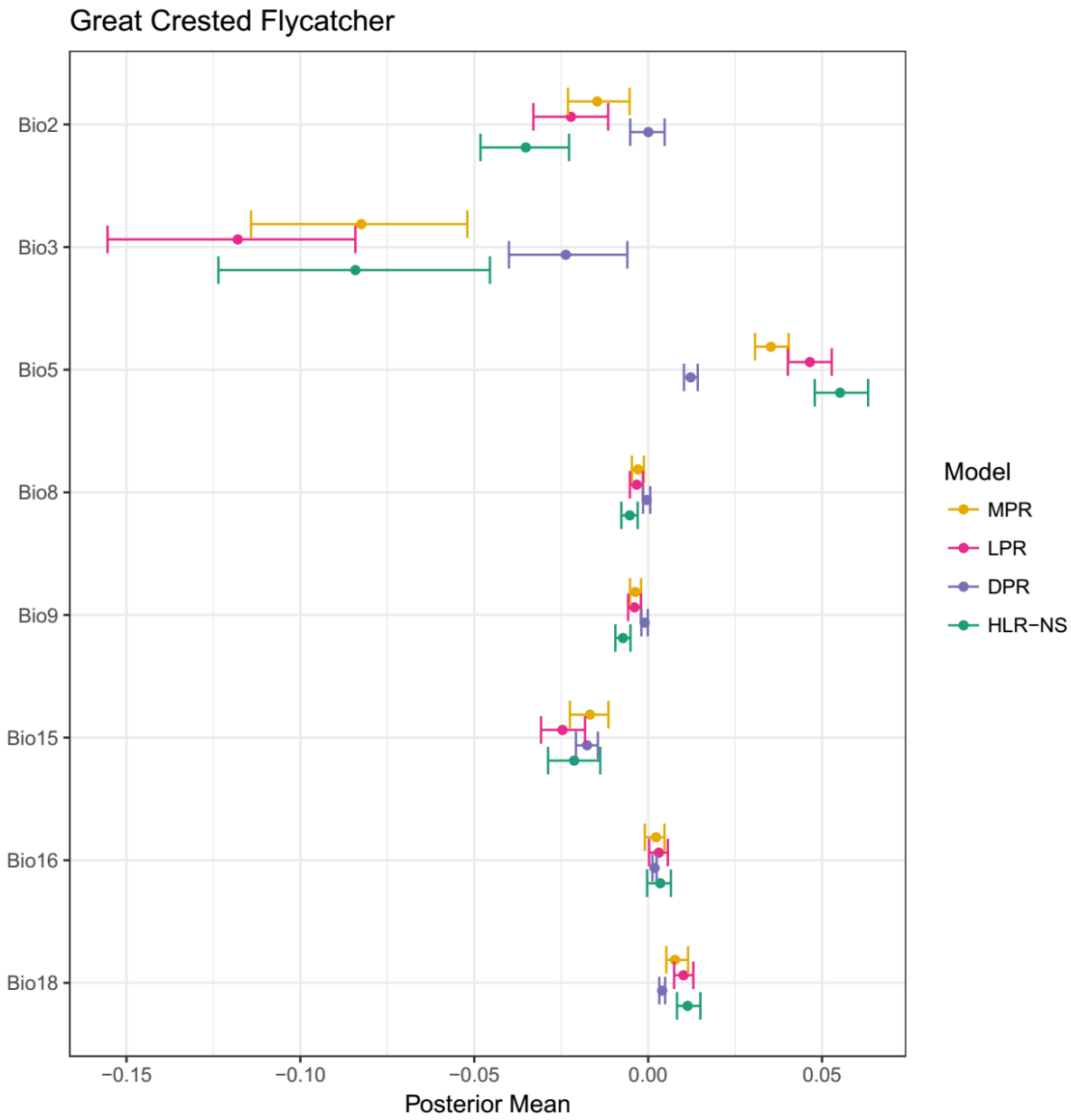


Figure 167: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the great crested flycatcher, *Myiarchus crinitus*.

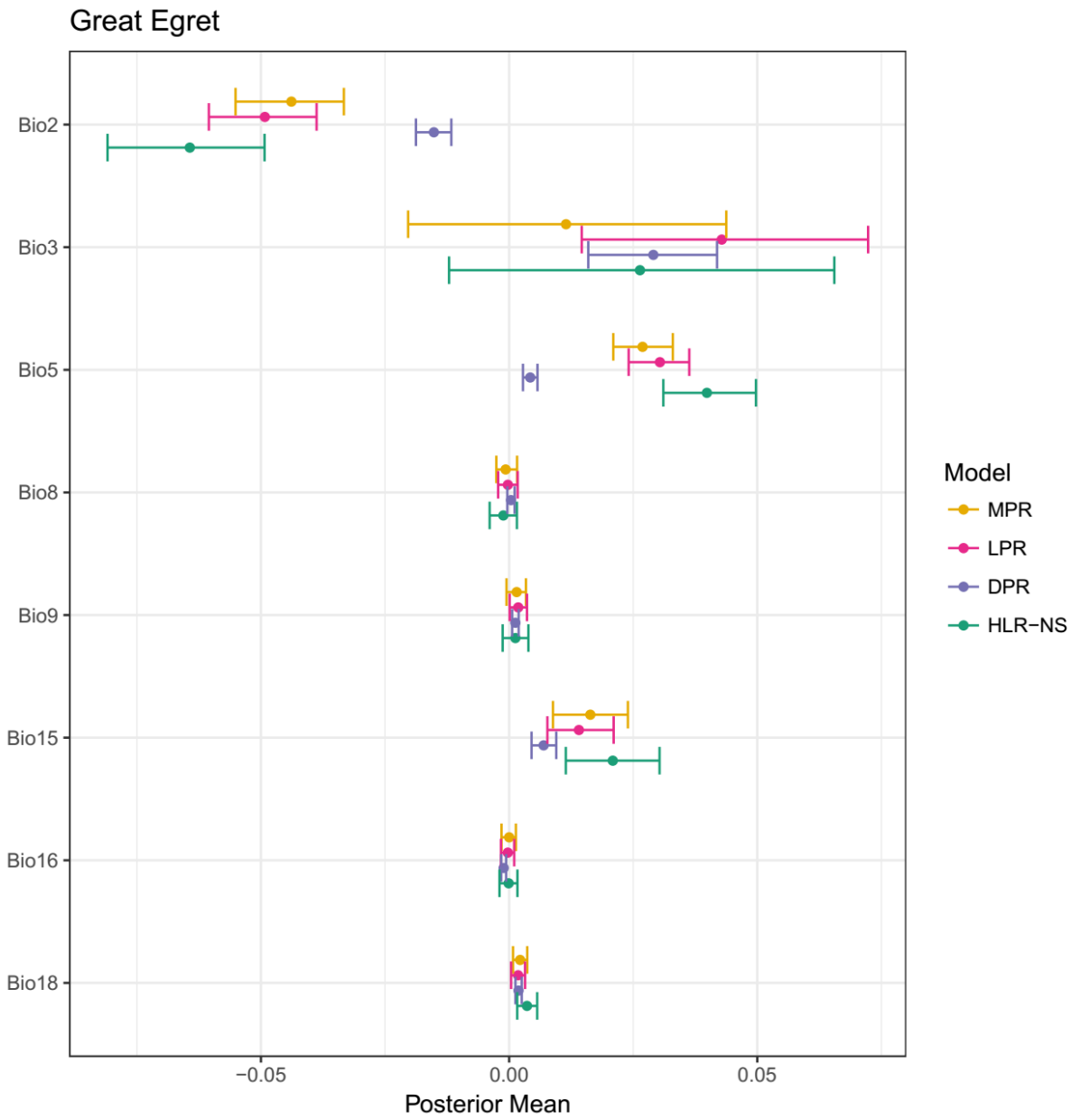


Figure 168: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the great egret, *Ardea alba*.

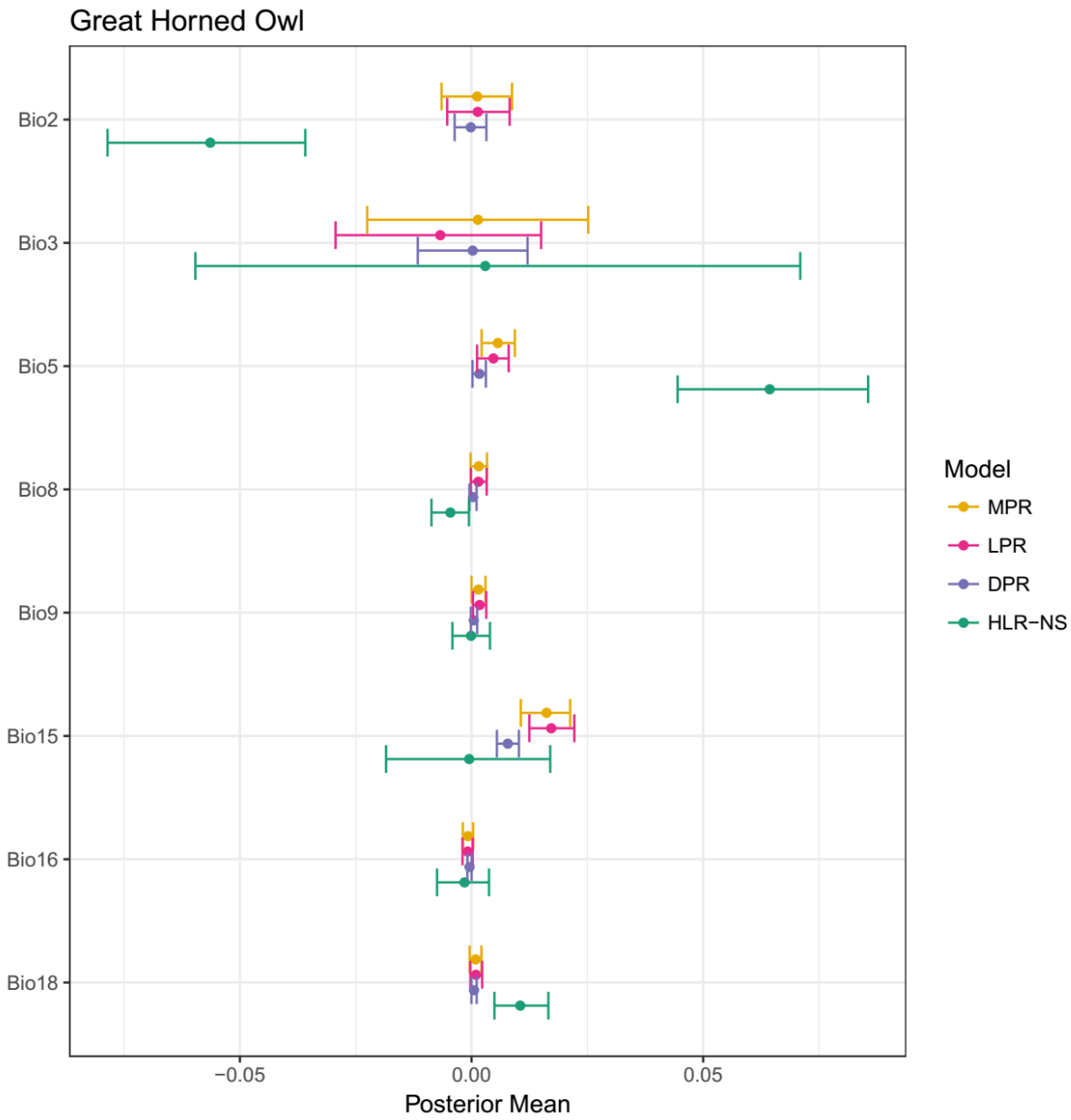


Figure 169: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the great horned owl, *Bubo virginianus*.

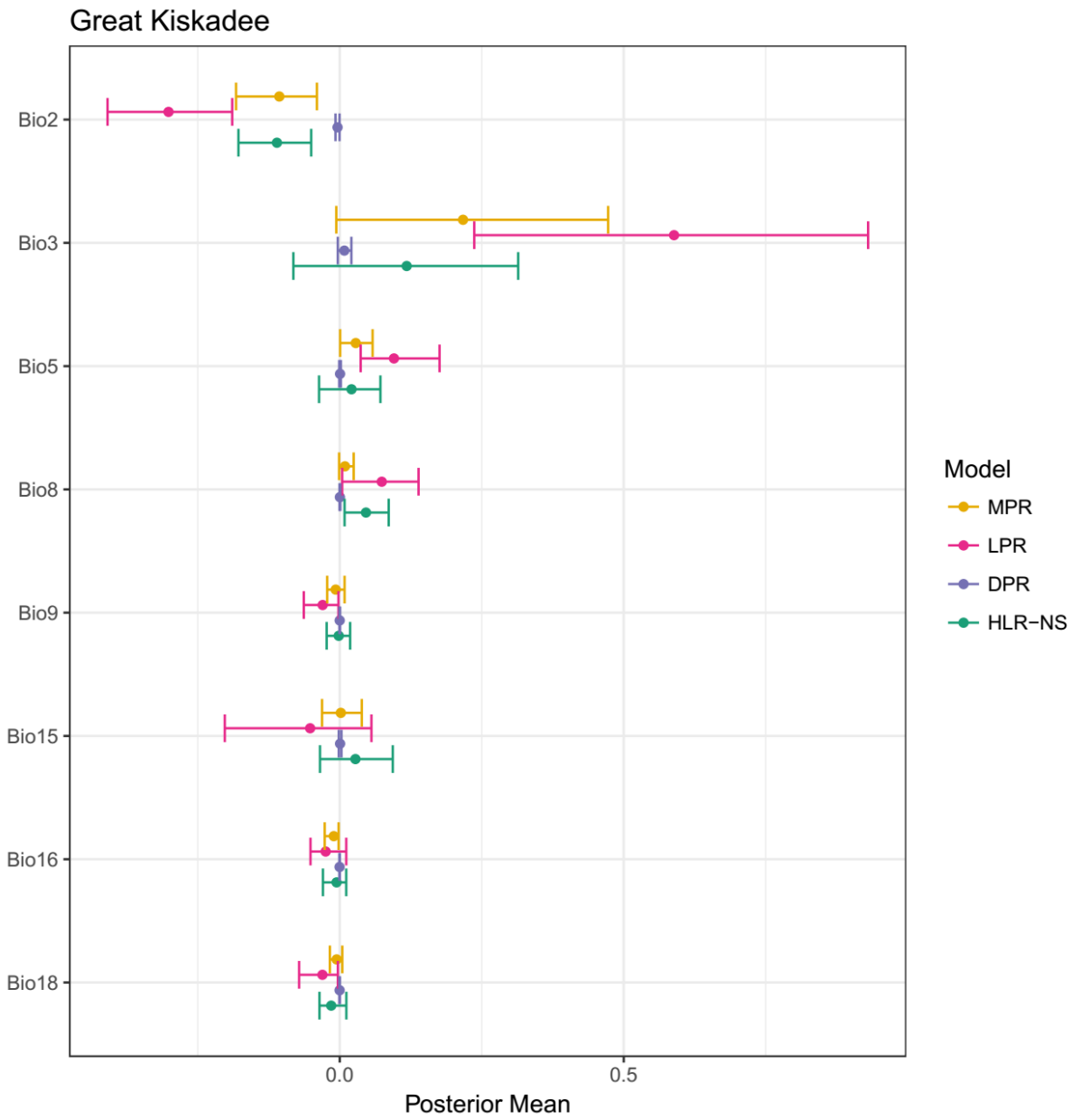


Figure 170: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the great kiskadee, *Pitangus sulphuratus*.

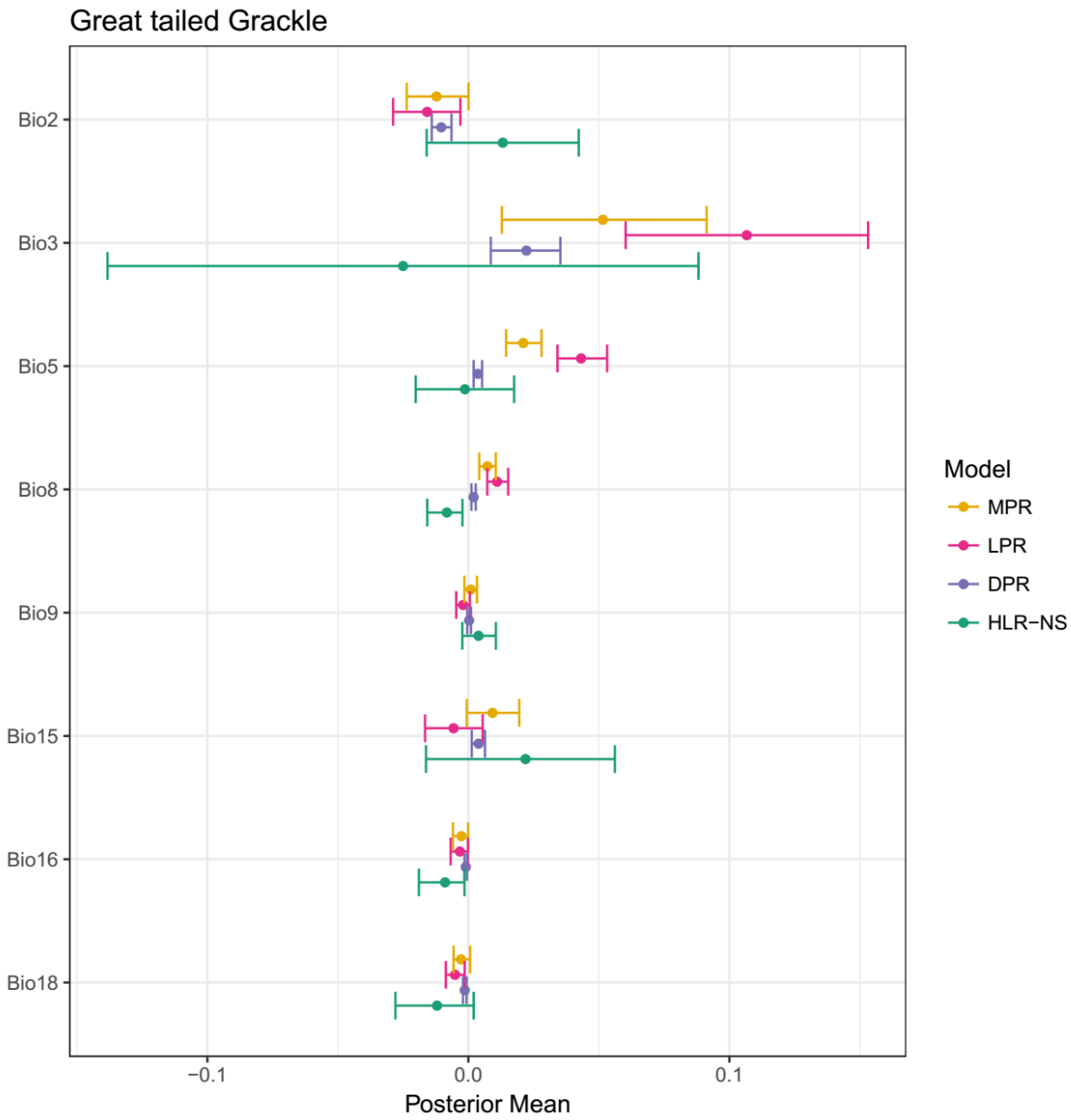


Figure 171: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the great-tailed grackle, *Quiscalus mexicanus*.

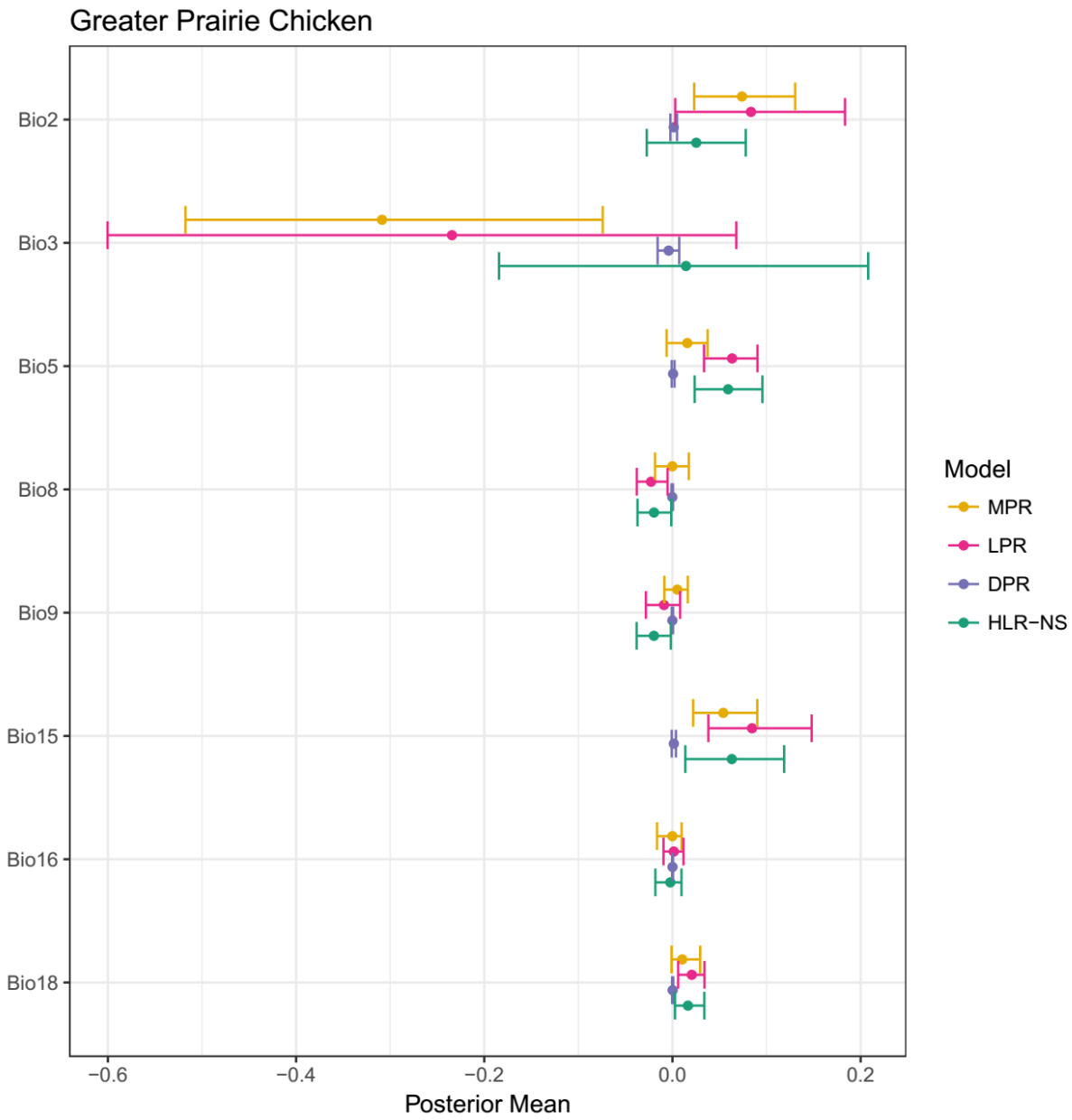


Figure 172: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the greater prairie chicken, *Tympanuchus cupido*.

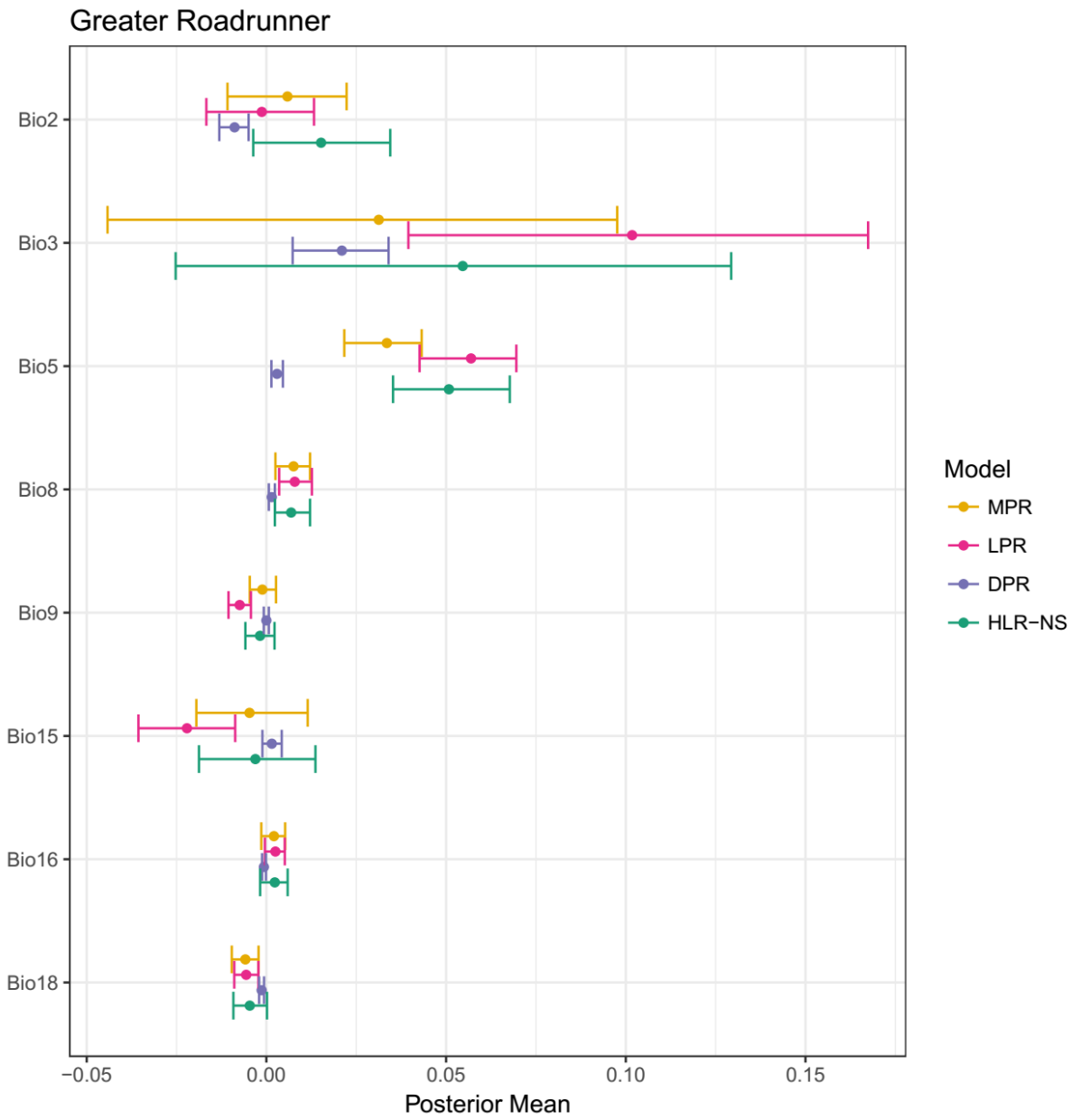


Figure 173: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the greater roadrunner, *Geococcyx californianus*.

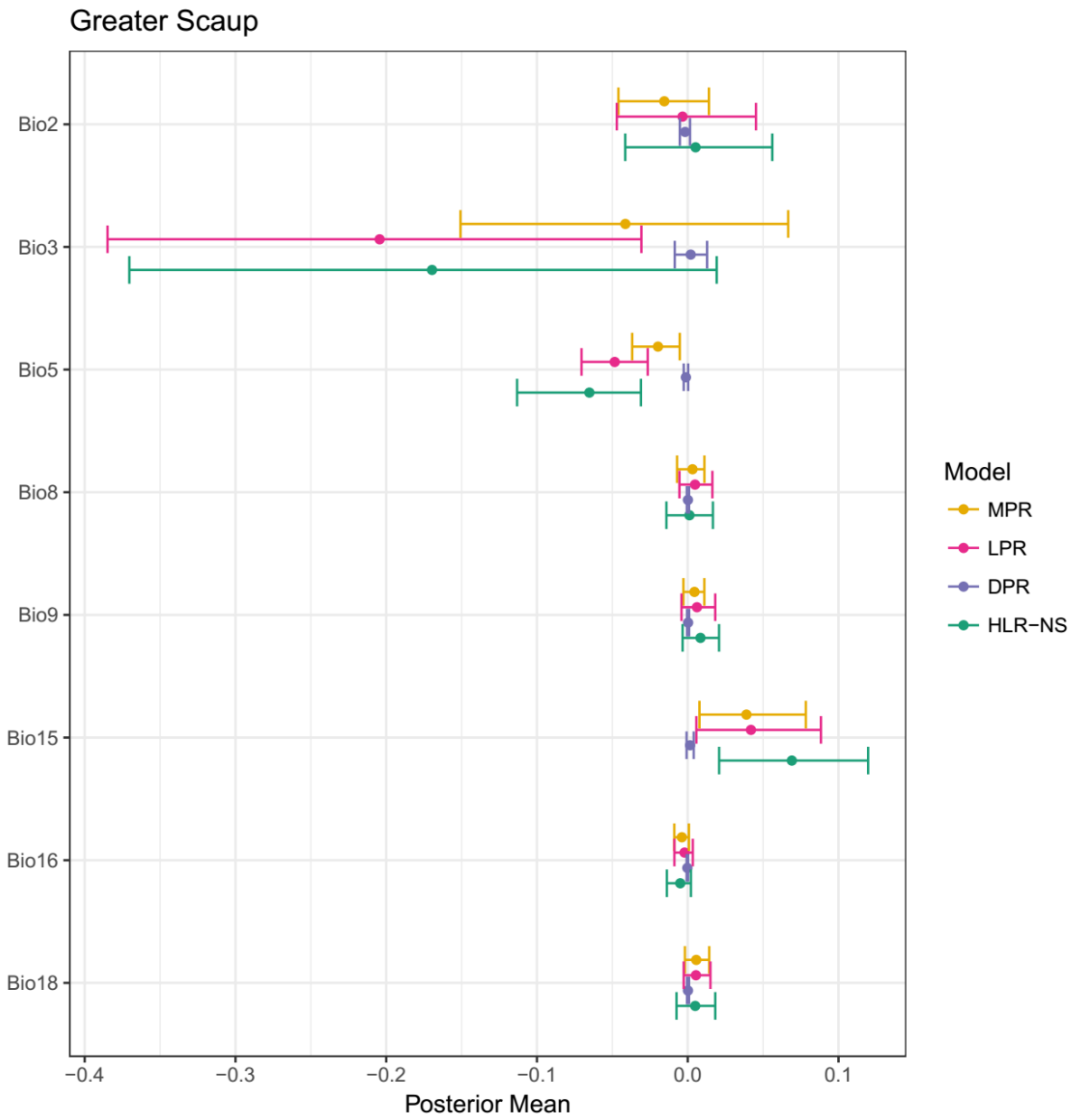


Figure 174: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the greater scaup, *Aythya marila*.

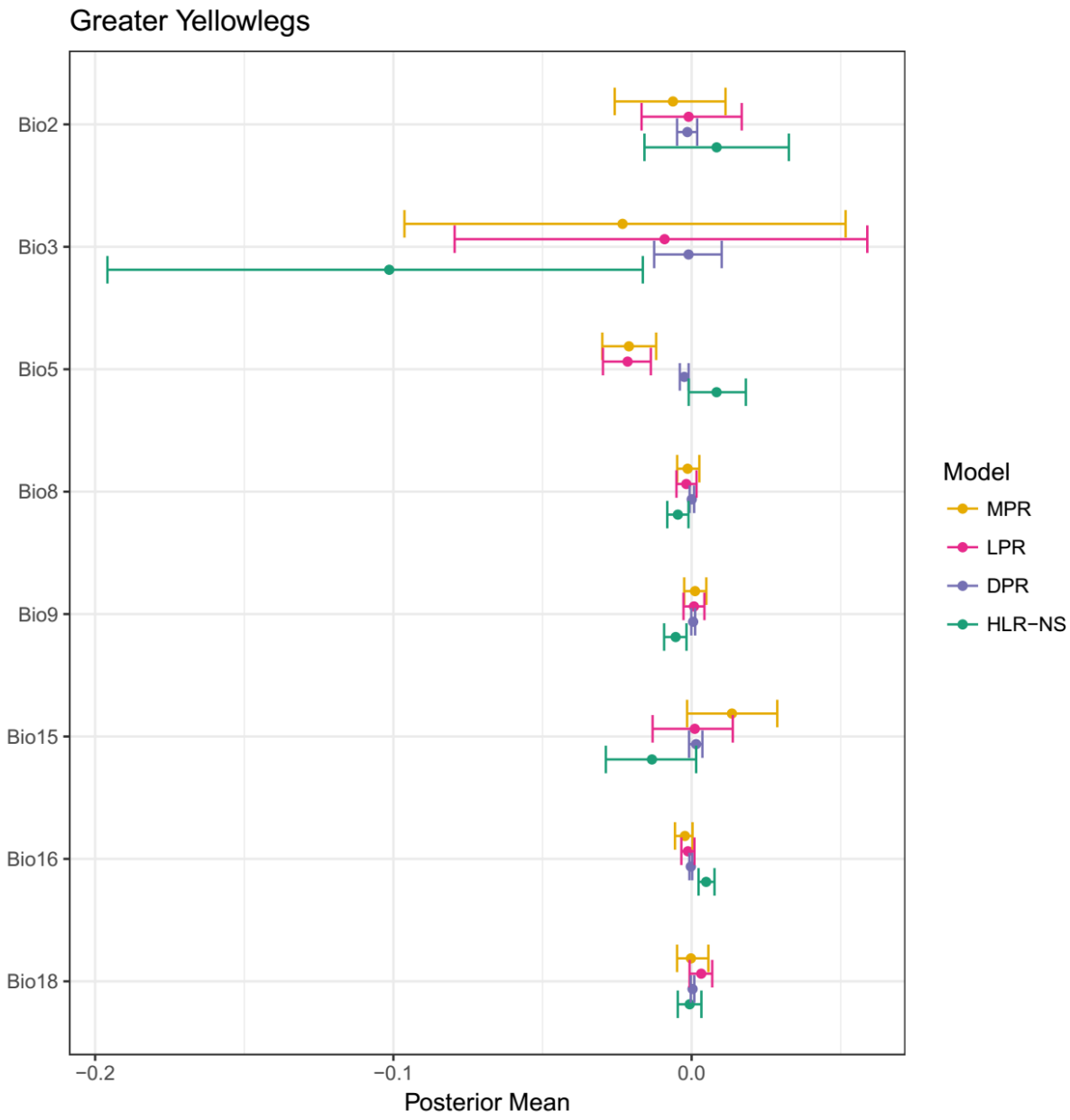


Figure 175: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the greater yellowlegs, *Tringa melanoleuca*.

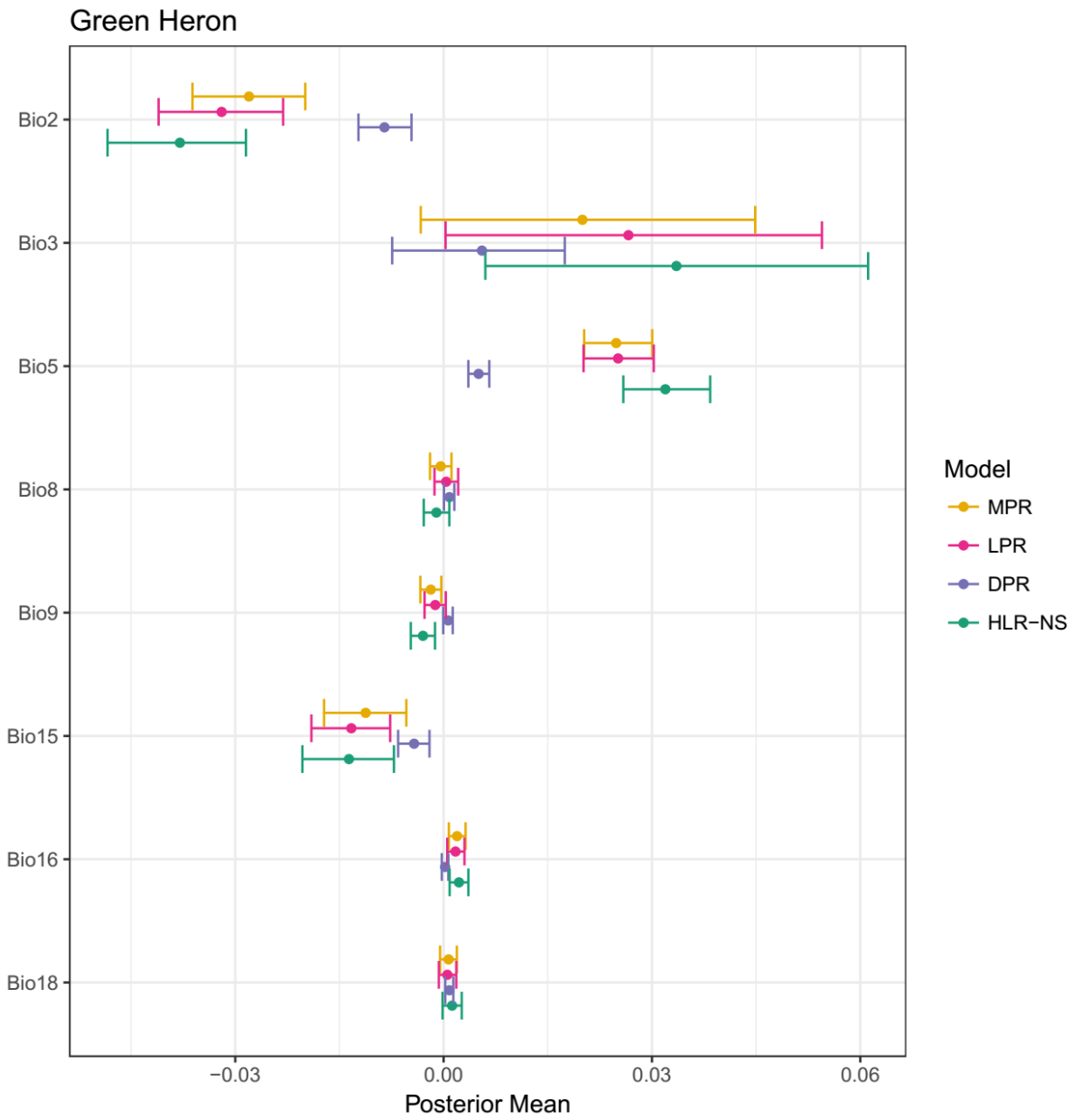


Figure 176: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the green heron, *Butorides virescens*.

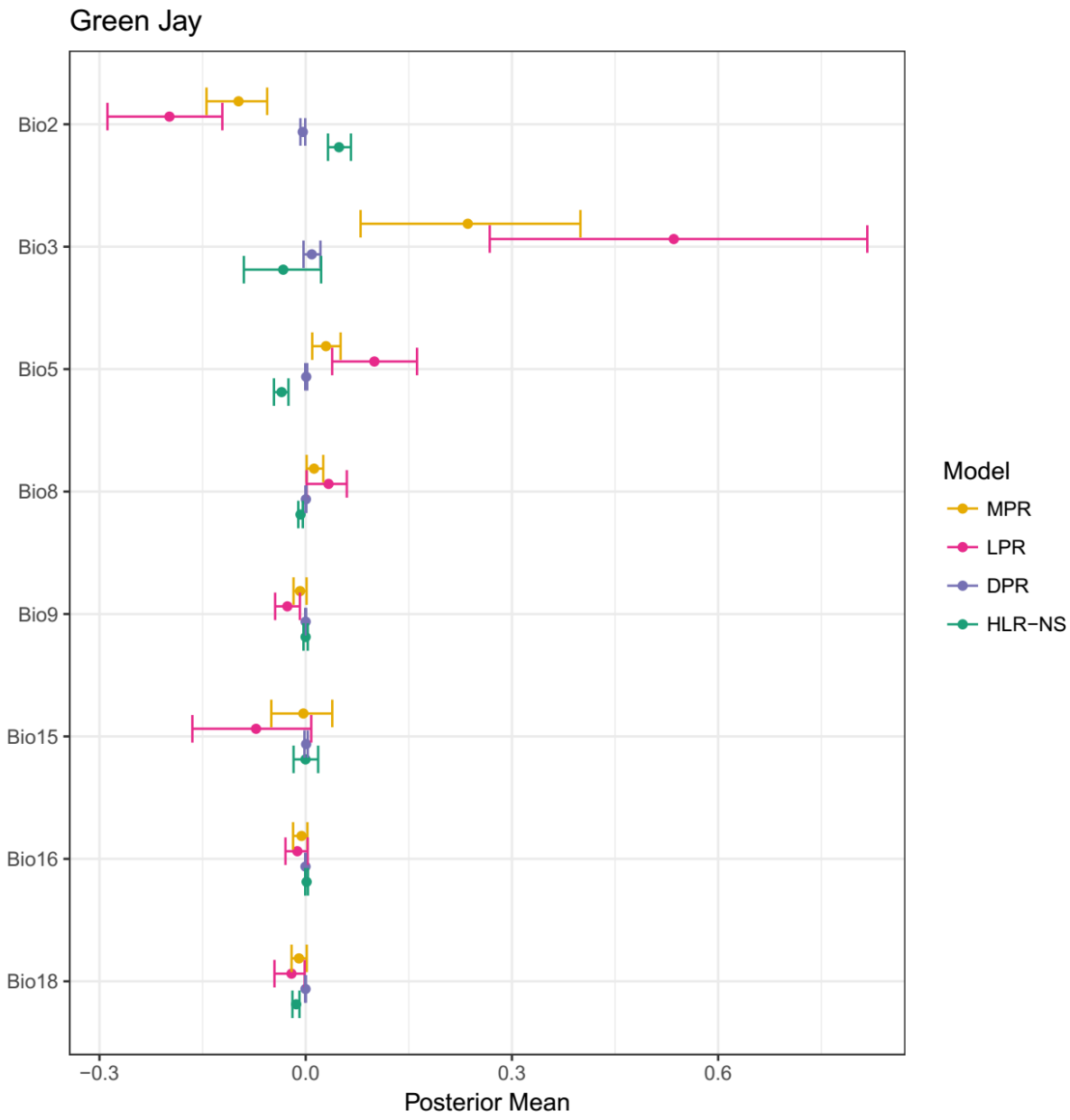


Figure 177: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the green jay, *Cyanocorax yncas*.

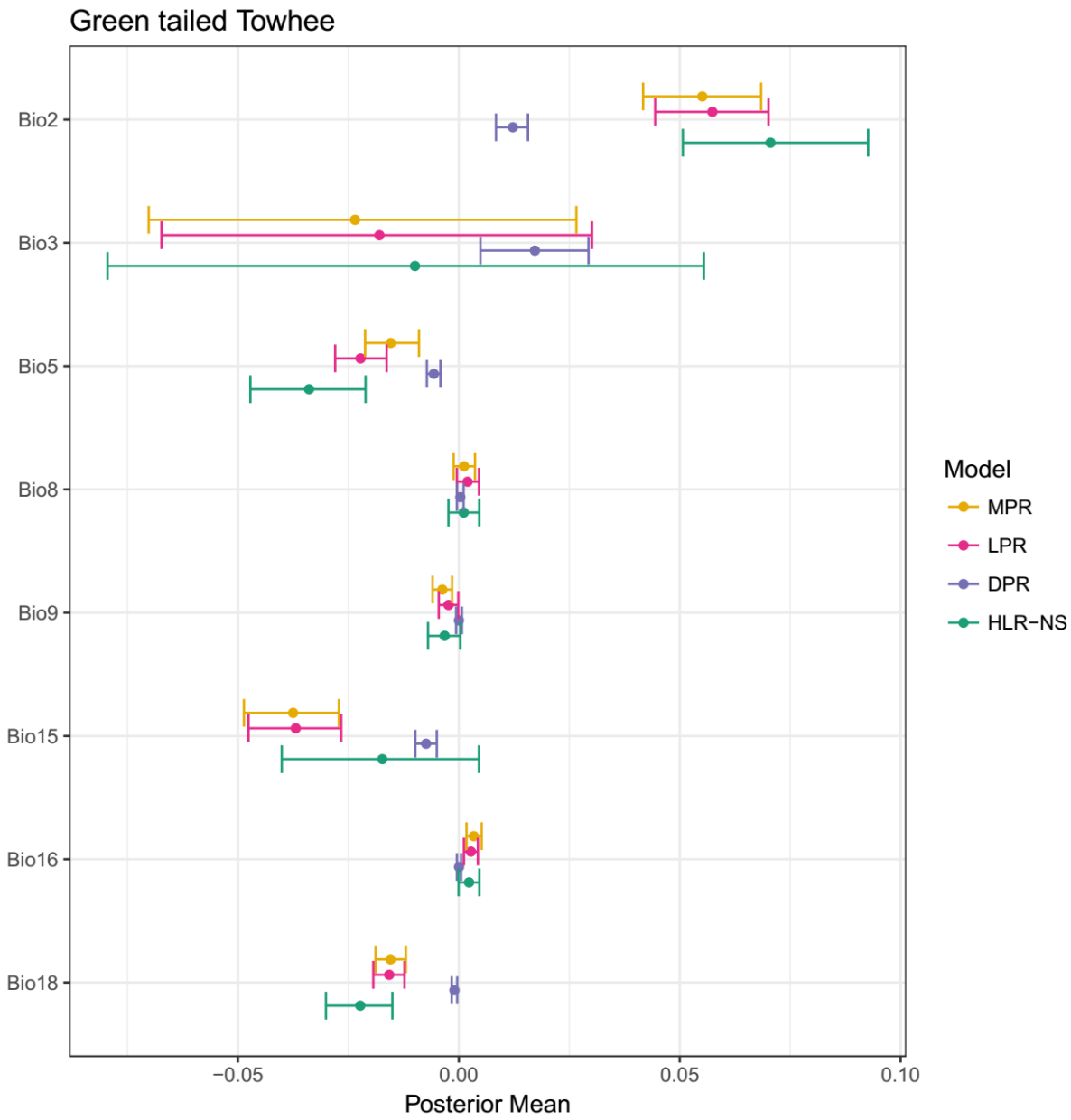


Figure 178: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the green-tailed towhee, *Pipilo chlorurus*.

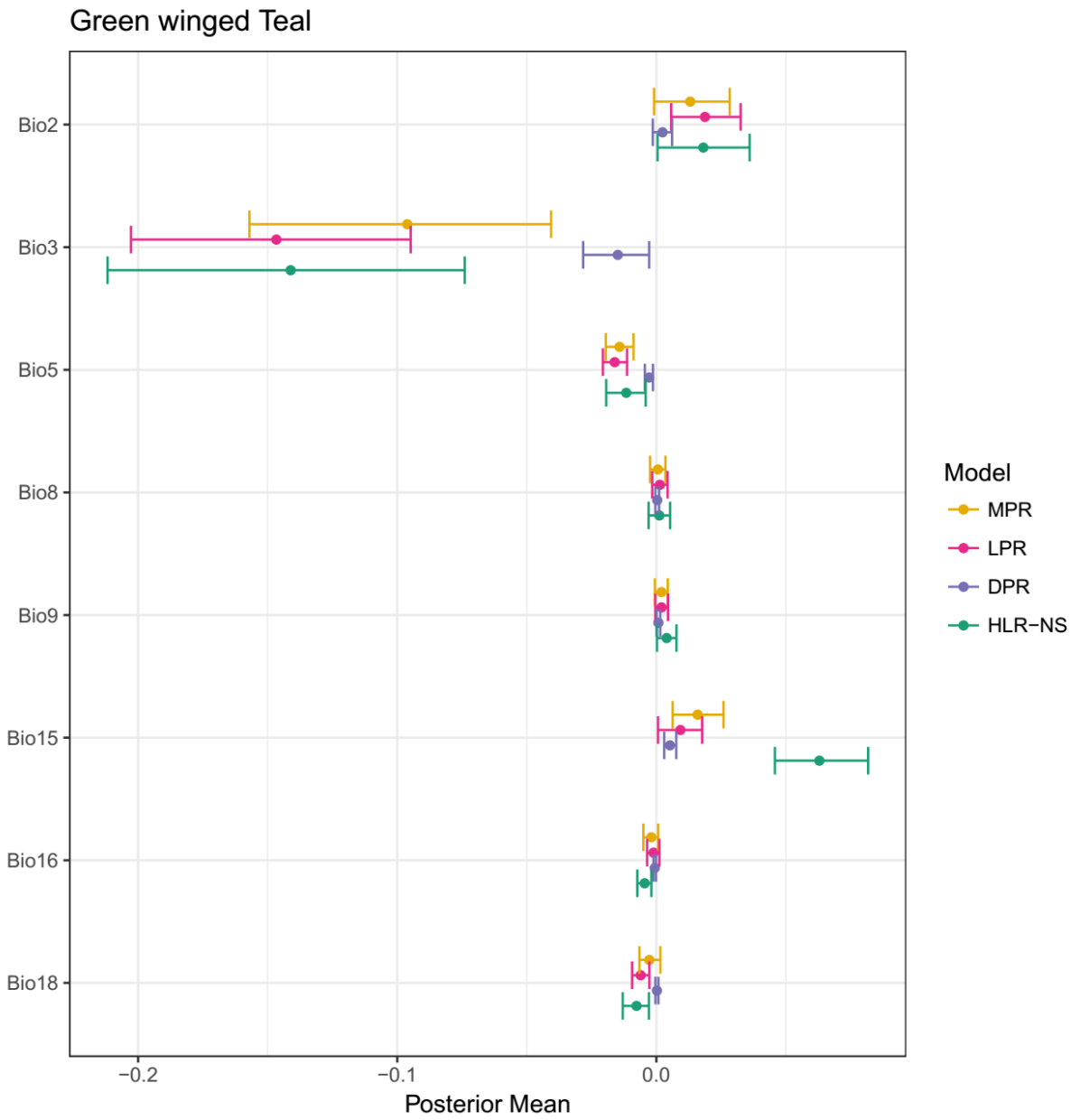


Figure 179: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the green-winged teal, *Anas carolinensis*.

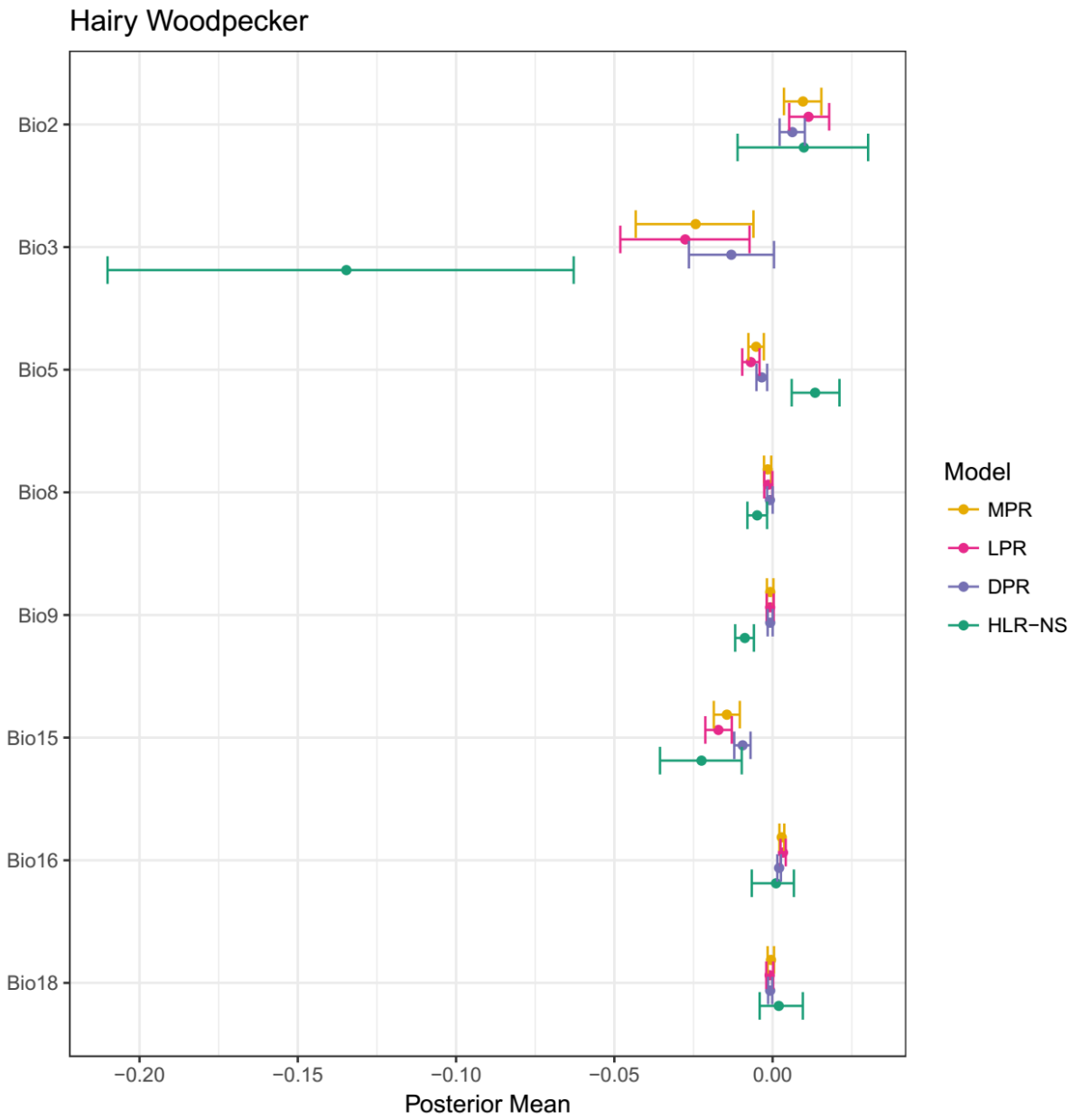


Figure 180: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hairy woodpecker, *Leuconotopicus villosus*.

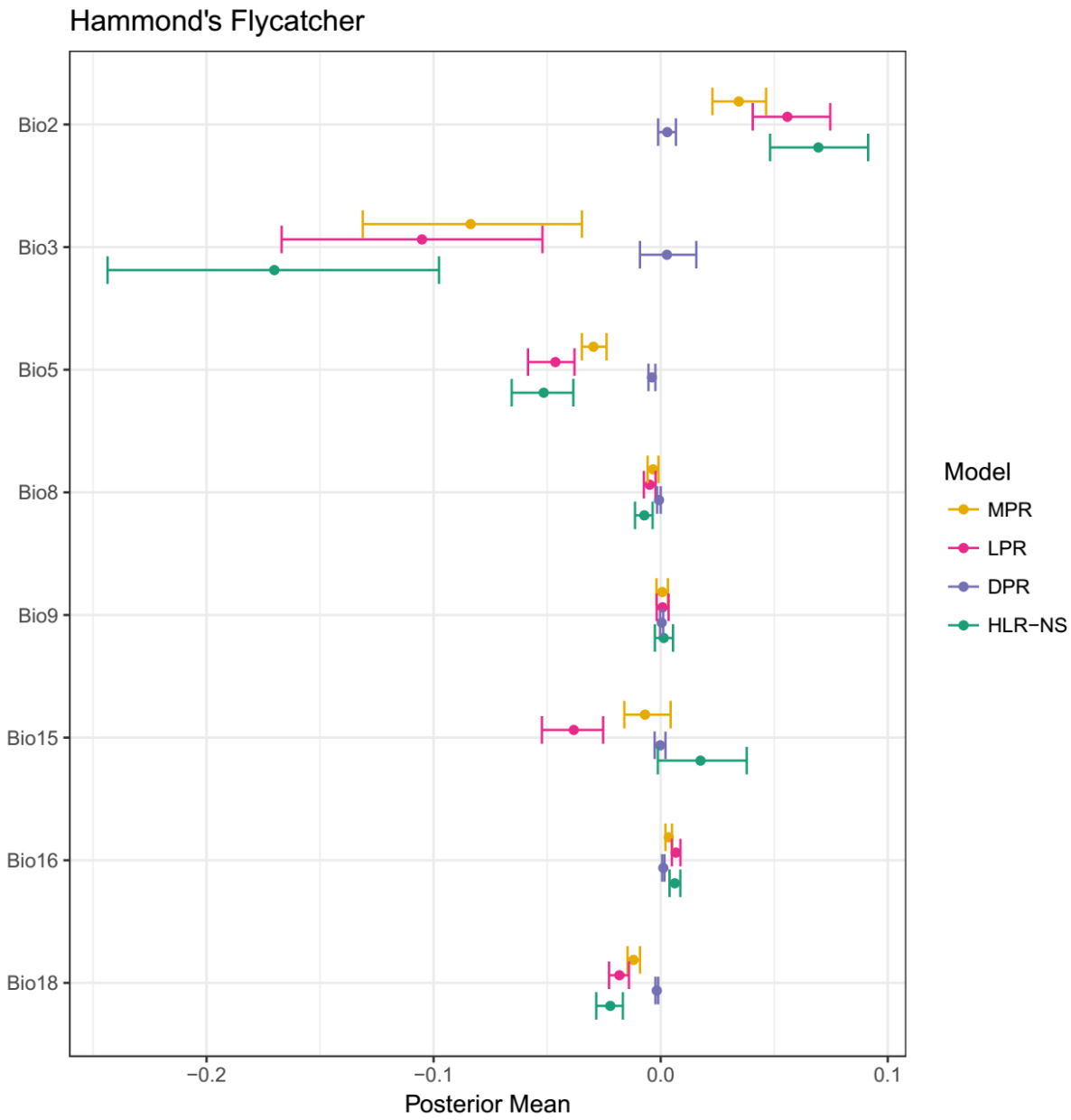


Figure 181: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Hammond's flycatcher, *Empidonax hammondii*.

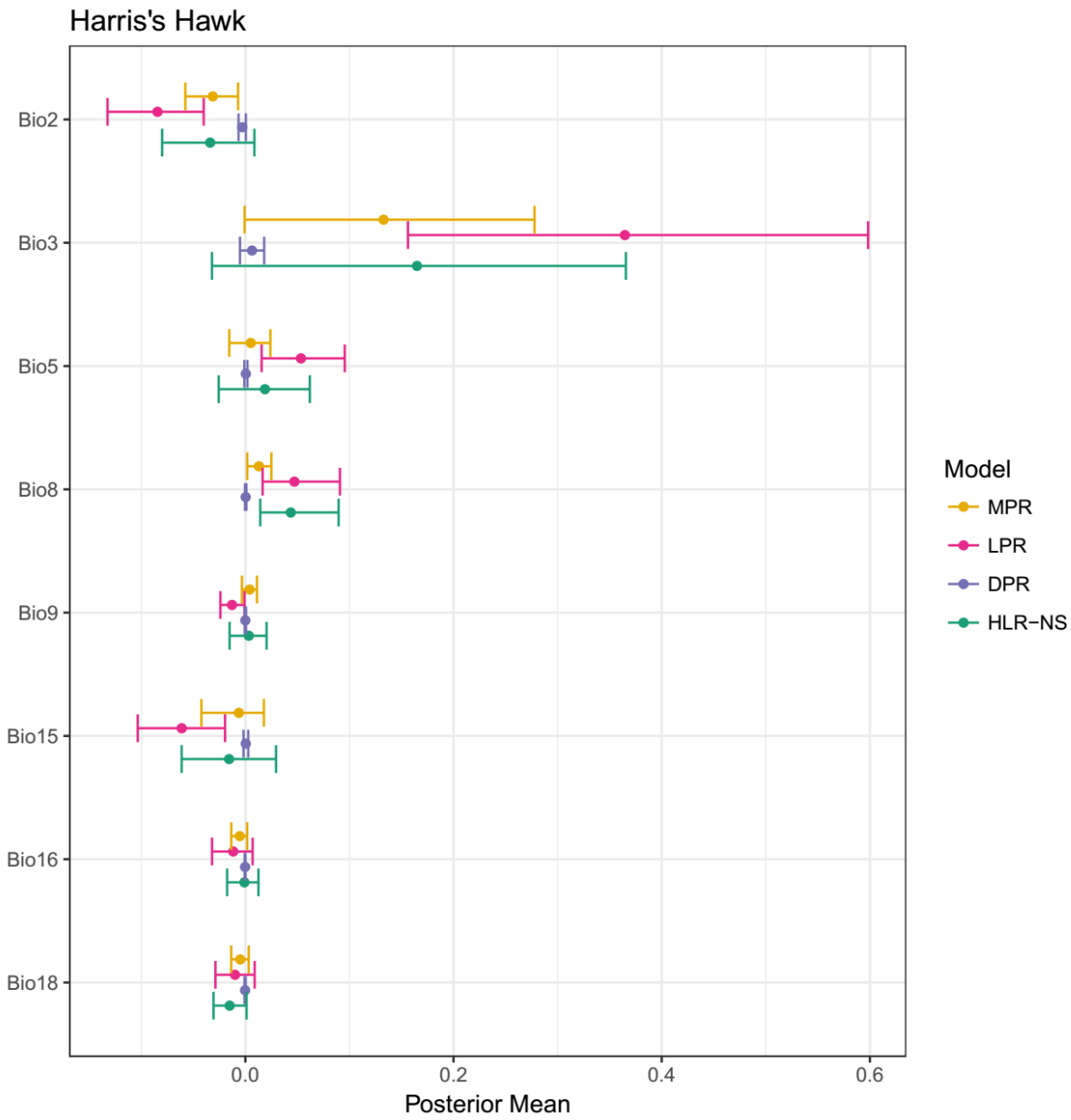


Figure 182: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Harris' hawk, *Parabuteo unicinctus*.

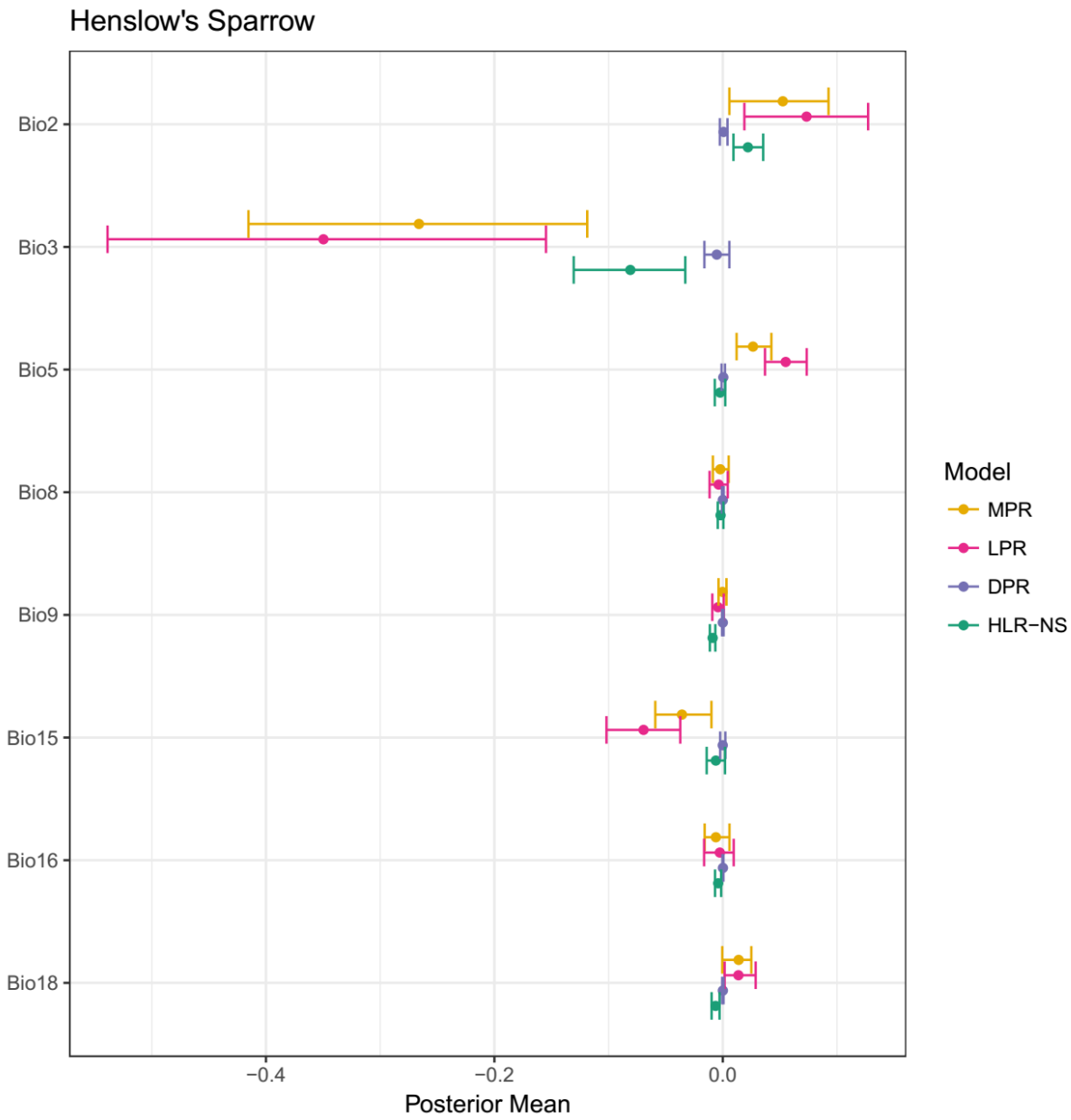


Figure 183: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Henslow's sparrow, *Ammodramus henslowii*.

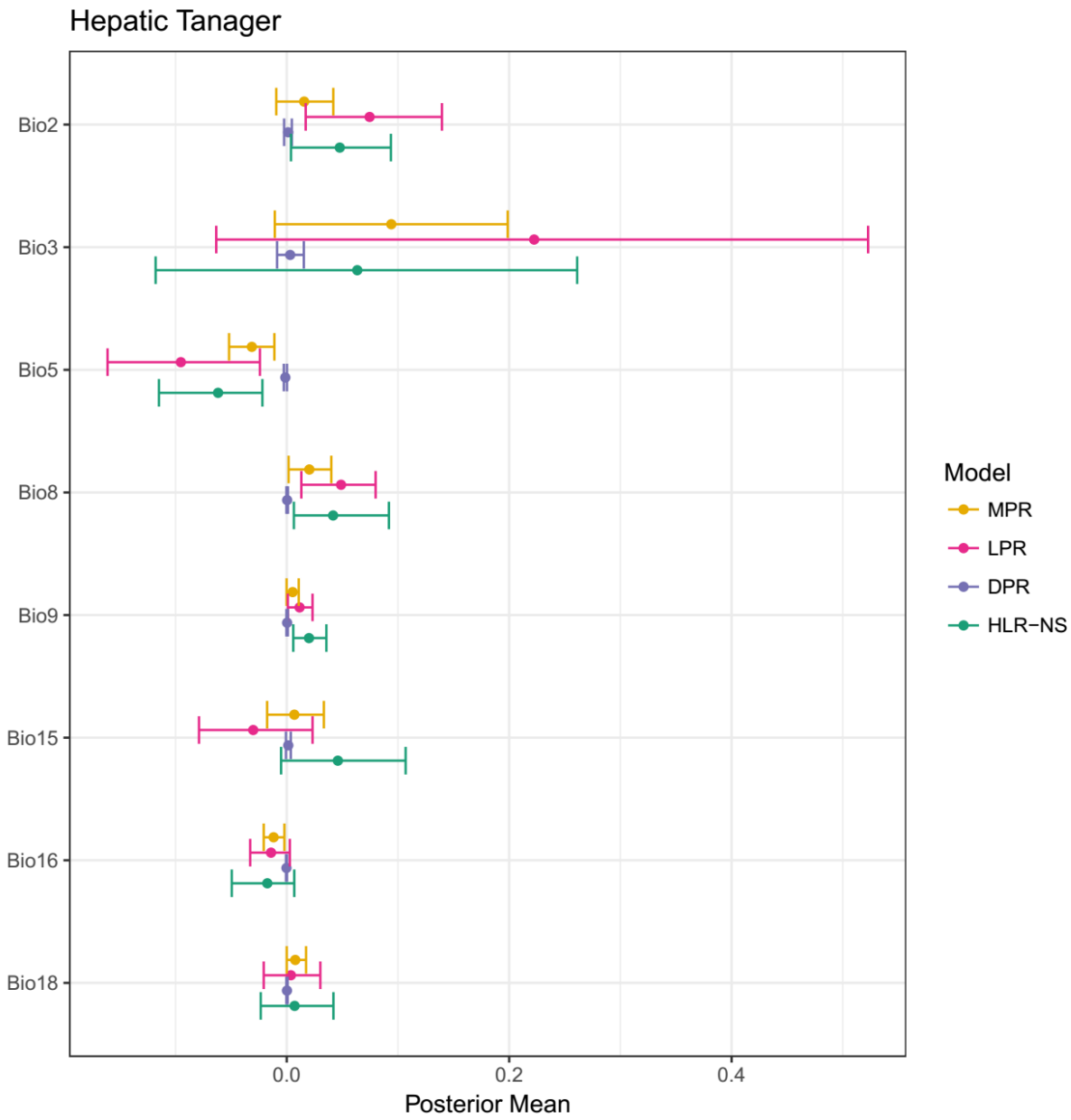


Figure 184: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hepatic tanager, *Piranga flava*.

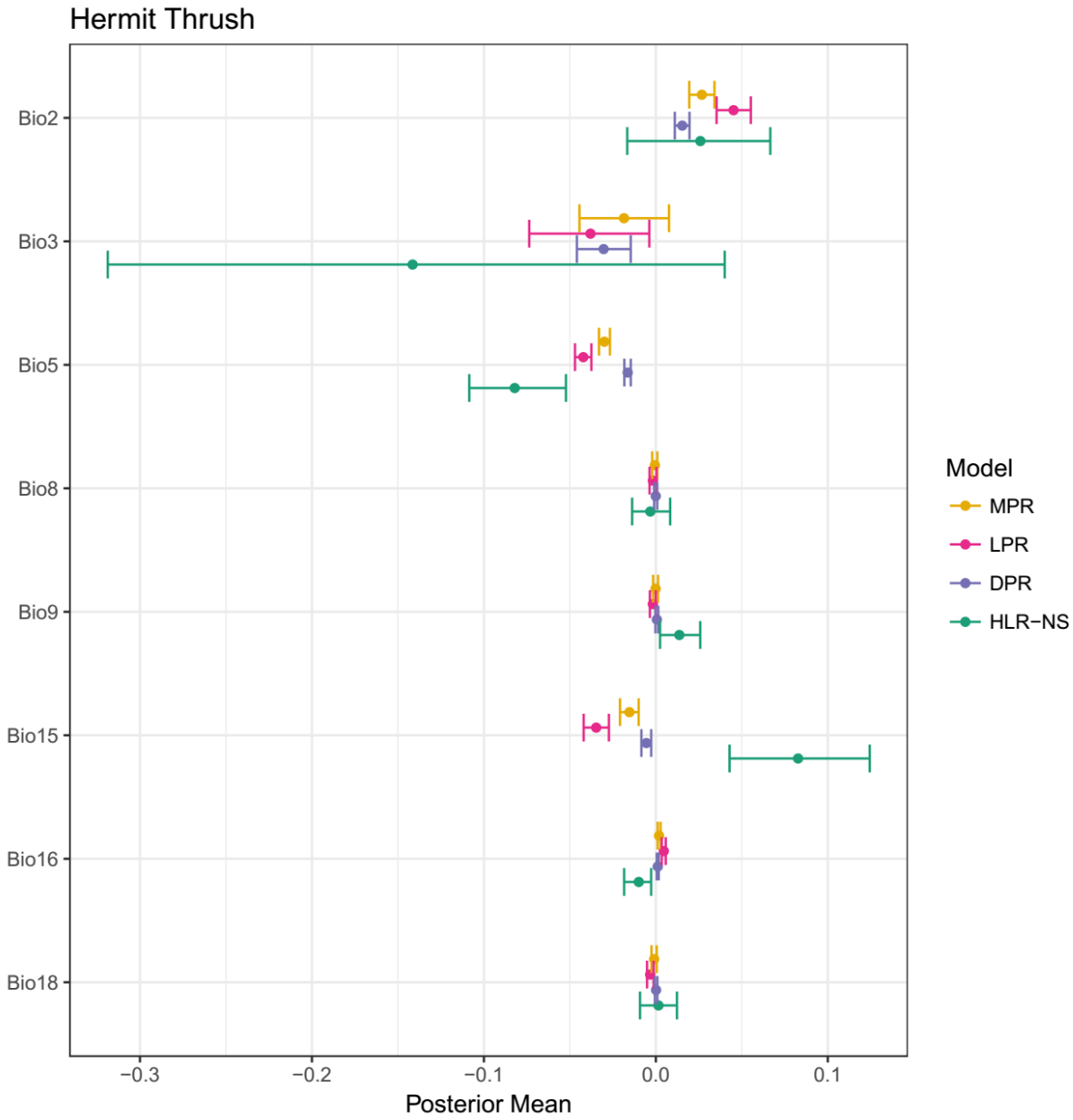


Figure 185: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hermit thrush, *Catharus guttatus*.

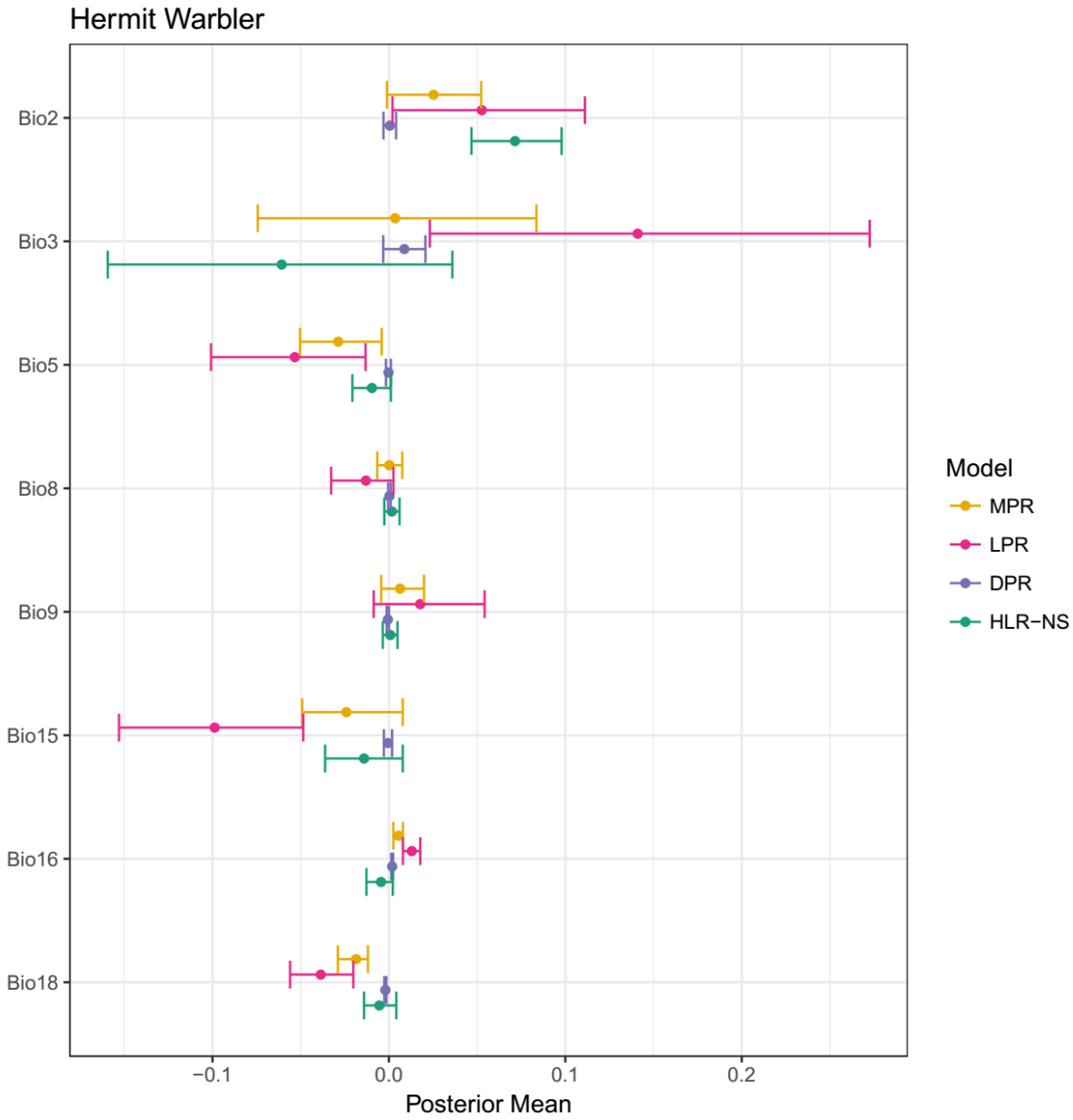


Figure 186: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hermit warbler, *Setophaga occidentalis*.

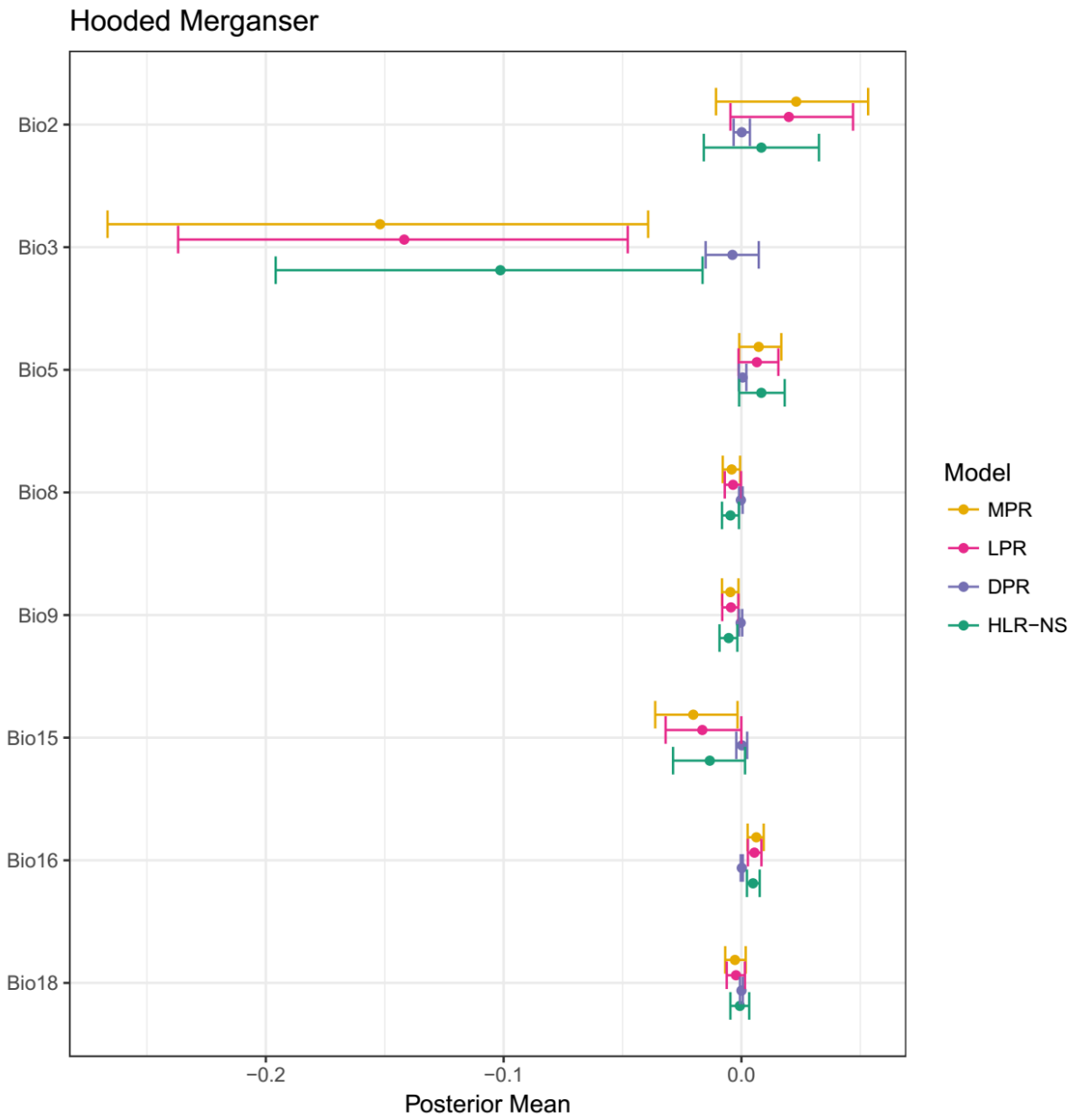


Figure 187: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hooded merganser, *Lophodytes cucullatus*.

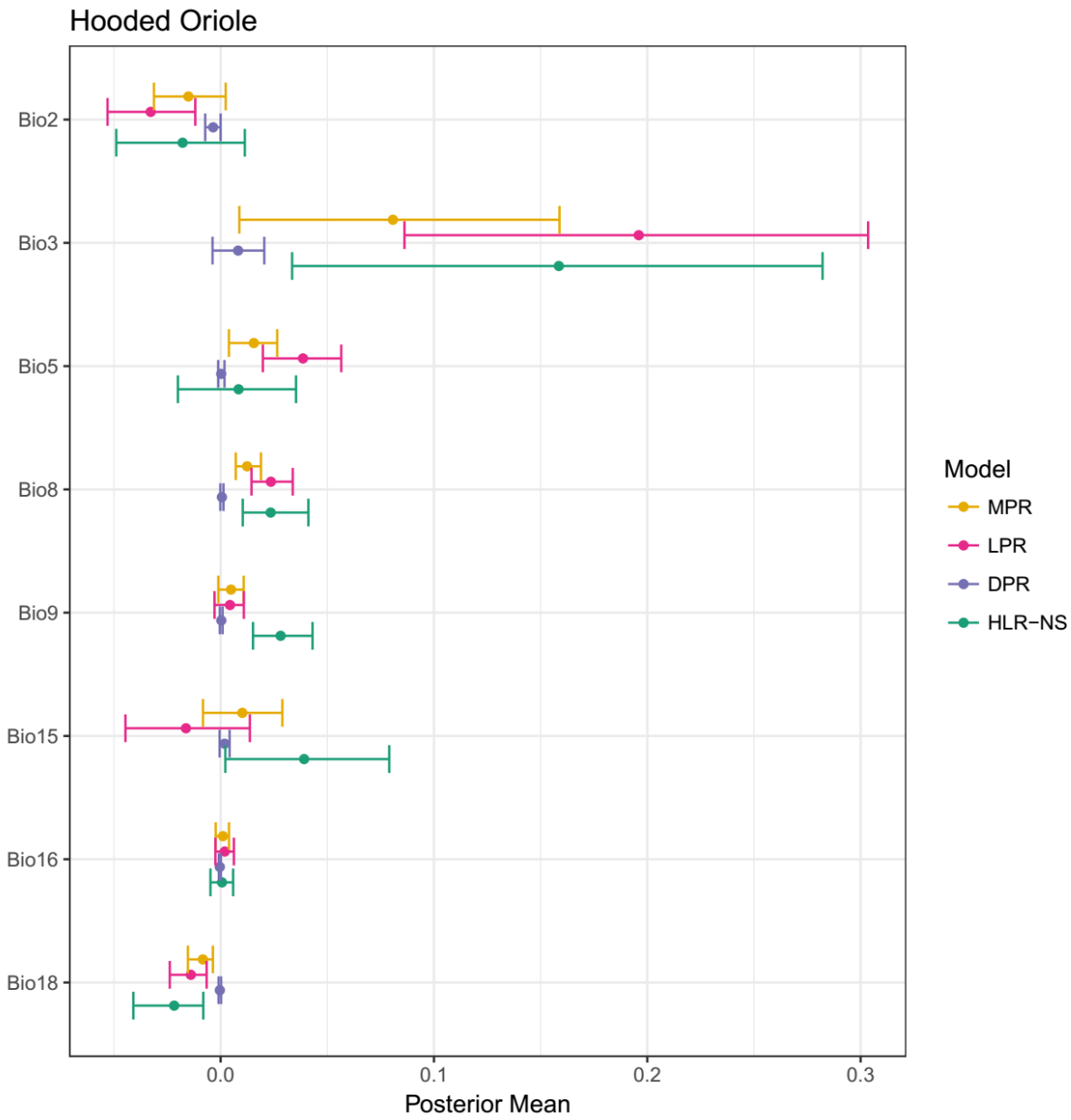


Figure 188: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hooded oriole, *Icterus cucullatus*.

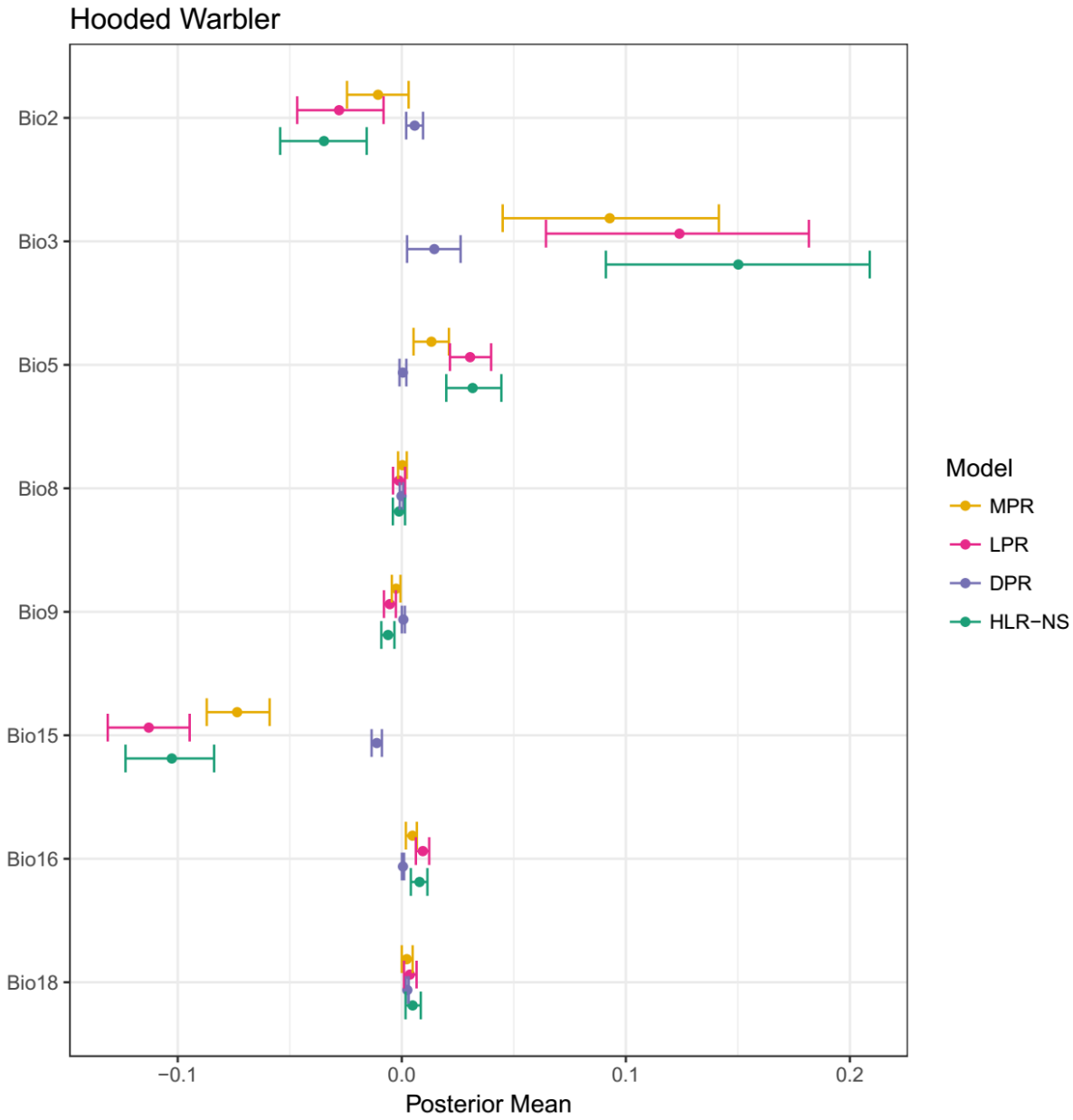


Figure 189: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the hooded warbler, *Setophaga citrina*.

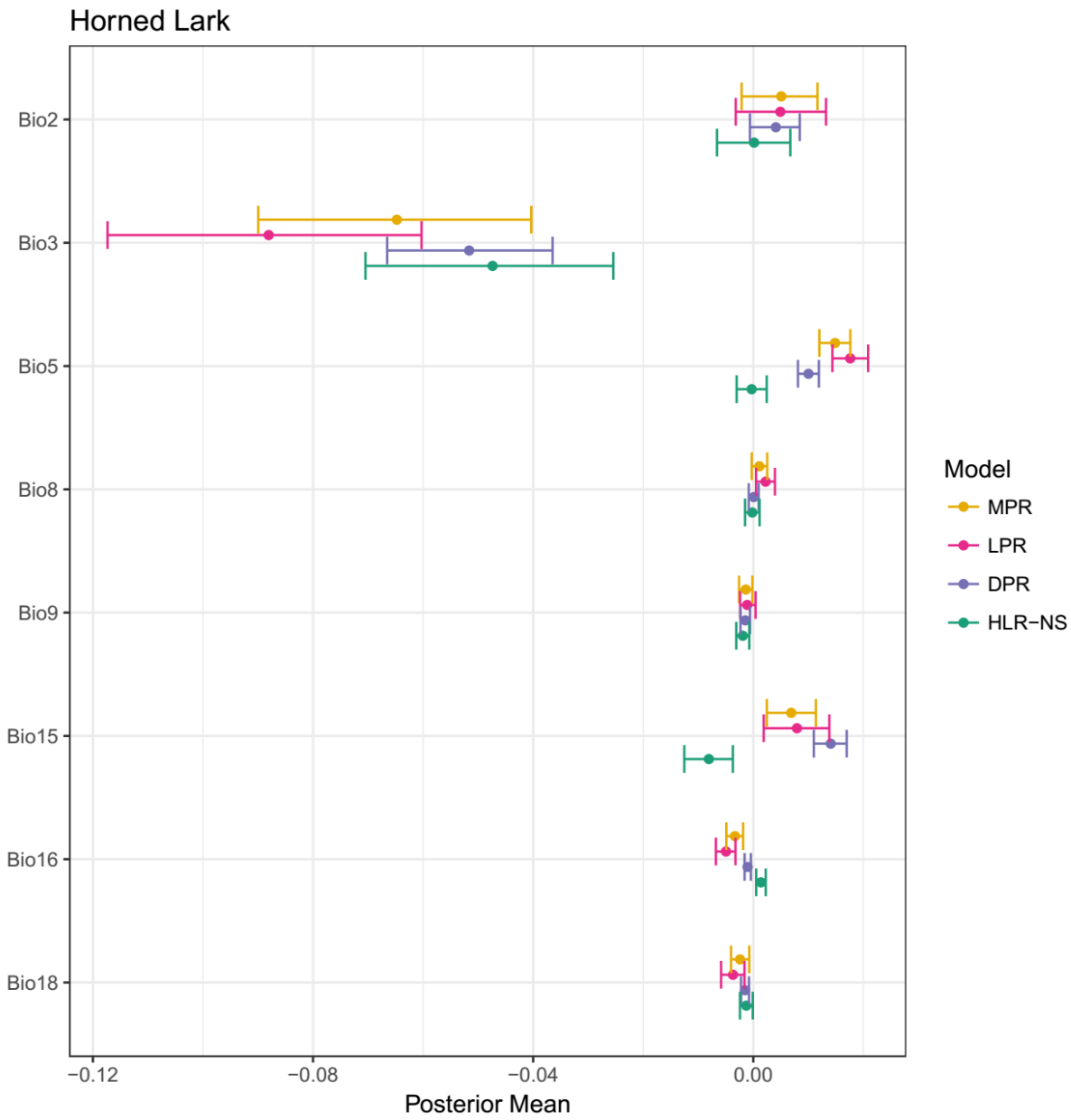


Figure 190: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the horned lark, *Eremophila alpestris*.

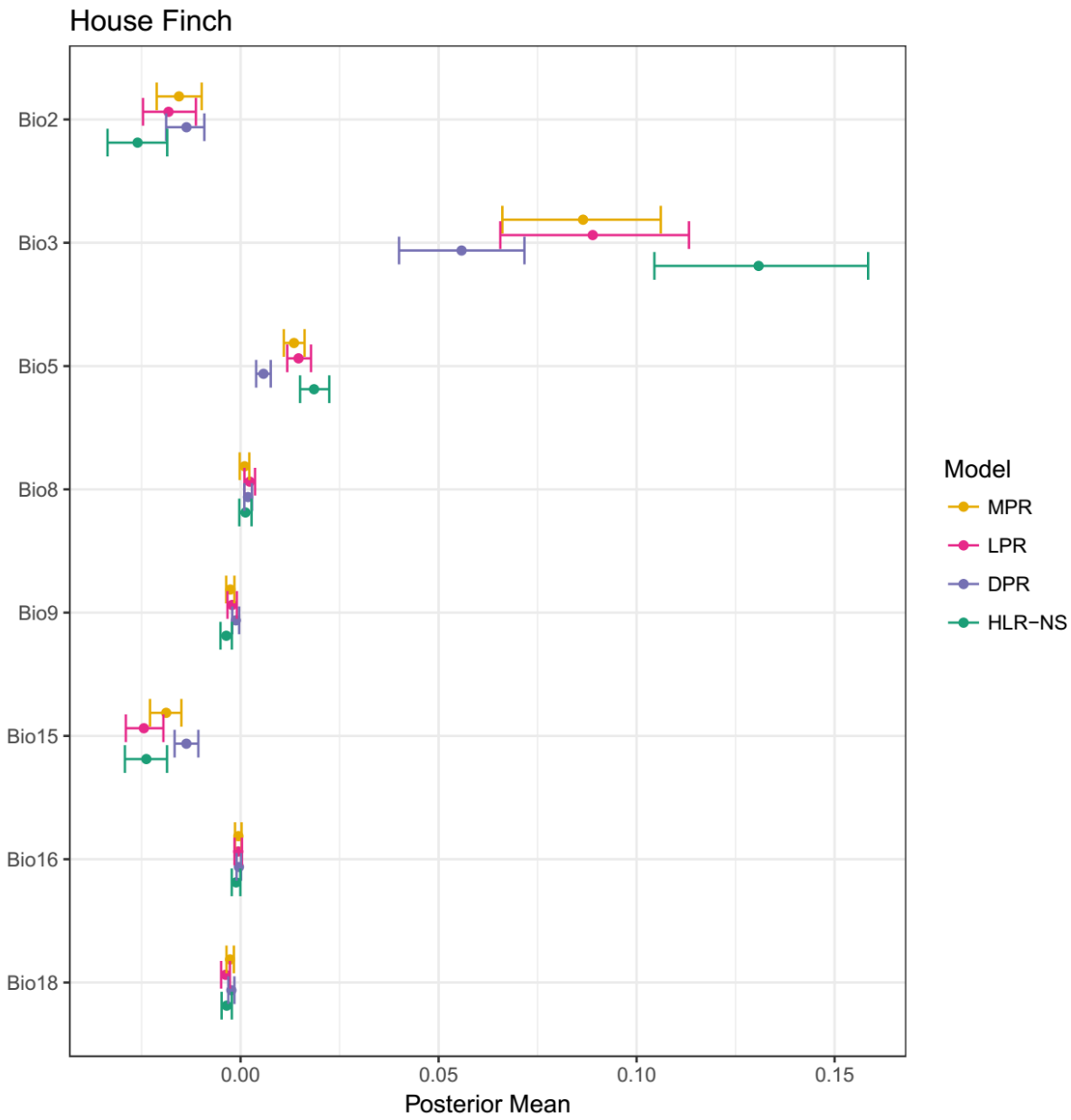


Figure 191: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the house finch, *Haemorhous mexicanus*.

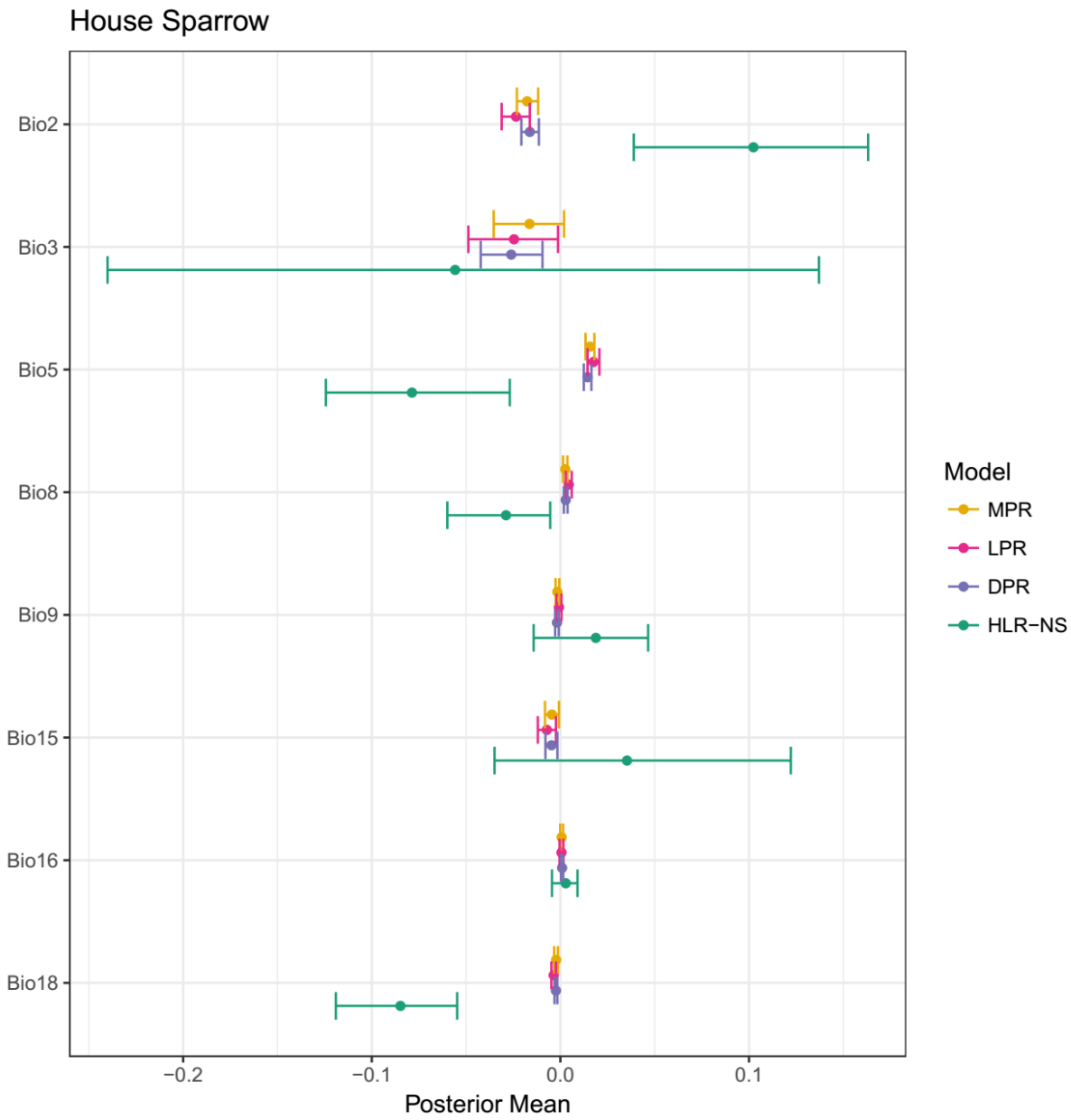


Figure 192: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the house sparrow, *Passer domesticus*.

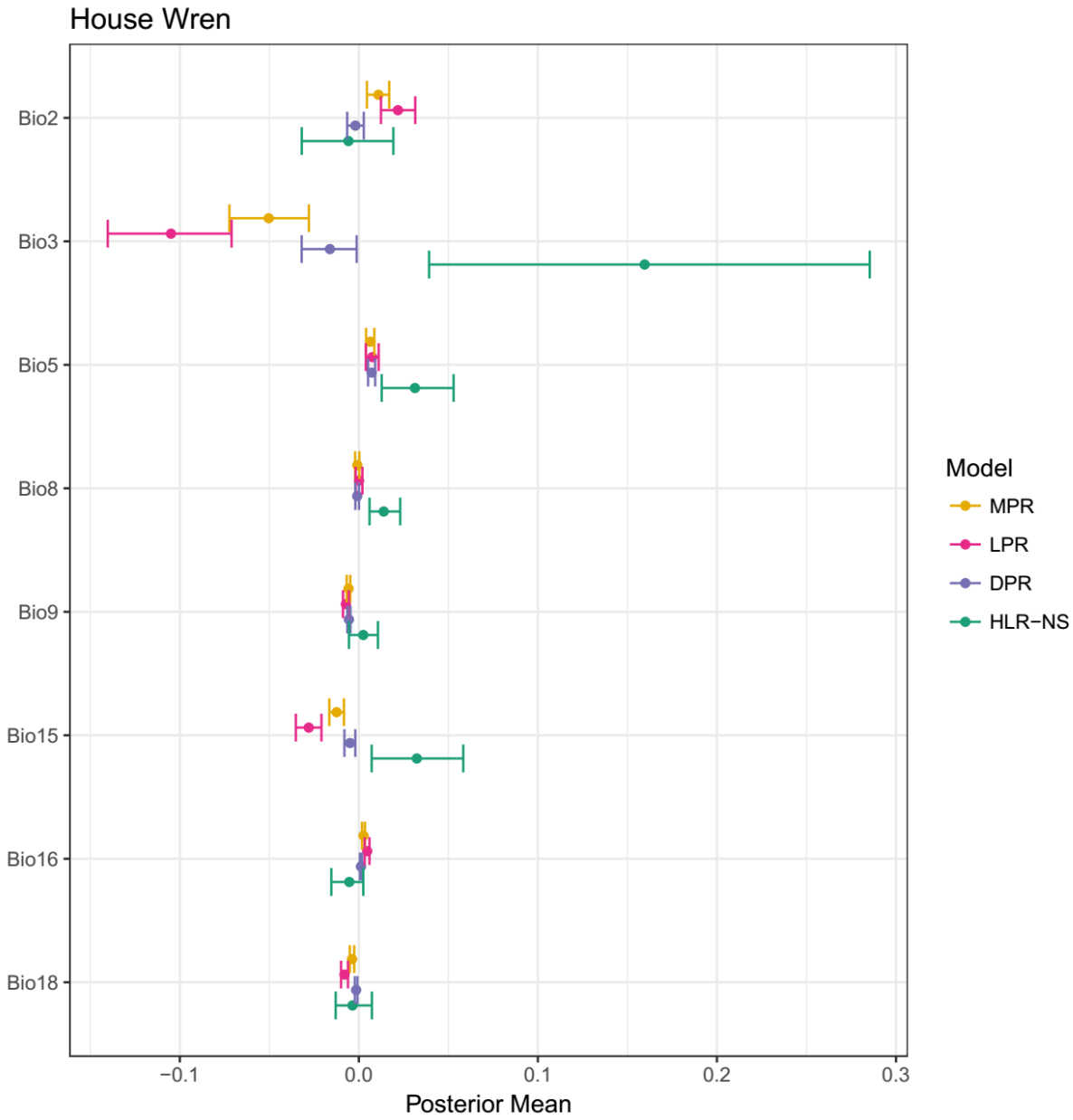


Figure 193: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the house wren, *Troglodytes aedon*.

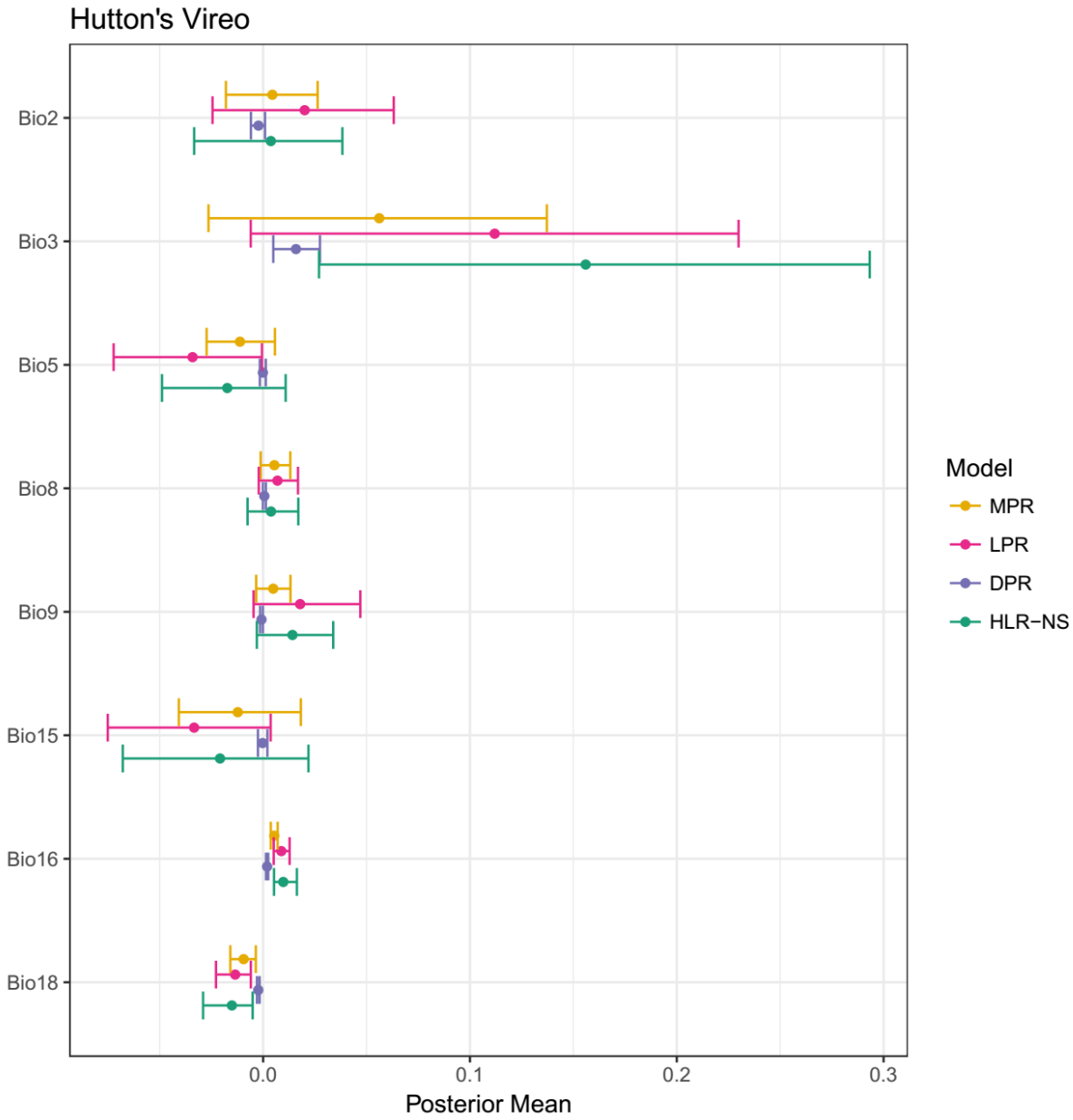


Figure 194: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Hutton's vireo, *Vireo huttoni*.

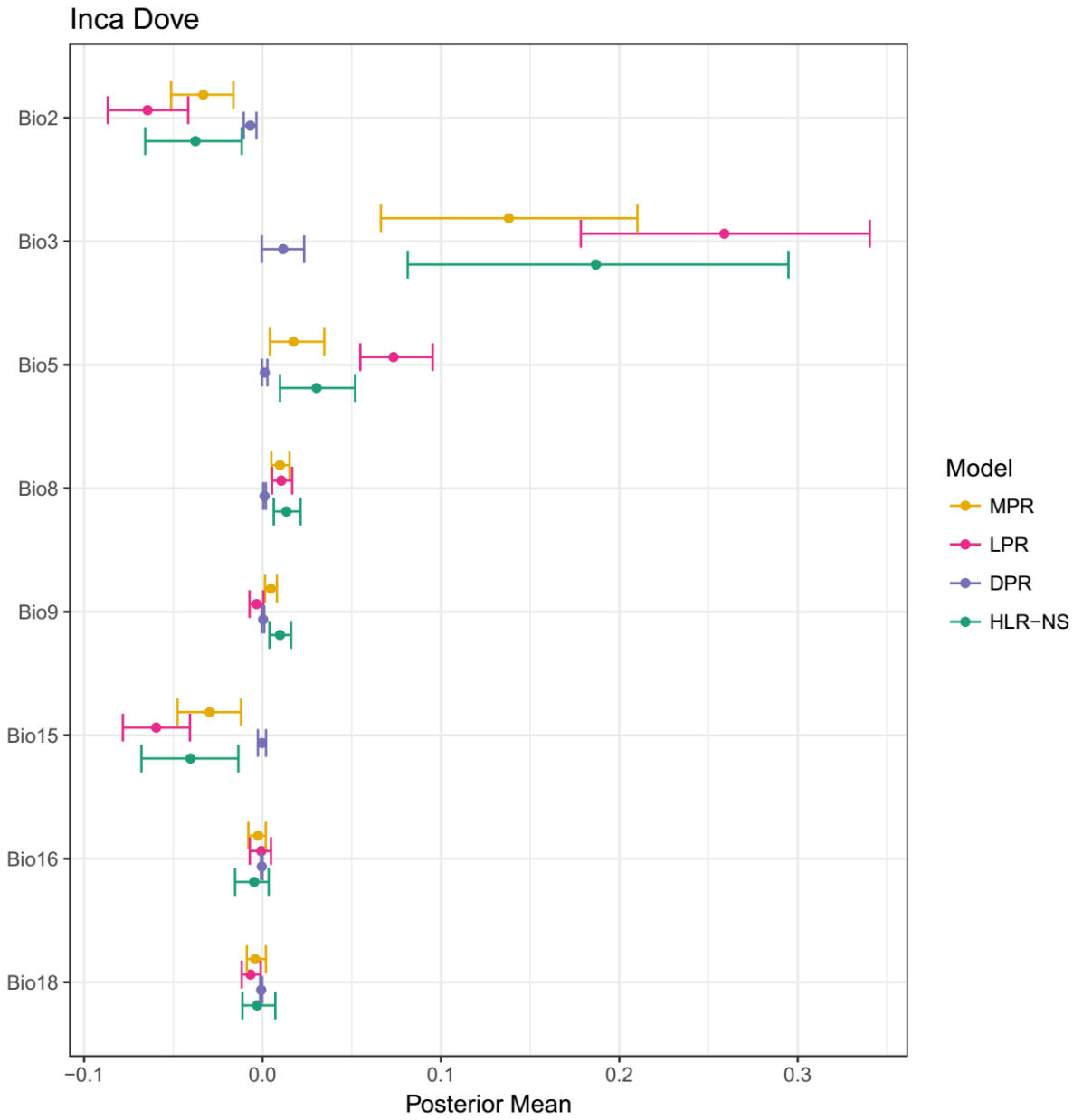


Figure 195: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Inca dove, *Columbina inca*.

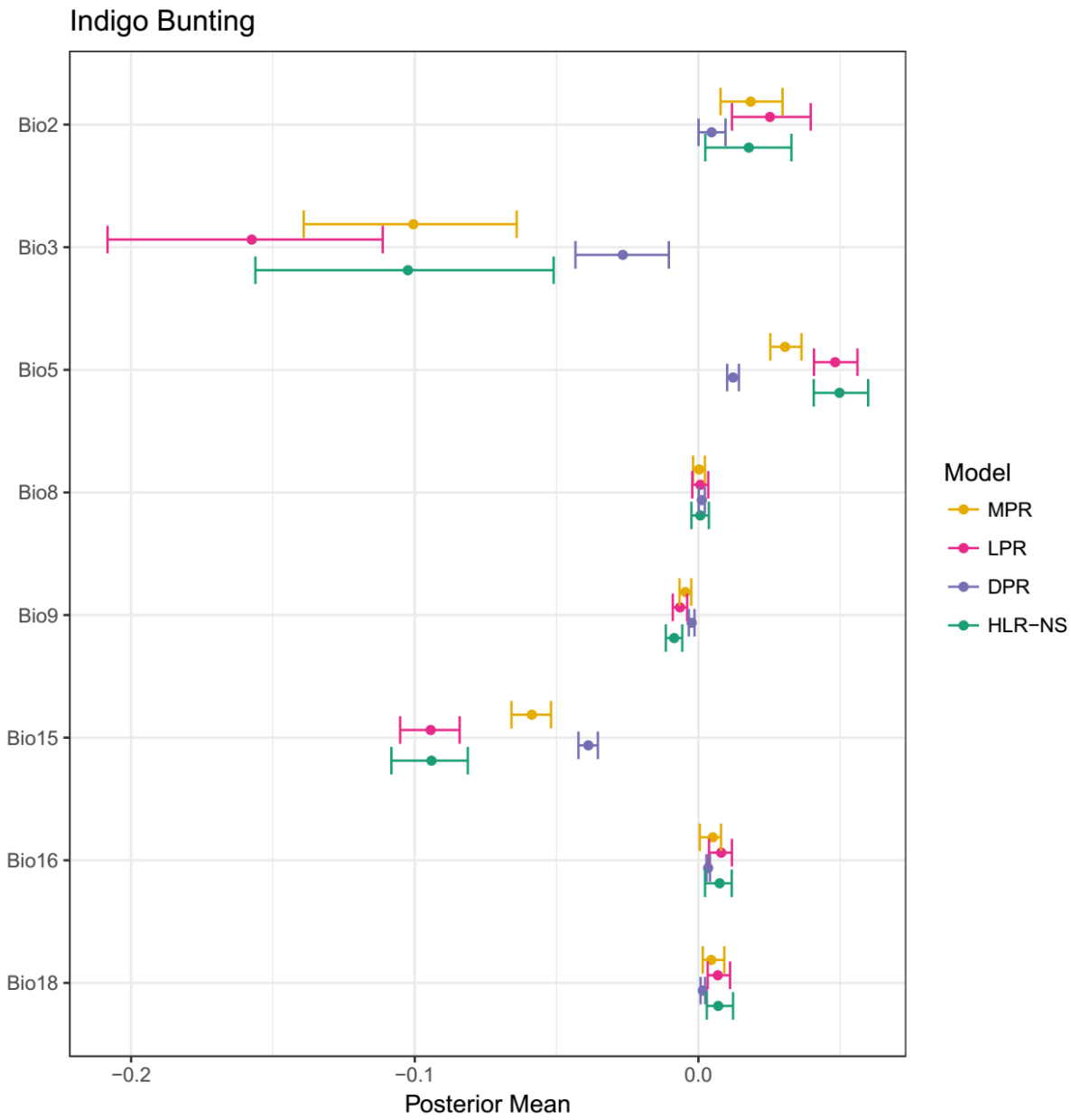


Figure 196: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the indigo bunting, *Passerina cyanea*.

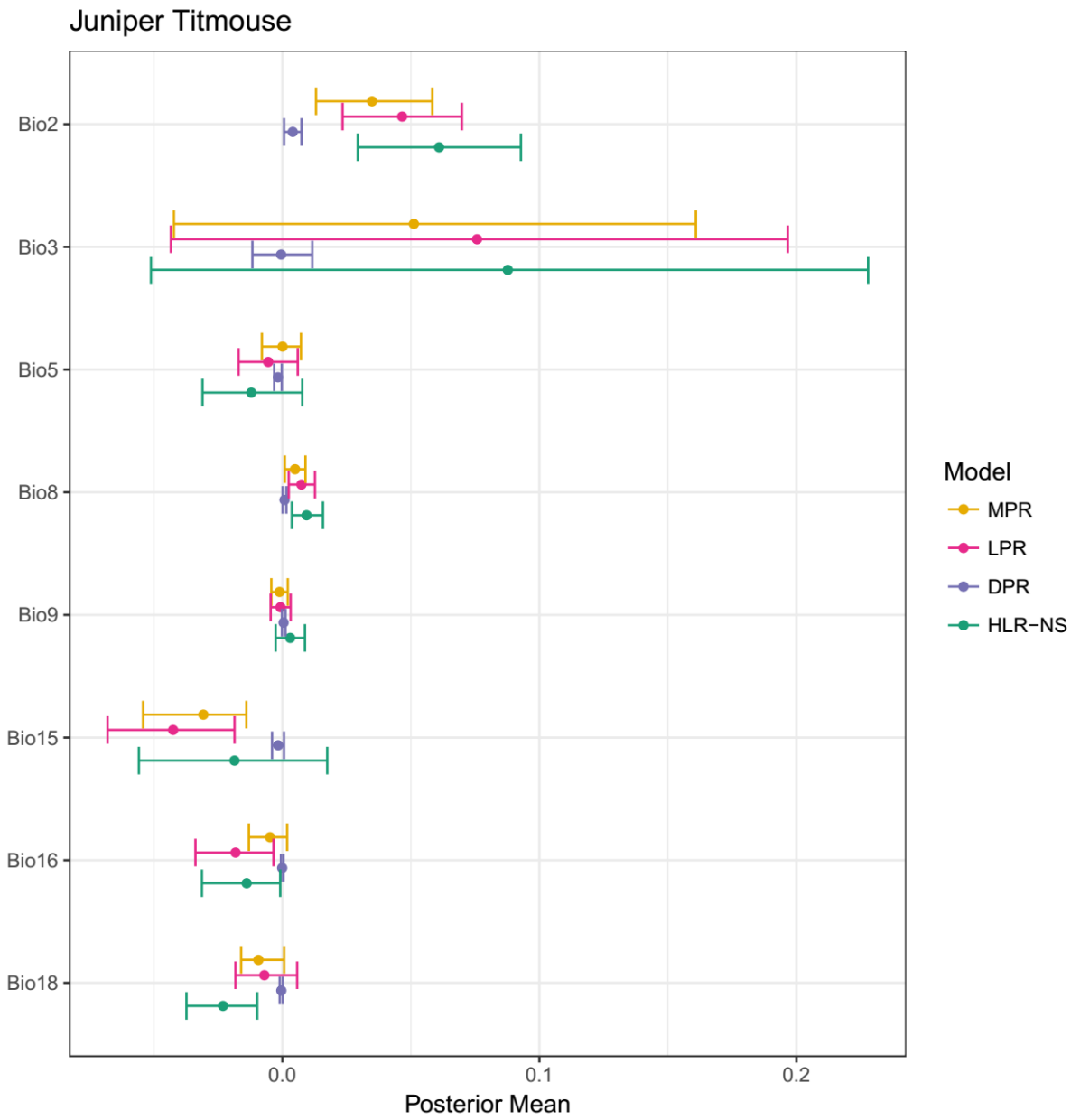


Figure 197: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the juniper titmouse, *Baeolophus ridgwayi*.

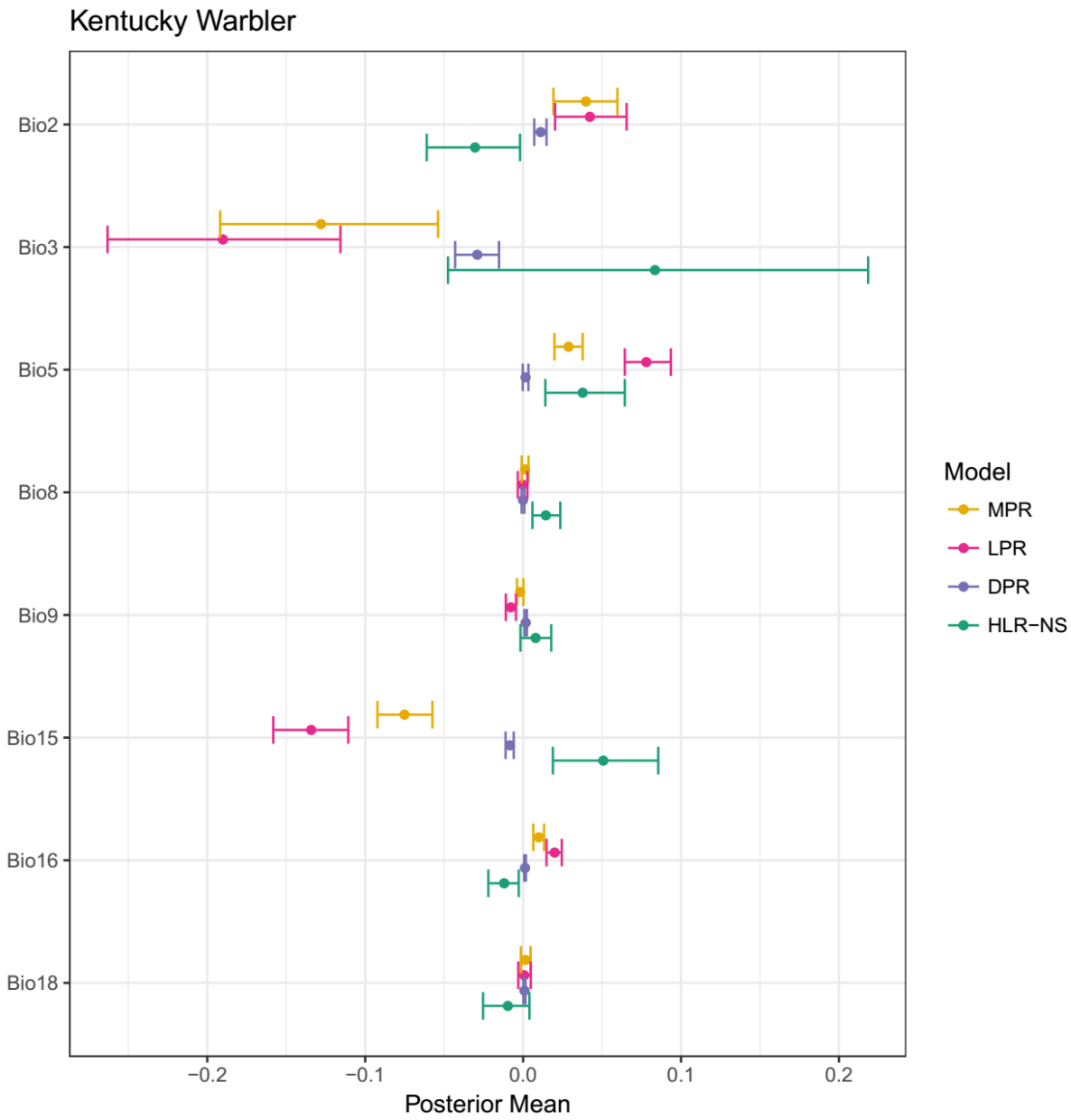


Figure 198: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Kentucky warbler, *Geothlypis formosa*.

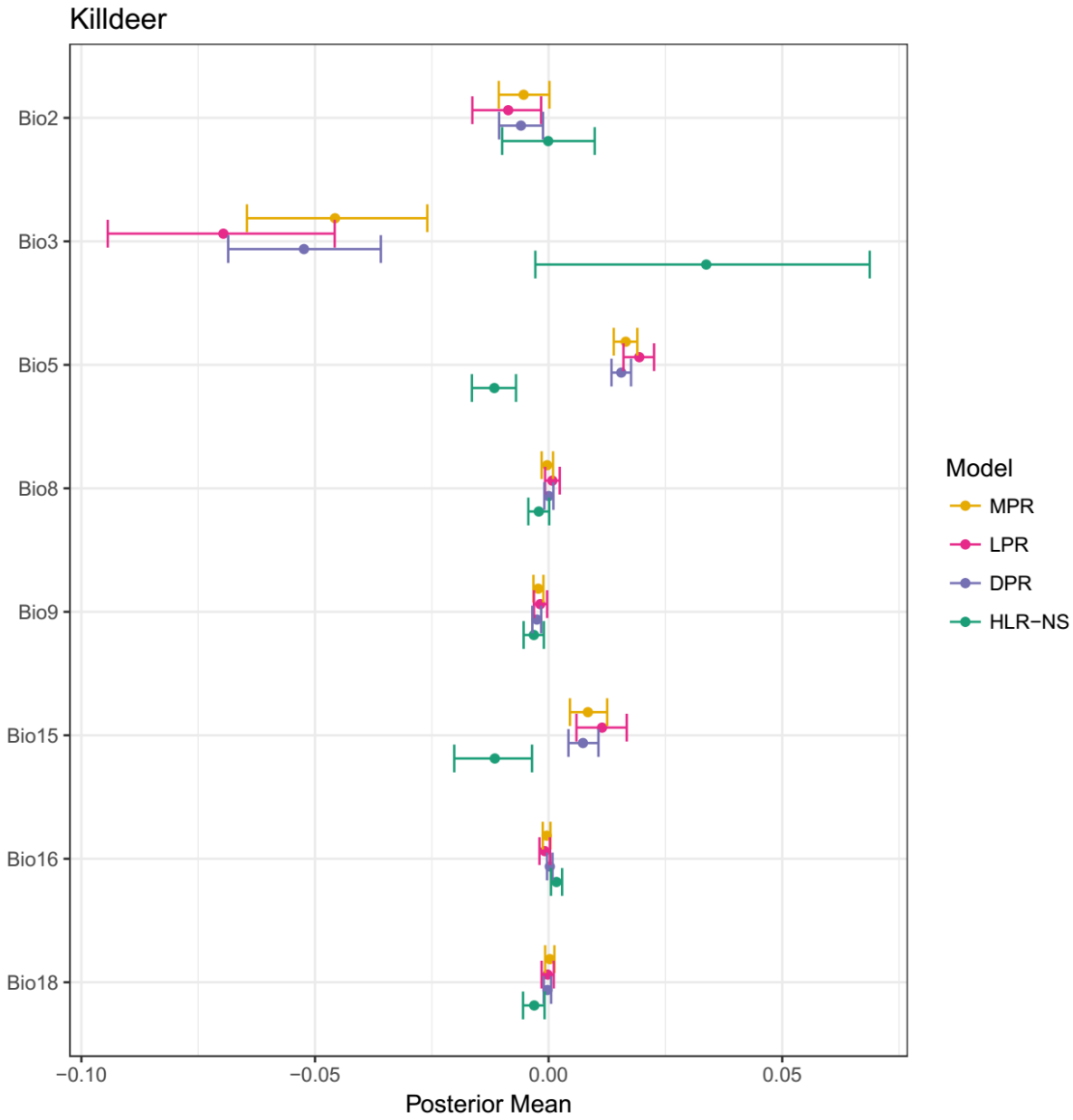


Figure 199: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the killdeer, *Charadrius vociferus*.

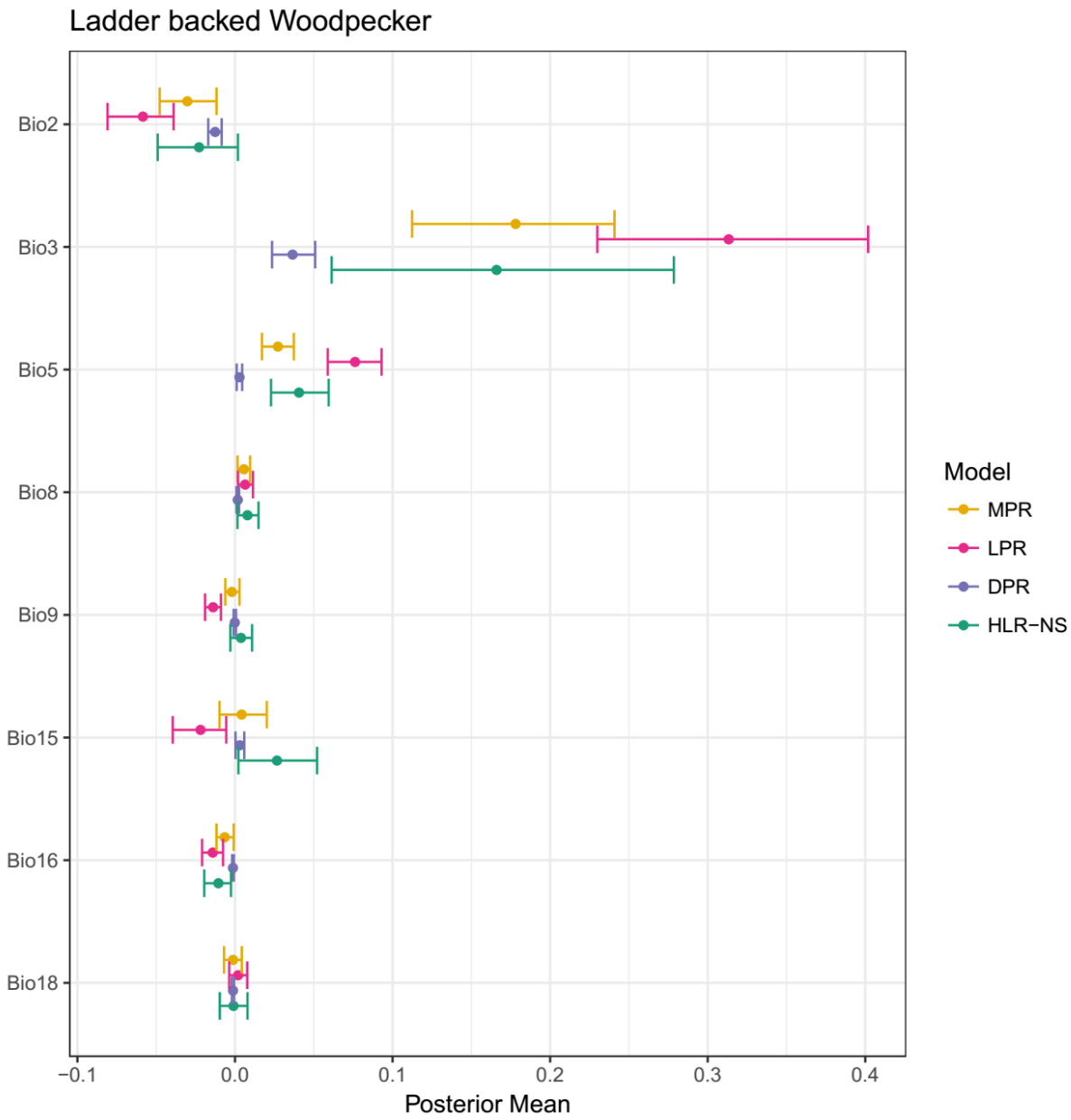


Figure 200: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ladder-backed woodpecker, *Dryobates scalaris*.

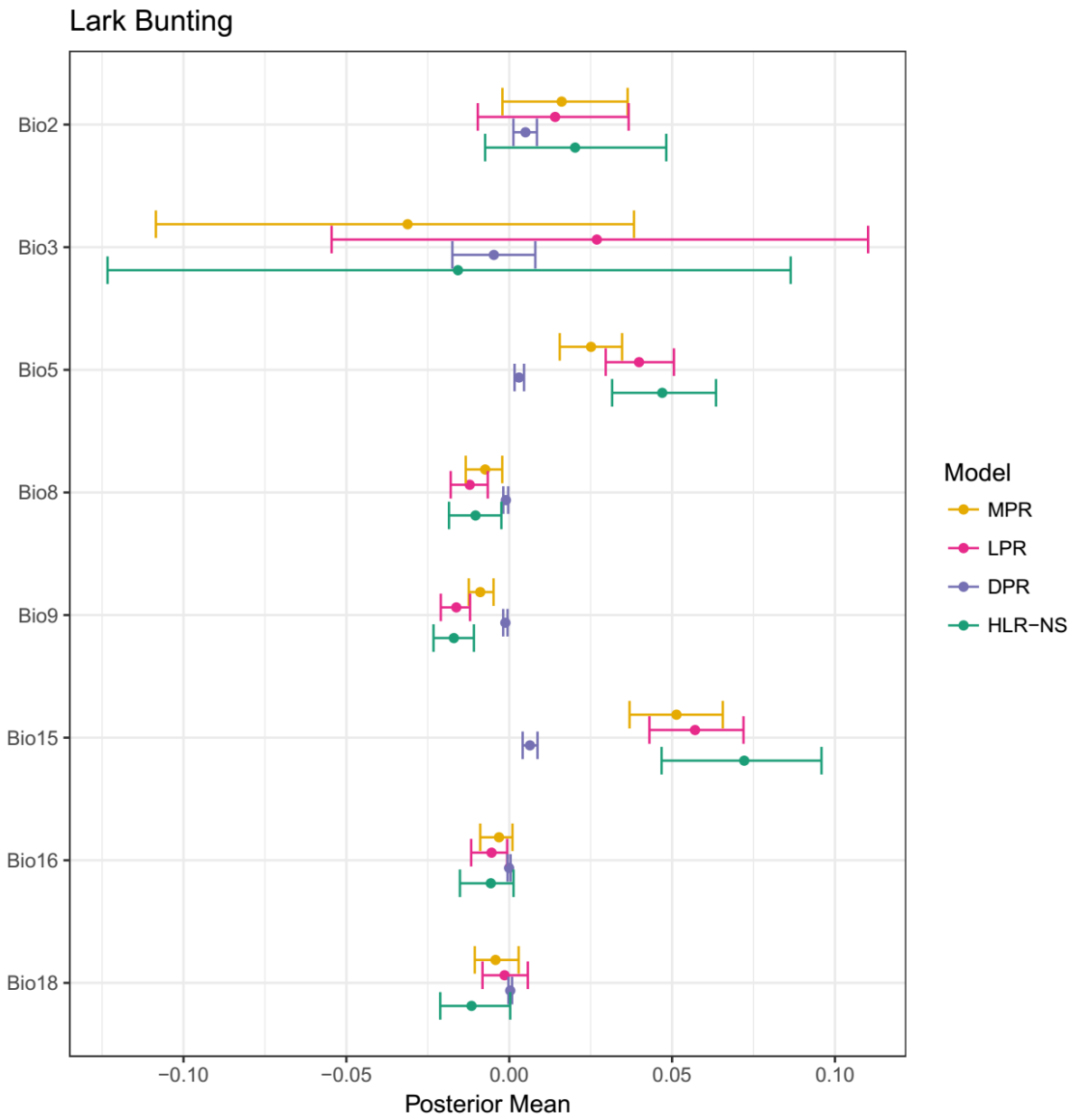


Figure 201: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lark bunting, *Calamospiza melanocorys*.

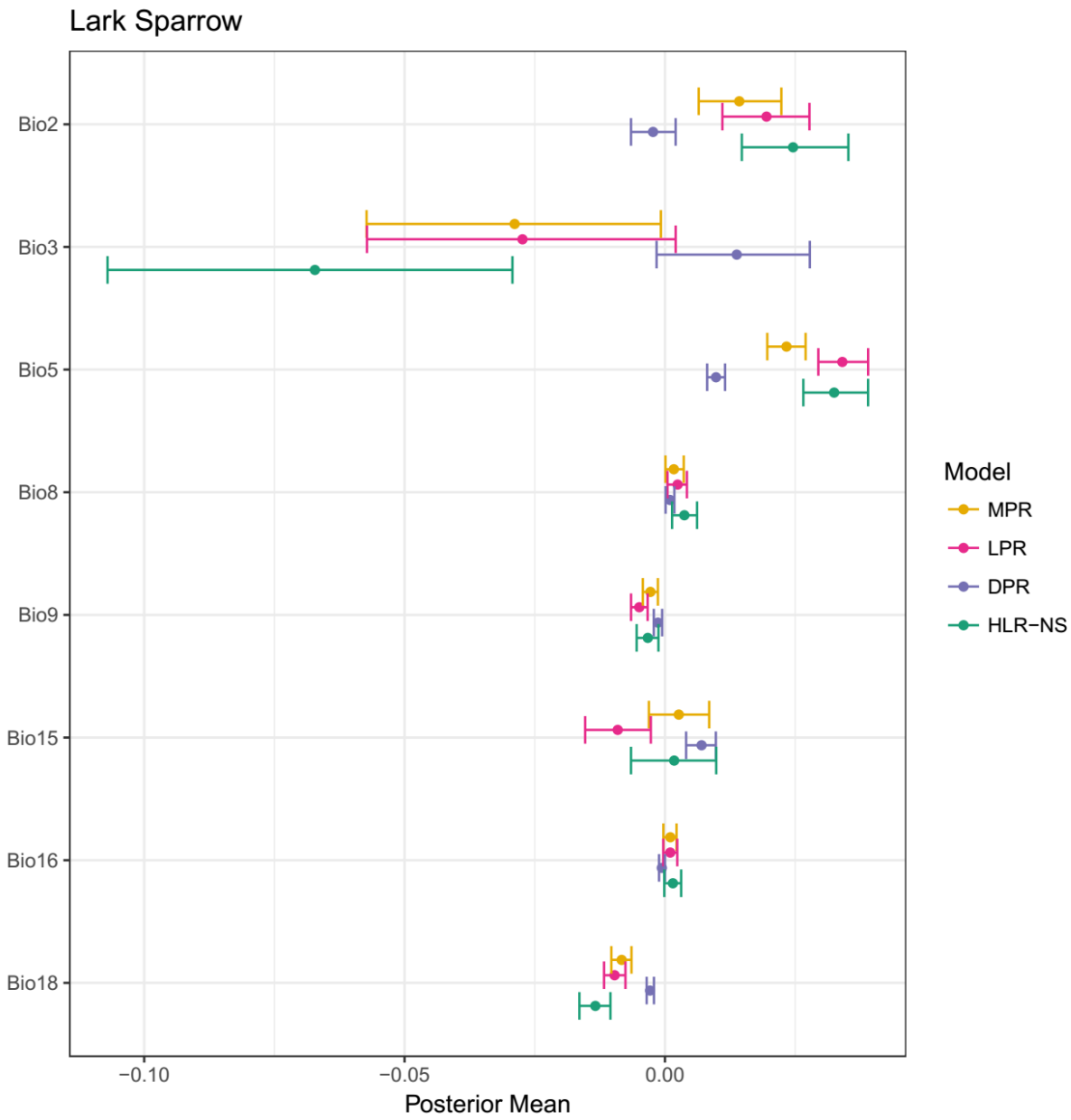


Figure 202: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lark sparrow, *Chondestes grammacus*.

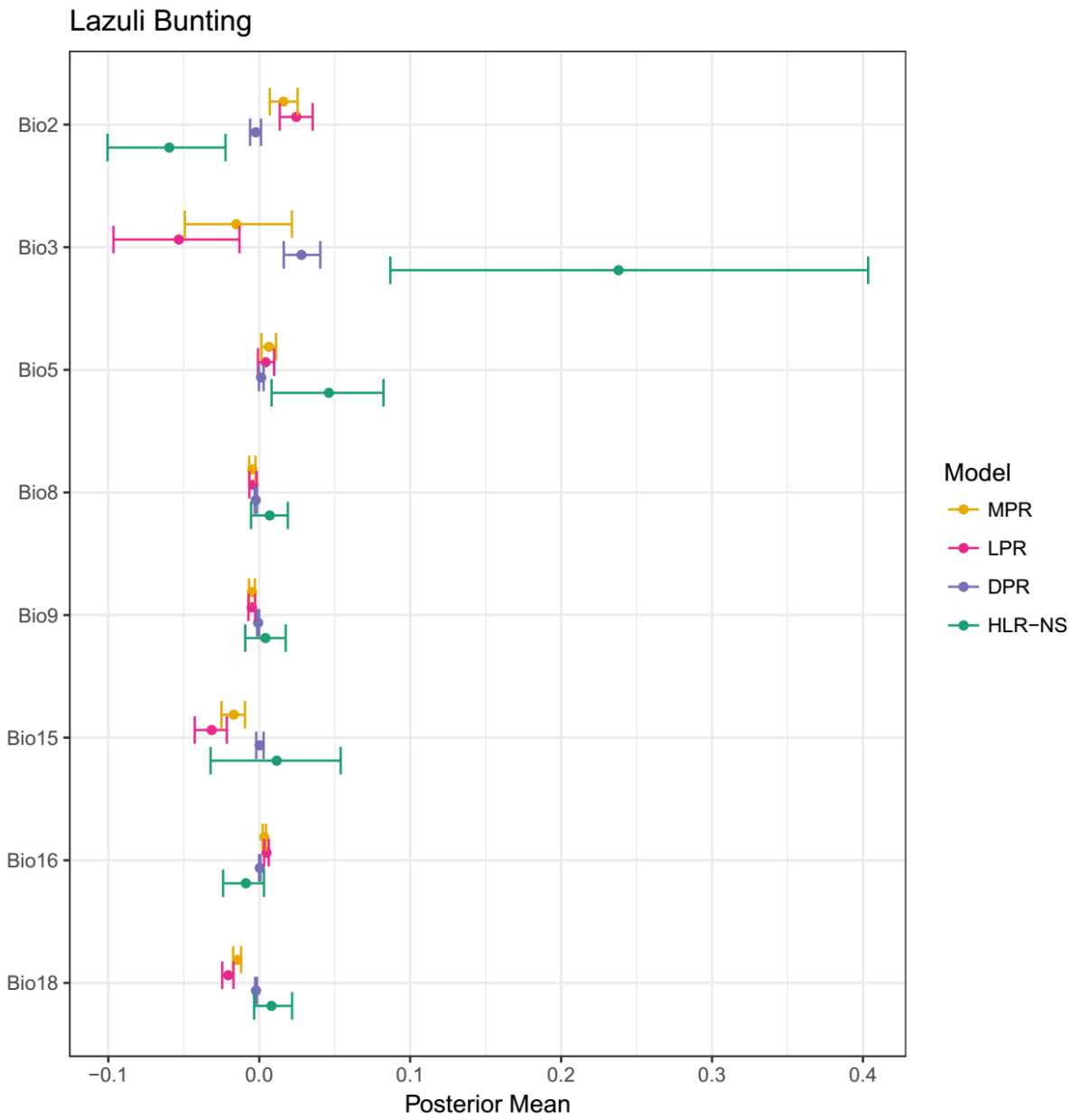


Figure 203: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lazuli bunting, *Passerina amoena*.

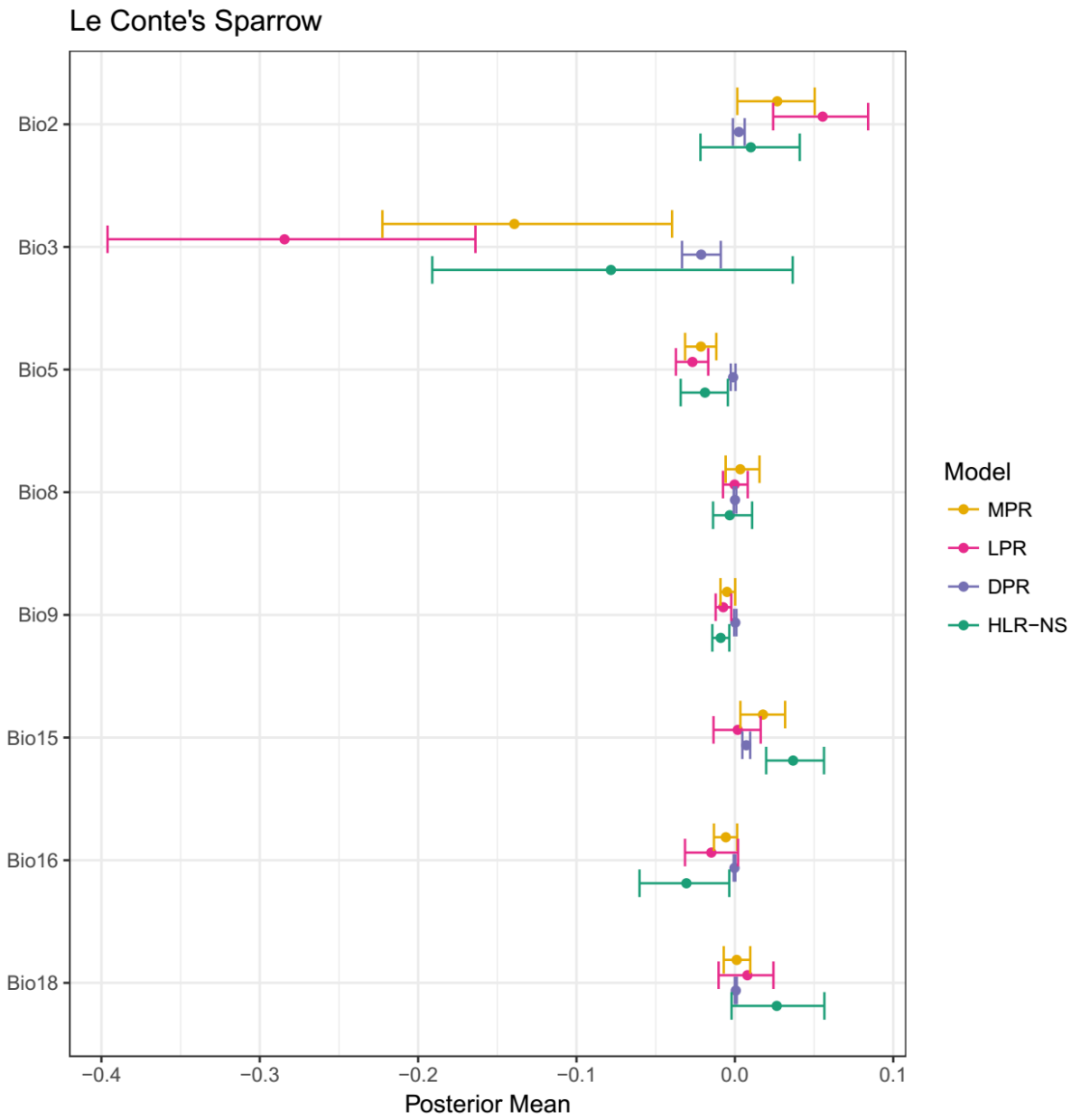


Figure 204: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the LeConte's sparrow, *Ammodramus leconteii*.

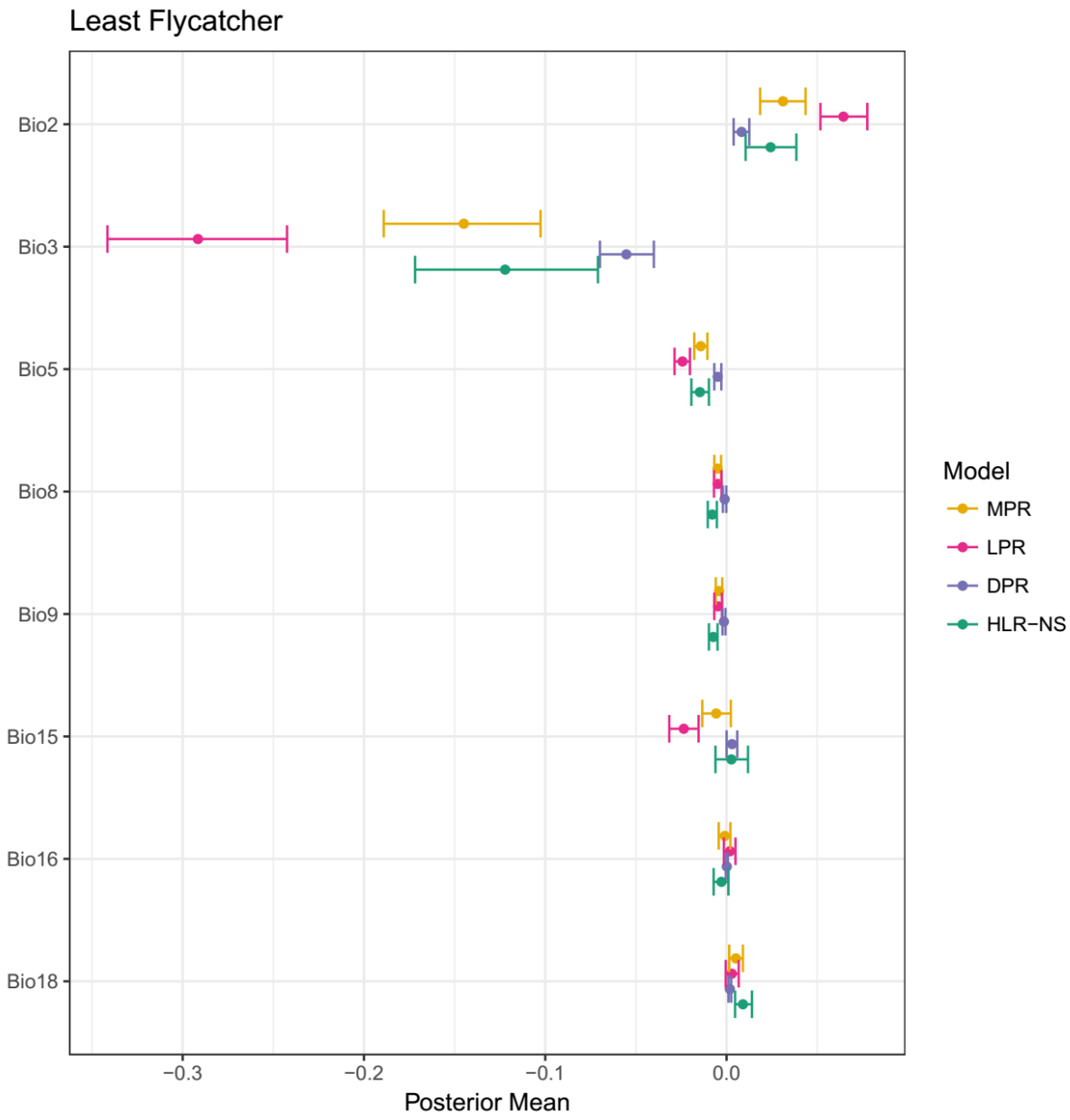


Figure 205: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the least flycatcher, *Empidonax minimus*.

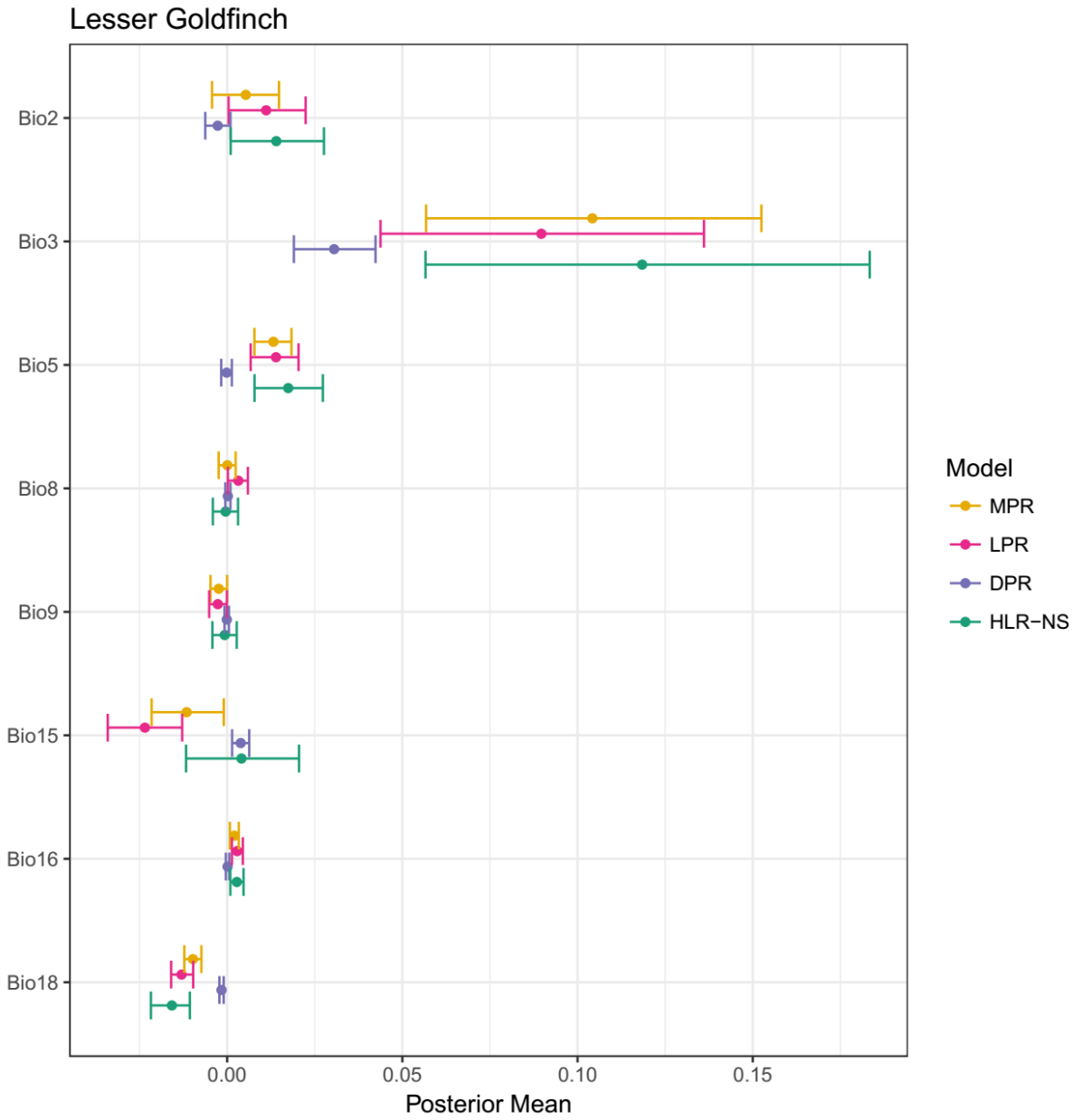


Figure 206: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lesser goldfinch, *Spinus psaltria*.

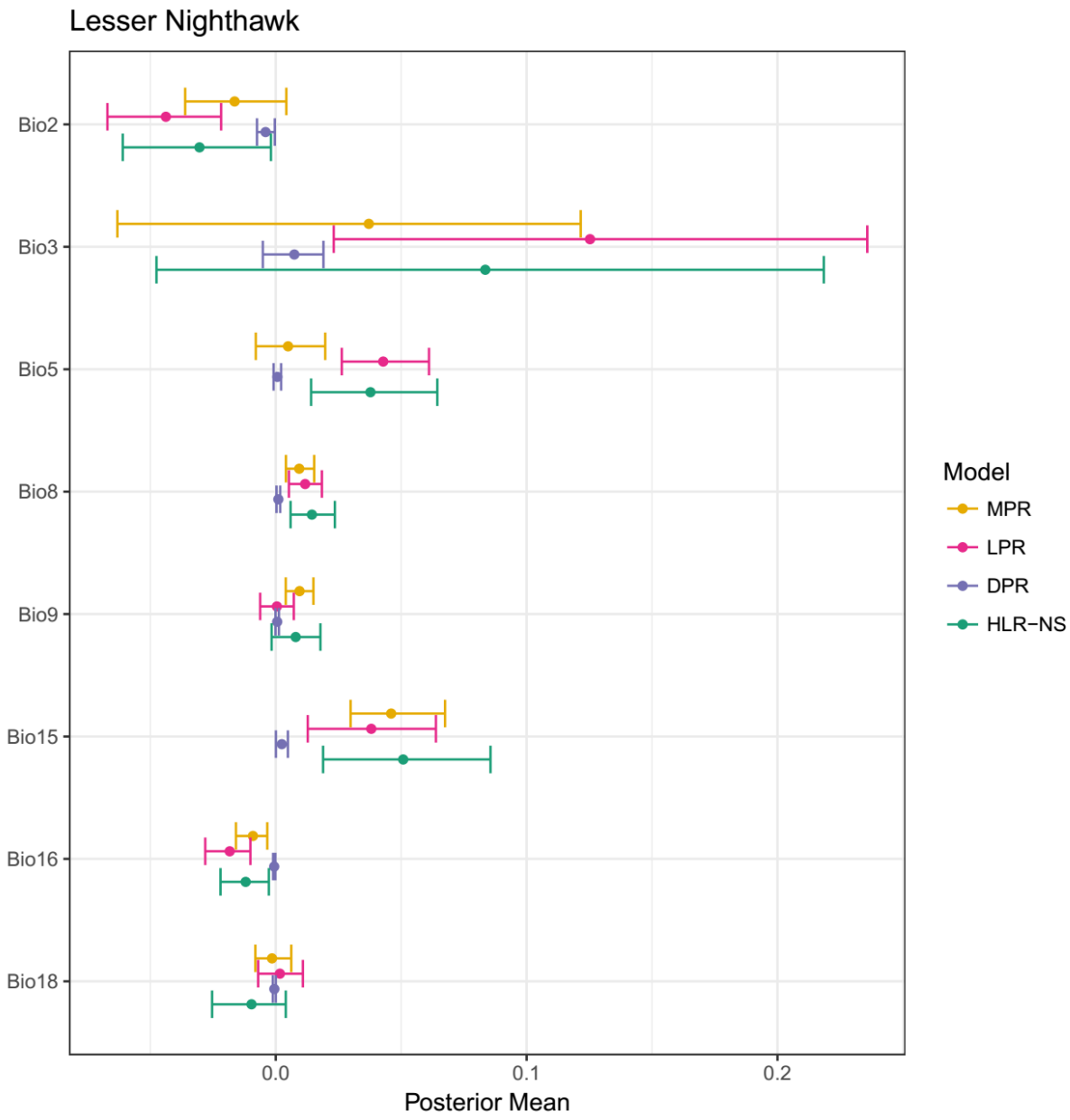


Figure 207: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lesser nighthawk, *Chordeiles acutipennis*.

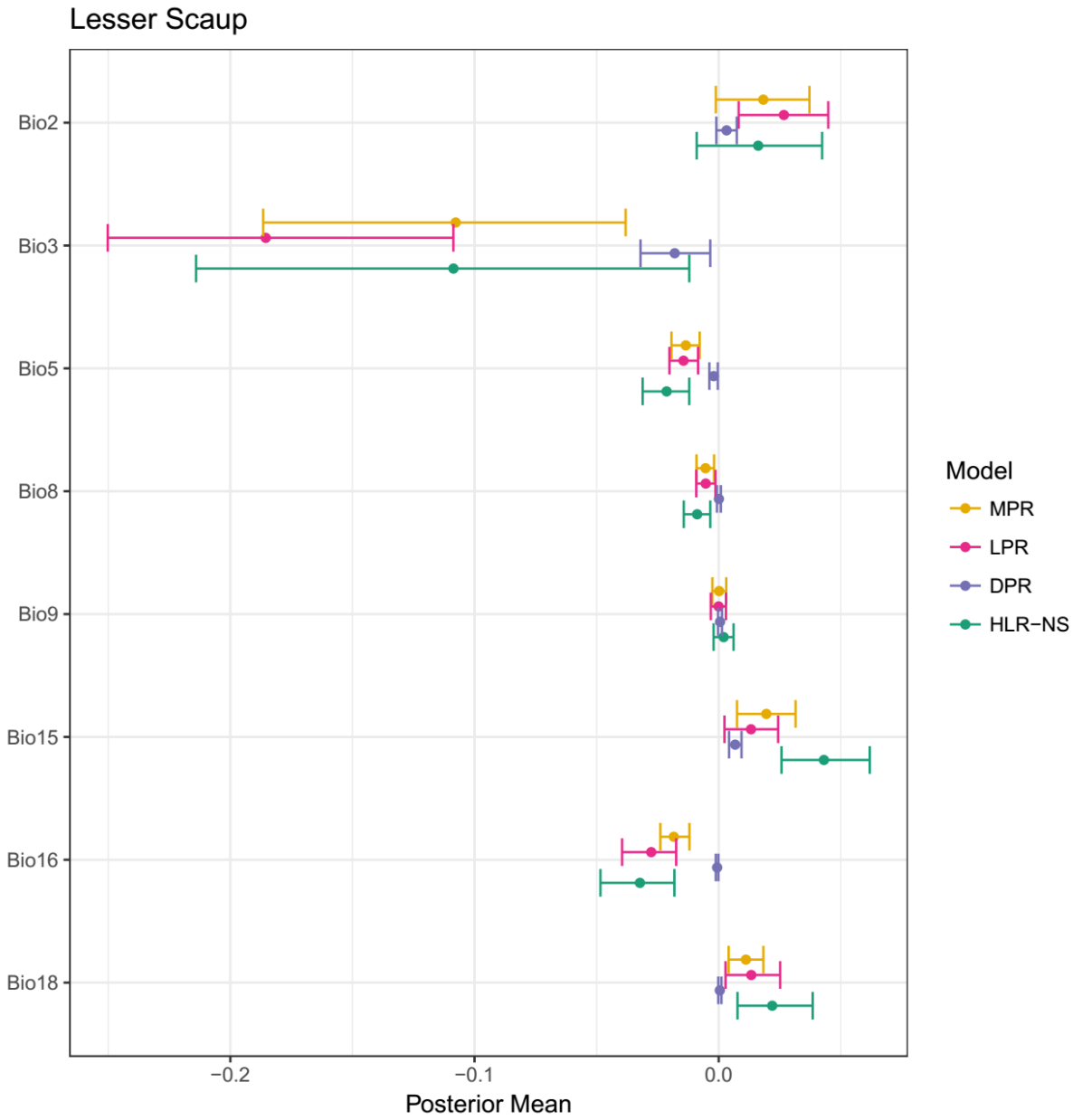


Figure 208: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lesser scaup, *Aythya affinis*.

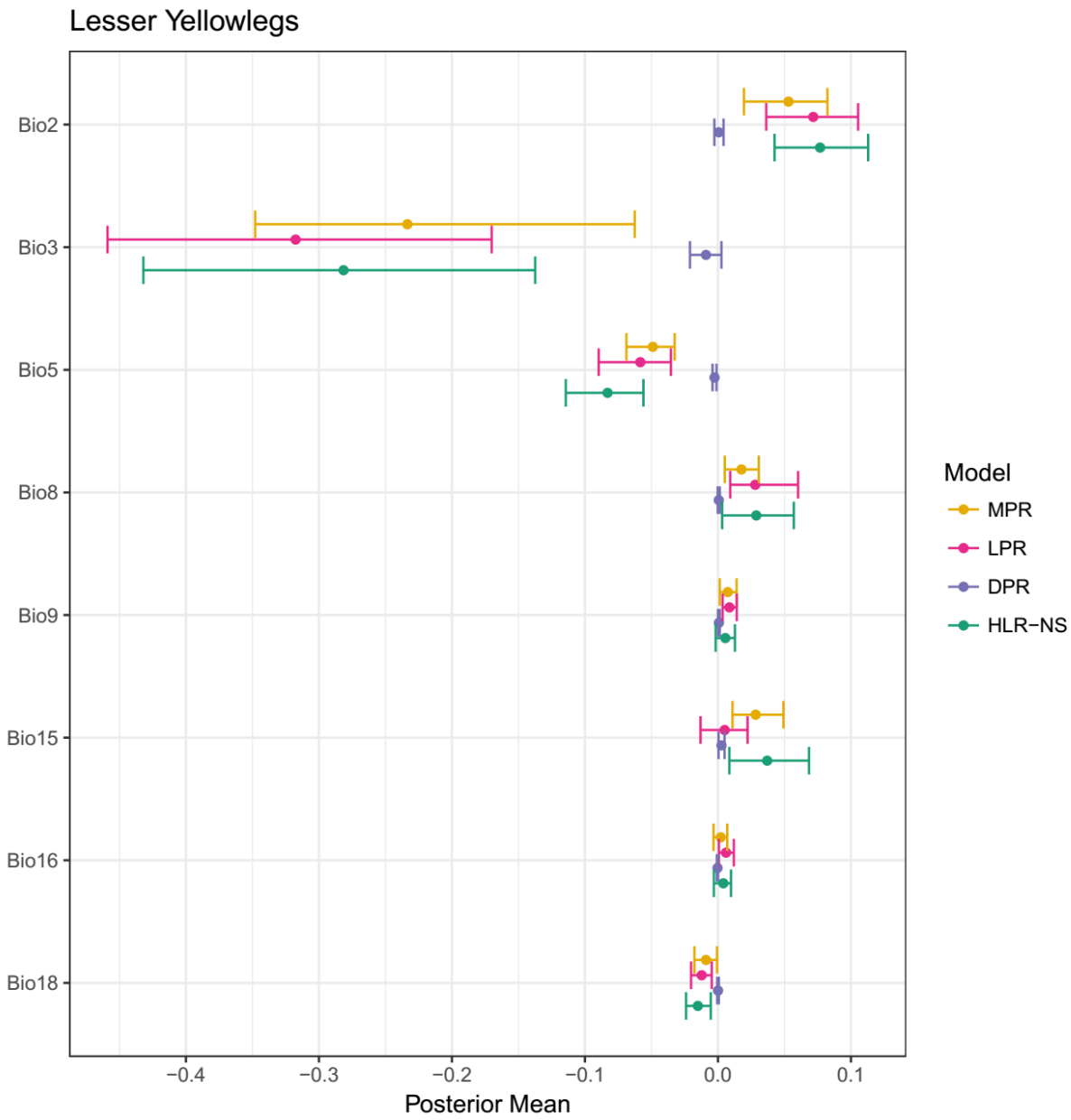


Figure 209: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the lesser yellowlegs, *Tringa flavipes*.

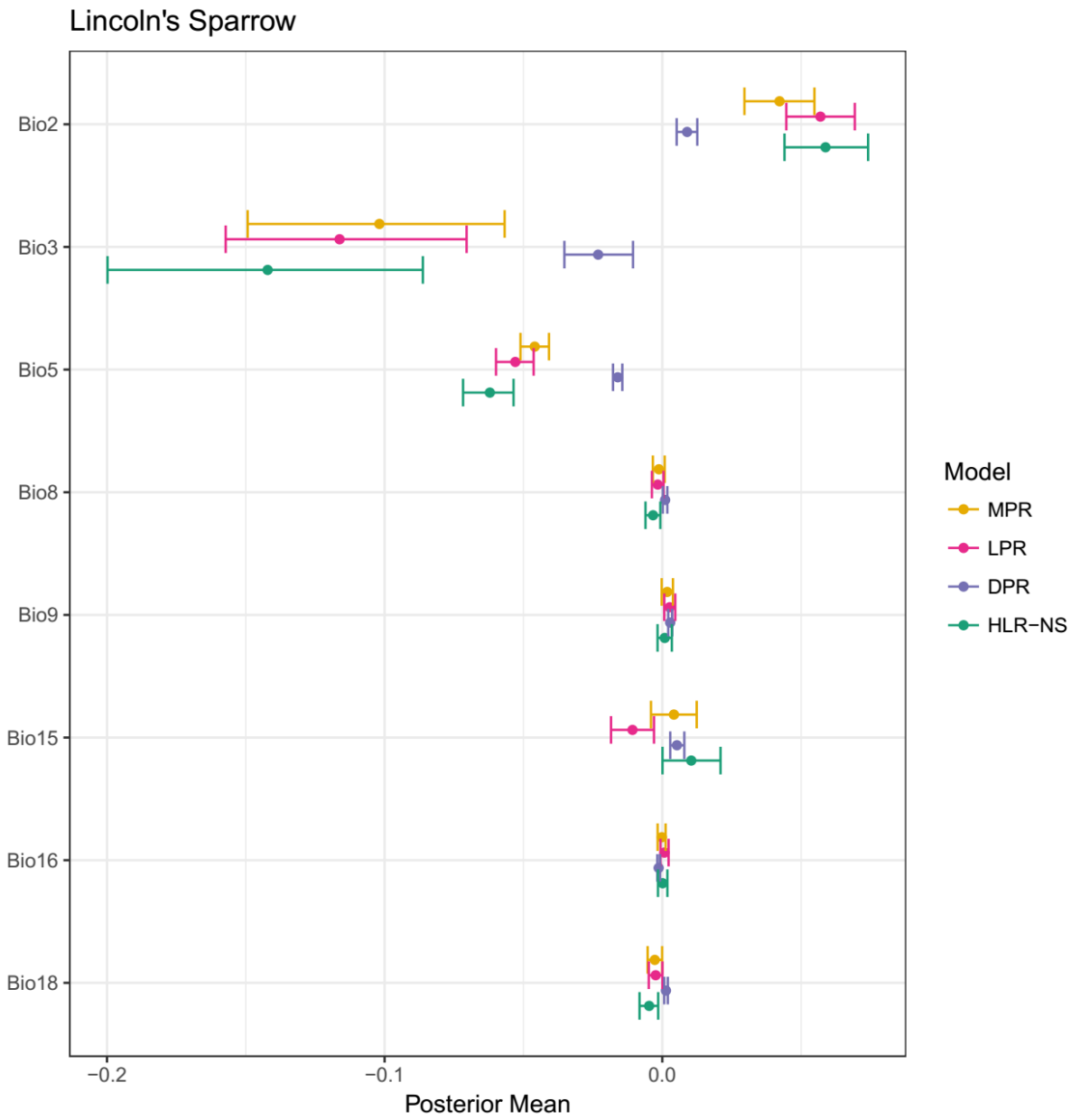


Figure 210: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Lincoln's sparrow, *Melospiza lincolnii*.

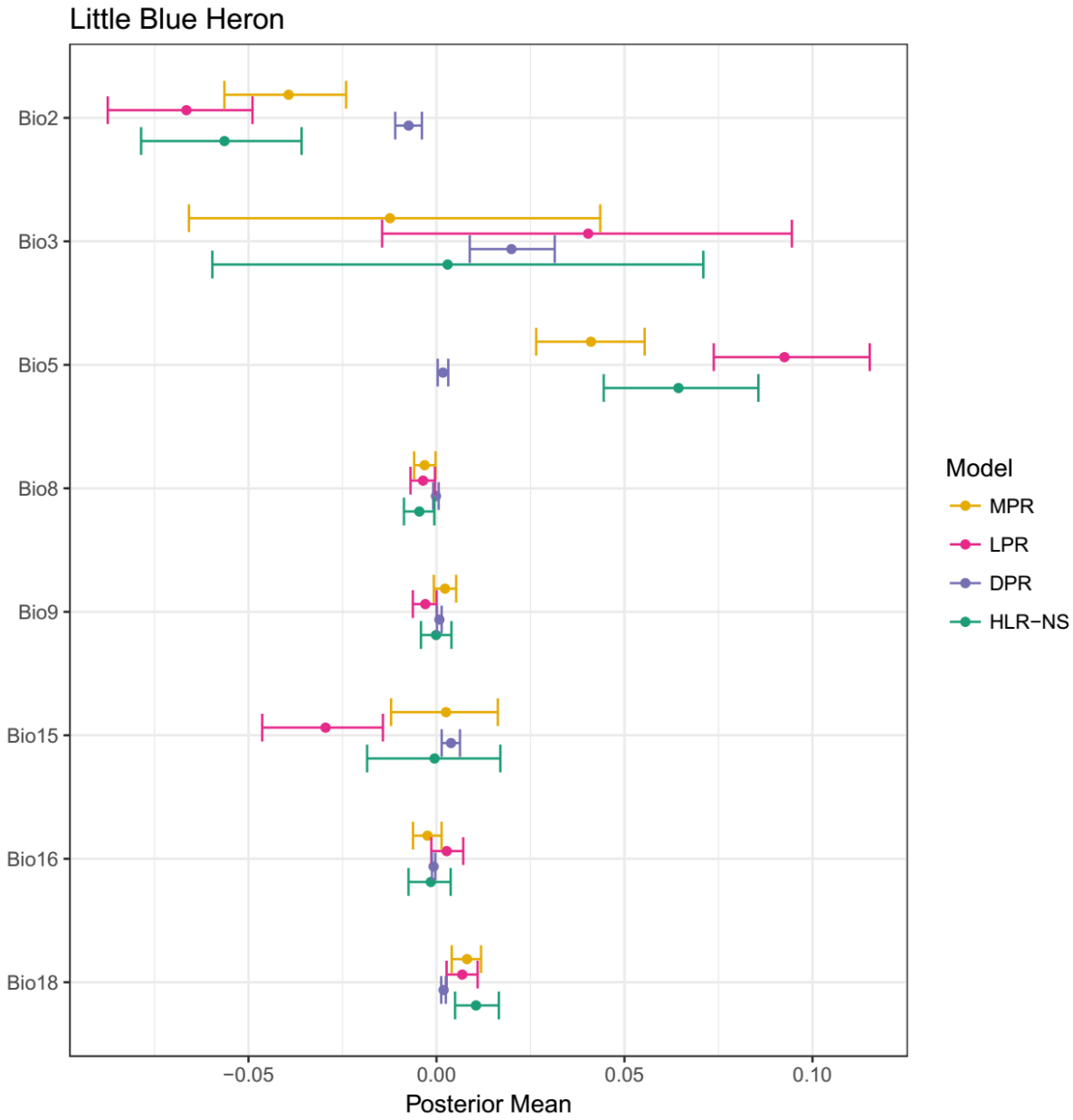


Figure 211: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the little blue heron, *Egretta caerulea*.

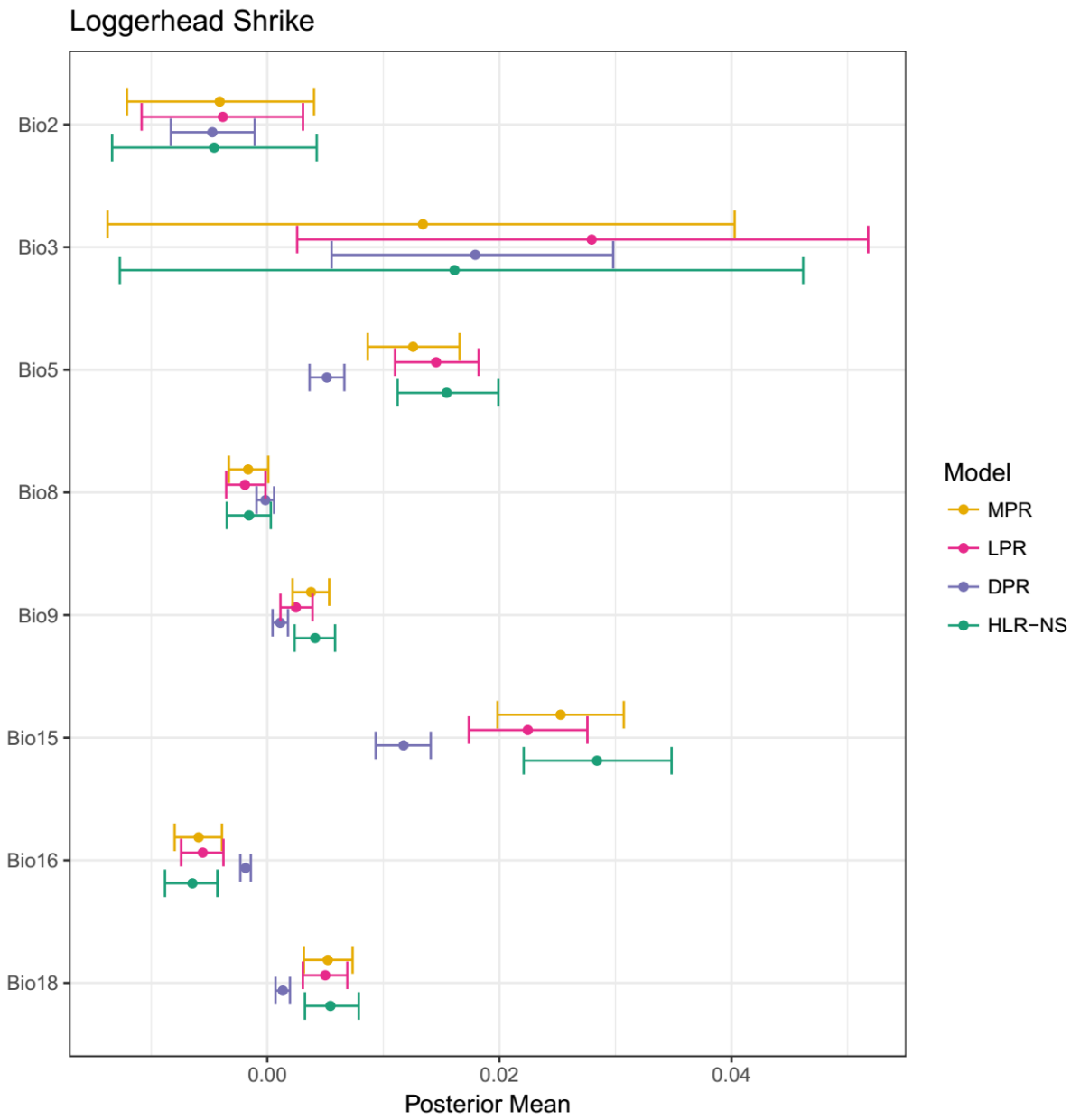


Figure 212: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the loggerhead shrike, *Lanius ludovicianus*.

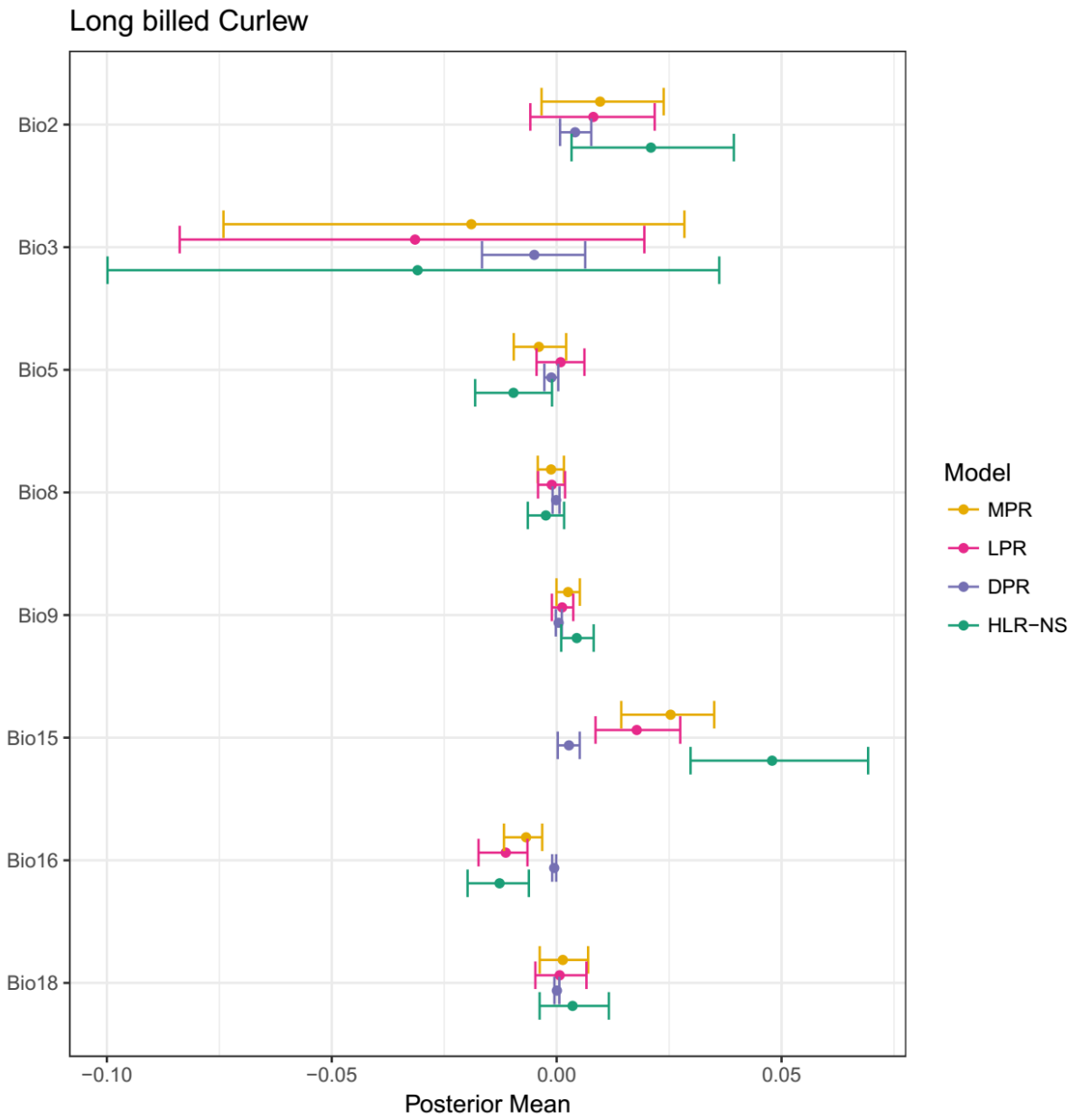


Figure 213: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the long-billed curlew, *Numenius americanus*.

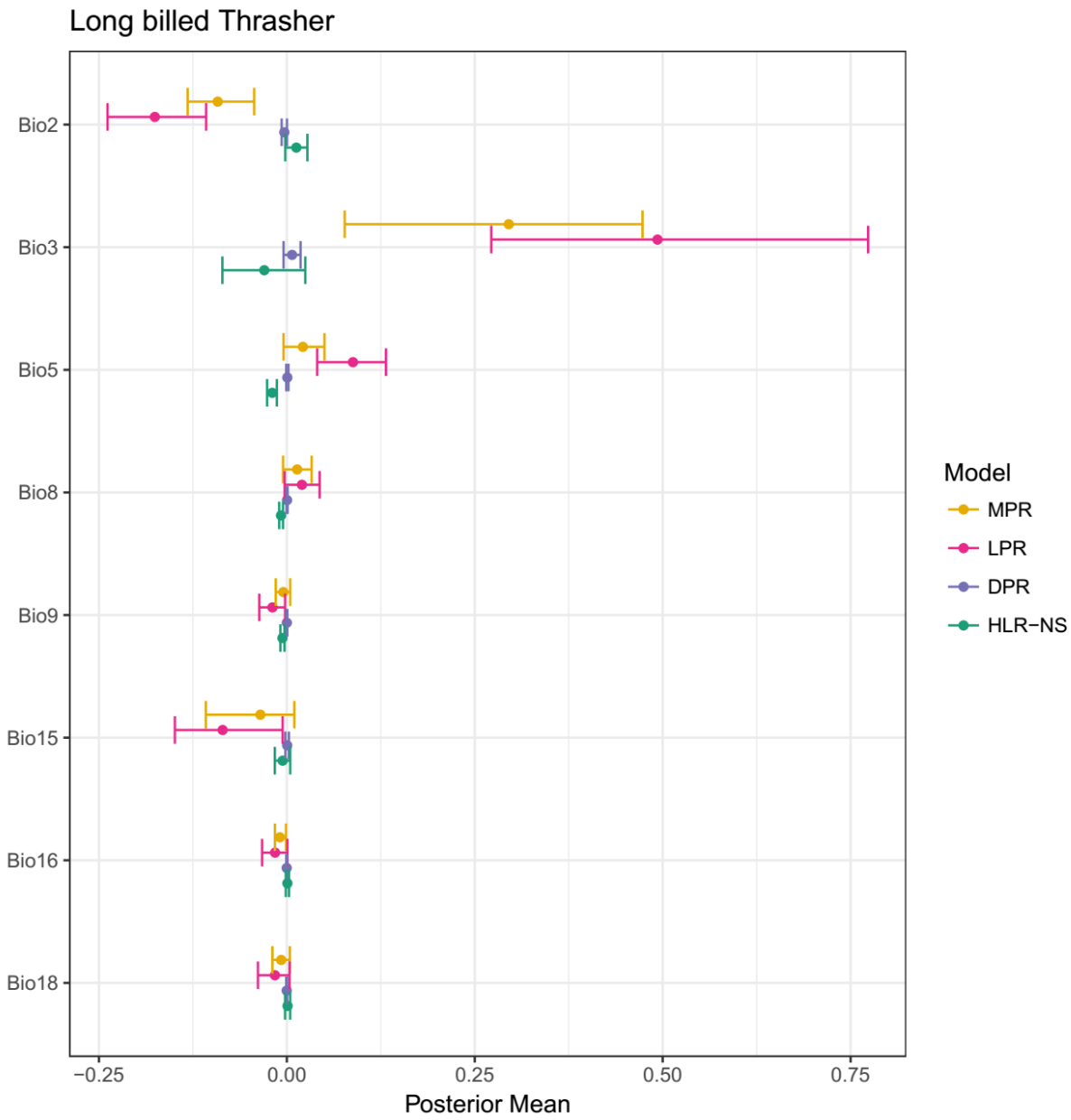


Figure 214: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the long-billed thrasher, *Toxostoma longirostre*.

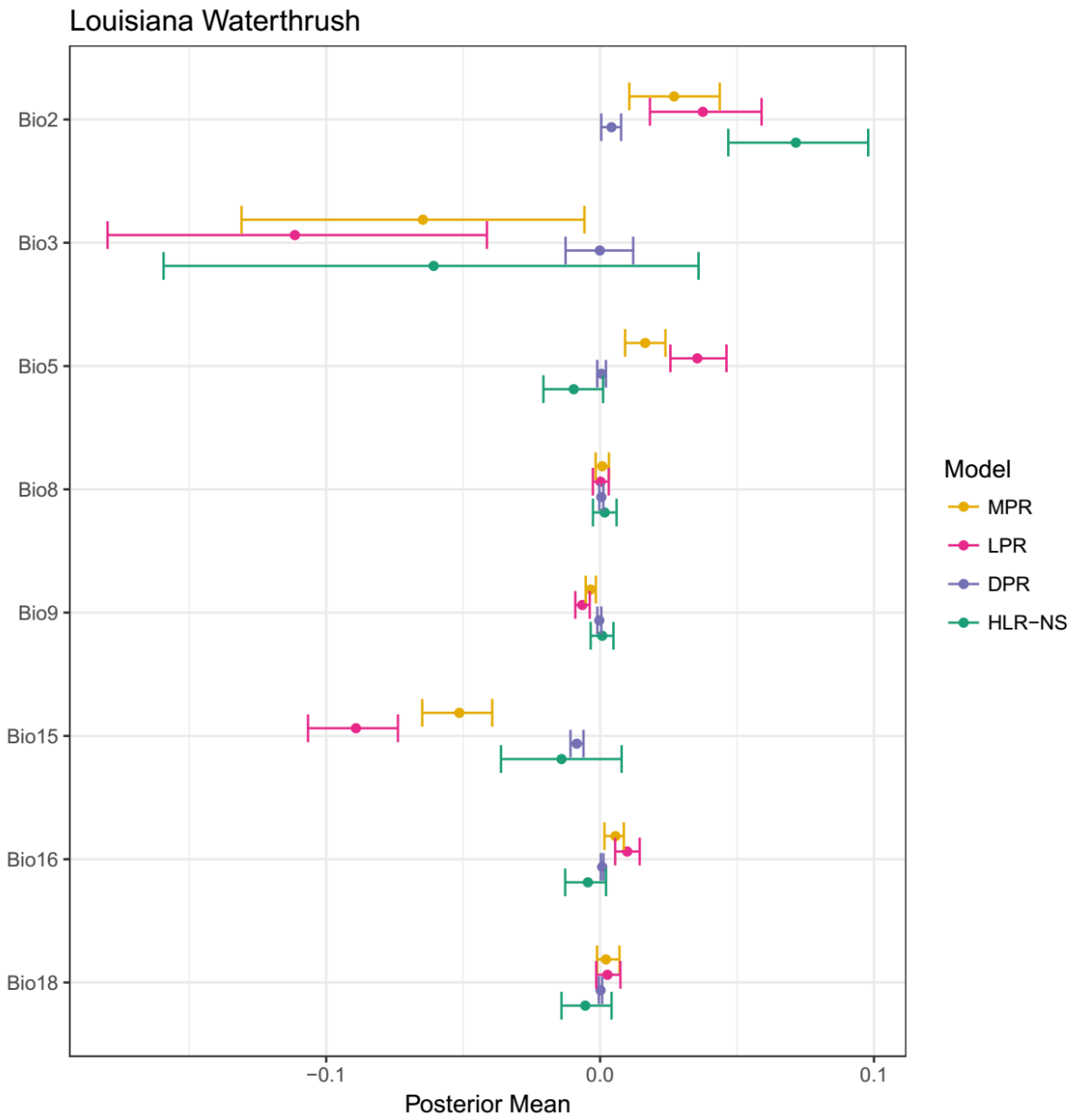


Figure 215: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Louisiana waterthrush, *Parkesia motacilla*.

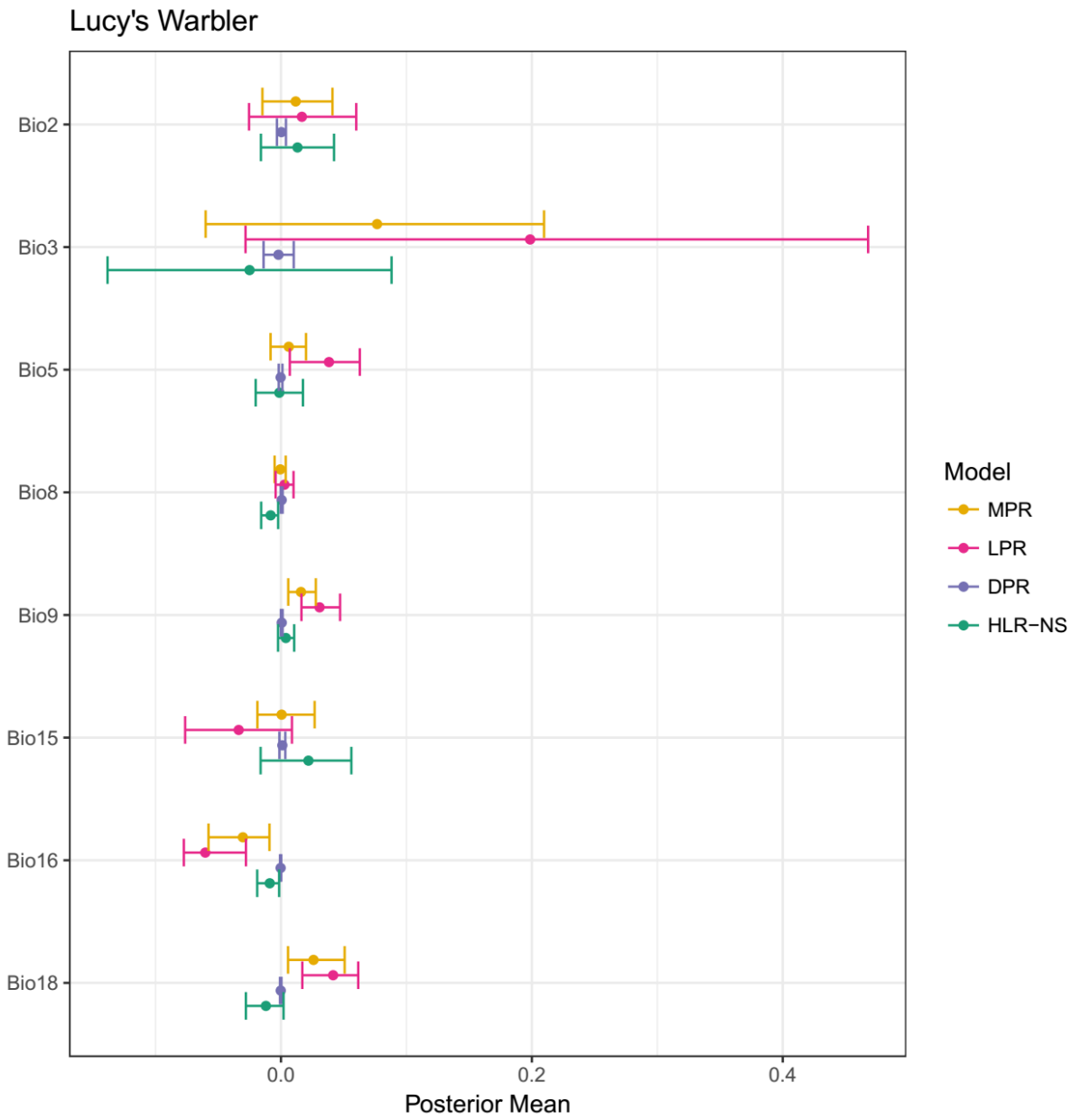


Figure 216: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Lucy's warbler, *Oreothlypis luciae*.

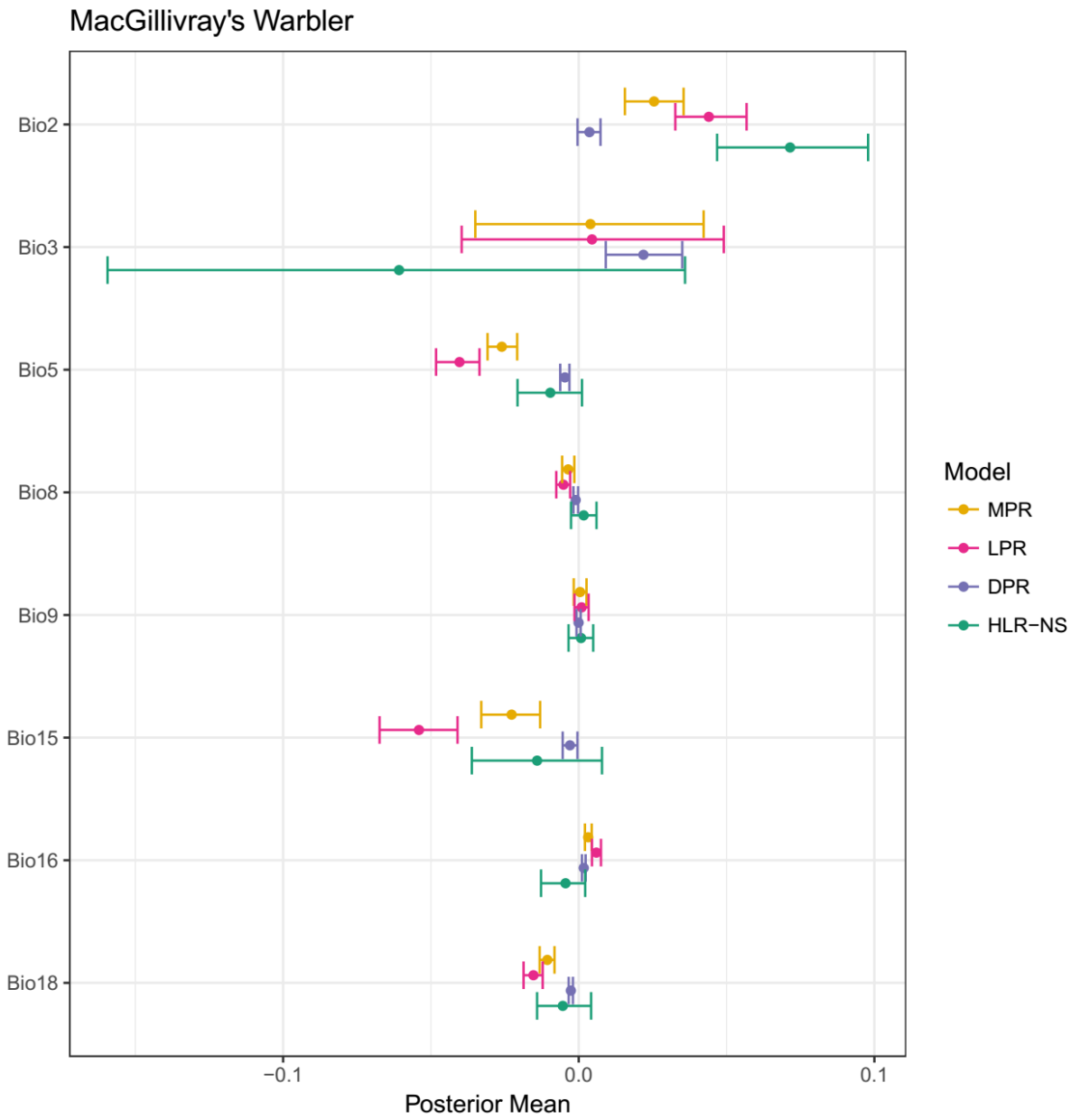


Figure 217: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the MacGillivray's warbler, *Geothlypis tolmiei*.

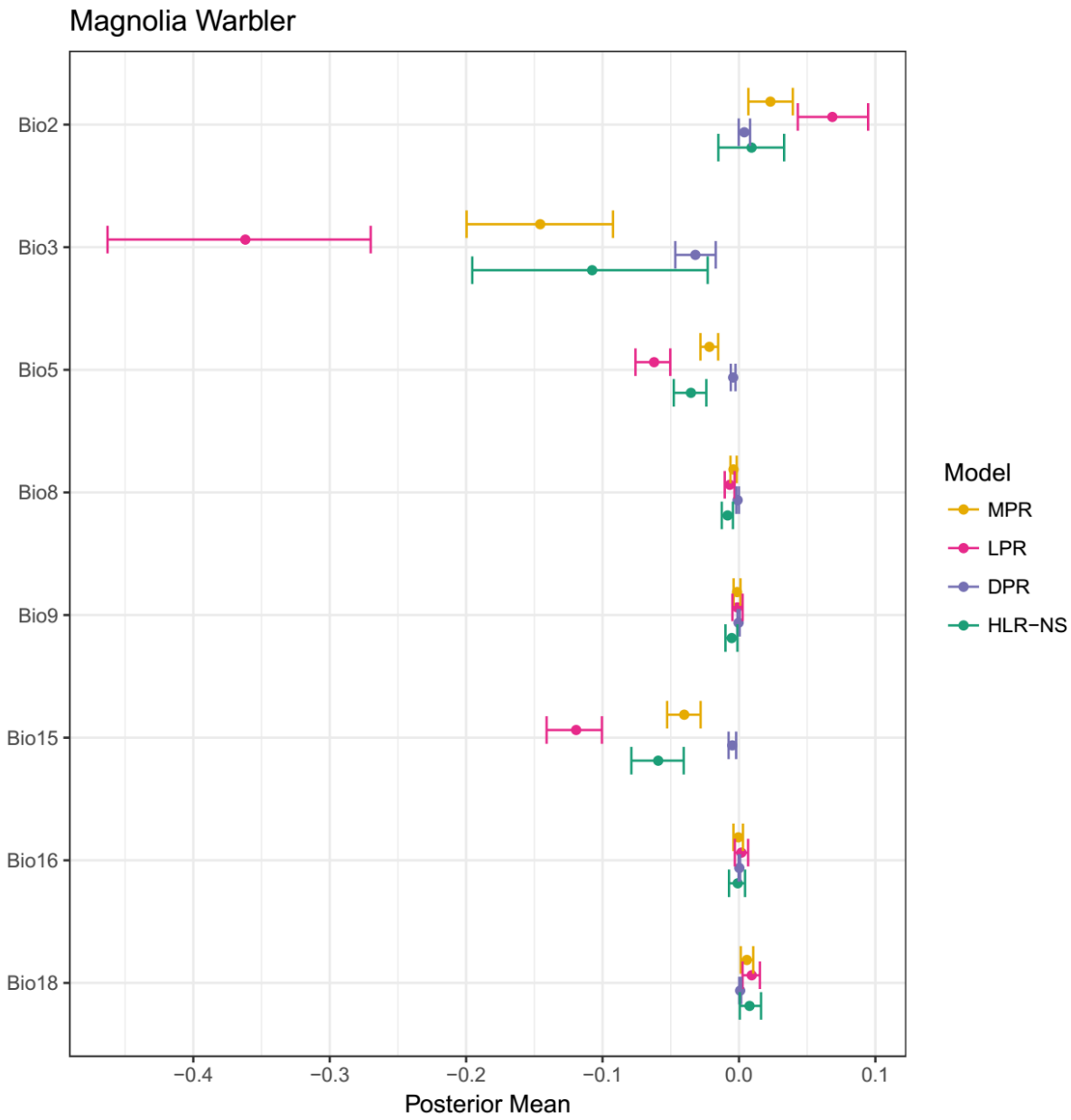


Figure 218: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the magnolia warbler, *Setophaga magnolia*.

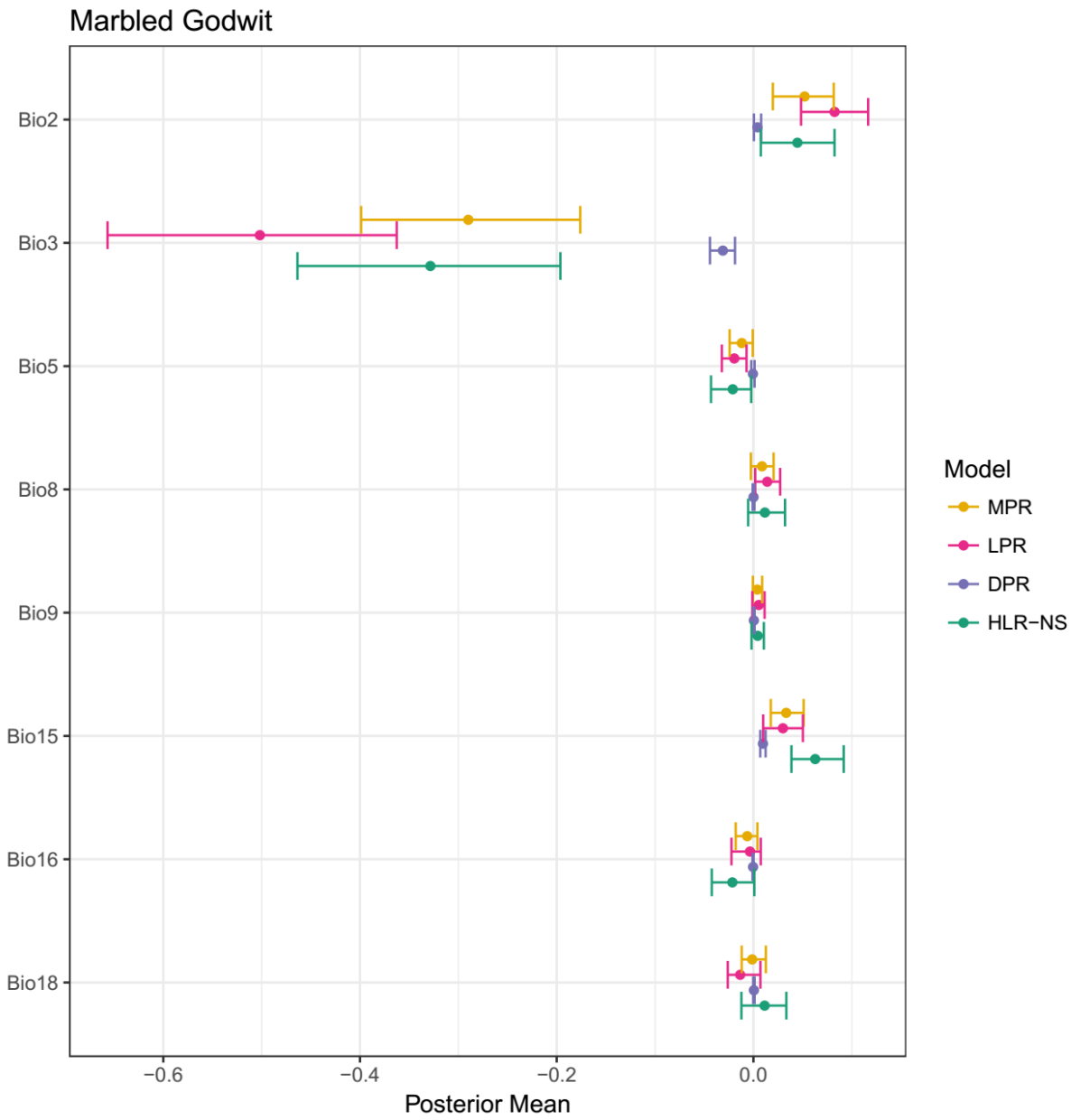


Figure 219: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the marbled godwit, *Limosa fedoa*.

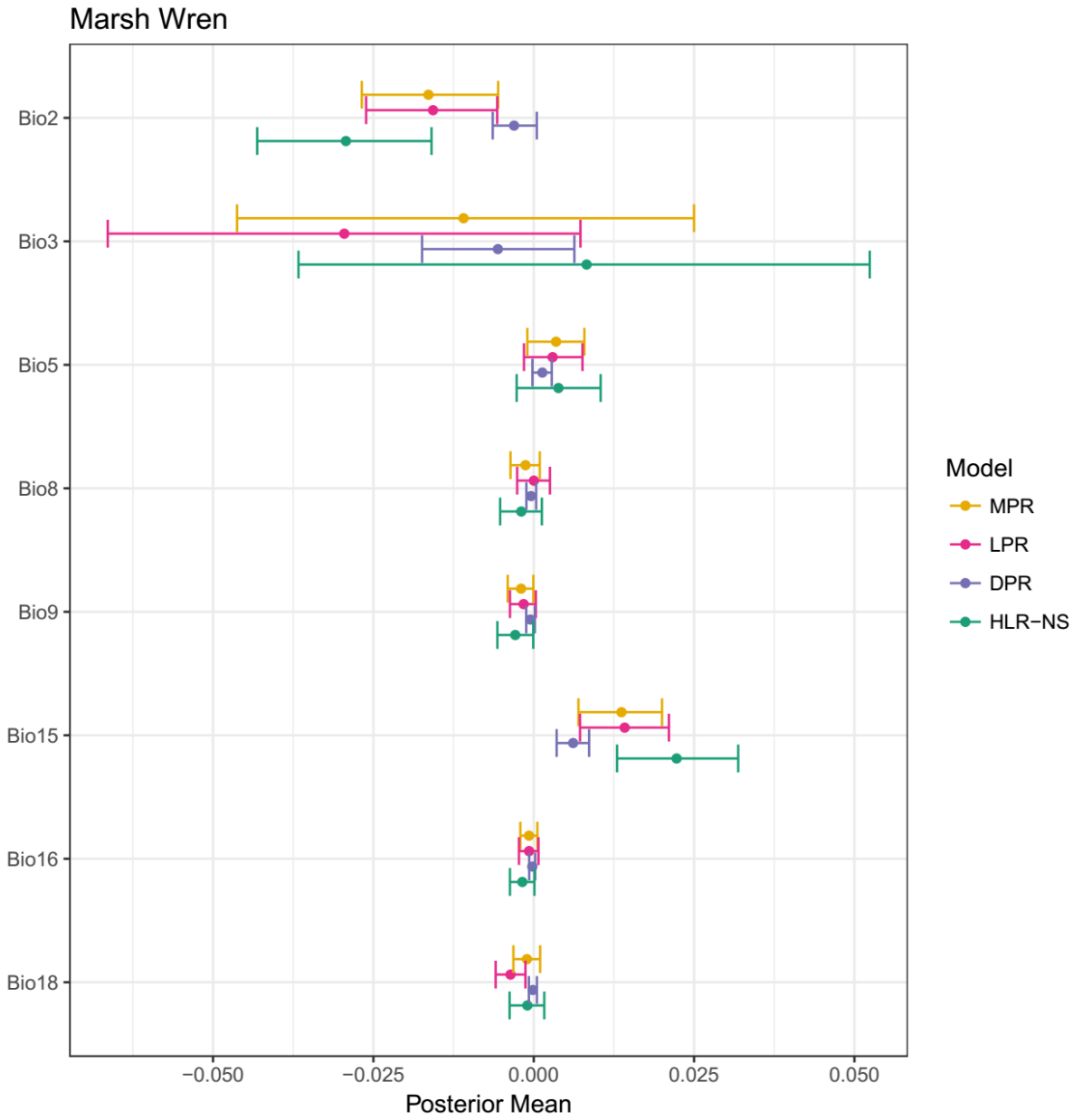


Figure 220: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the marsh wren, *Cistothorus palustris*.

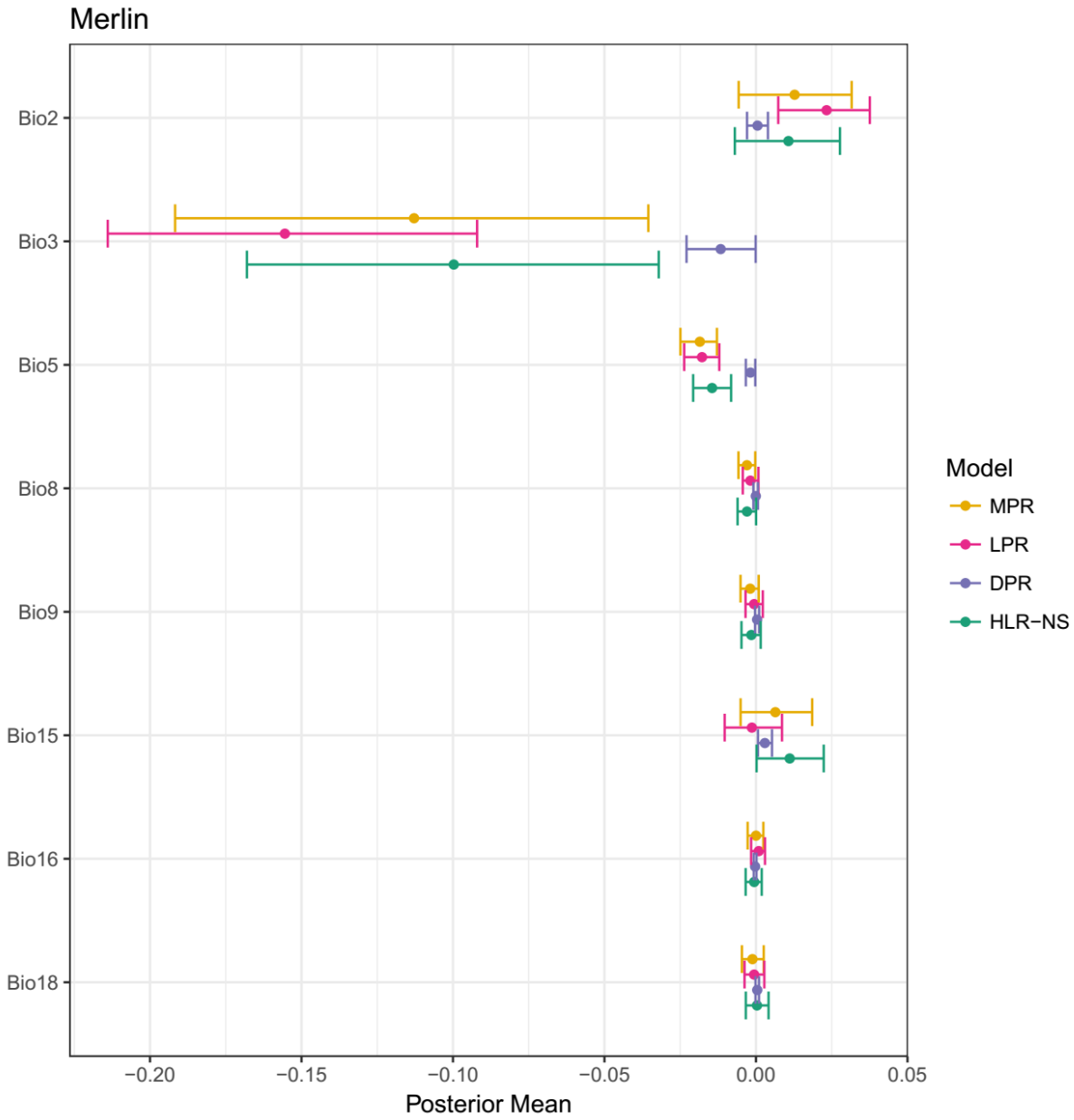


Figure 221: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Merlin, *Falco columbarius*.

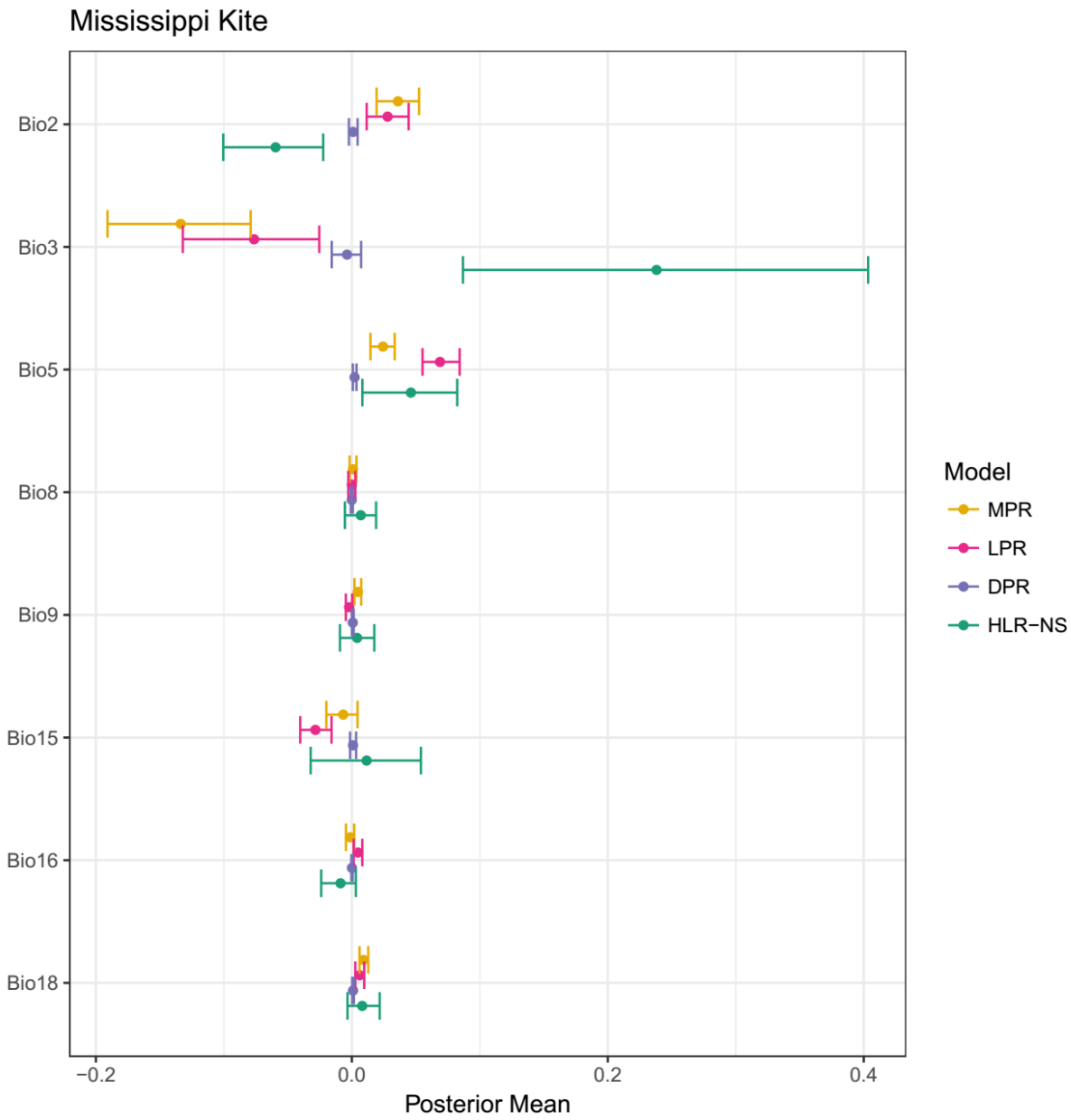


Figure 222: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Mississippi kite, *Ictinia mississippiensis*.

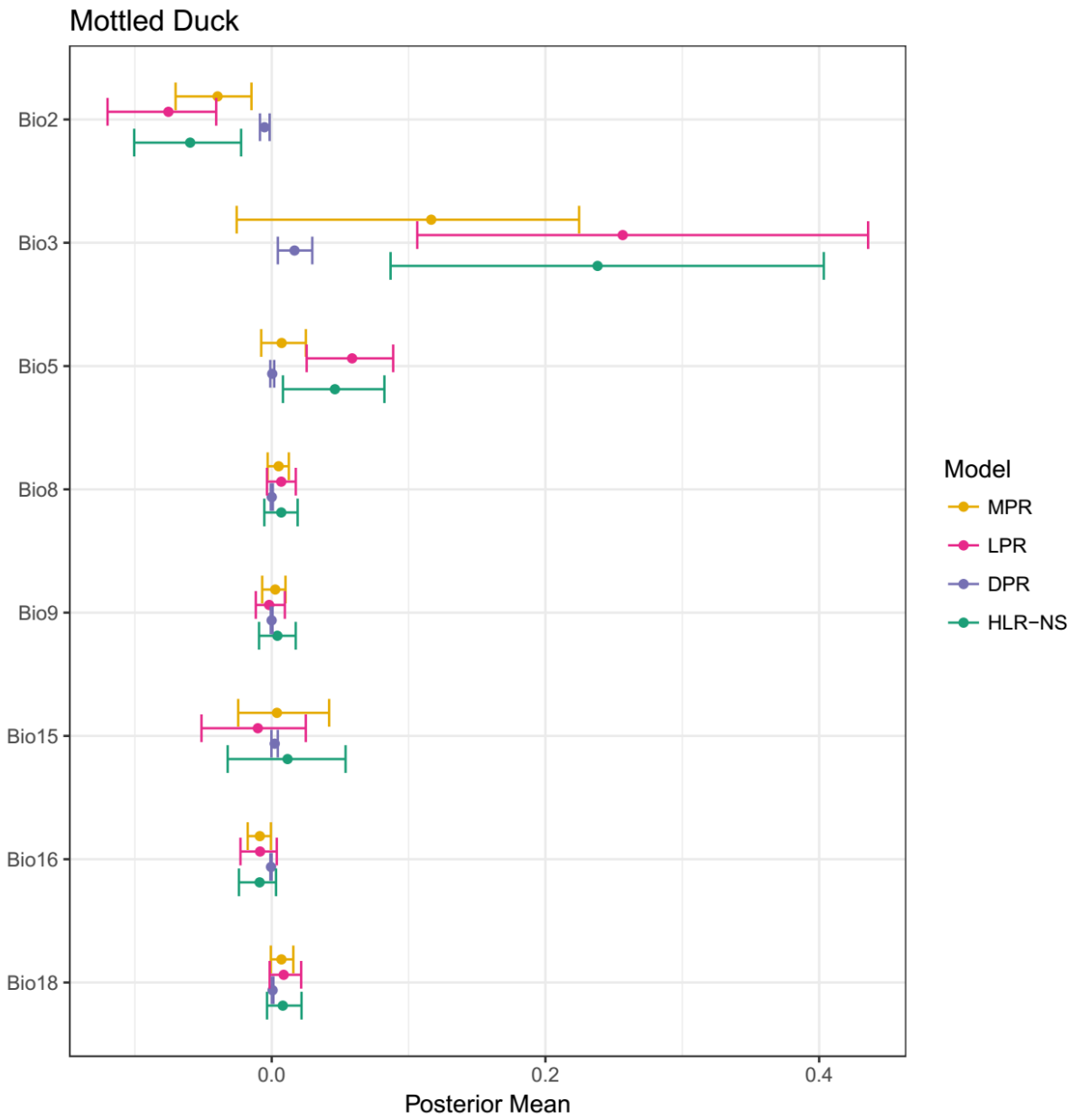


Figure 223: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mottled duck, *Anas fulvigula*.

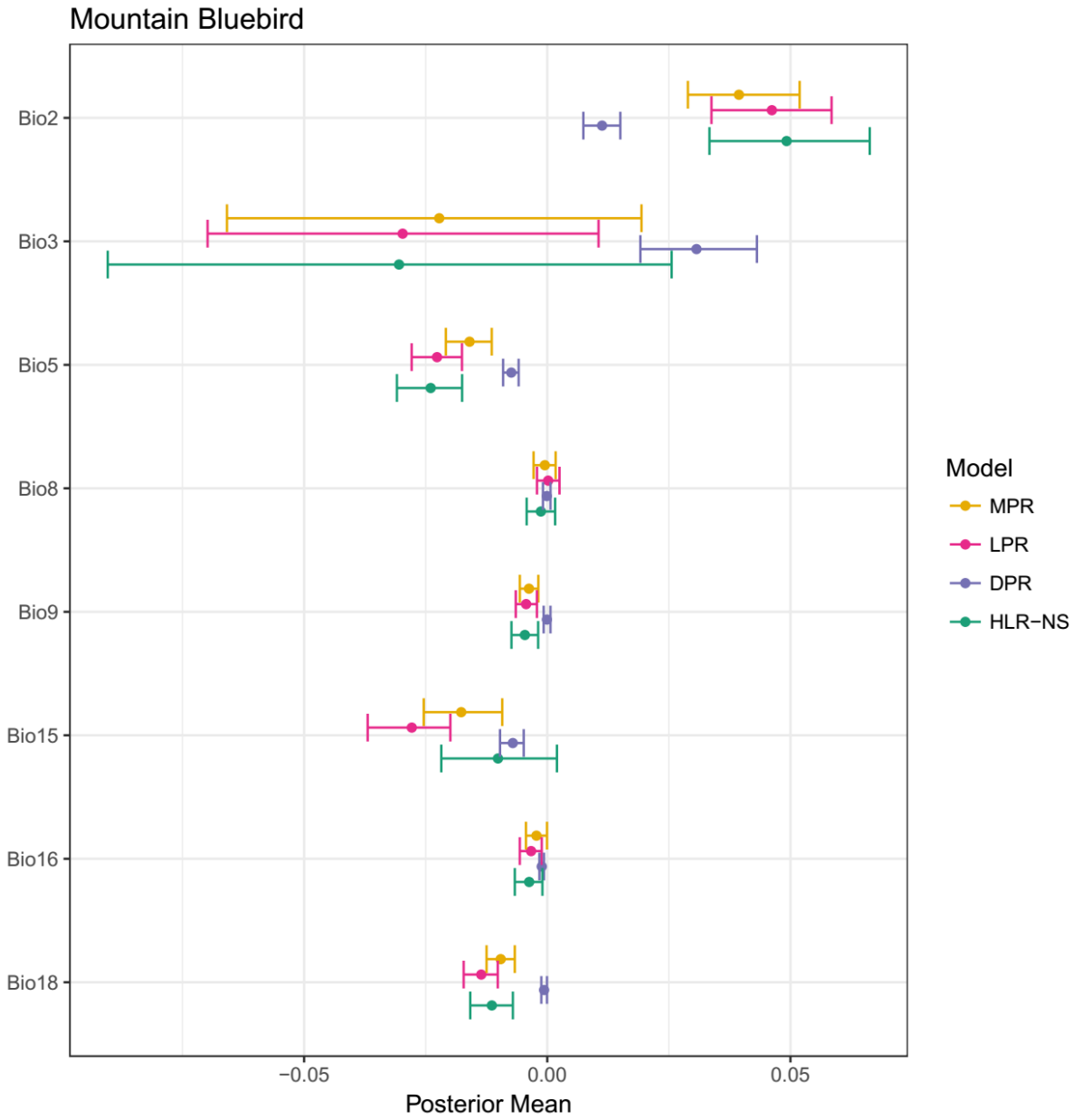


Figure 224: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mountain bluebird, *Sialia currucoides*.

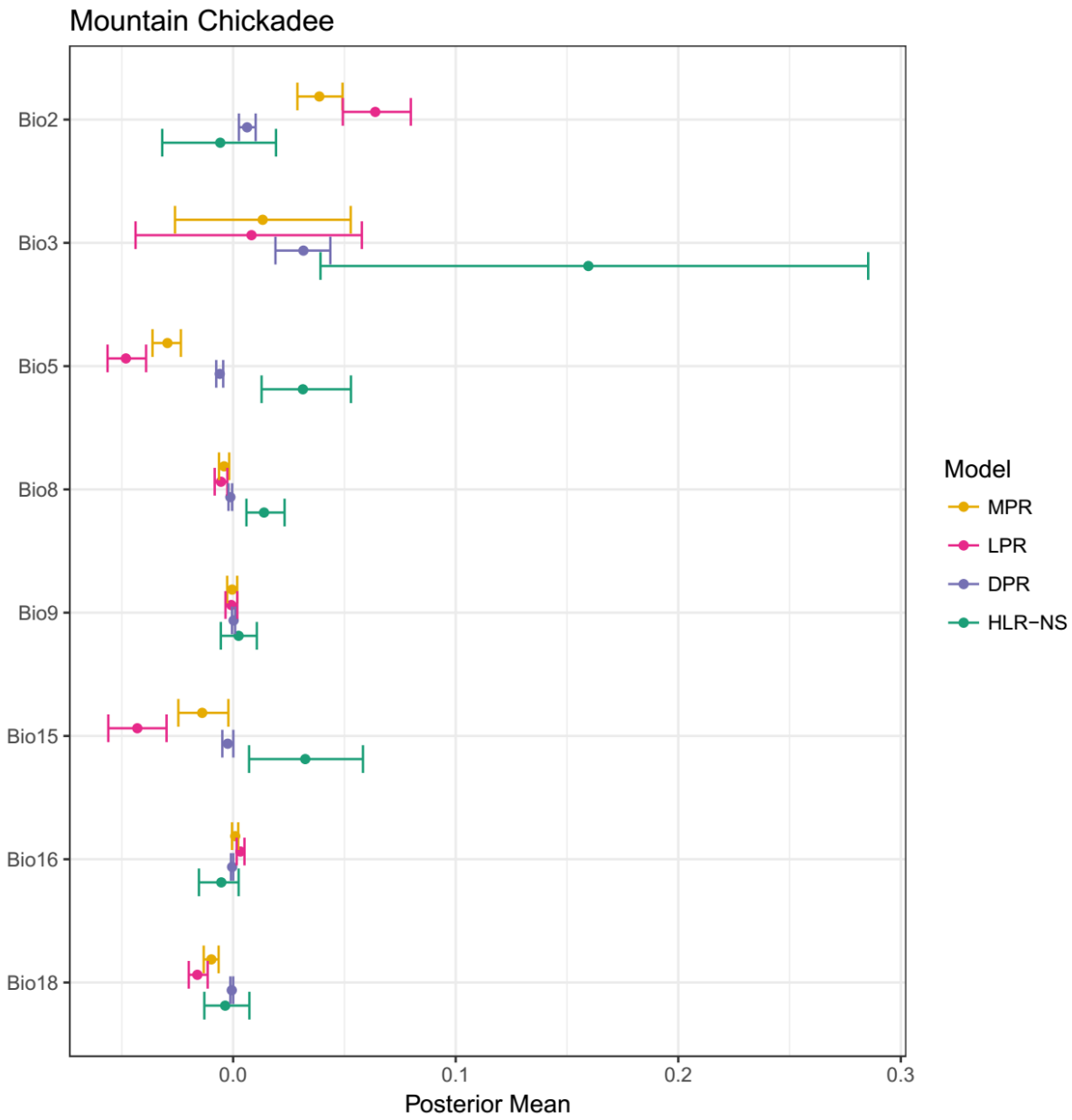


Figure 225: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mountain chickadee, *Poecile gambeli*.

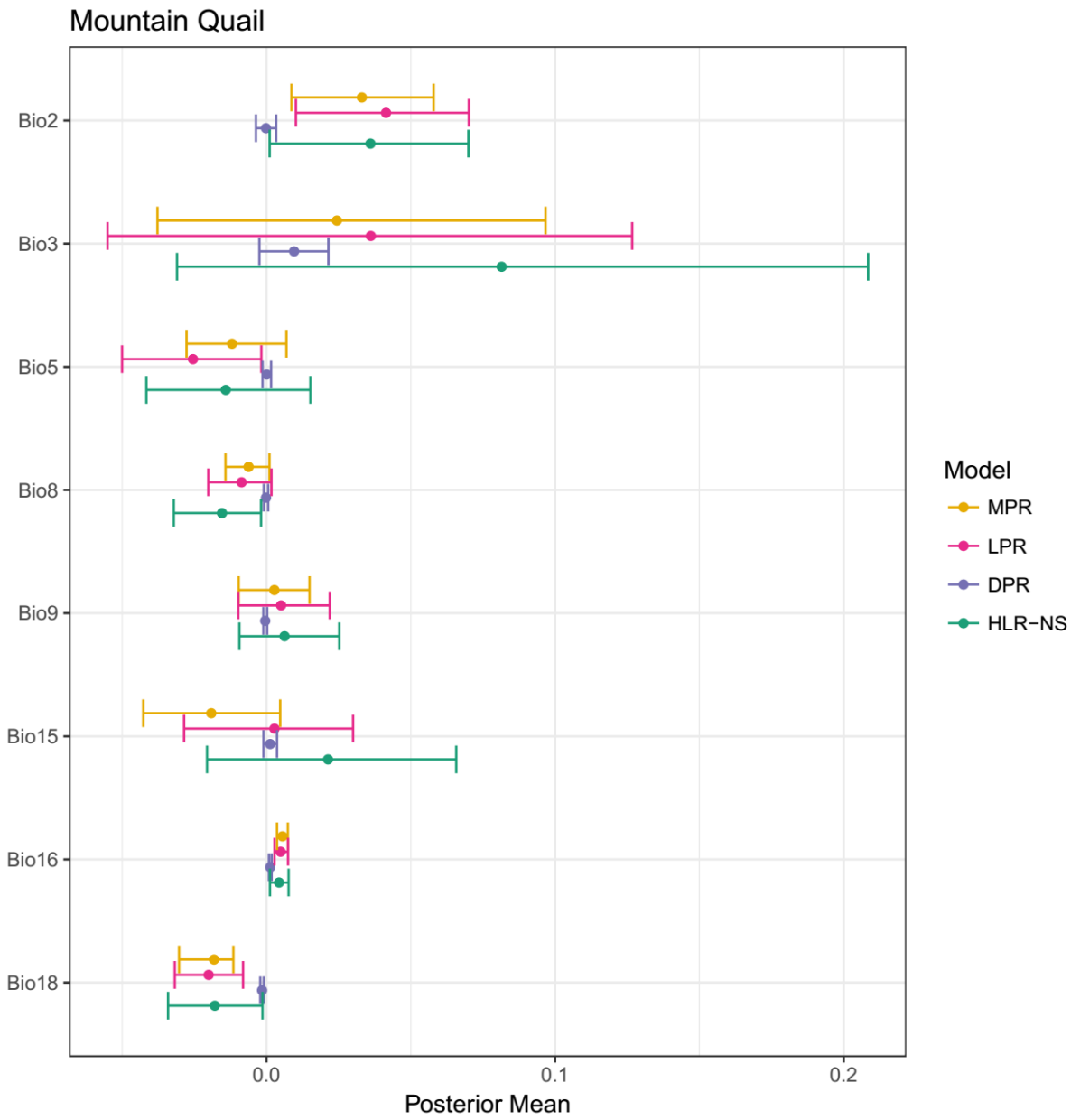


Figure 226: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mountain quail, *Oreortyx pictus*.

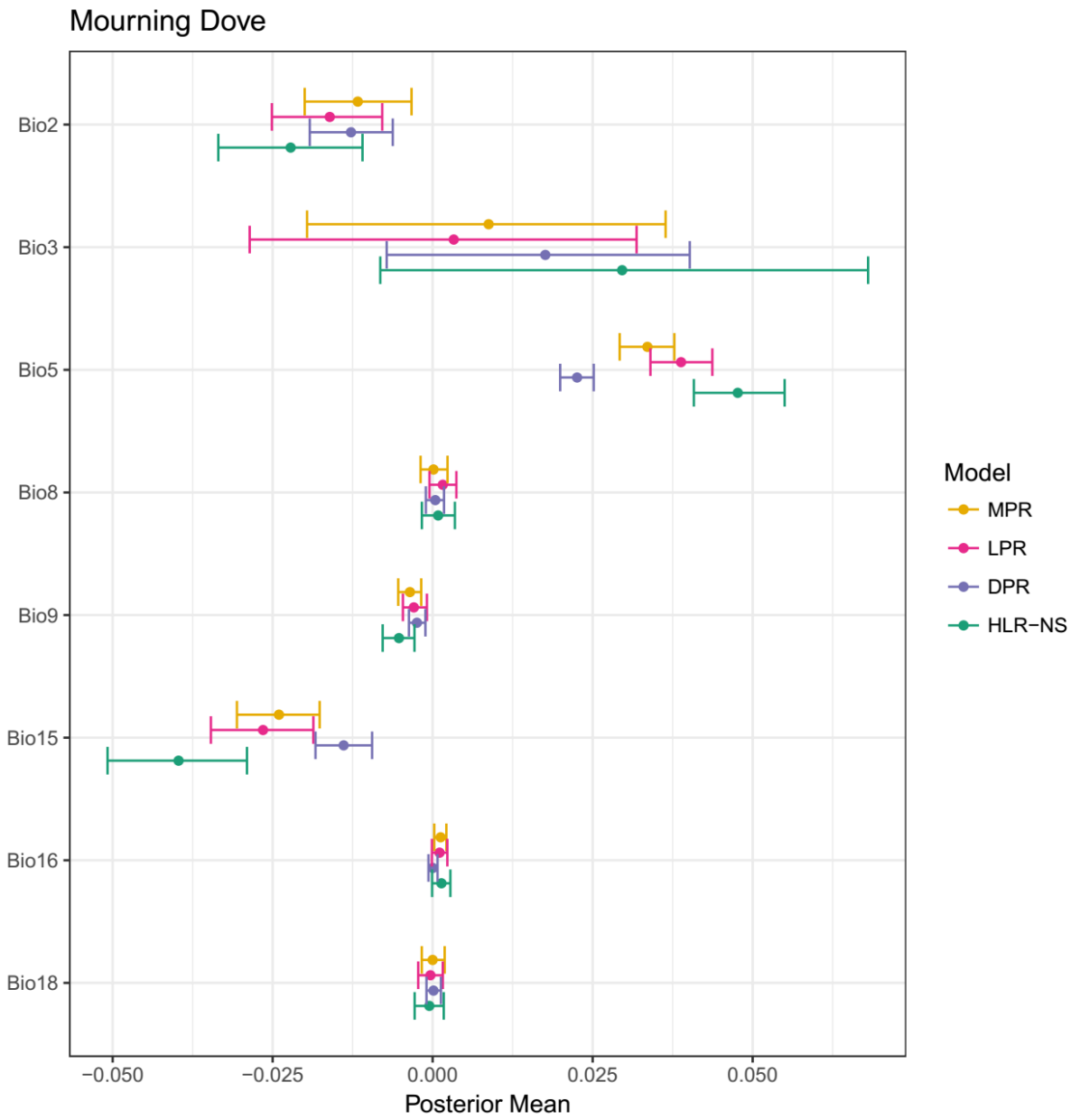


Figure 227: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mourning dove, *Zenaida macroura*.

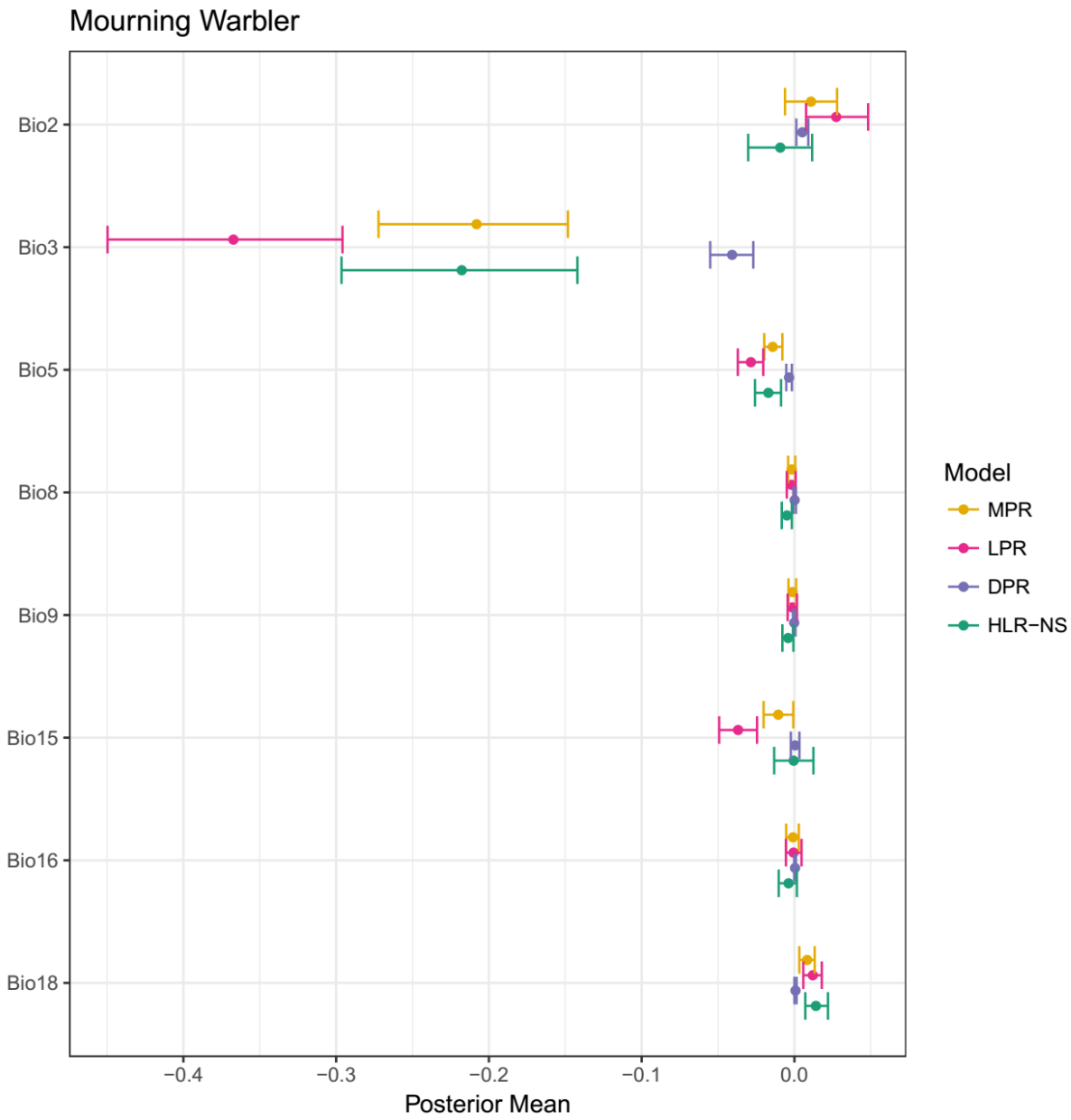


Figure 228: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mourning warbler, *Geothlypis philadelphia*.

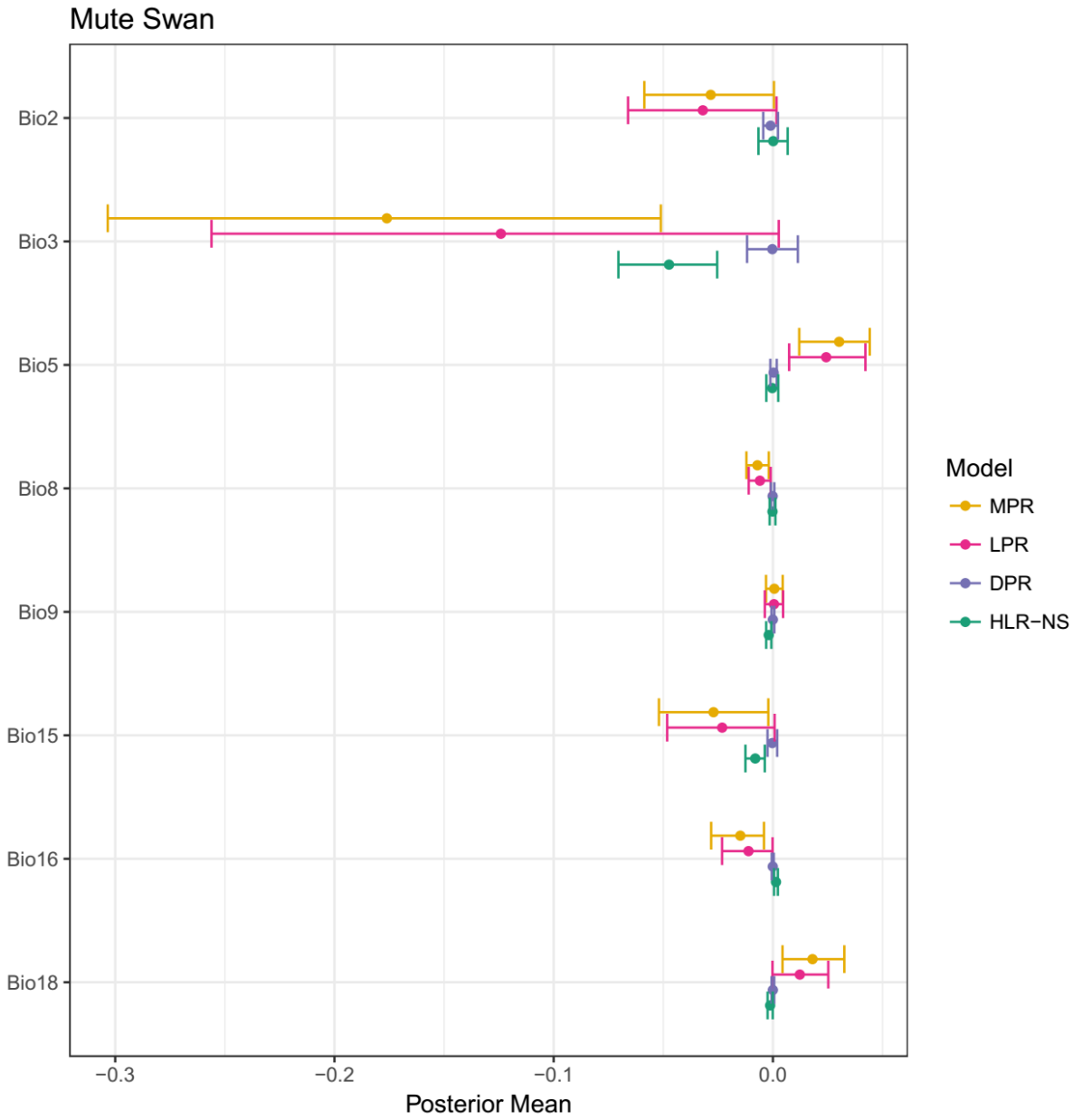


Figure 229: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Mute swan, *Cygnus olor*.

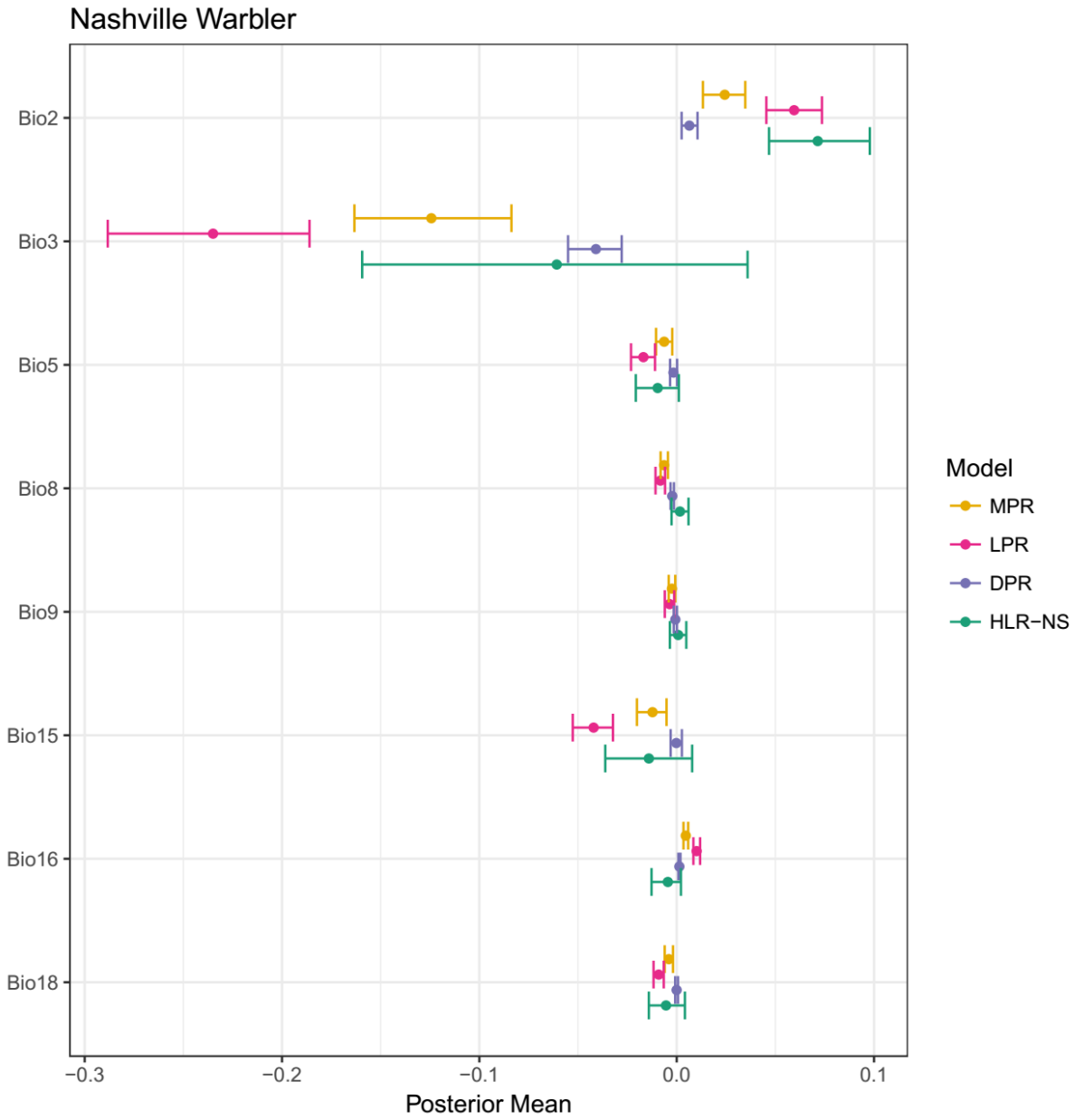


Figure 230: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Nashville warbler, *Leiothlypis ruficapilla*.

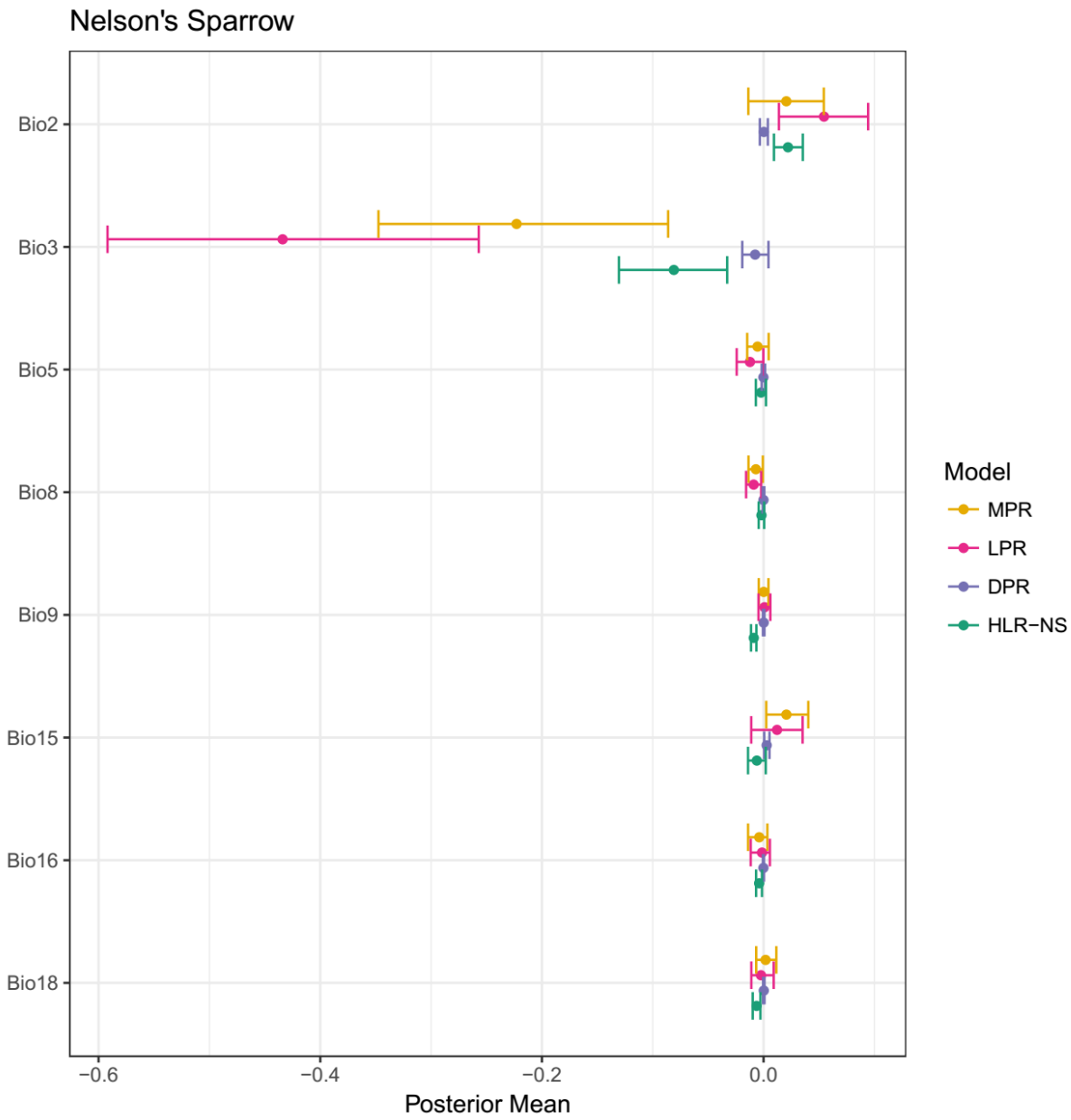


Figure 231: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Nelson's sparrow, *Ammodramus nelsoni*.

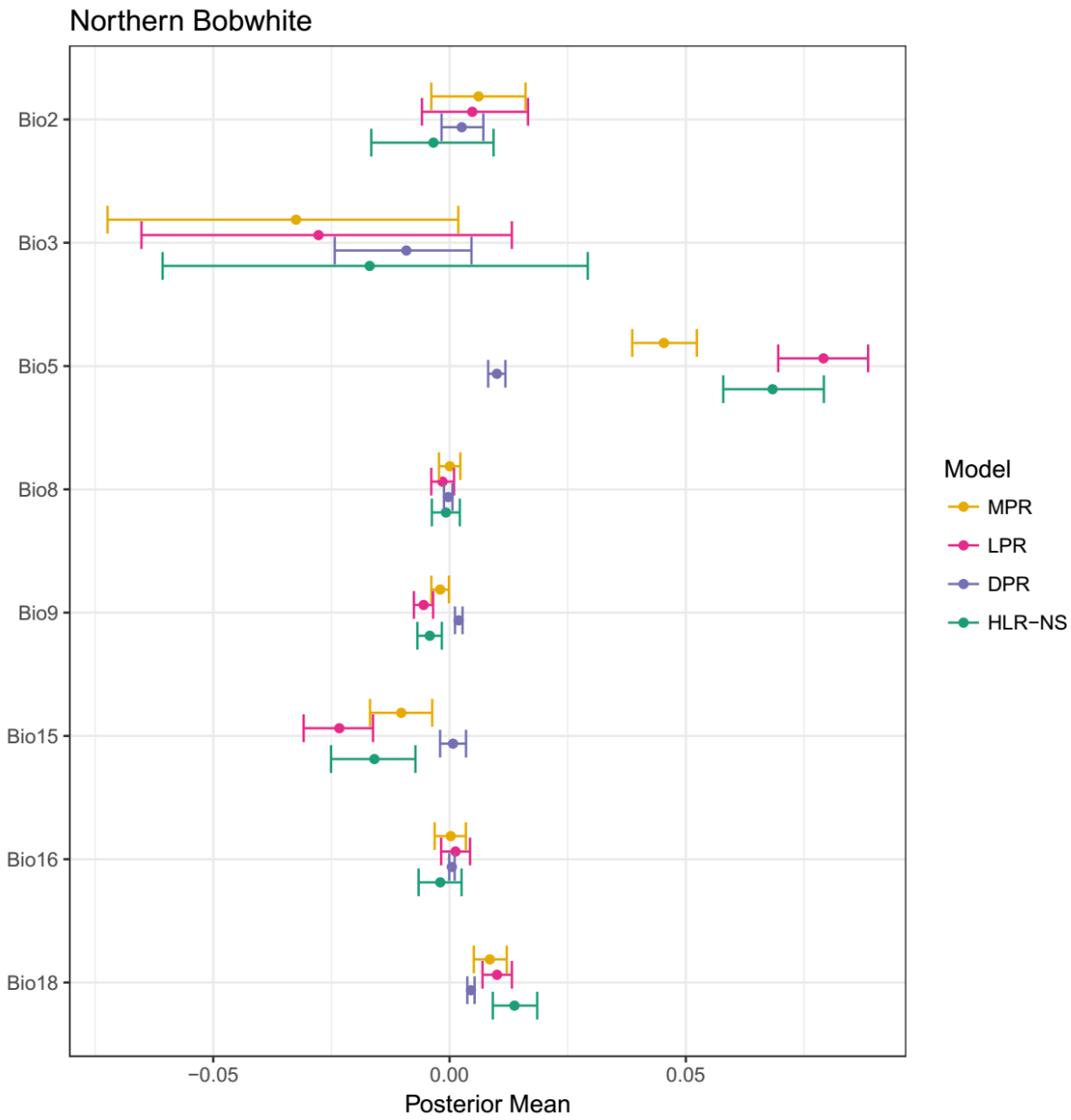


Figure 232: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern bobwhite, *Colinus virginianus*.

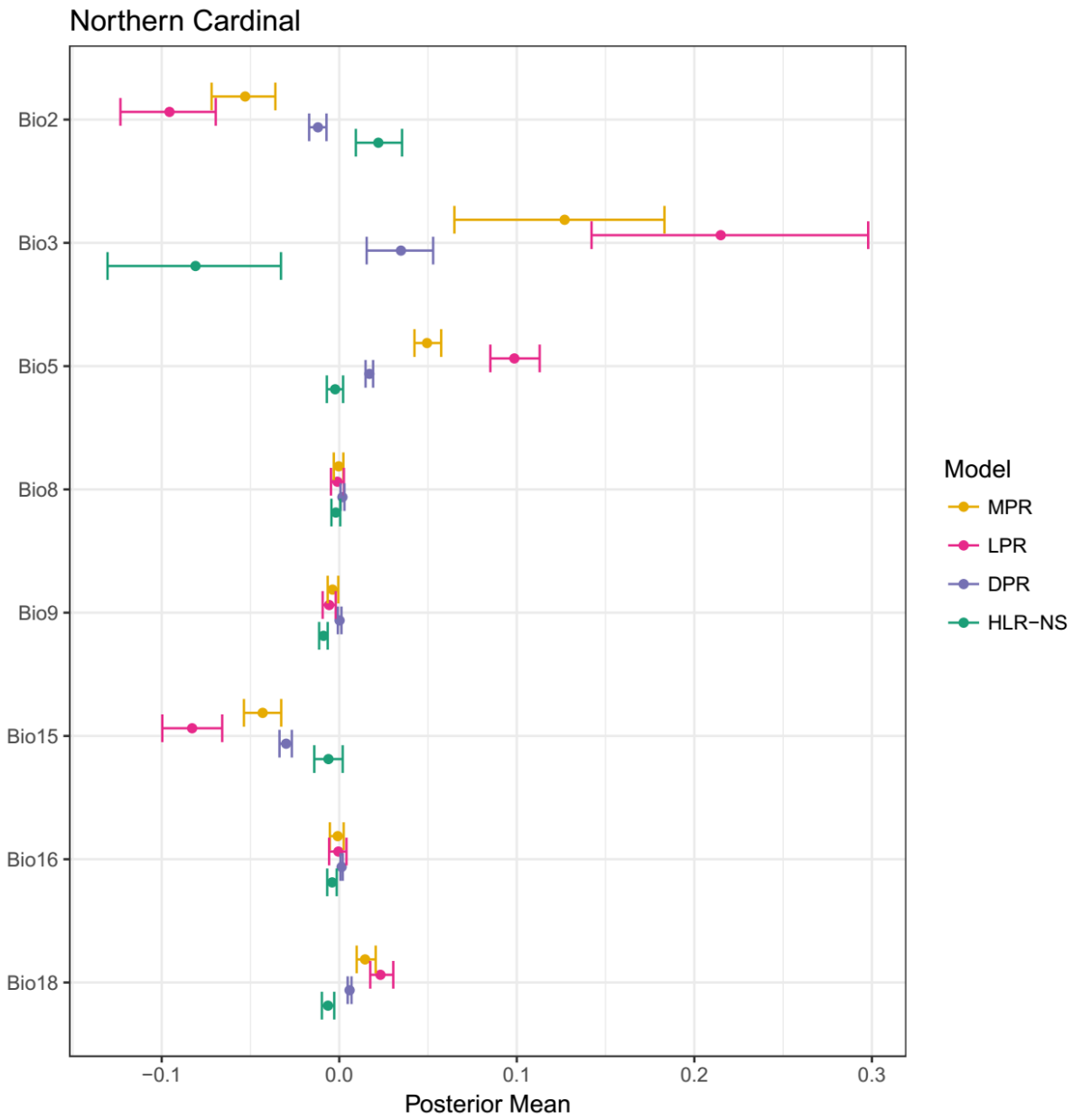


Figure 233: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern cardinal, *Cardinalis cardinalis*.

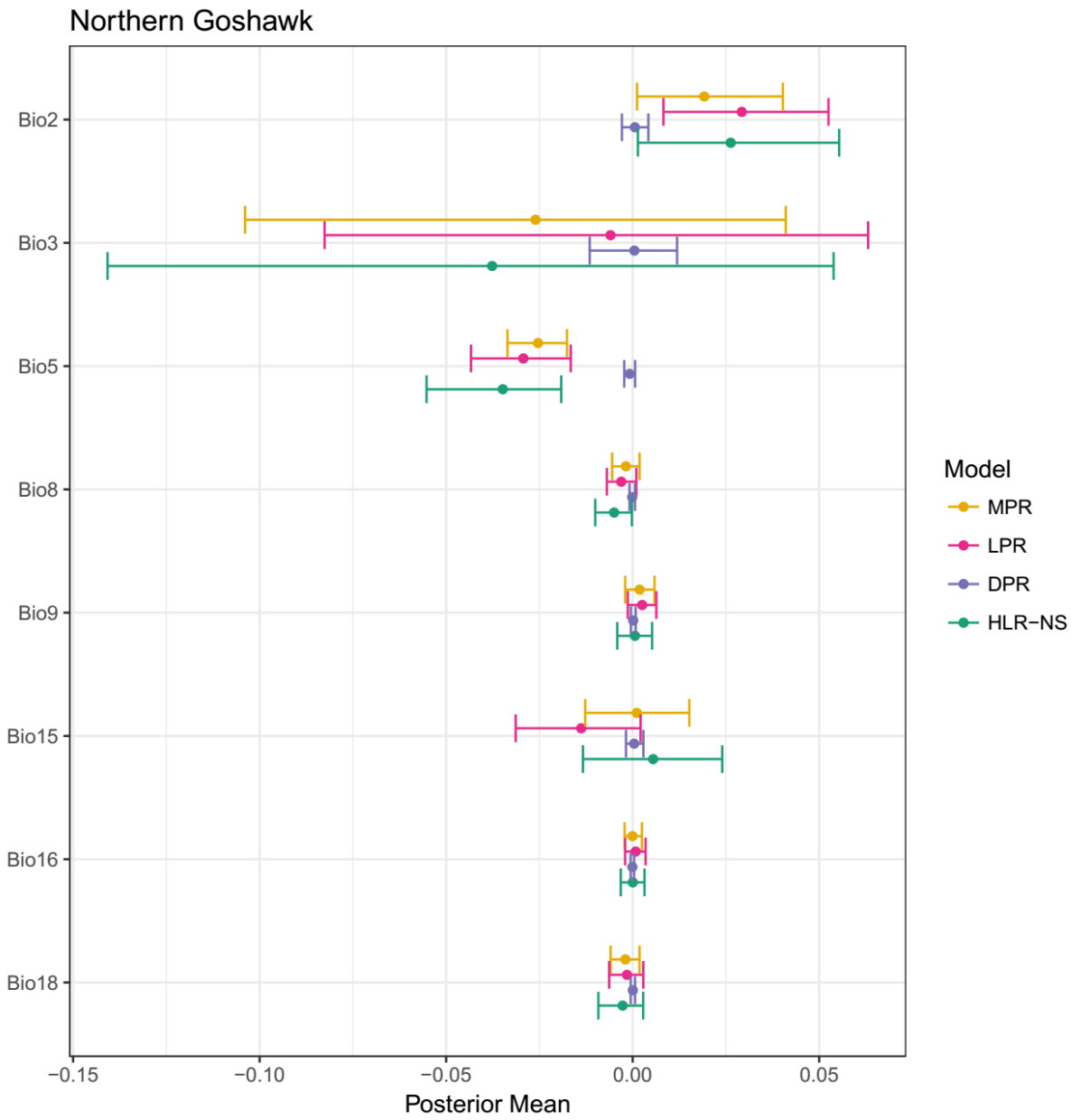


Figure 234: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern goshawk, *Accipiter gentilis*.

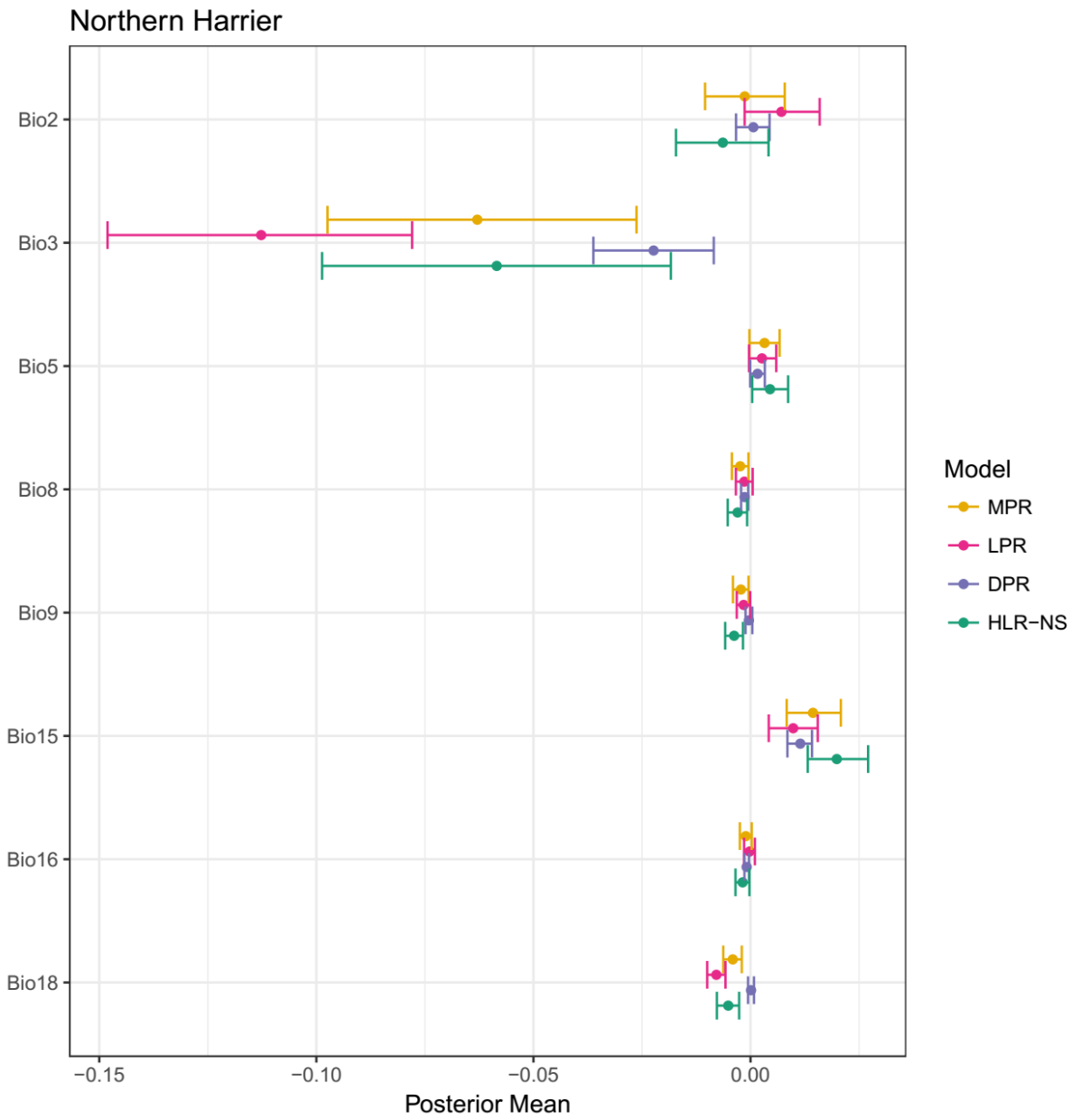


Figure 235: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern harrier, *Circus cyaneus*.

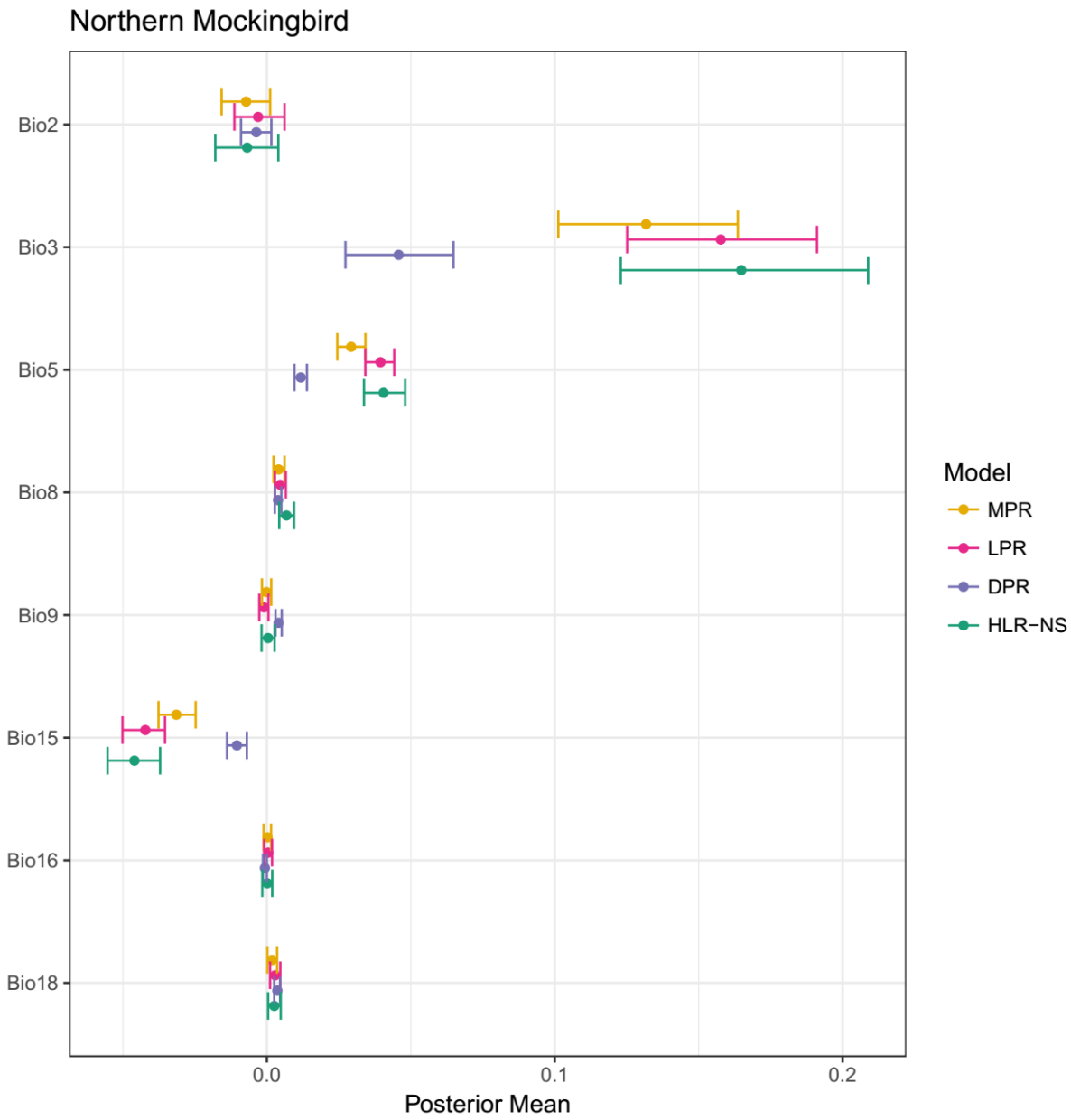


Figure 236: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern mockingbird, *Mimus polyglottos*.

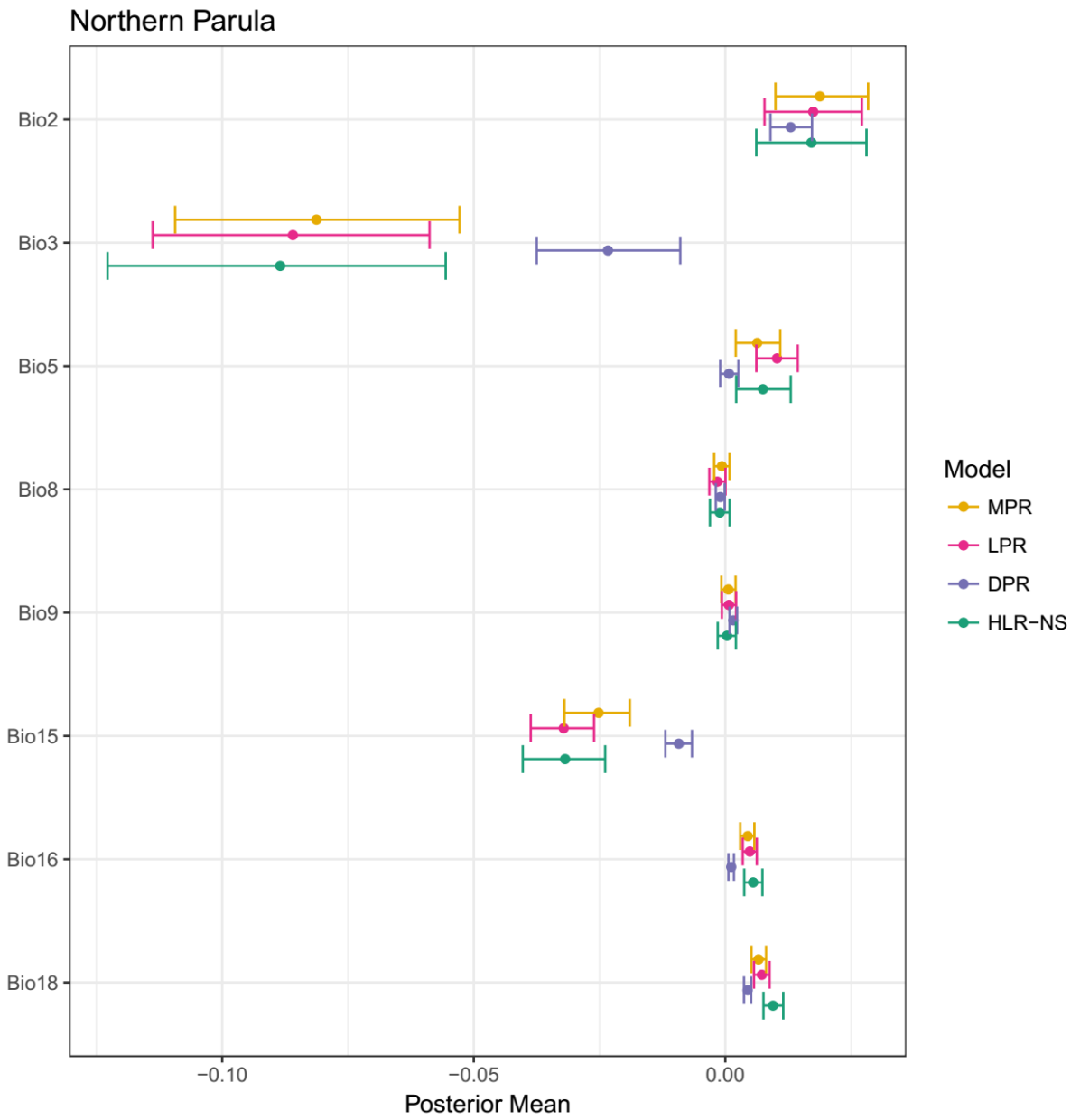


Figure 237: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern parula, *Setophaga americana*.

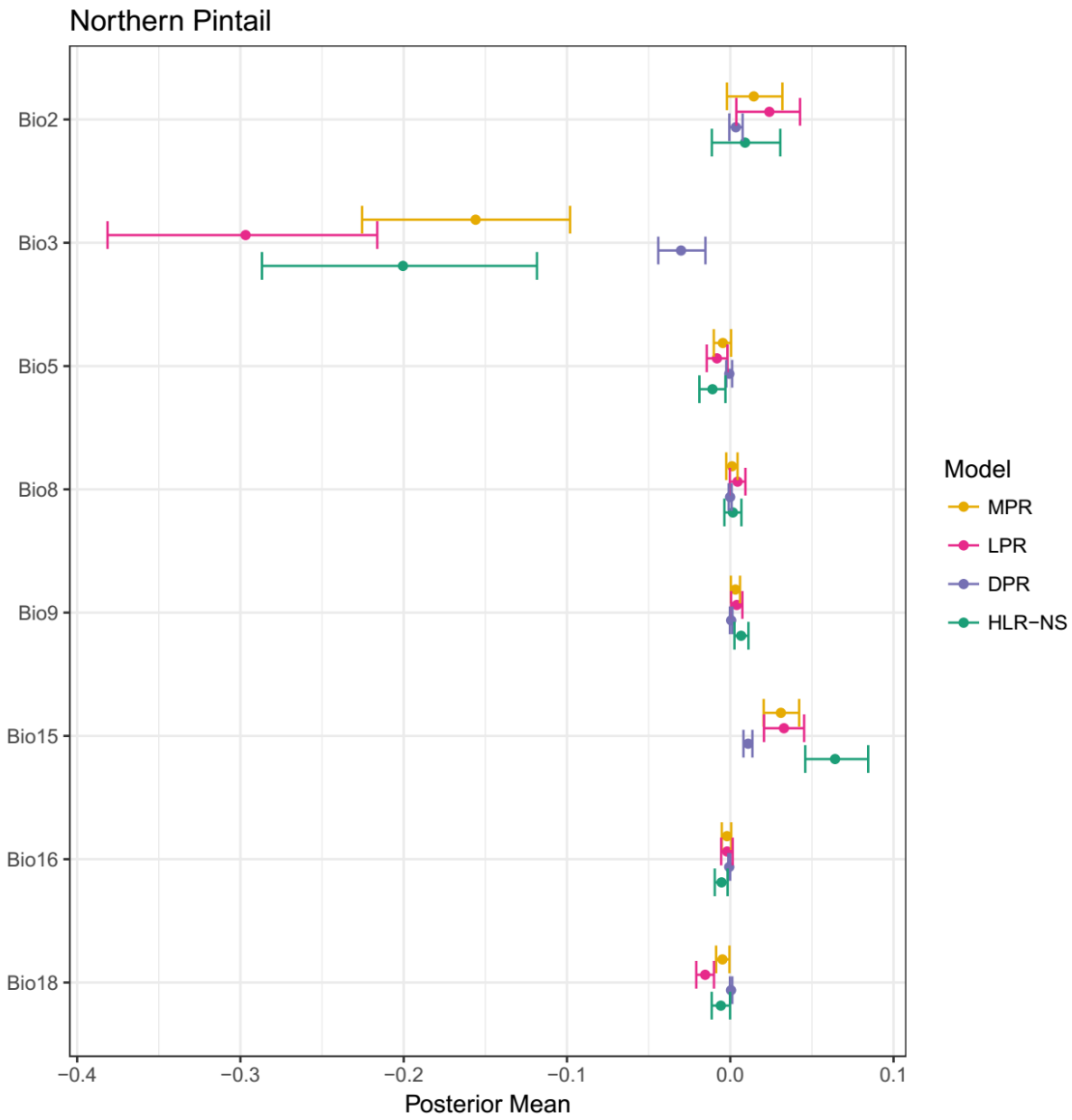


Figure 238: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern pintail, *Anas acuta*.

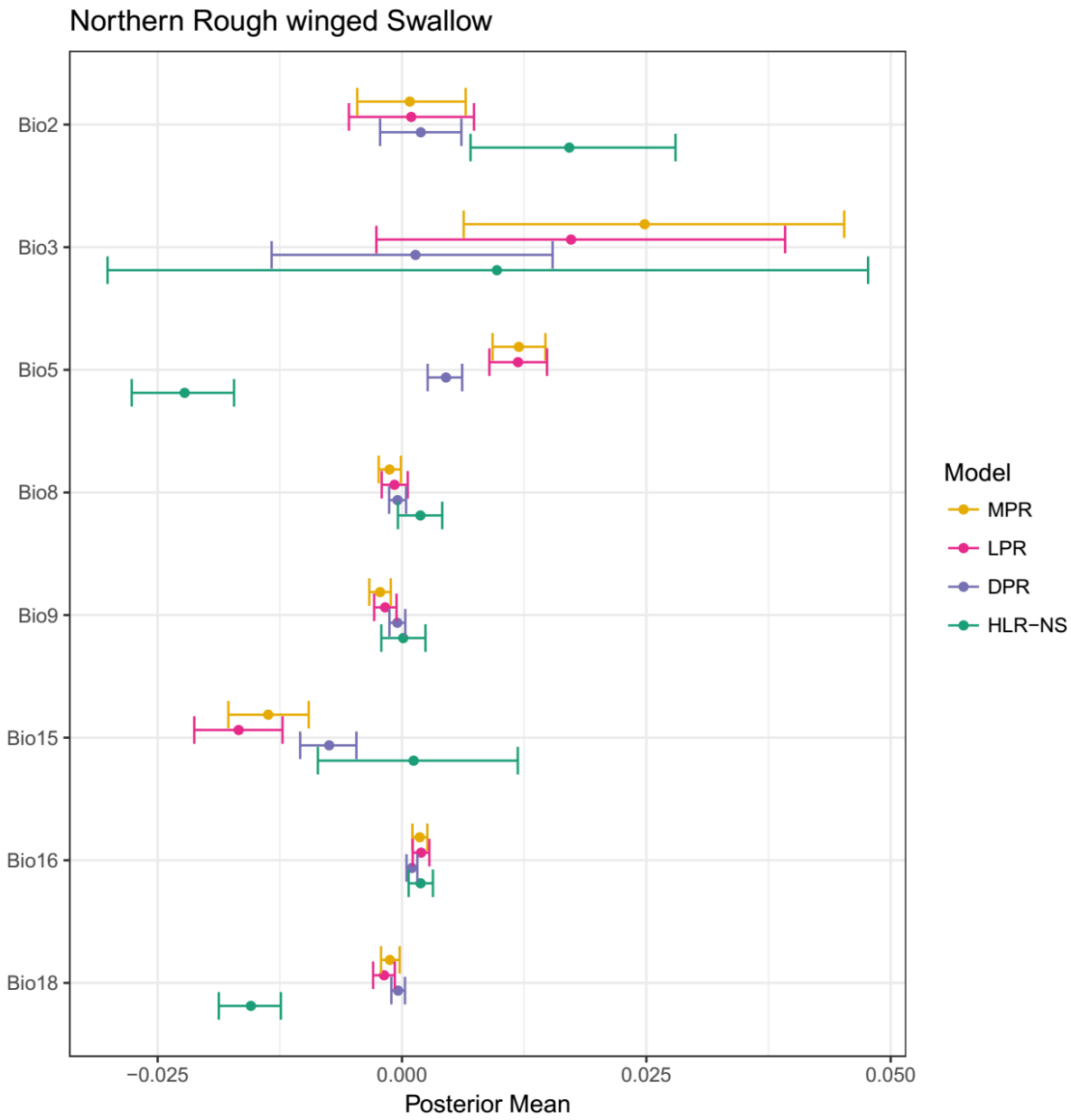


Figure 239: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern rough-winged swallow, *Stelgidopteryx serripennis*.

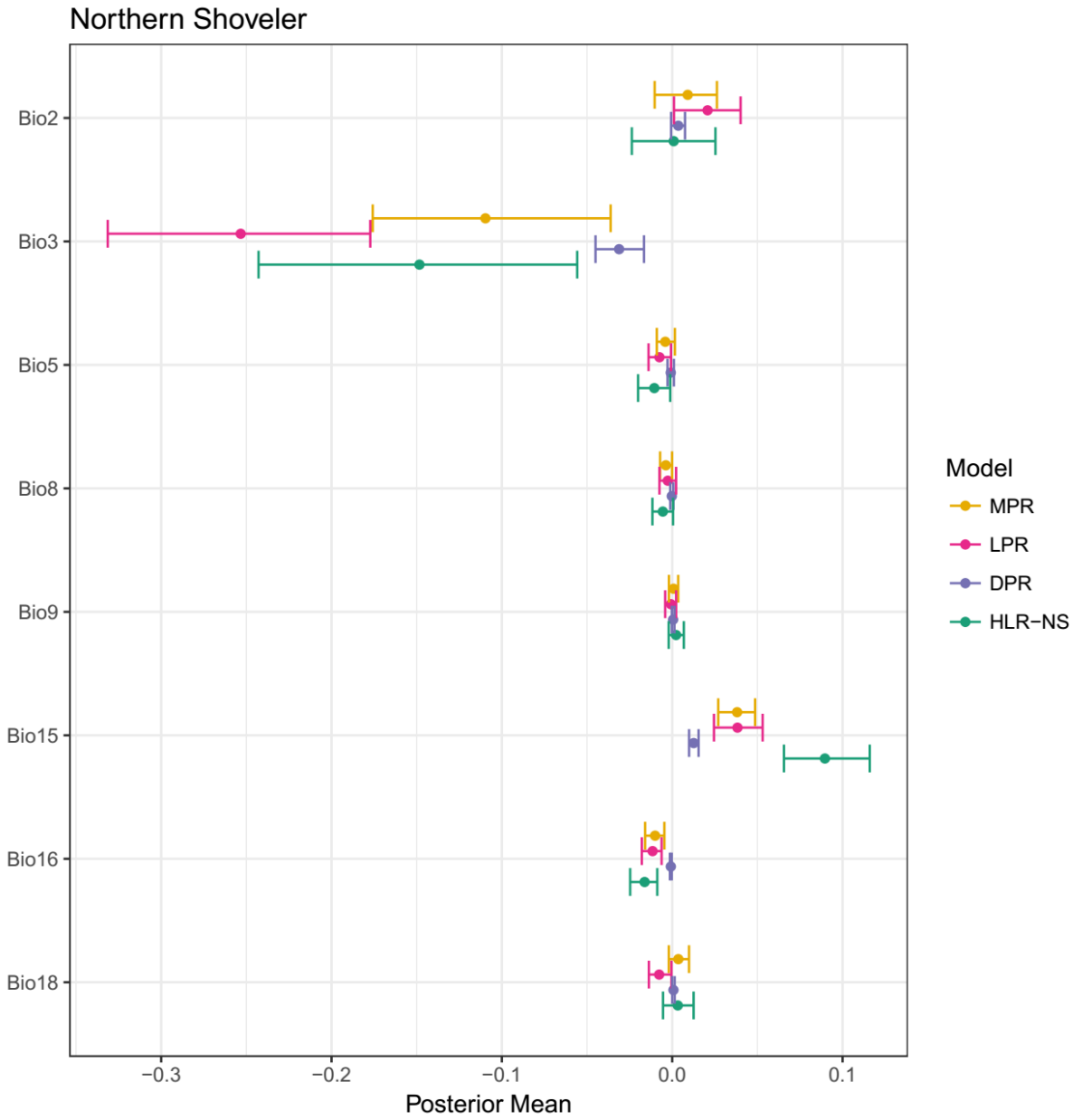


Figure 240: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern shoveler, *Anas clypeata*.

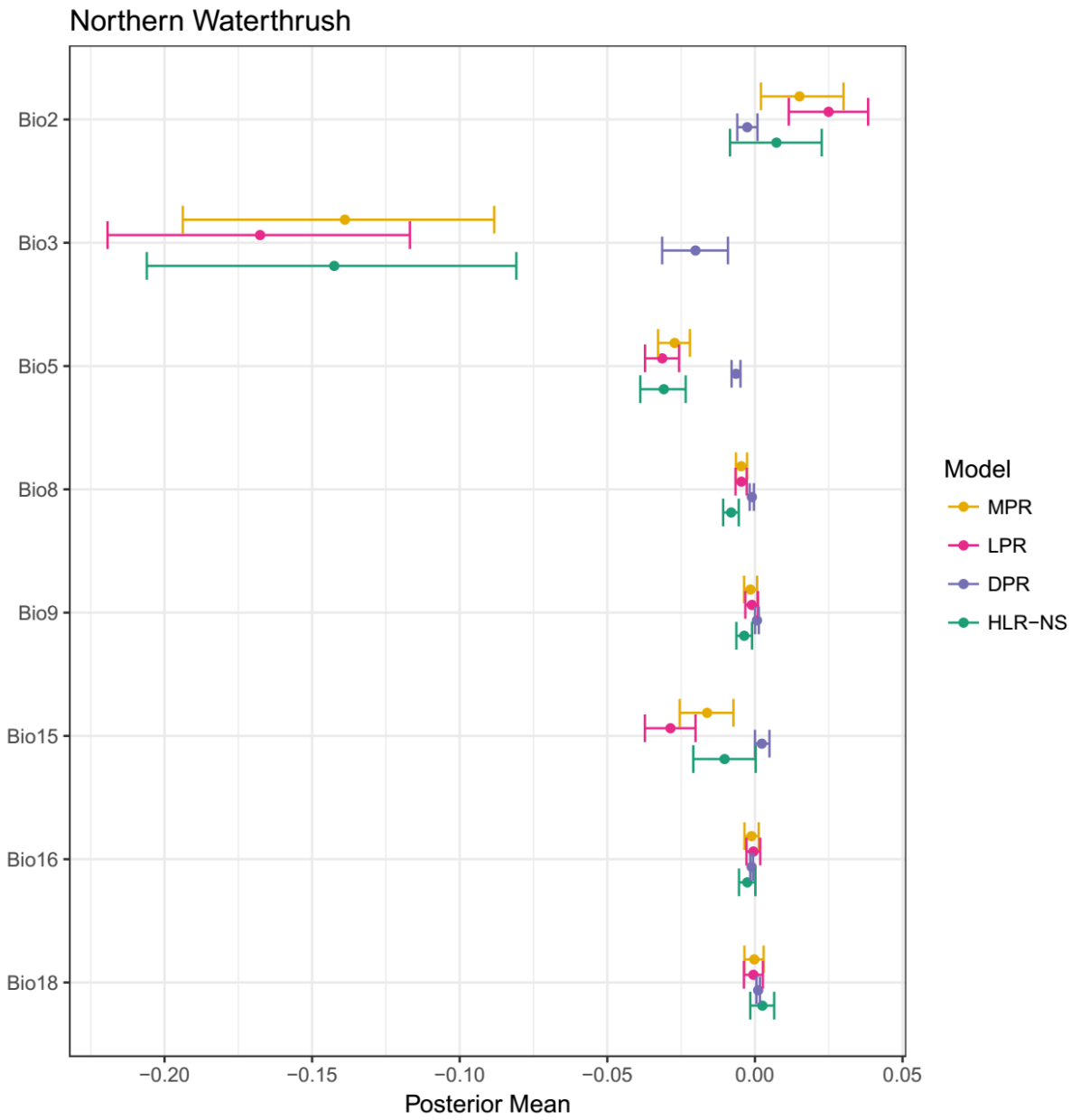


Figure 241: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern waterthrush, *Parkesia noveboracensis*.

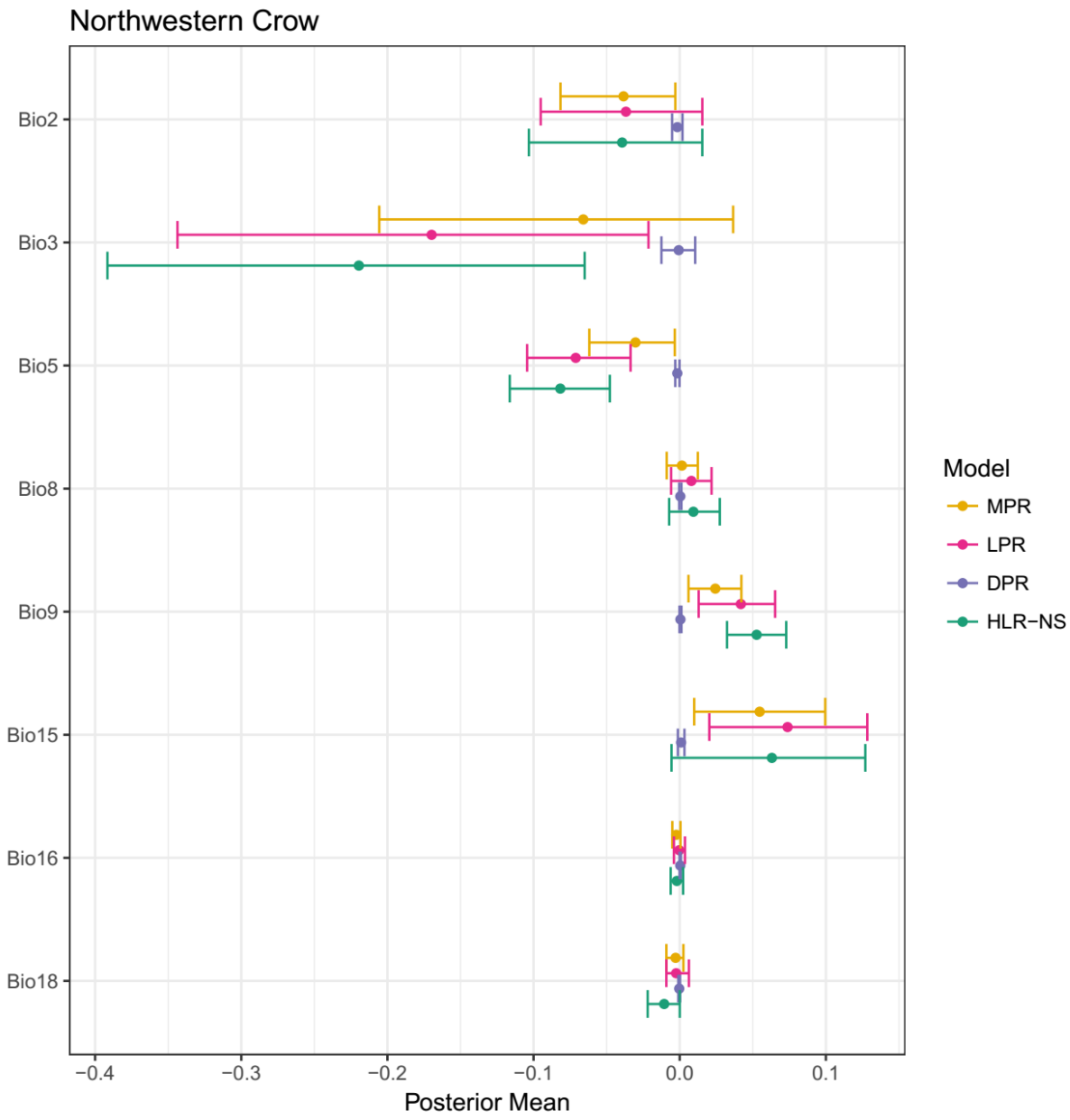


Figure 242: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northwestern crow, *Corvus caurinus*.

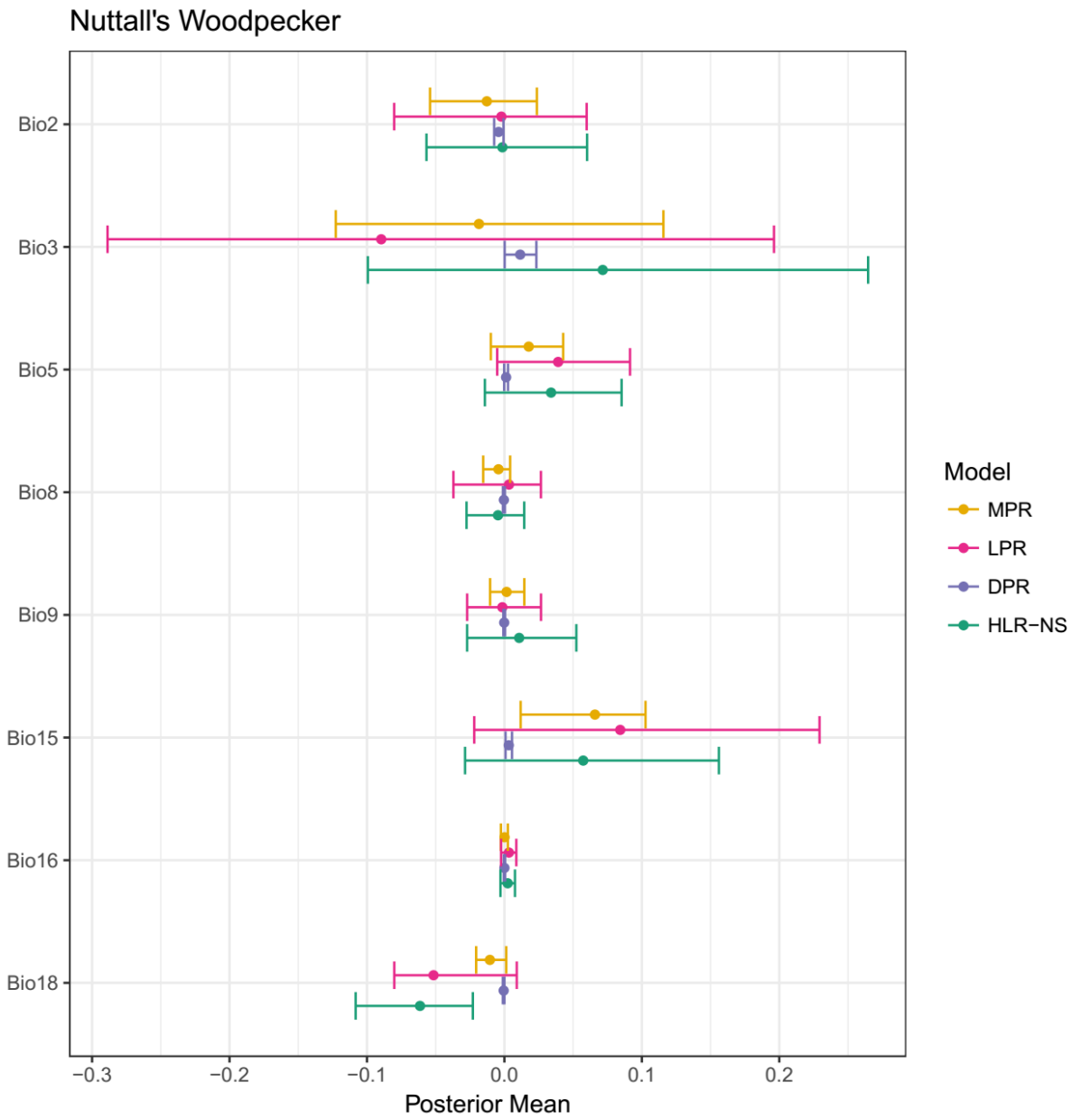


Figure 243: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Nuttall's woodpecker, *Picoides nuttallii*.

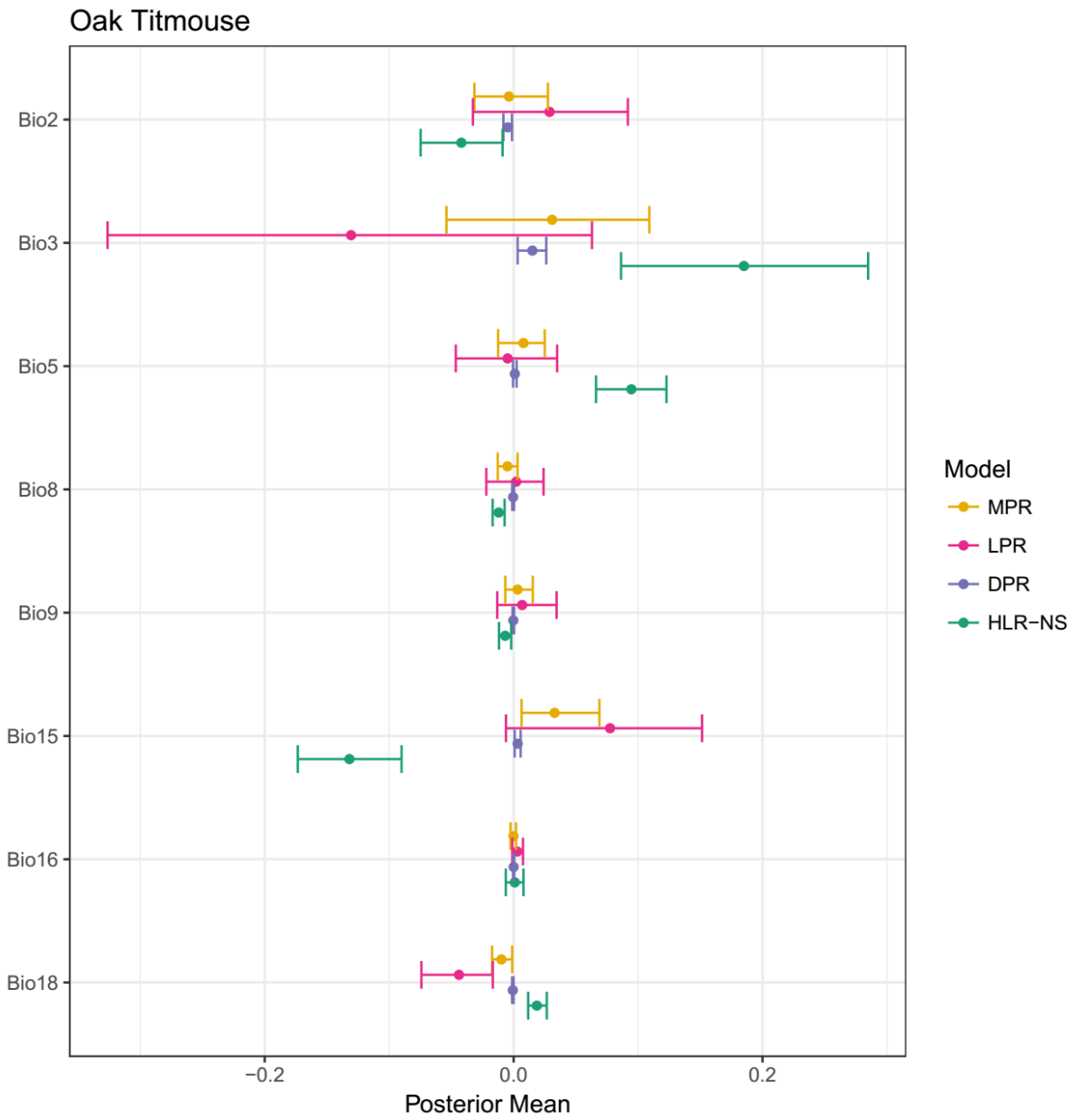


Figure 244: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the oak titmouse, *Baeolophus inornatus*.

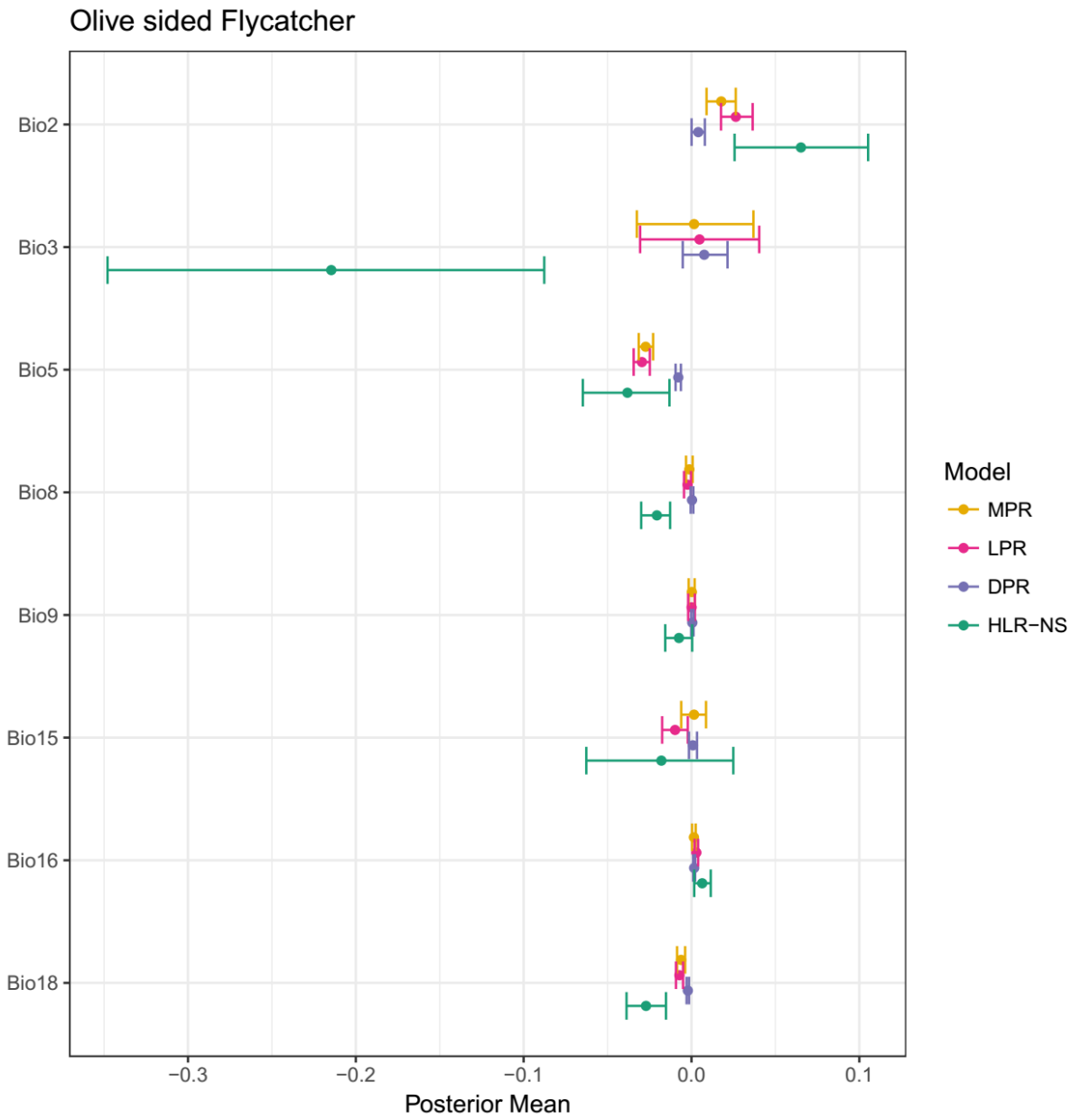


Figure 245: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the olive-sided flycatcher, *Contopus cooperi*.

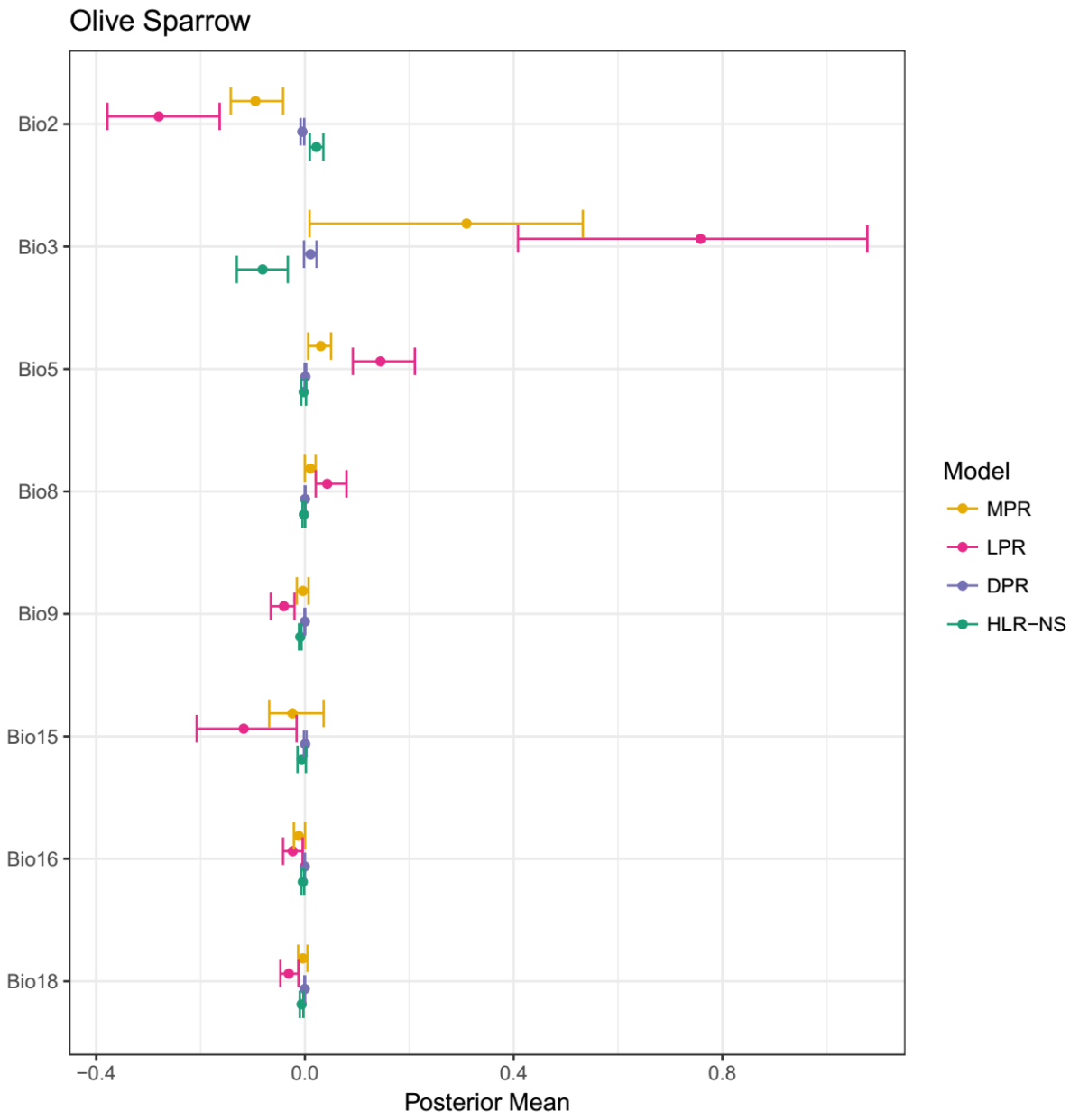


Figure 246: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the olive sparrow, *Arremonops rufivirgatus*.

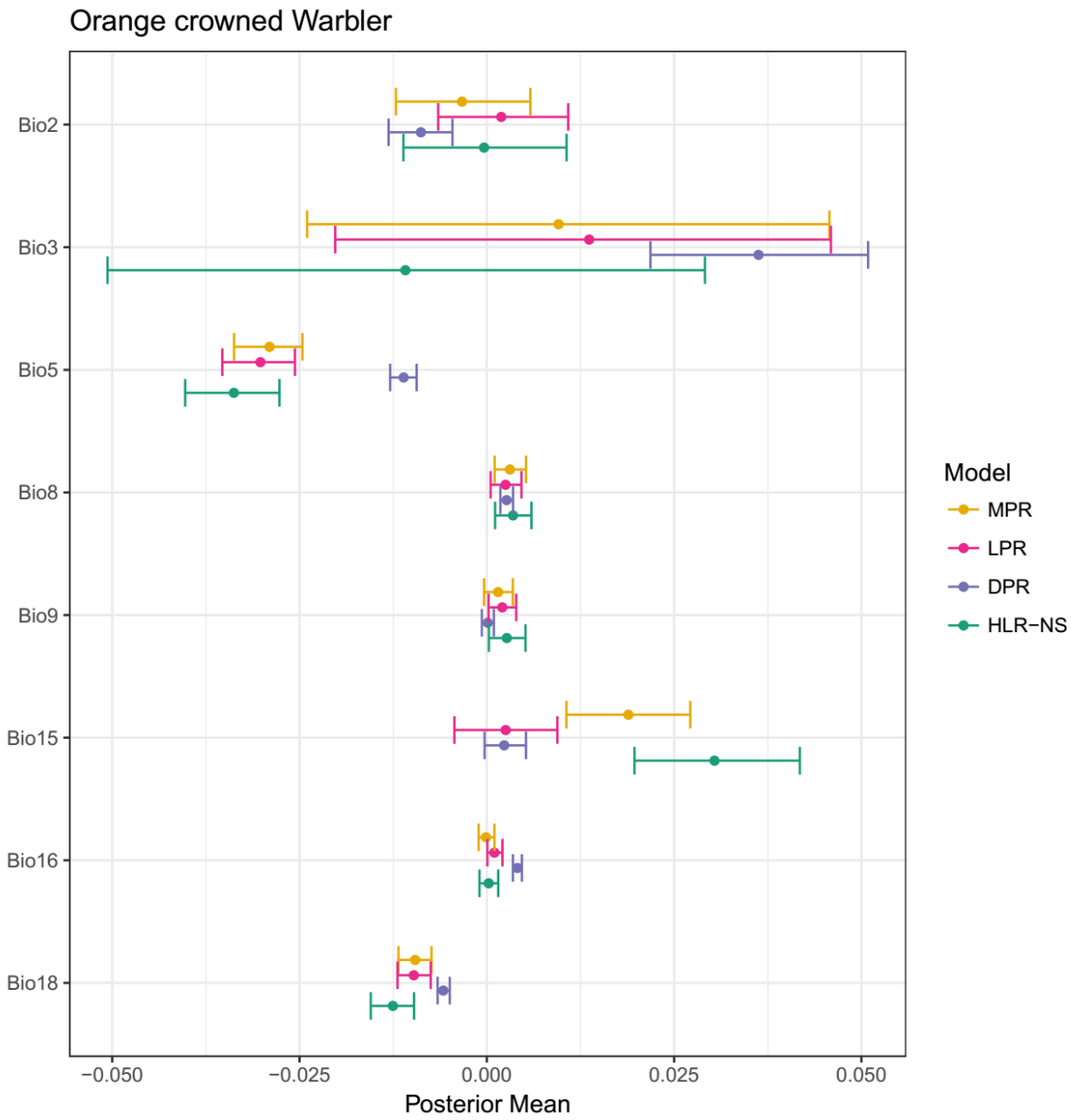


Figure 247: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the orange-crowned warbler, *Vermivora celata*.

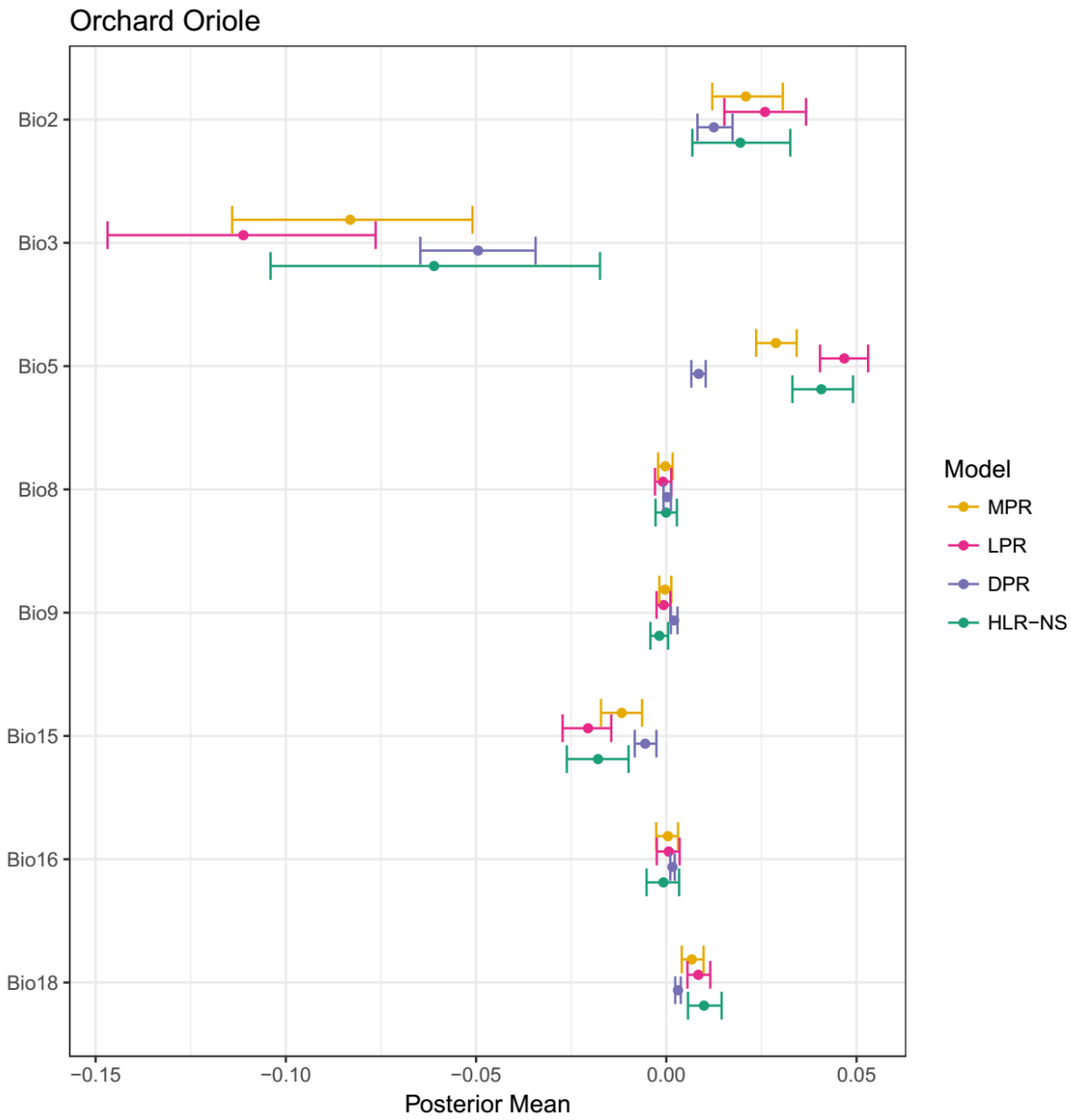


Figure 248: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the orchard oriole, *Icterus spurius*.

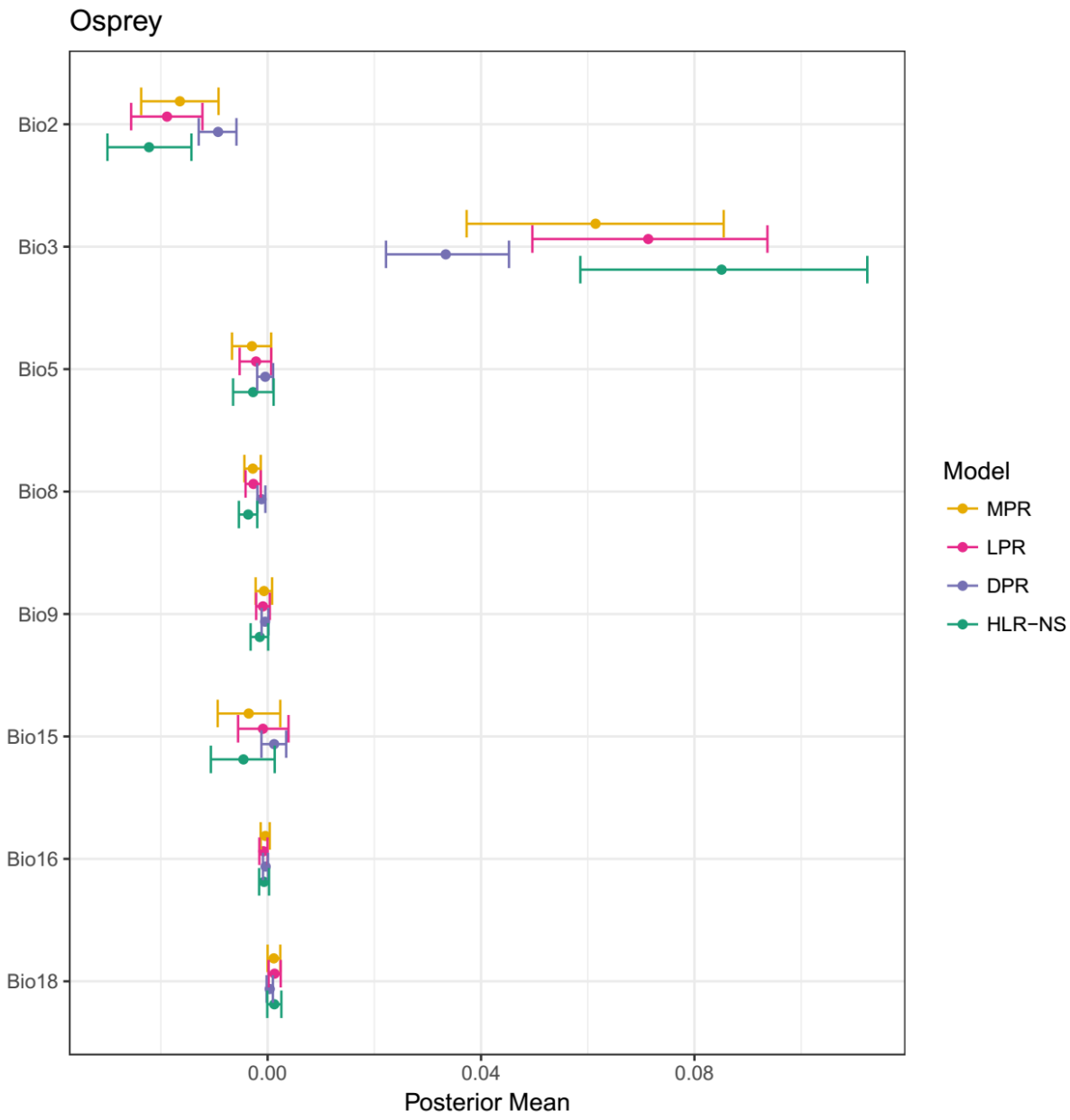


Figure 249: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the osprey, *Pandion haliaetus*.

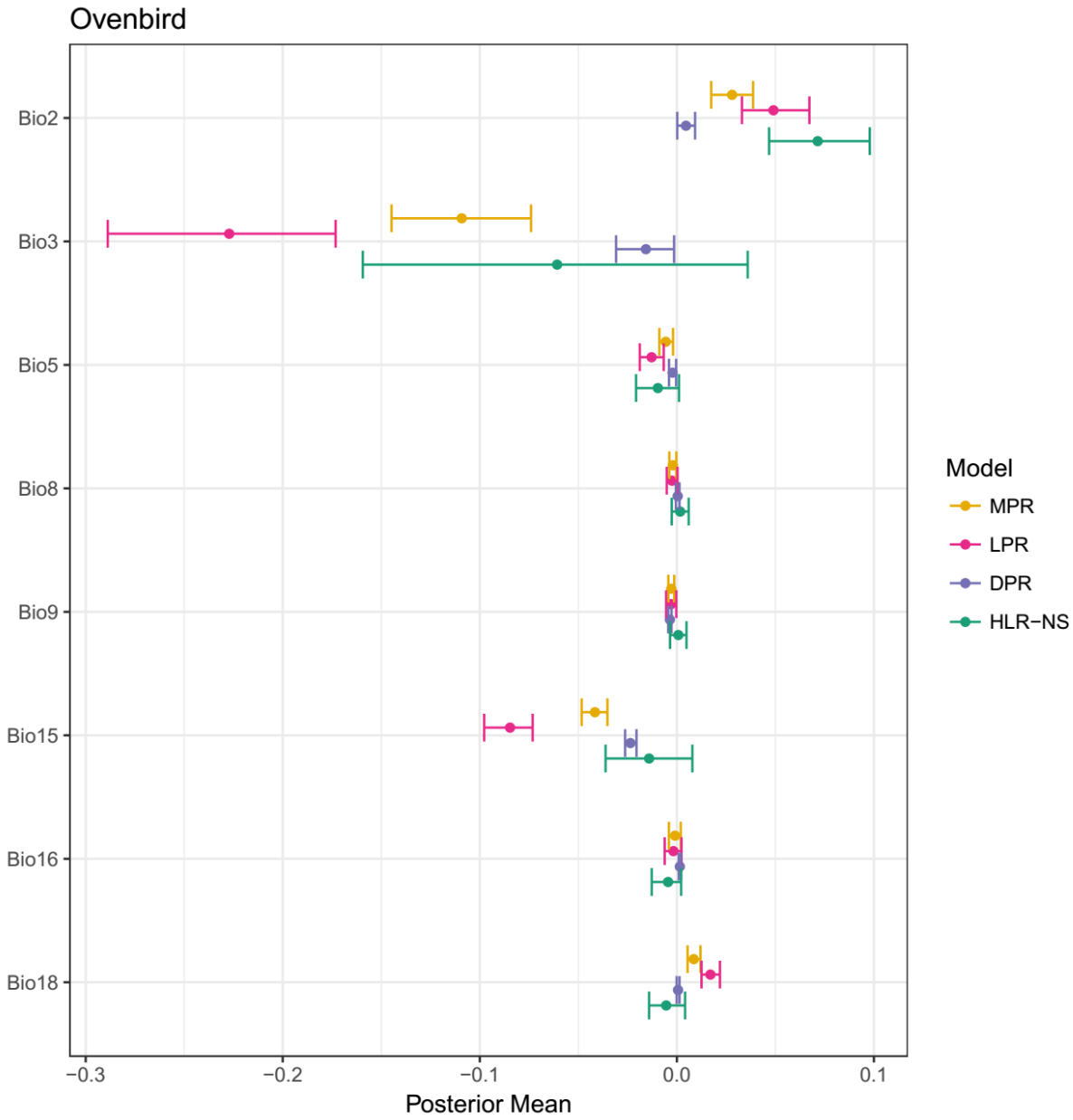


Figure 250: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ovenbird, *Seiurus aurocapillus*.

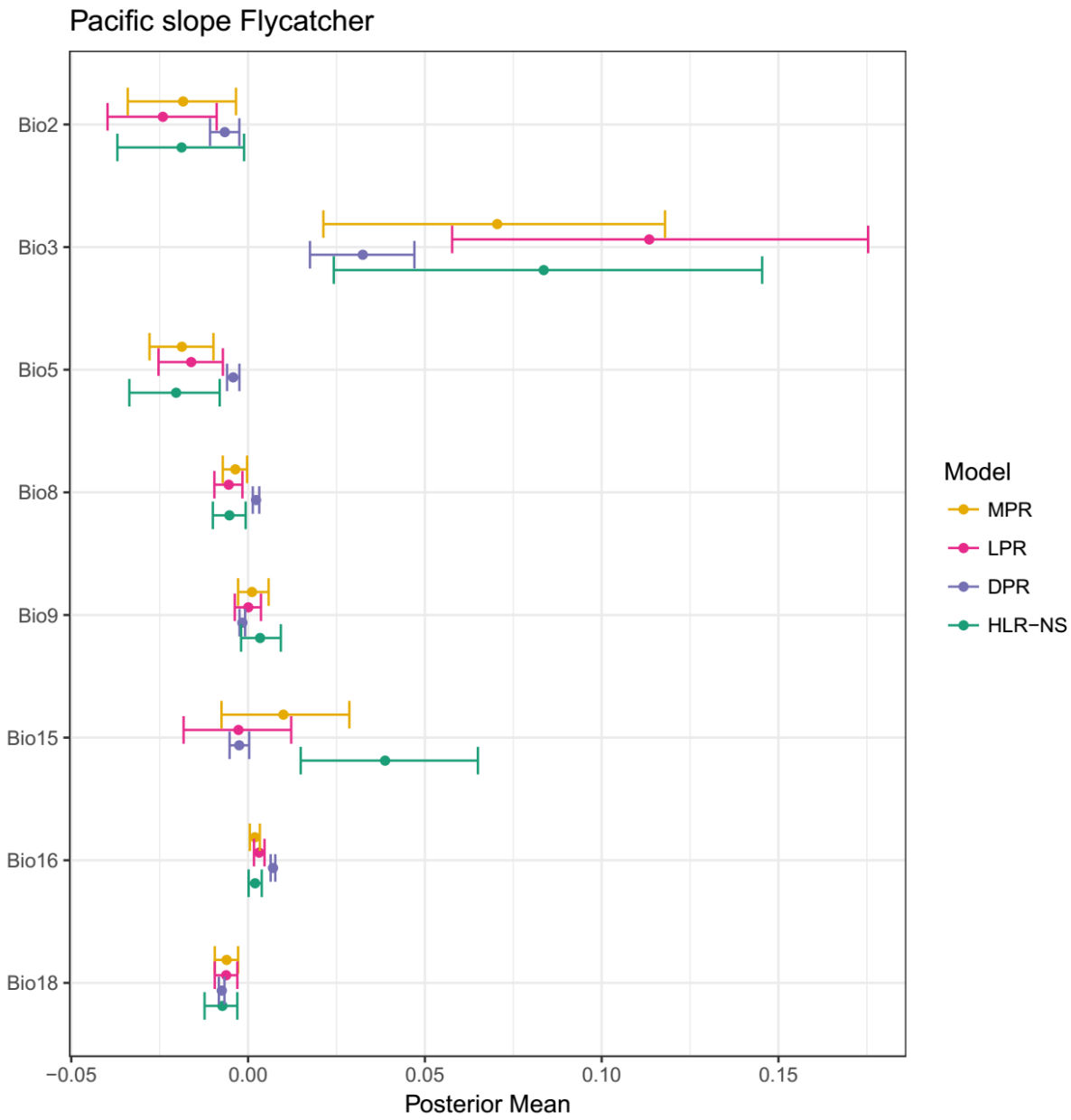


Figure 251: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Pacific-slope flycatcher, *Empidonax difficilis*.

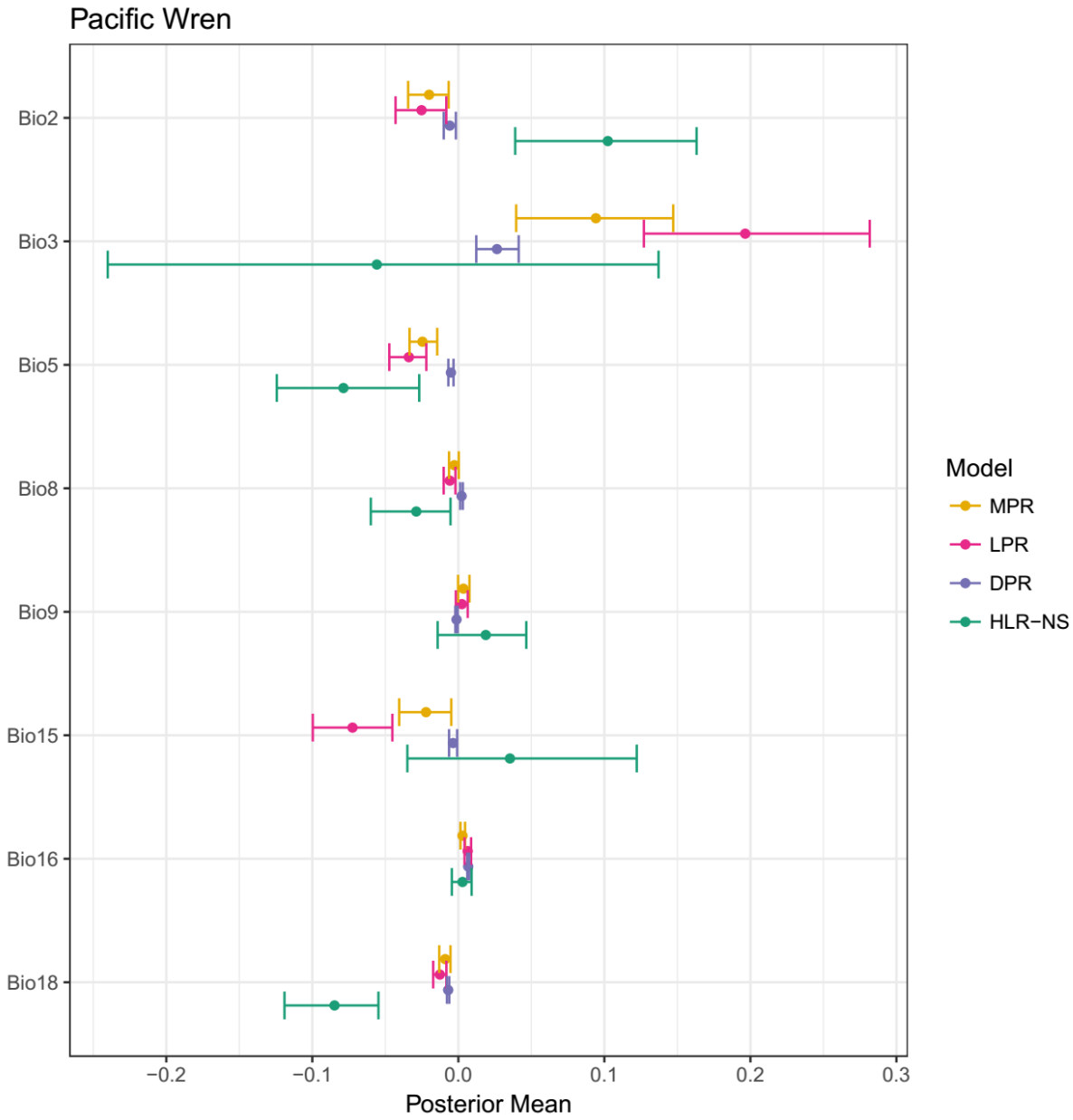


Figure 252: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Pacific wren, *Troglodytes pacificus*.

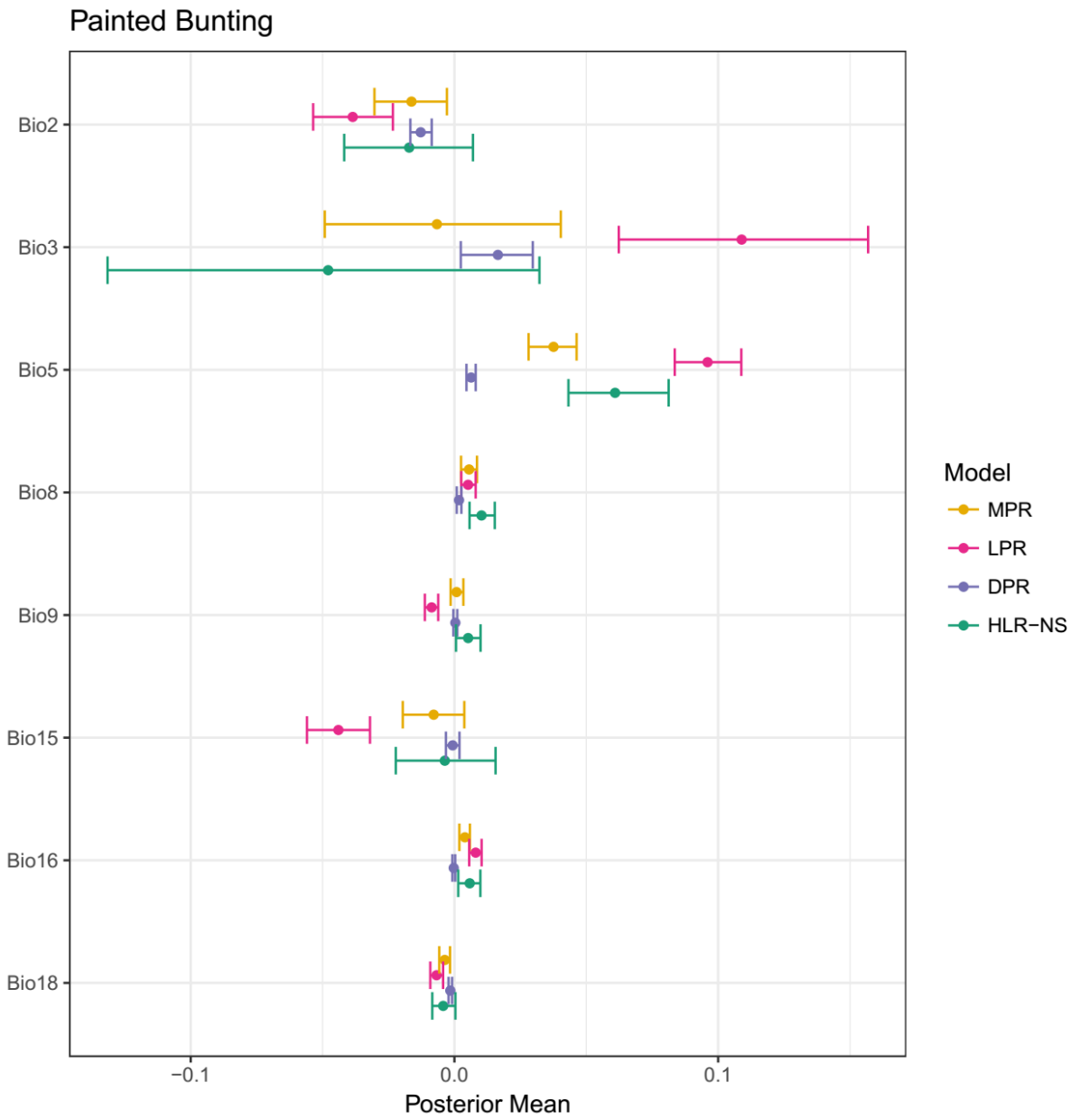


Figure 253: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the painted bunting, *Passerina ciris*.

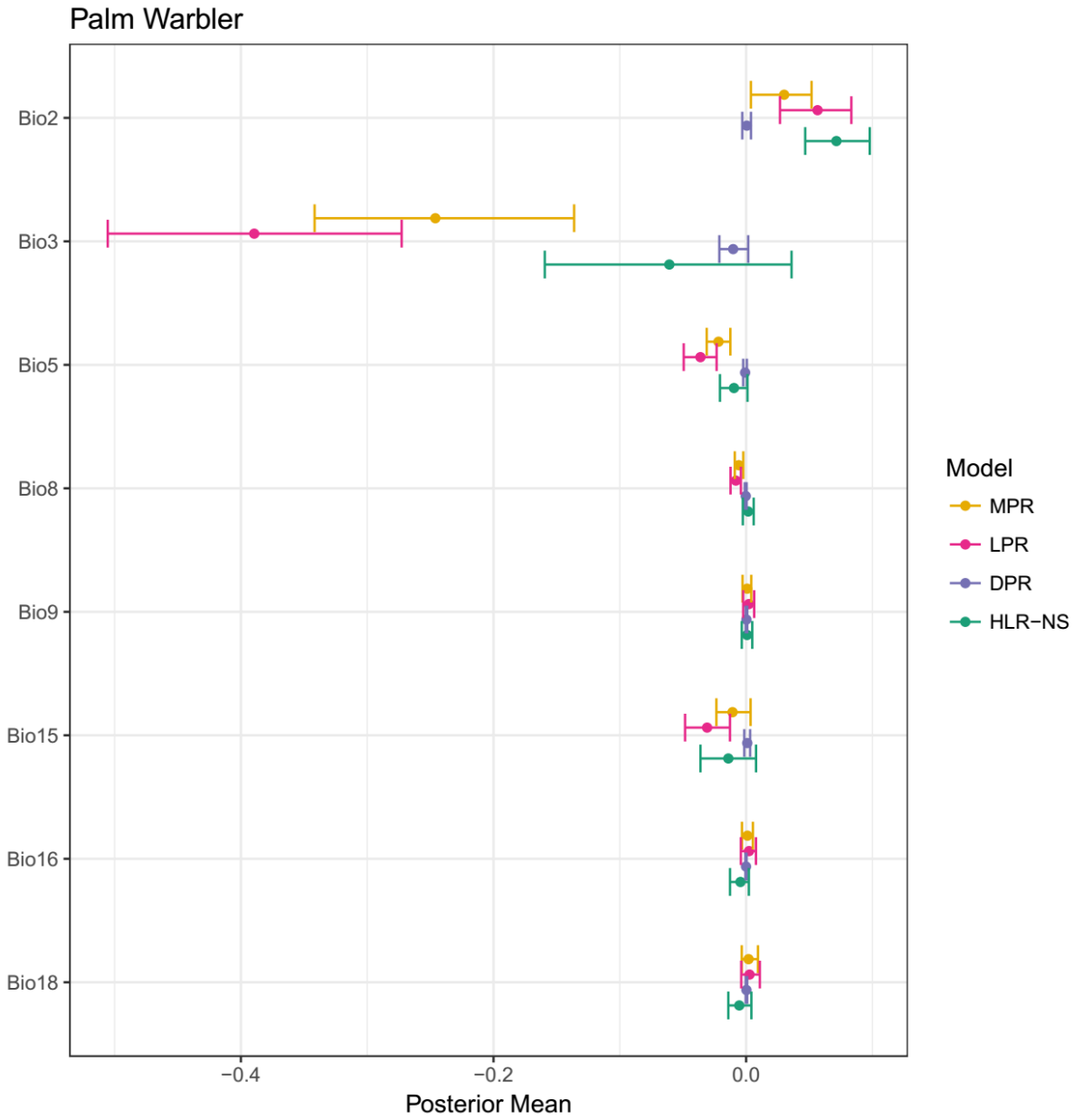


Figure 254: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the palm warbler, *Setophaga palmarum*.

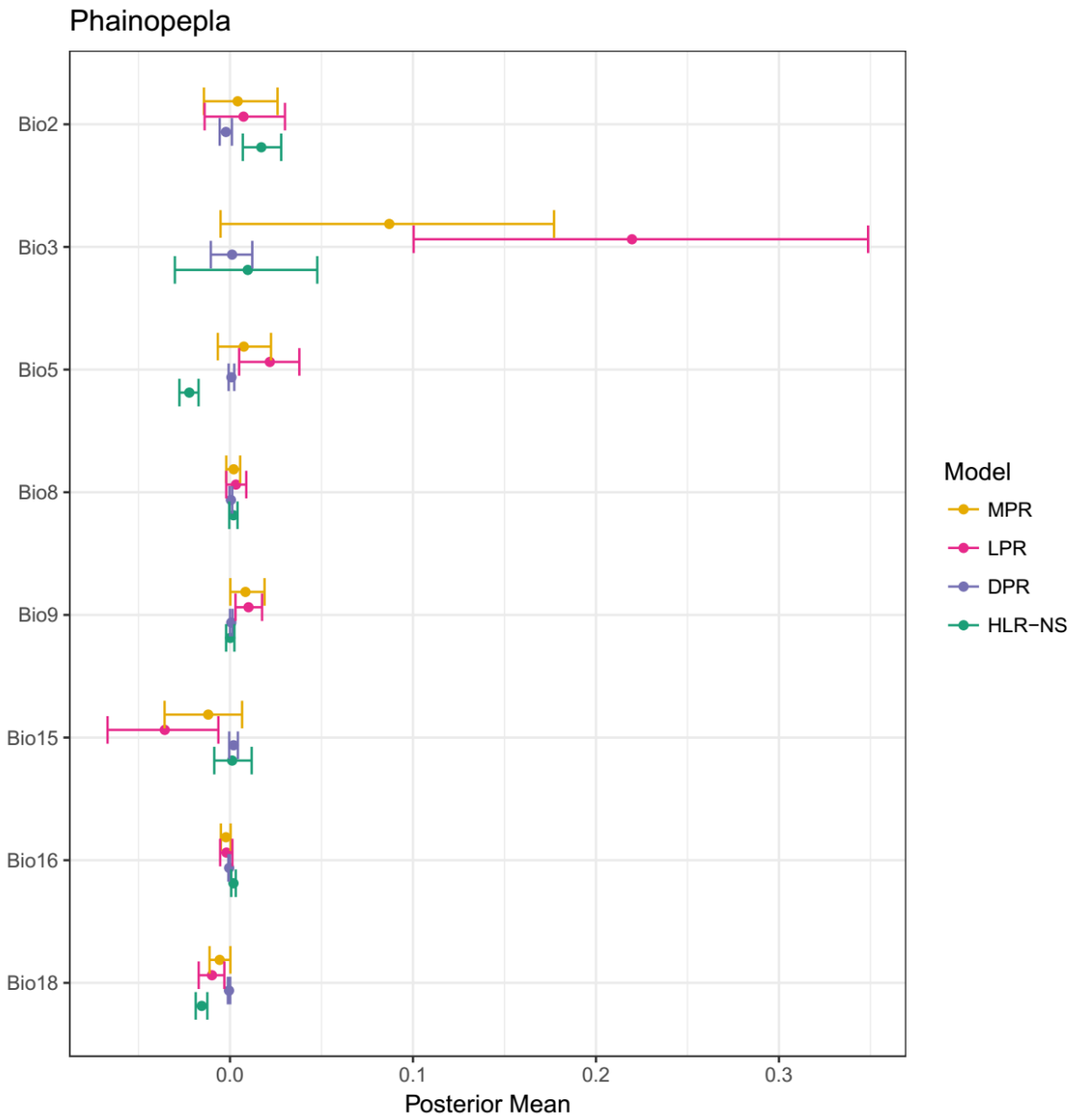


Figure 255: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Phainopepla, *Phainopepla nitens*.

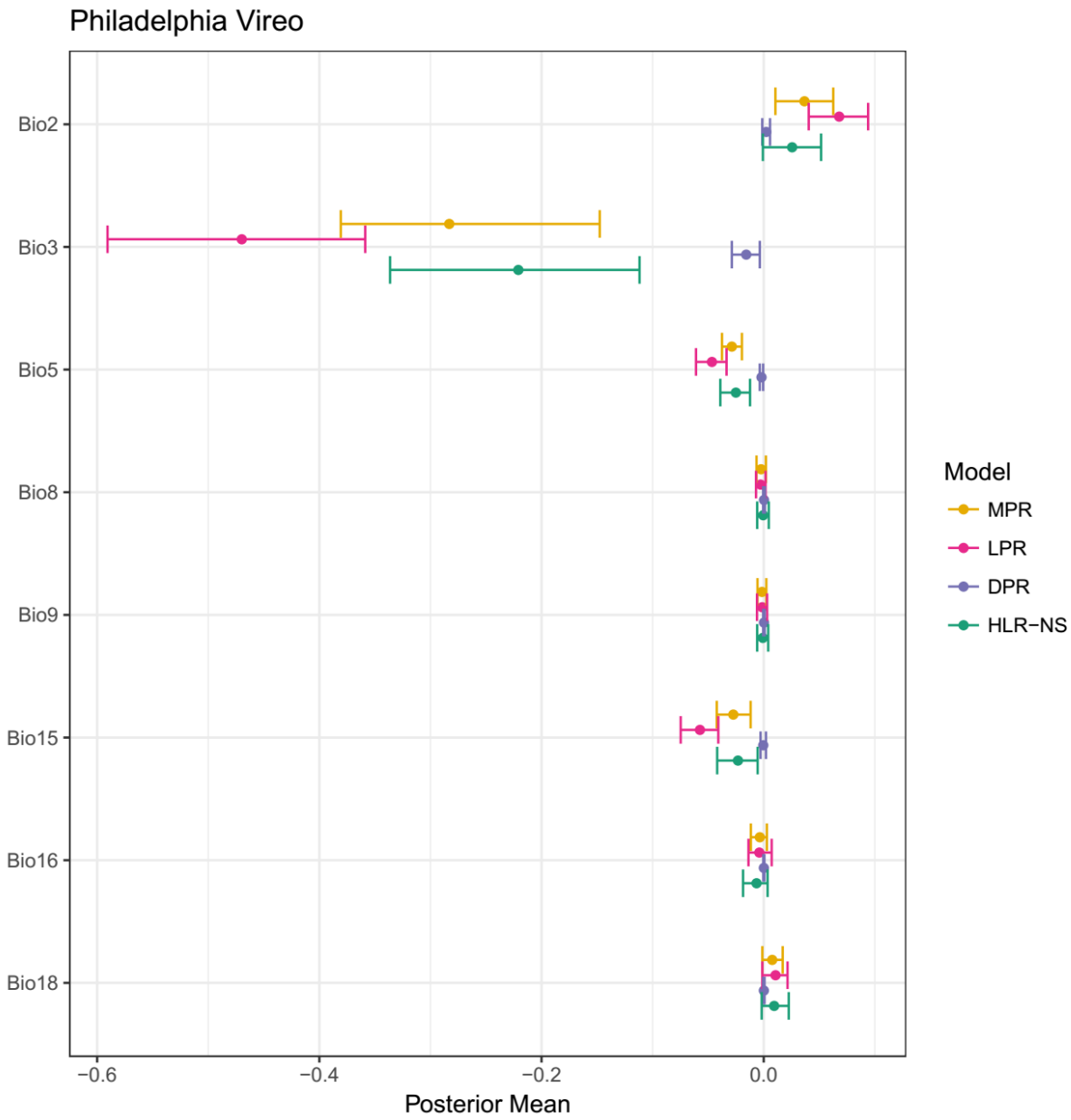


Figure 256: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Philadelphia vireo, *Vireo philadelphicus*.

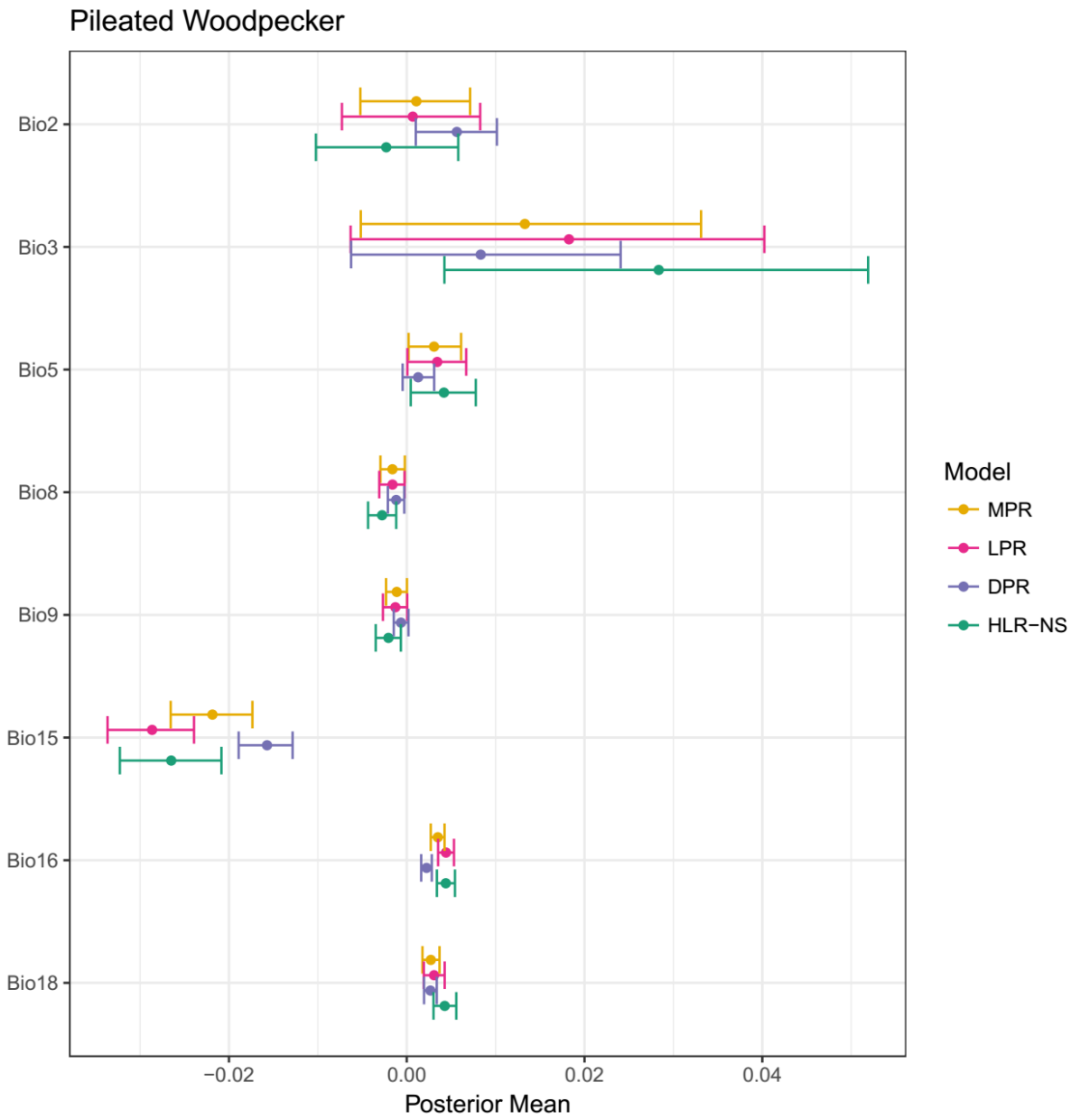


Figure 257: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pileated woodpecker, *Hylatomus pileatus*.

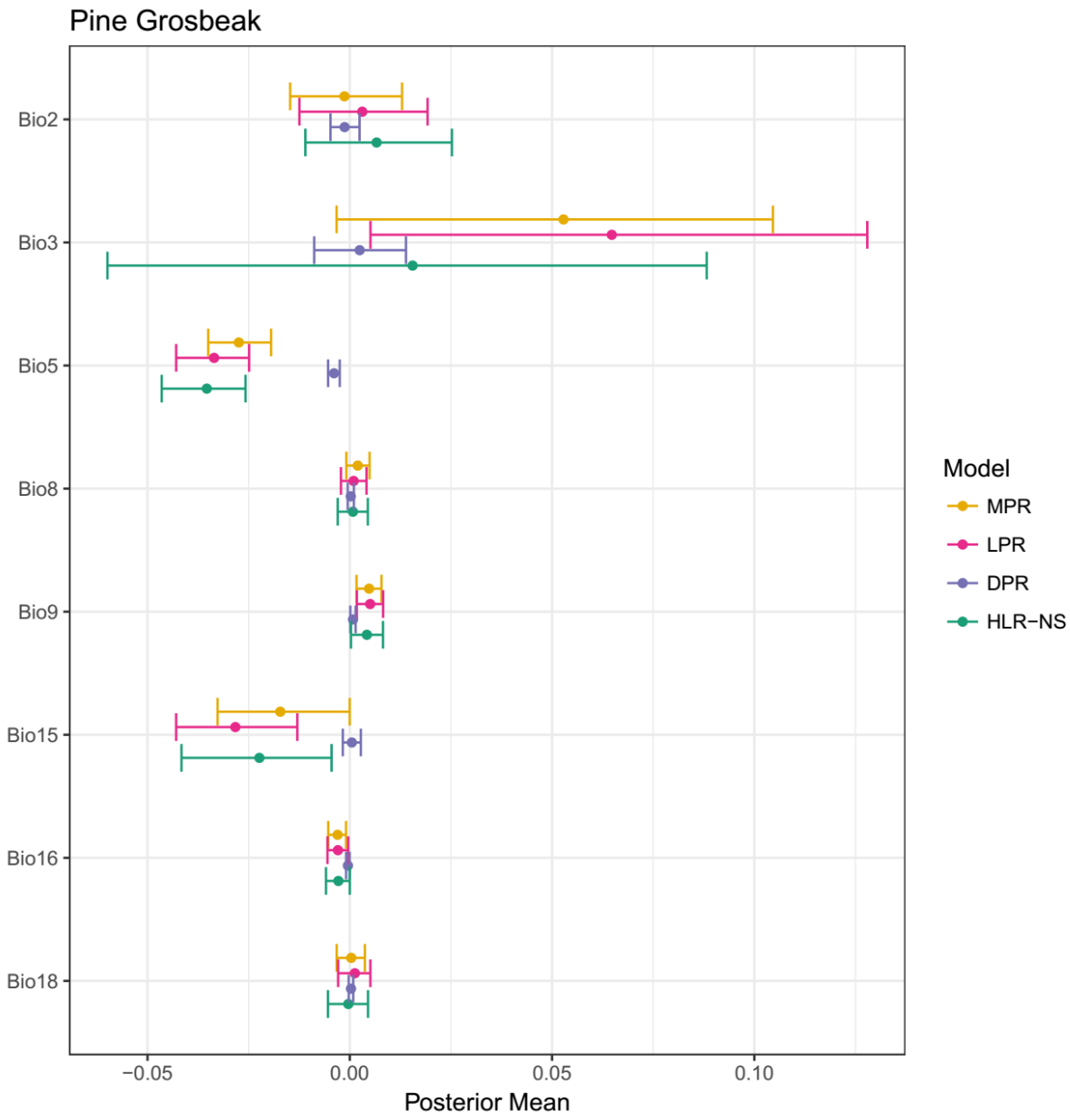


Figure 258: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pine grosbeak, *Pinicola enucleator*.

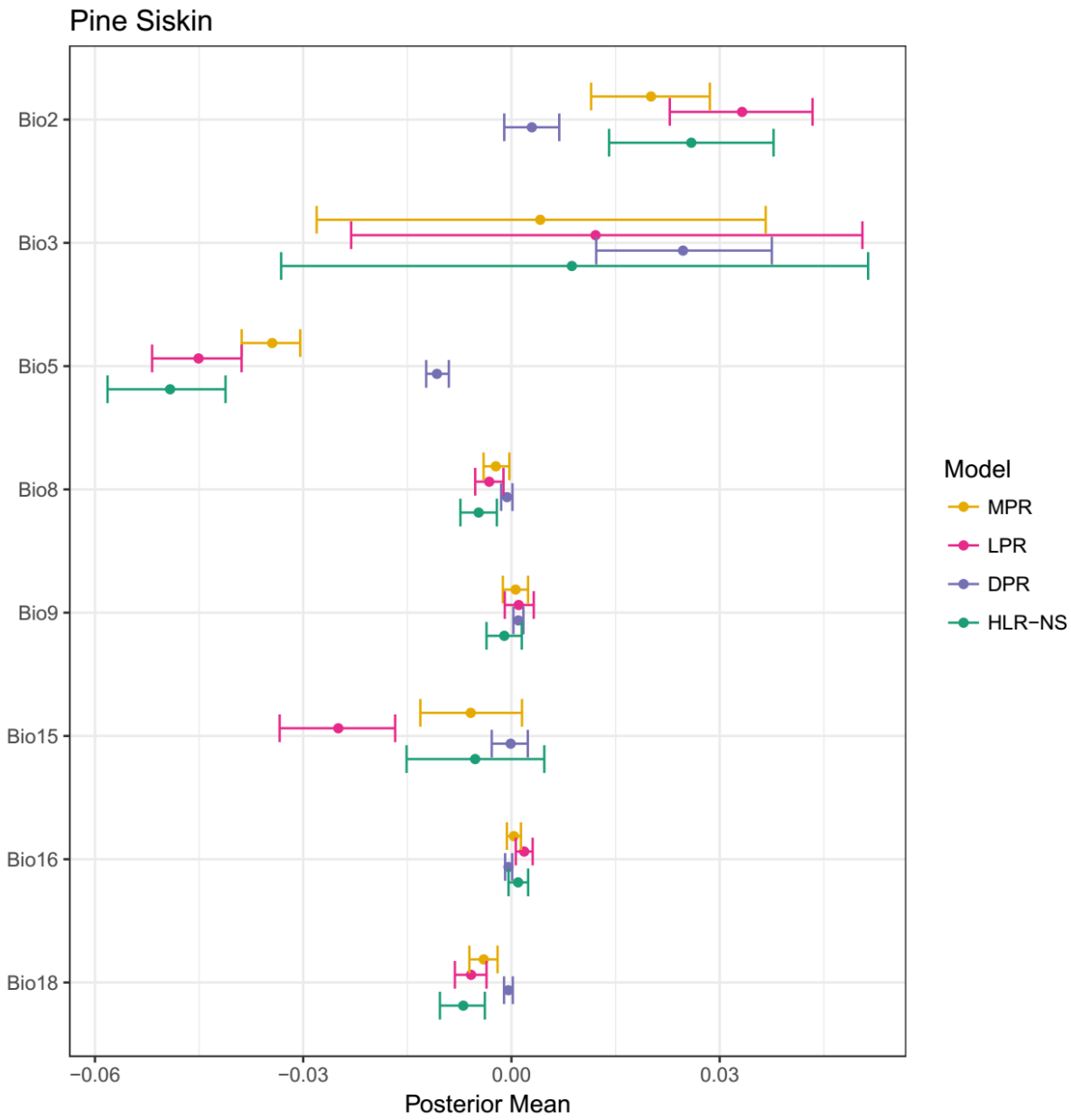


Figure 259: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pine siskin, *Spinus pinus*.

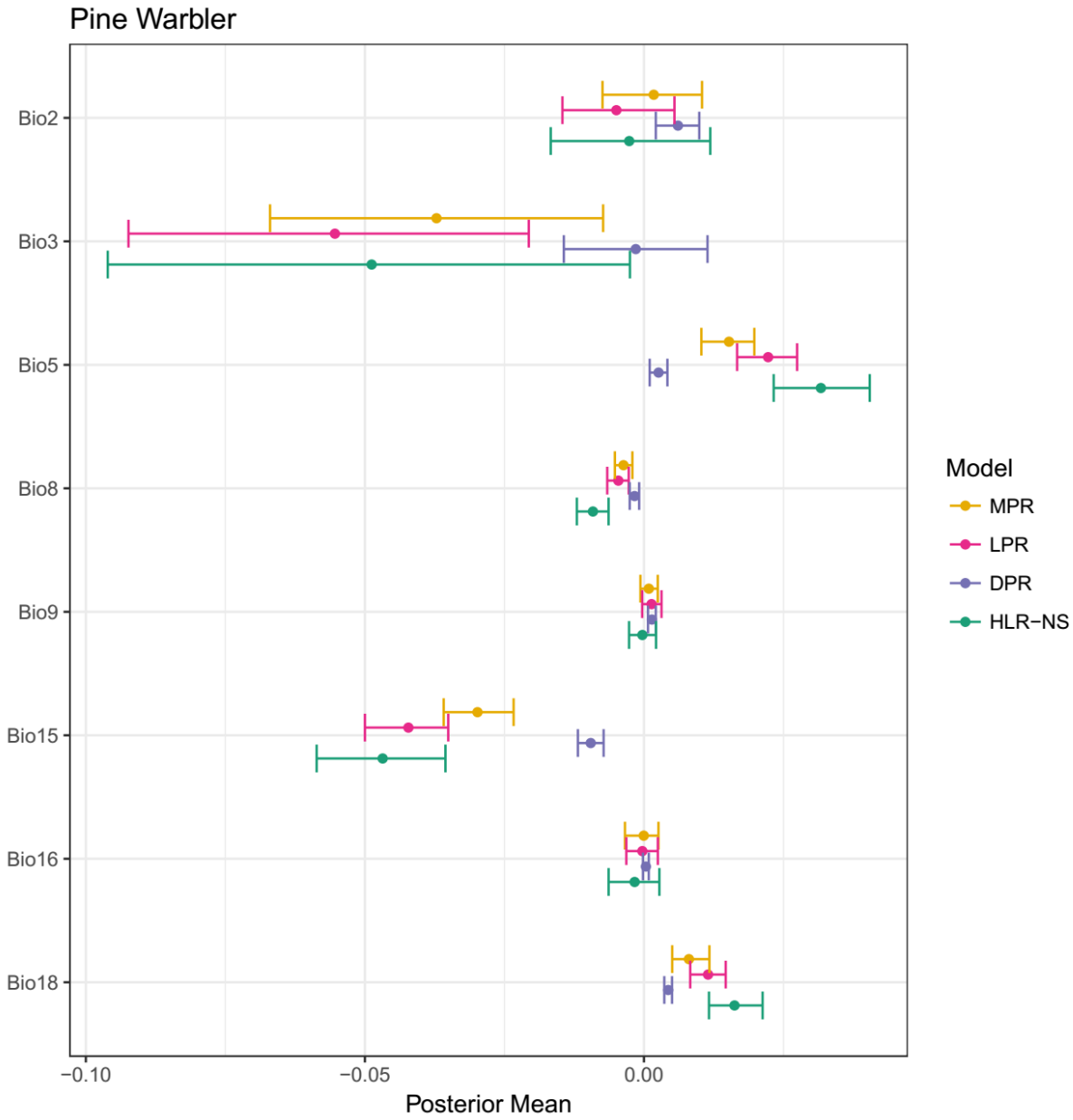


Figure 260: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pine warbler, *Setophaga pinus*.

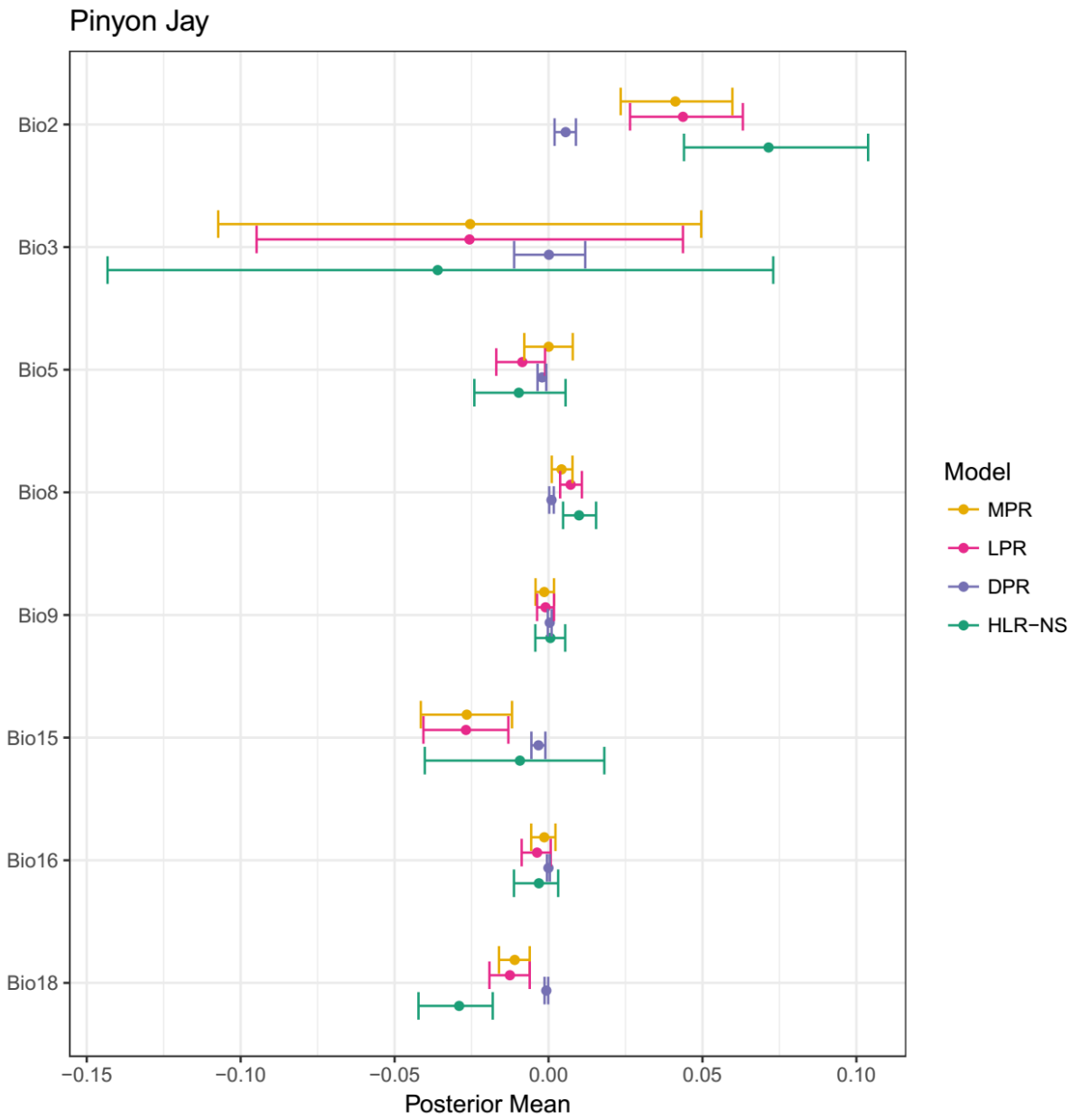


Figure 261: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pinyon jay, *Gymnorhinus cyanocephalus*.

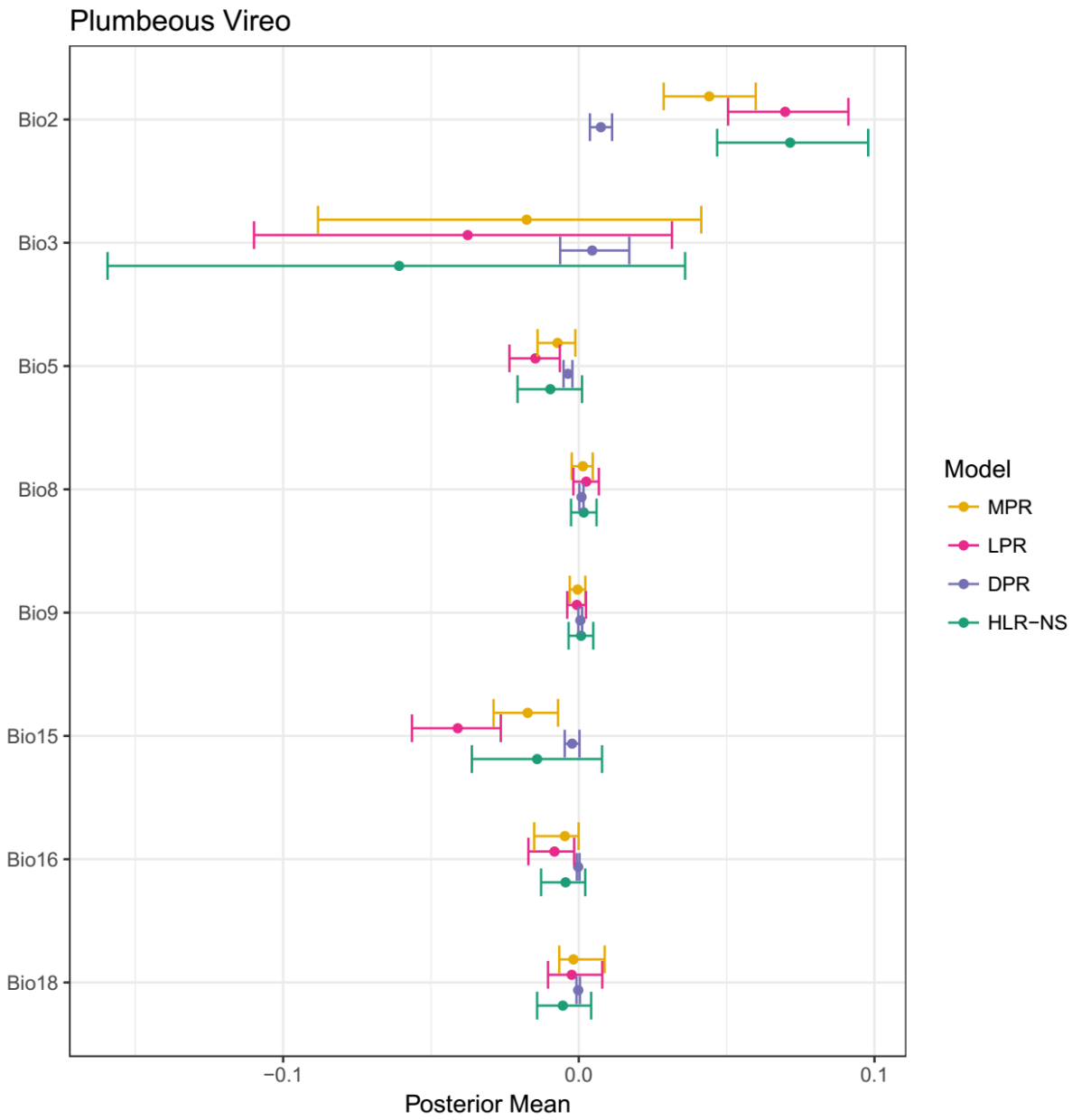


Figure 262: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the plumbeous vireo, *Vireo plumbeus*.

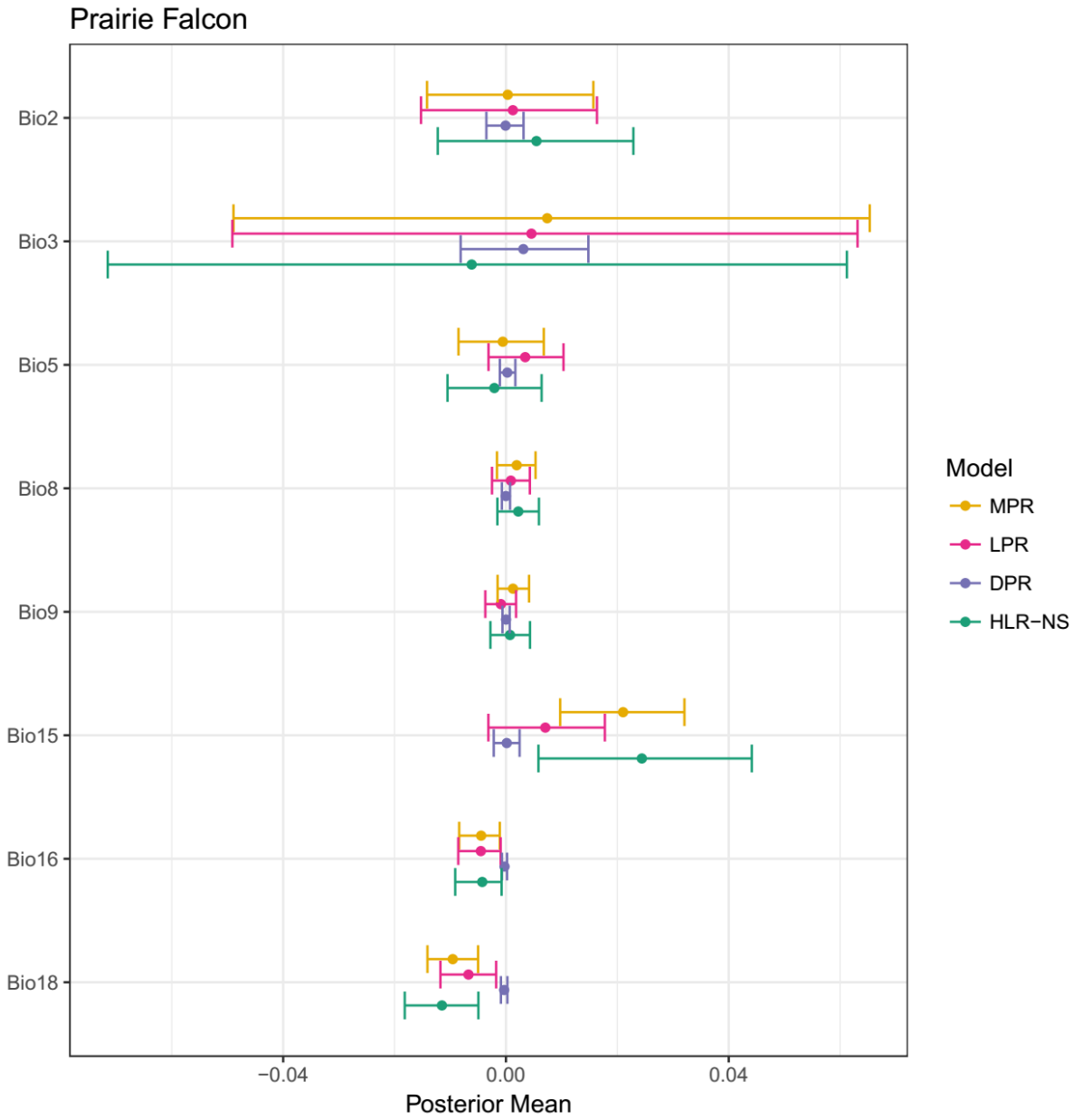


Figure 263: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the prairie falcon, *Falco mexicanus*.

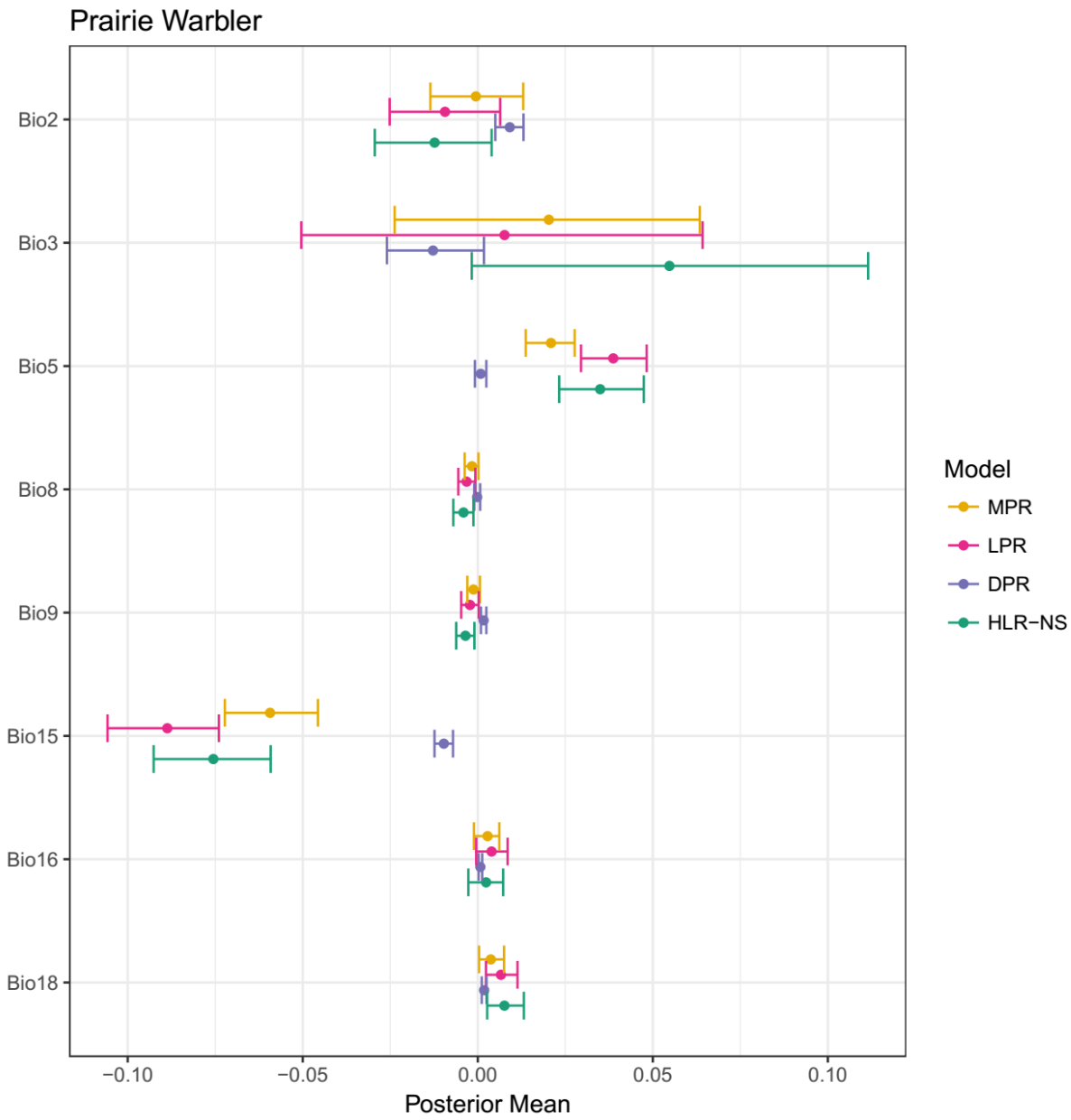


Figure 264: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the prairie warbler, *Setophaga discolor*.

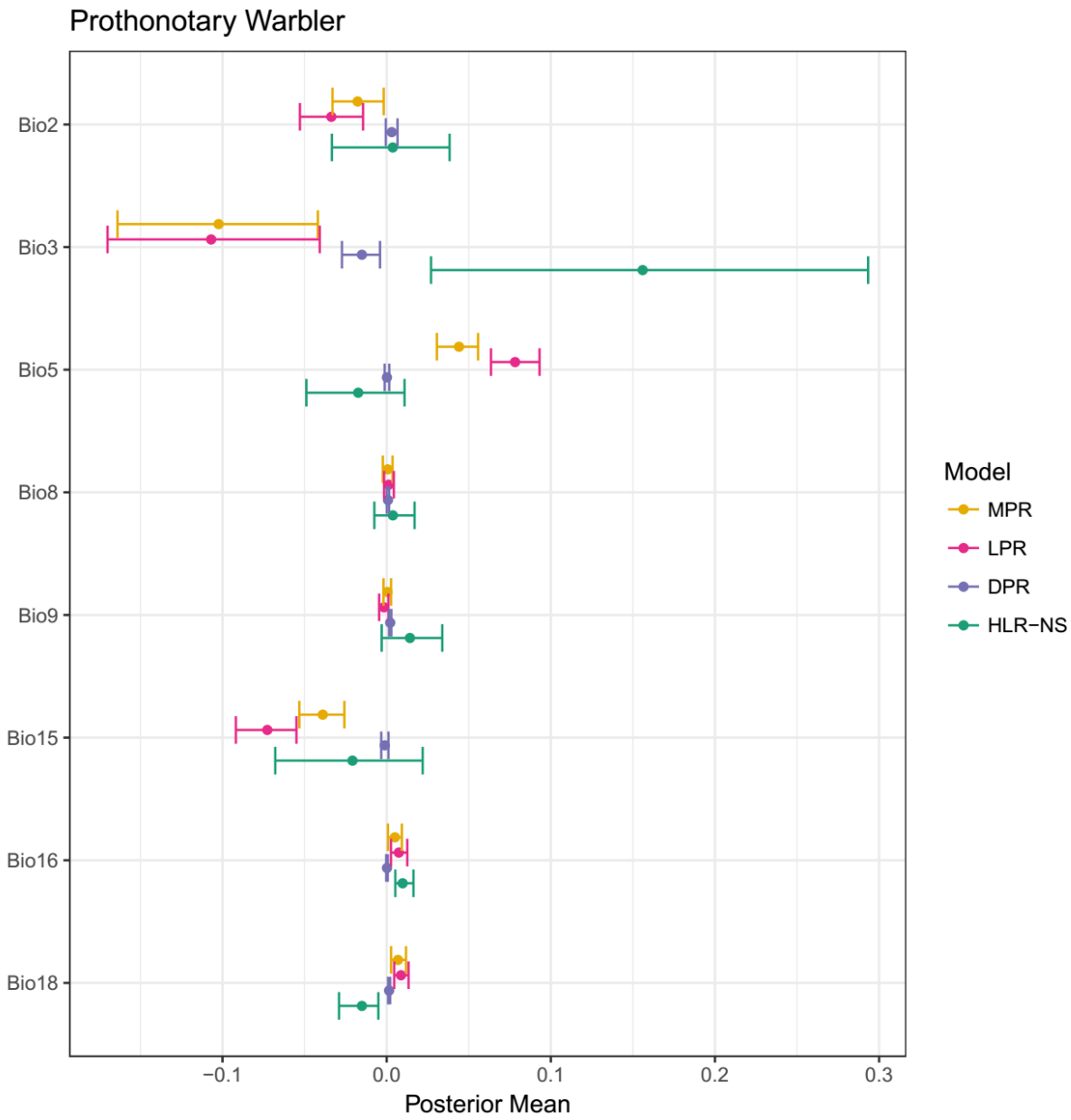


Figure 265: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the prothonotary warbler, *Protonotaria citrea*.

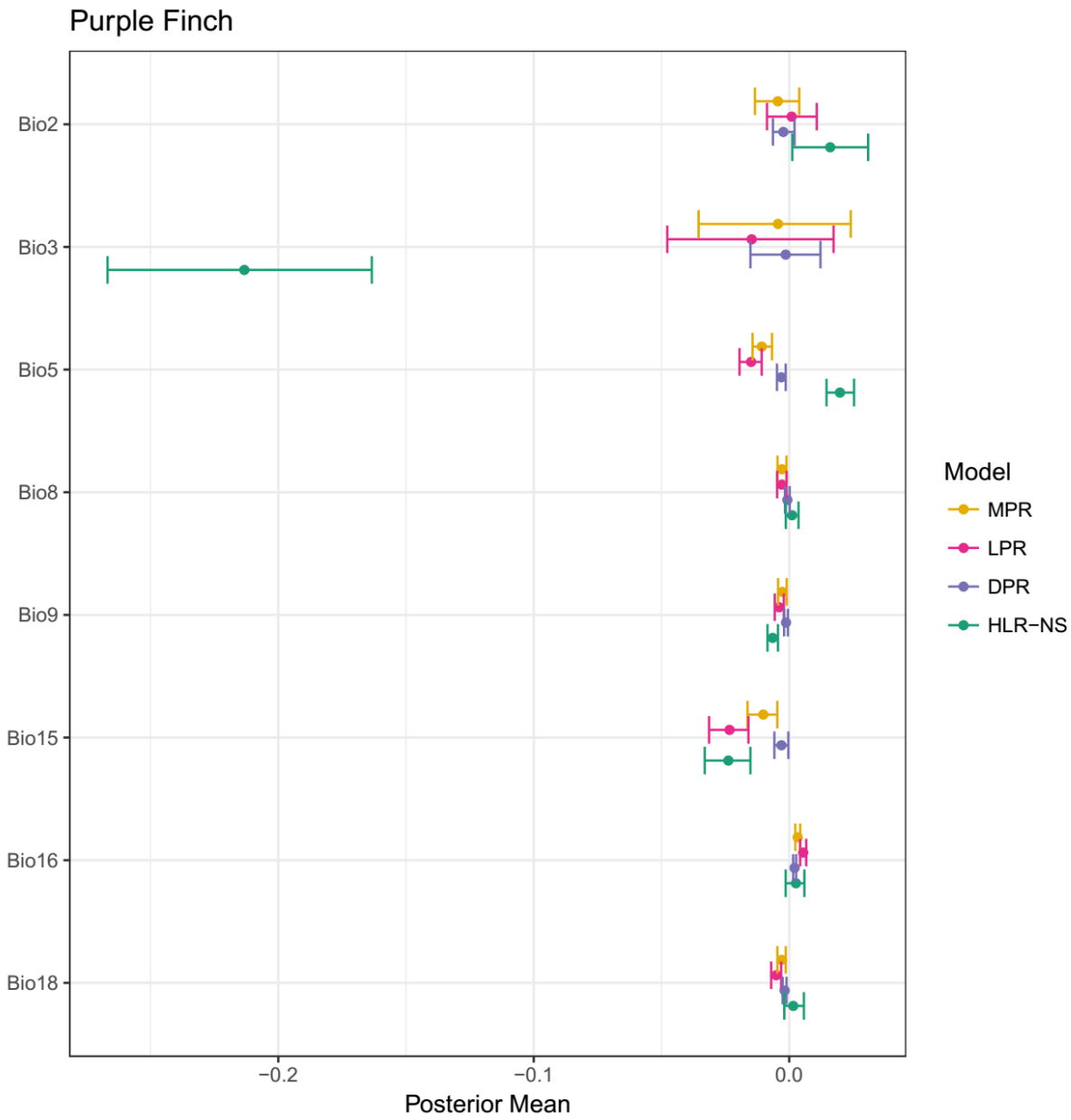


Figure 266: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the purple finch, *Haemorhous purpureus*.

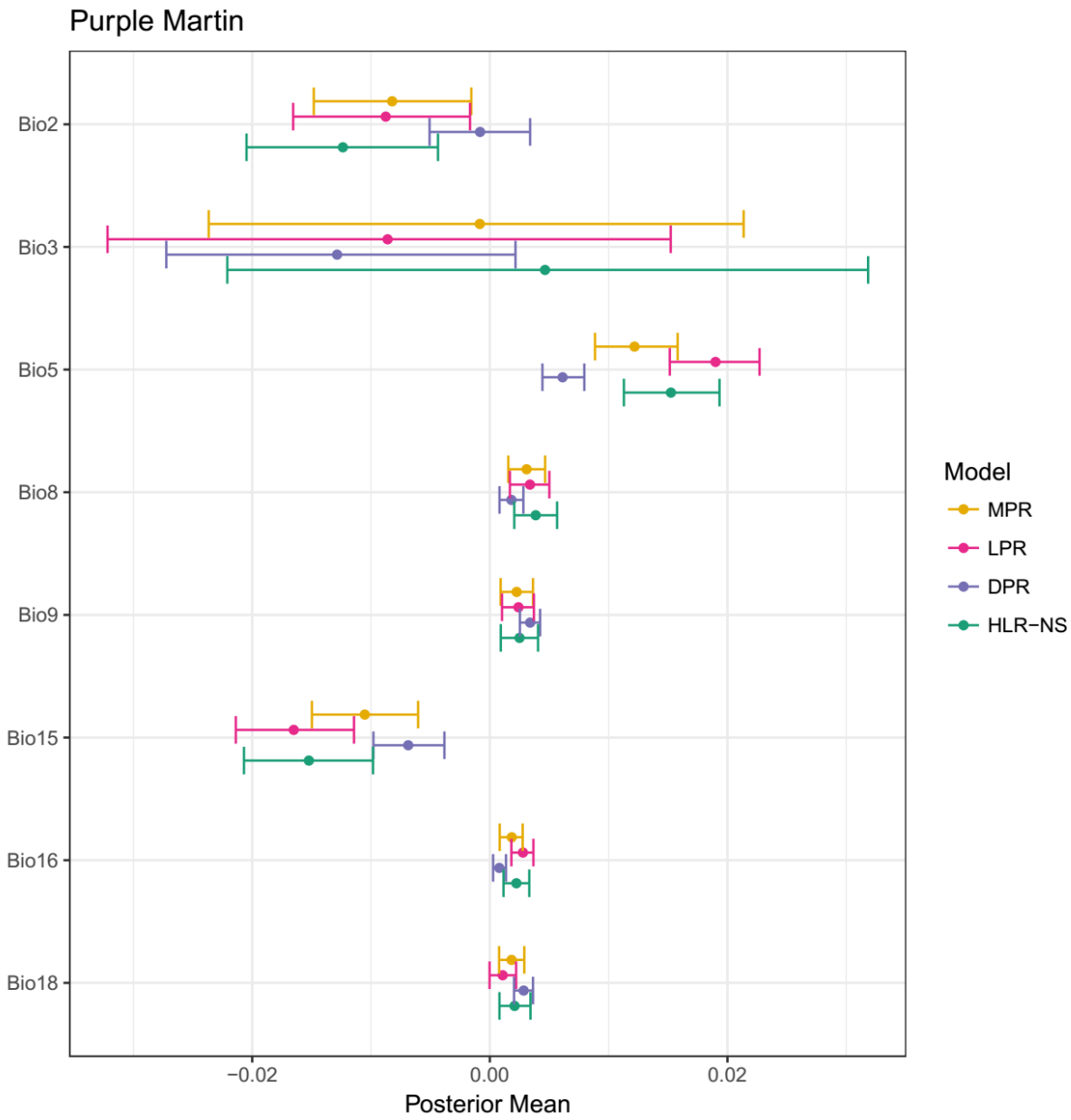


Figure 267: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the purple martin, *Progne subis*.

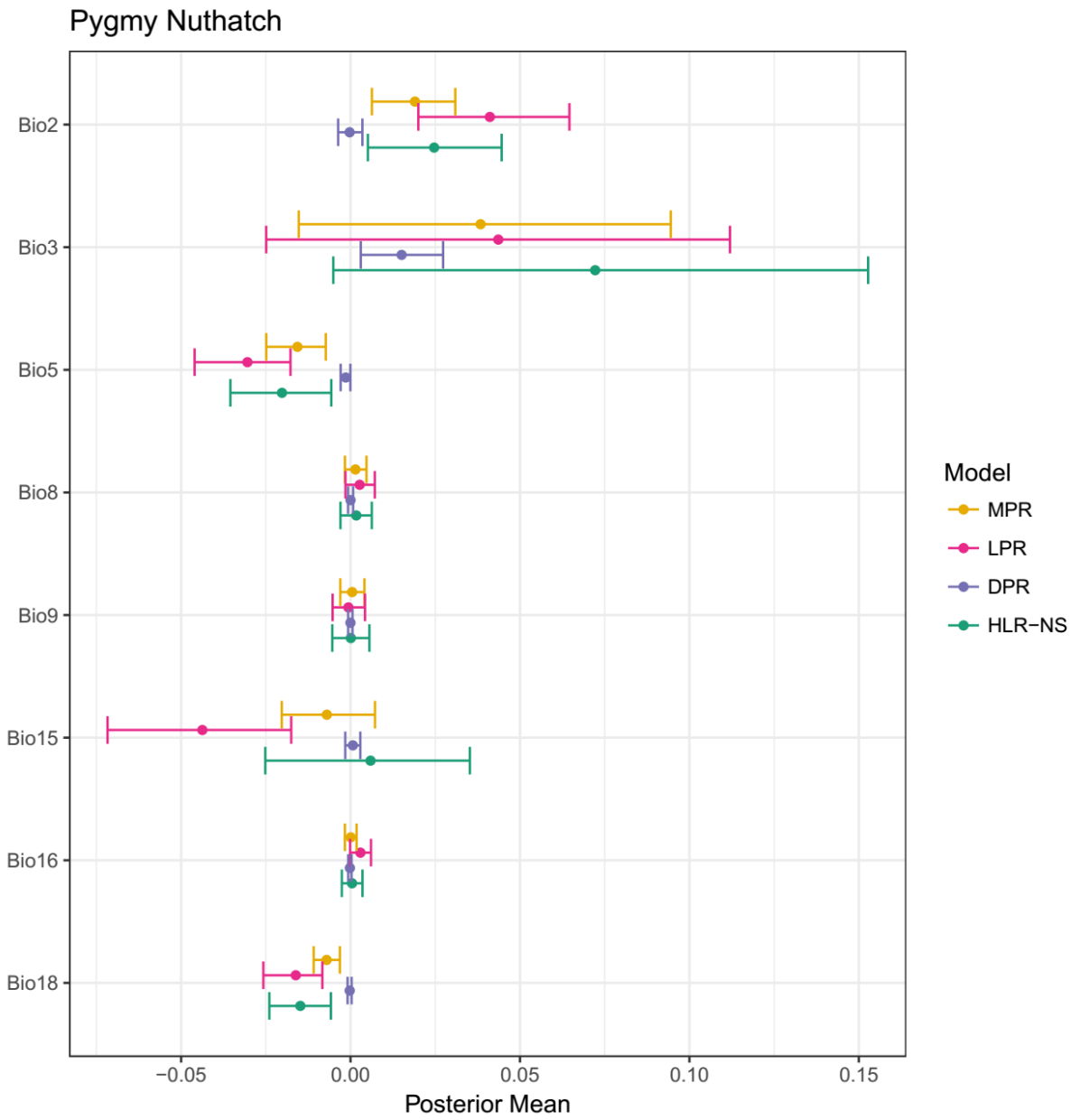


Figure 268: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pygmy nuthatch, *Sitta pygmaea*.

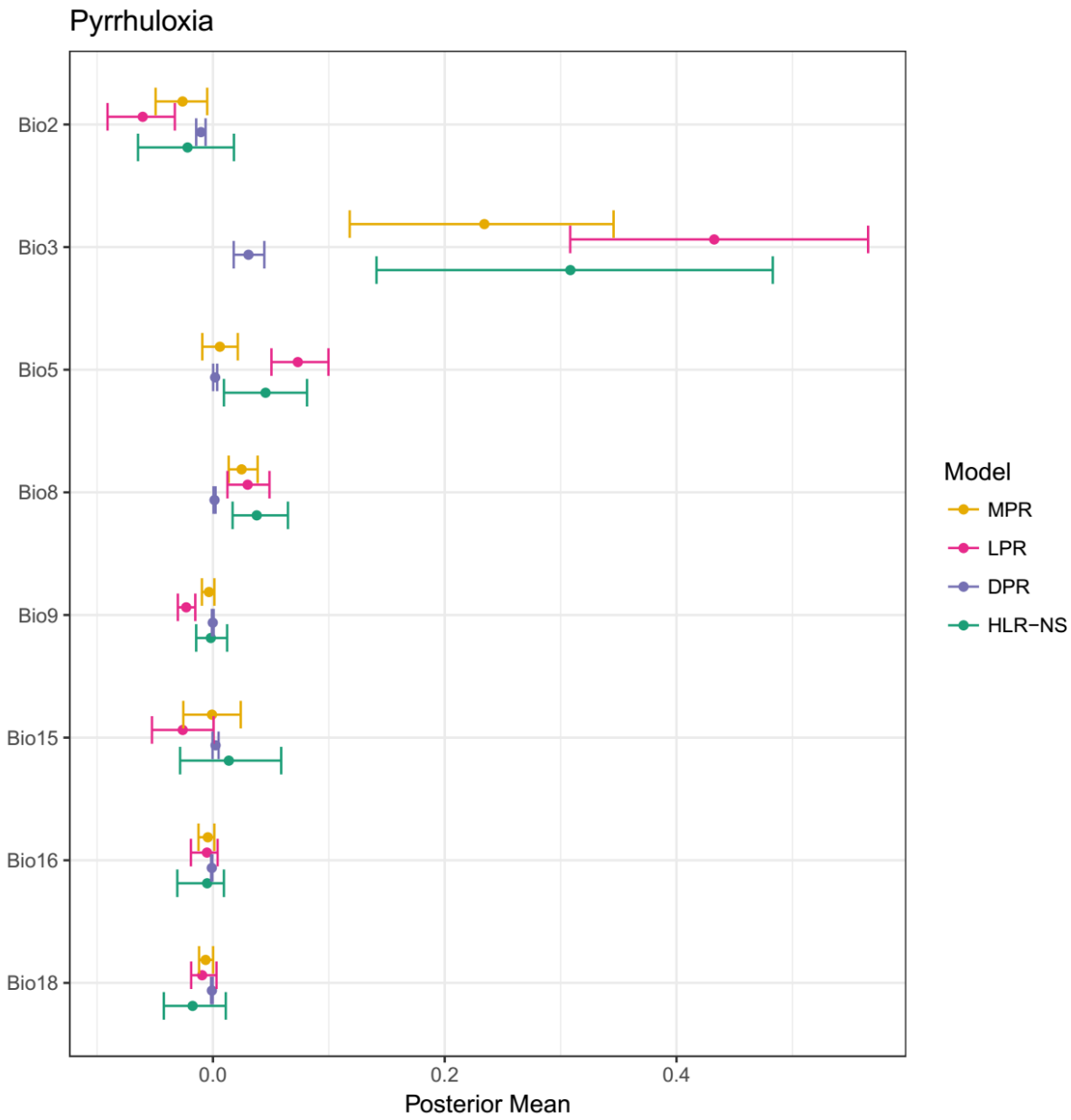


Figure 269: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the phyrruloxia, *Cardinalis sinuatus*.

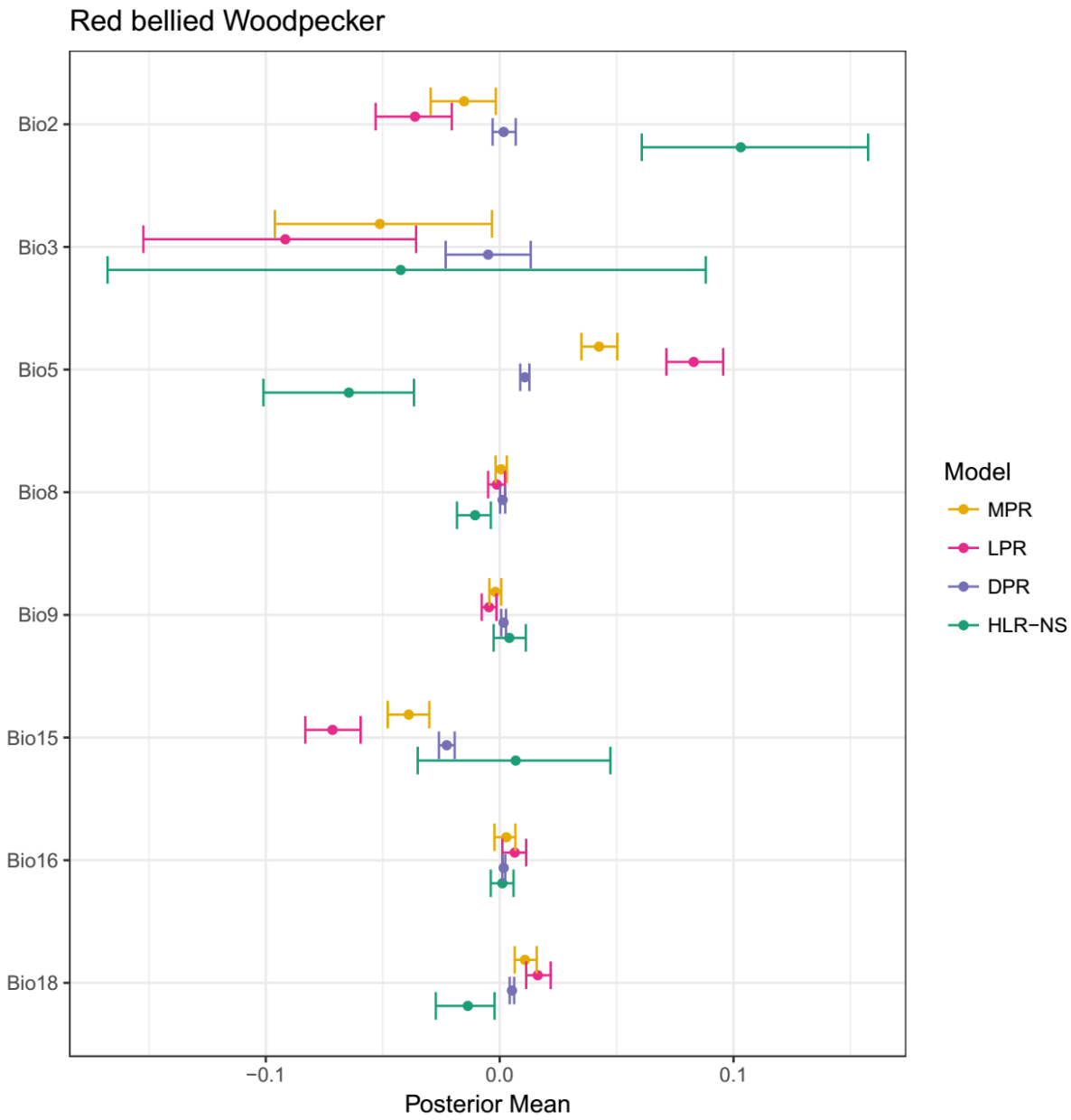


Figure 270: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-bellied woodpecker, *Melanerpes carolinus*.

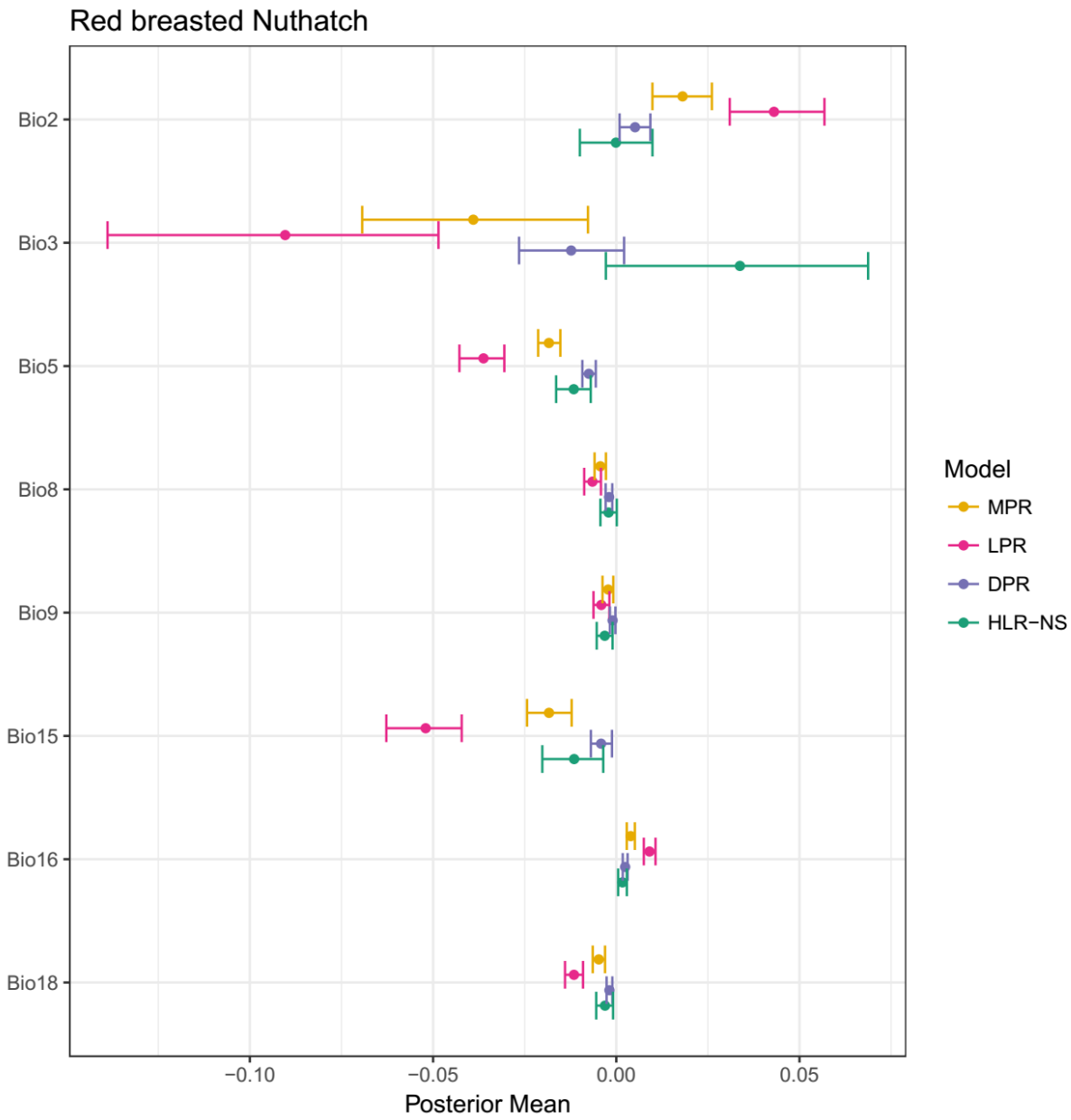


Figure 271: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-breasted nuthatch, *Sitta canadensis*.

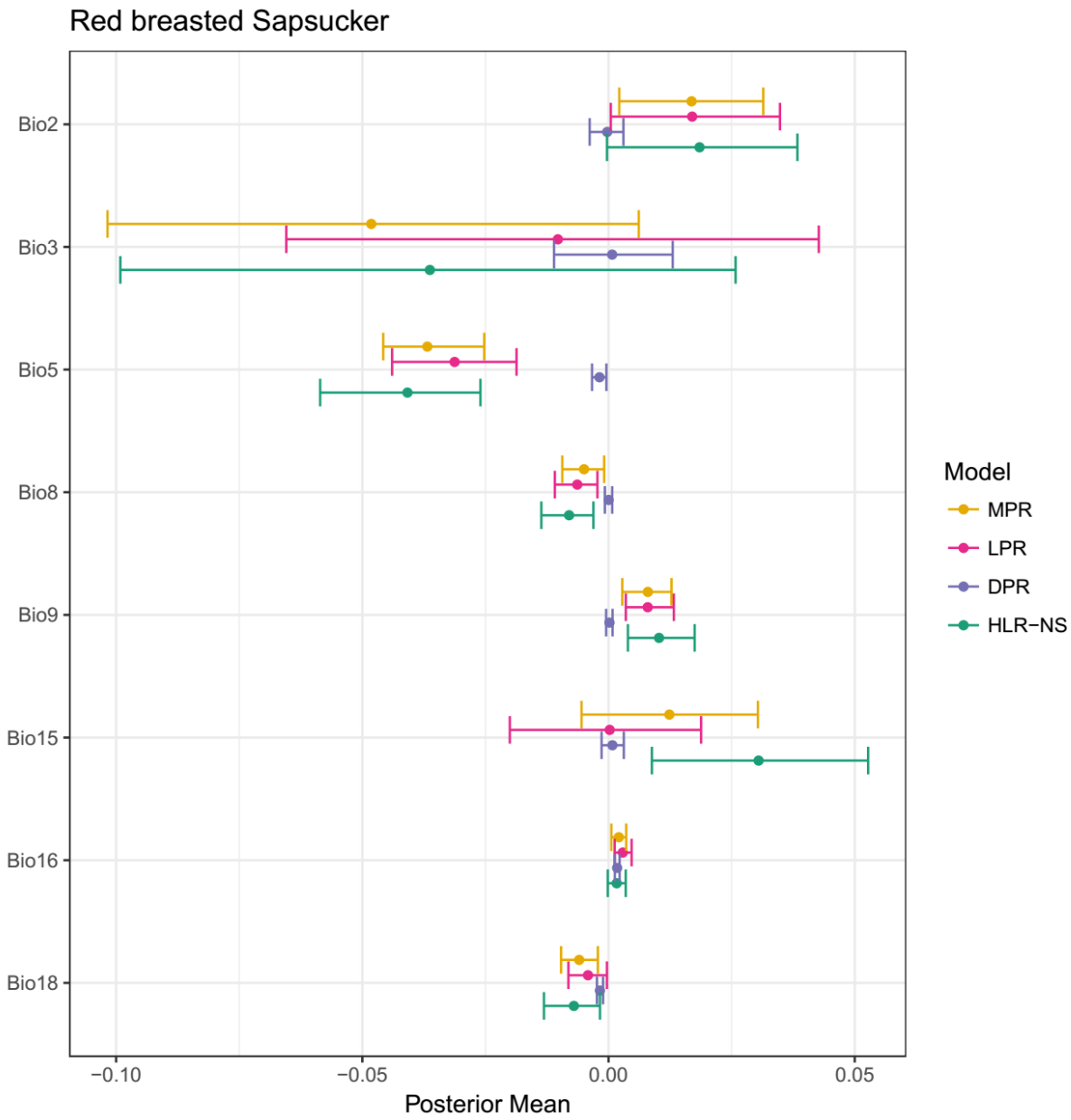


Figure 272: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-breasted sapsucker, *Sphyrapicus ruber*.

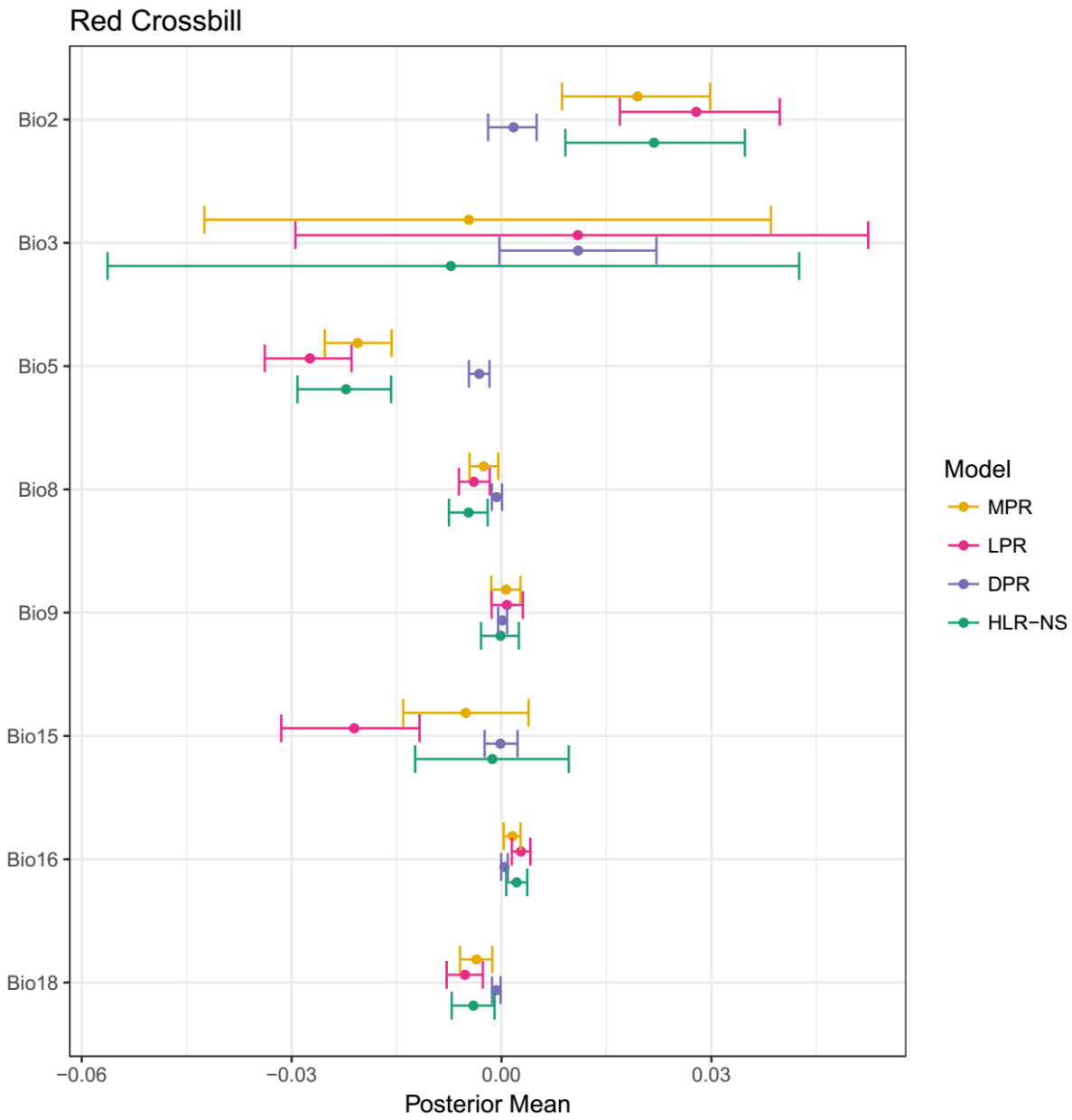


Figure 273: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red crossbill, *Loxia curvirostra*.

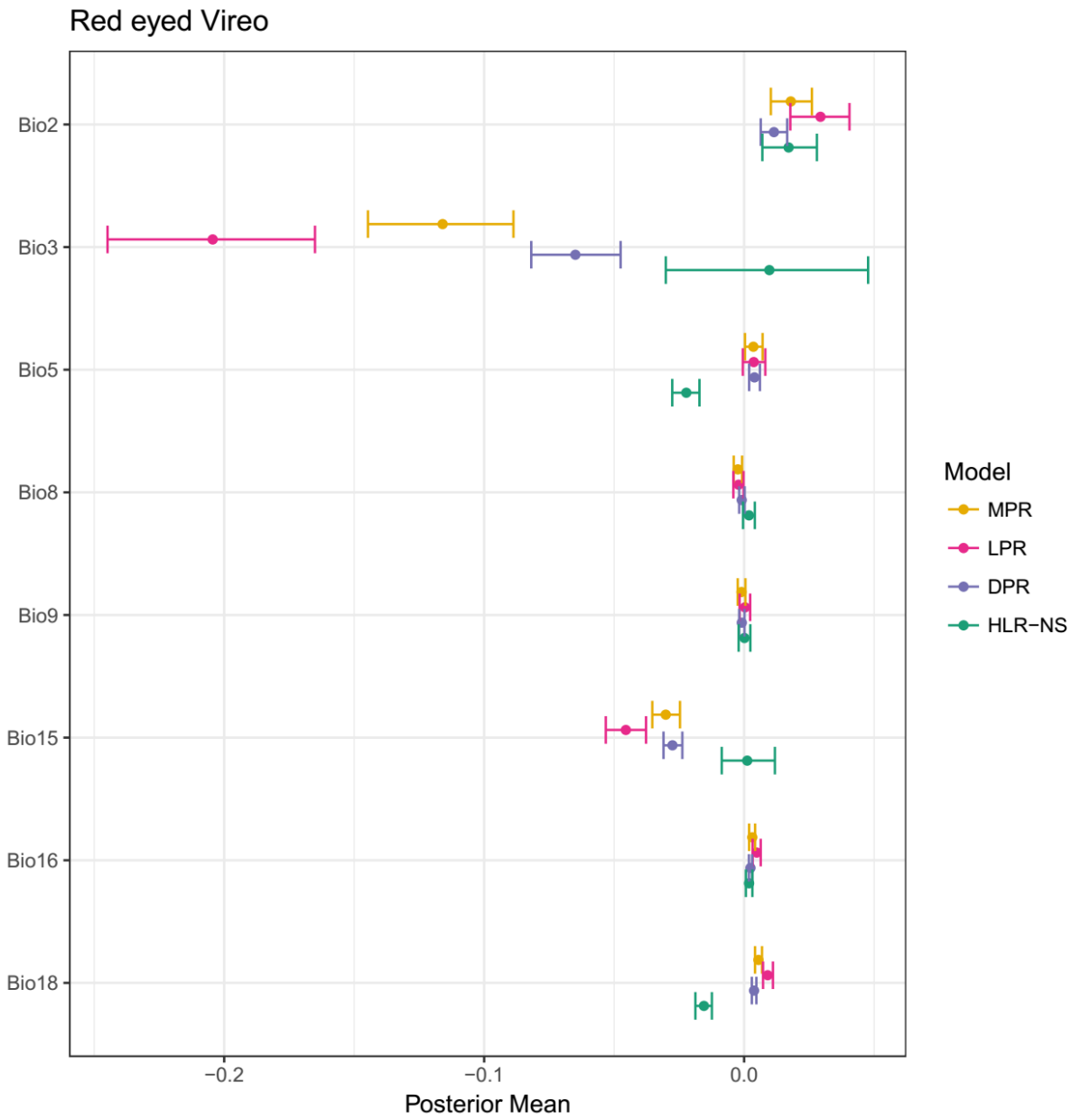


Figure 274: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-eyed vireo, *Vireo olivaceus*.

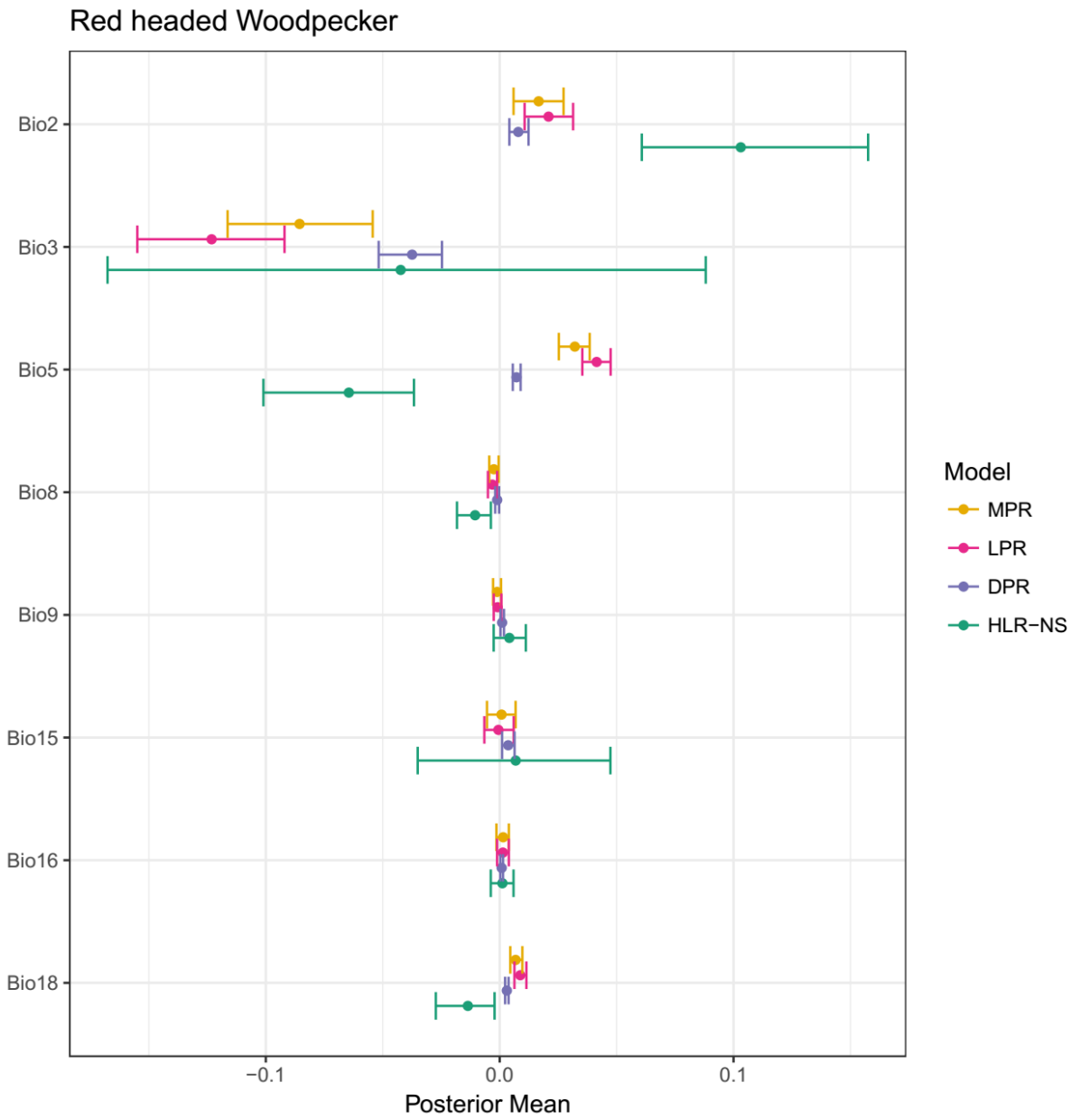


Figure 275: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-headed woodpecker, *Melanerpes erythrocephalus*.

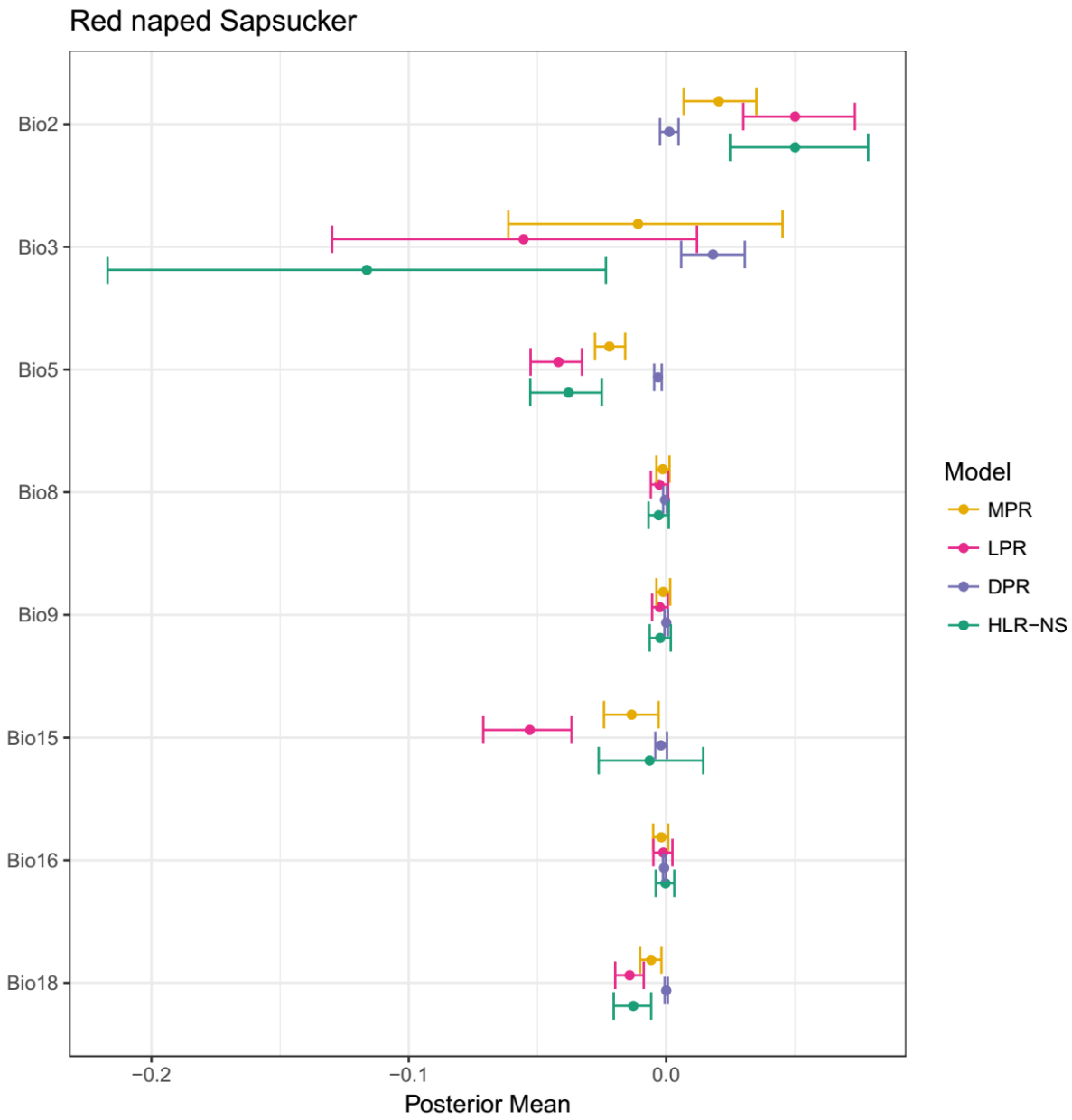


Figure 276: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-naped sapsucker, *Sphyrapicus nuchalis*.

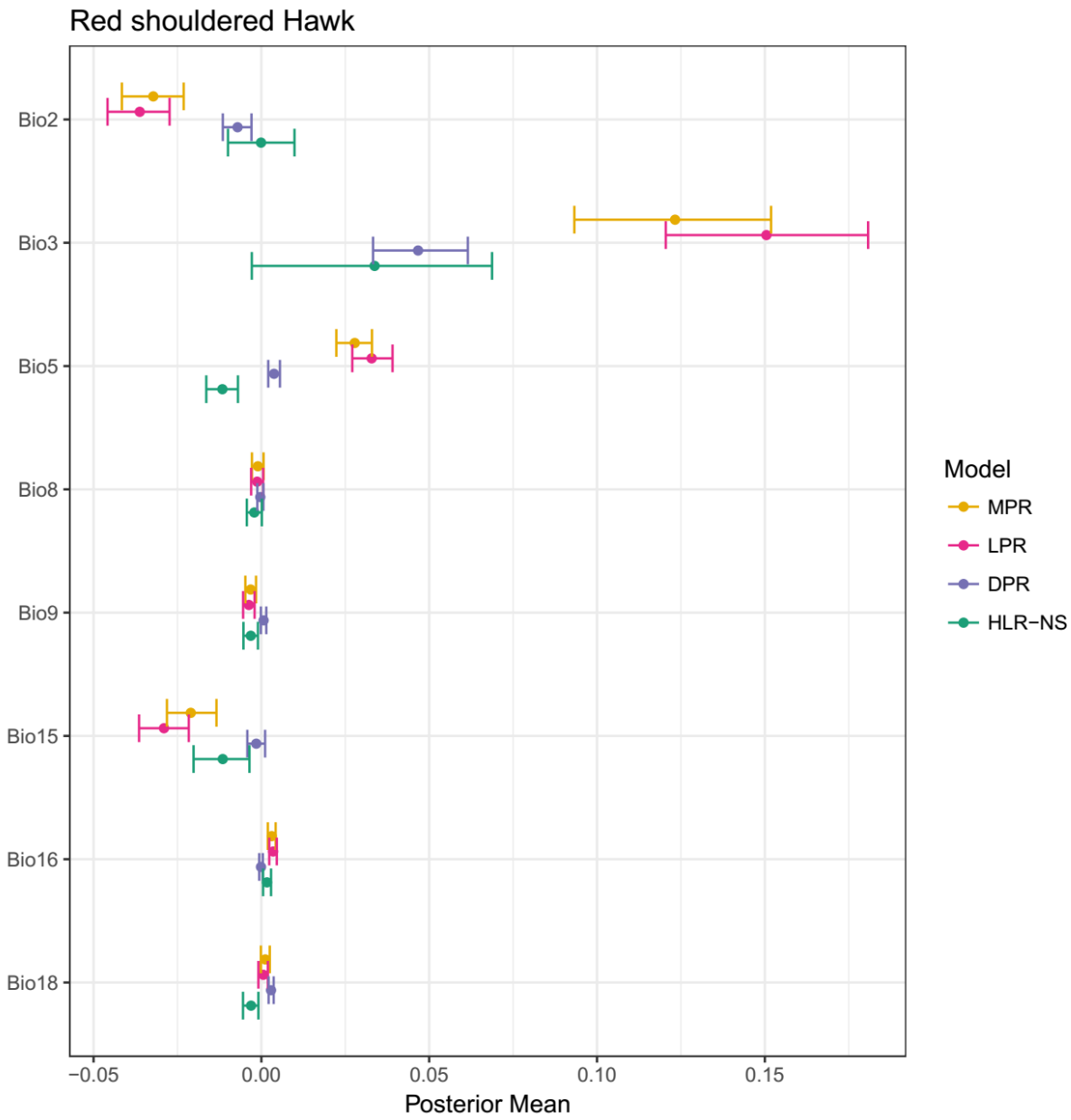


Figure 277: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-shouldered hawk, *Buteo lineatus*.

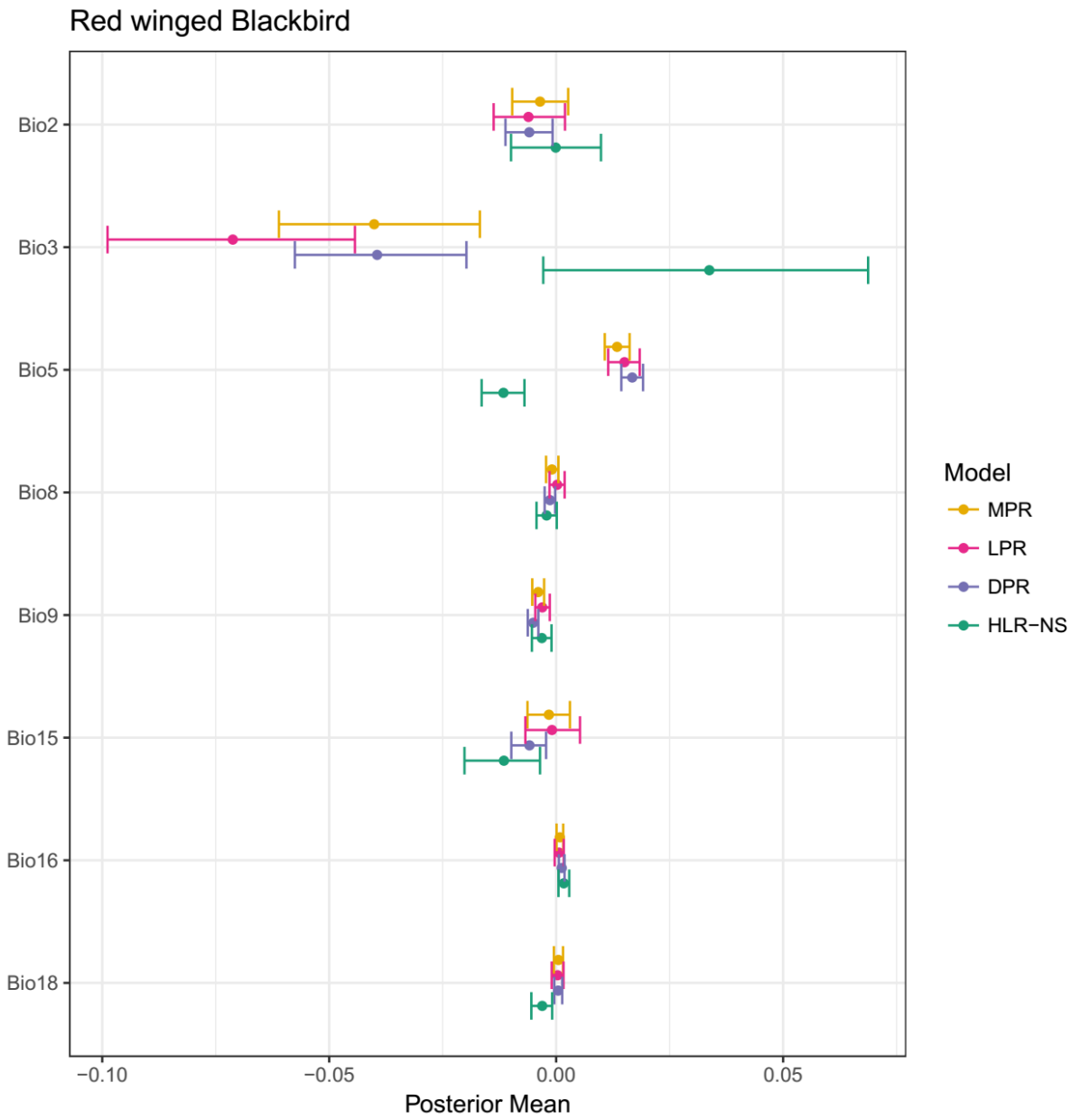


Figure 278: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red-winged blackbird, *Agelaius phoeniceus*.

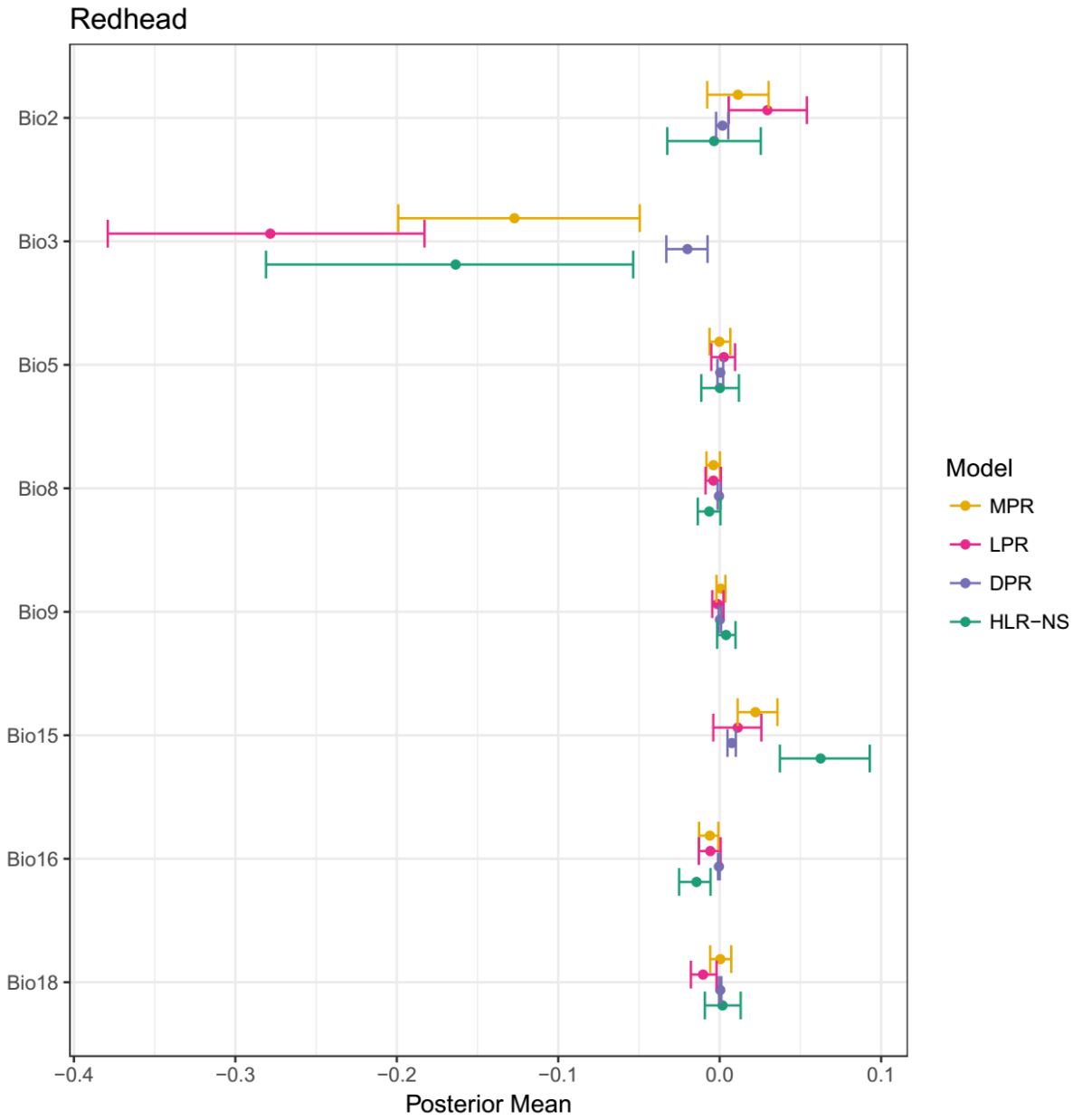


Figure 279: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the redhead, *Aythya americana*.

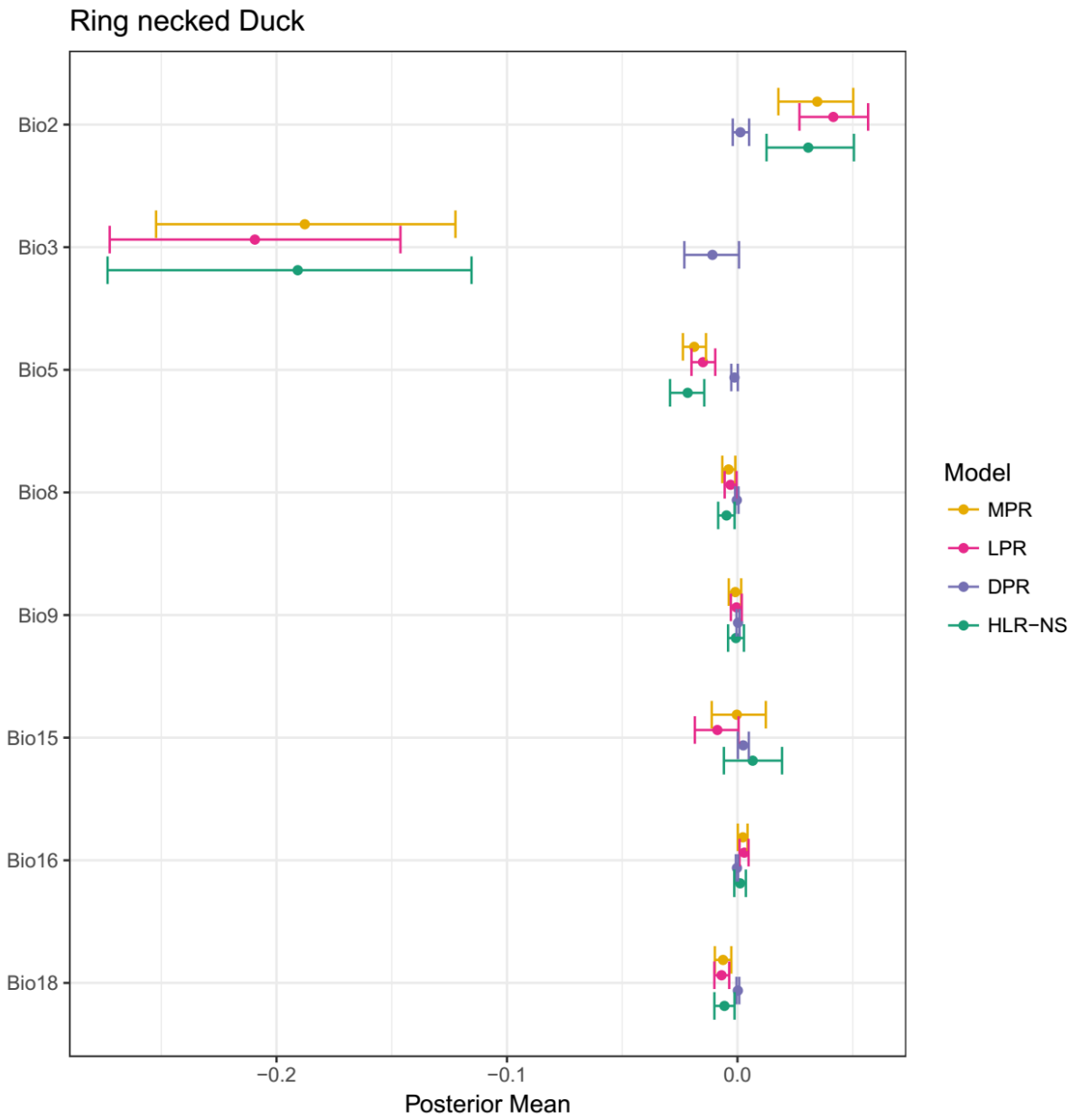


Figure 280: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ring-necked duck, *Aythya collaris*.

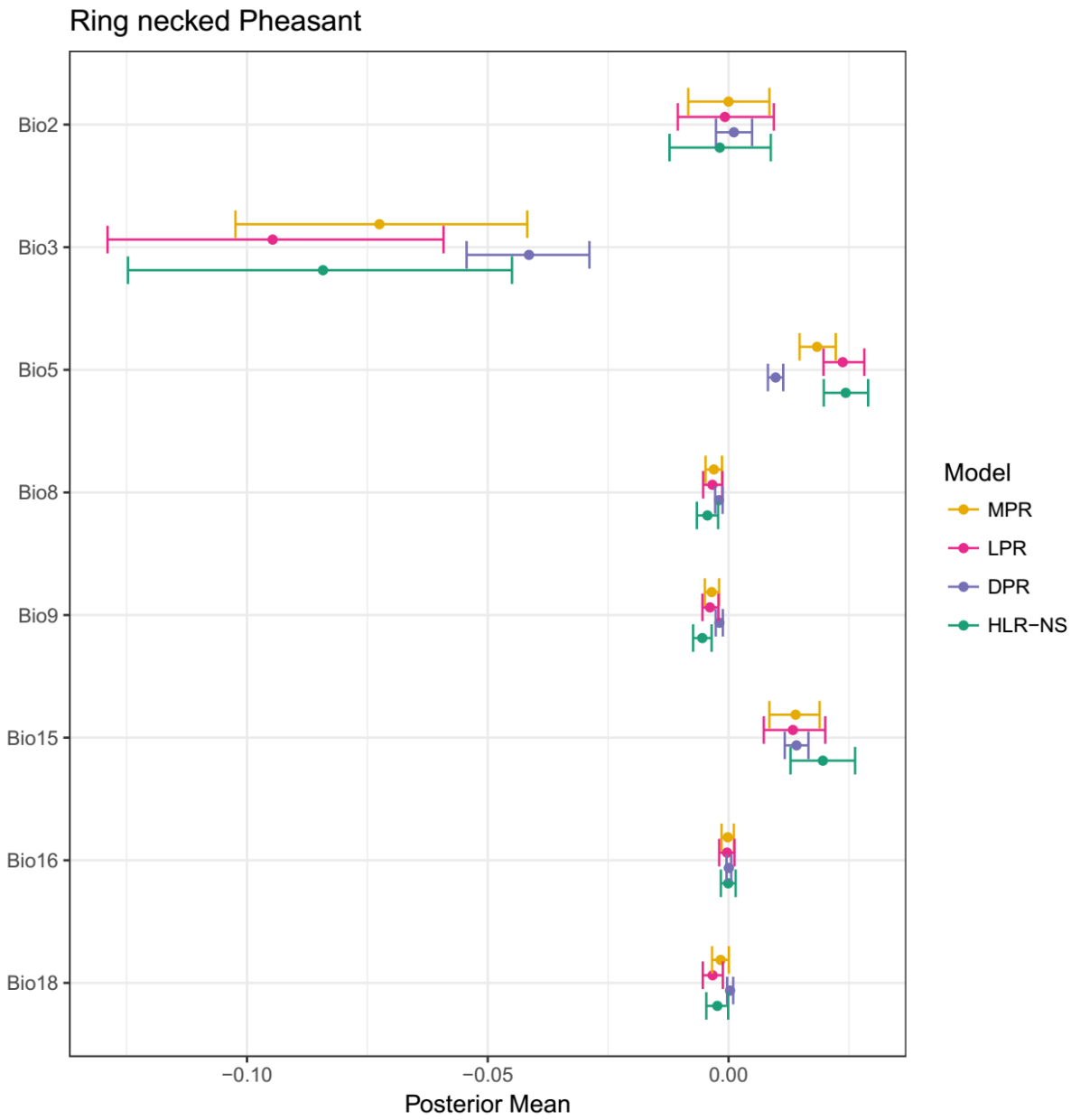


Figure 281: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ring-necked pheasant, *Phasianus colchicus*.

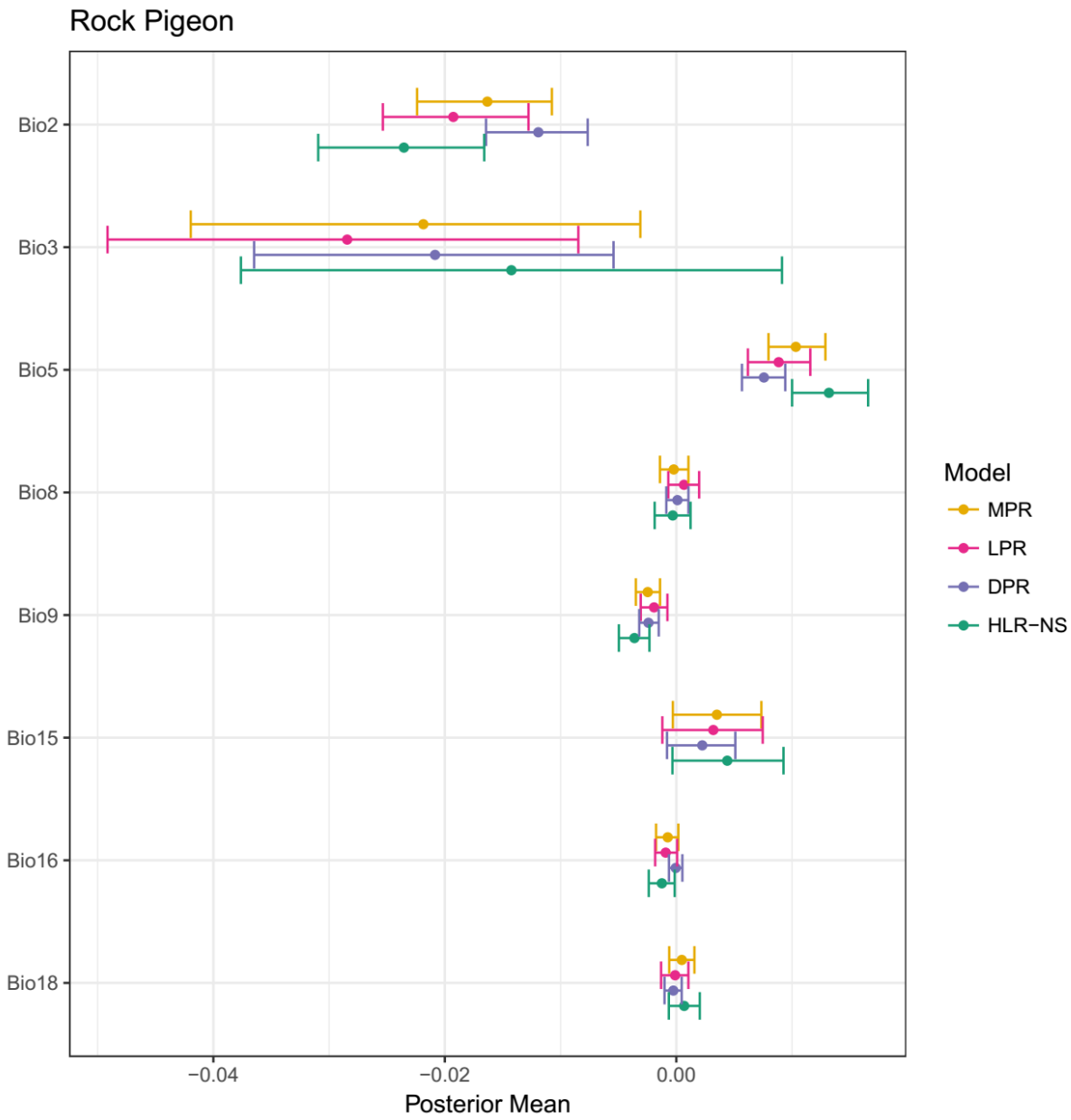


Figure 282: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the rock pigeon, *Columba livia*.

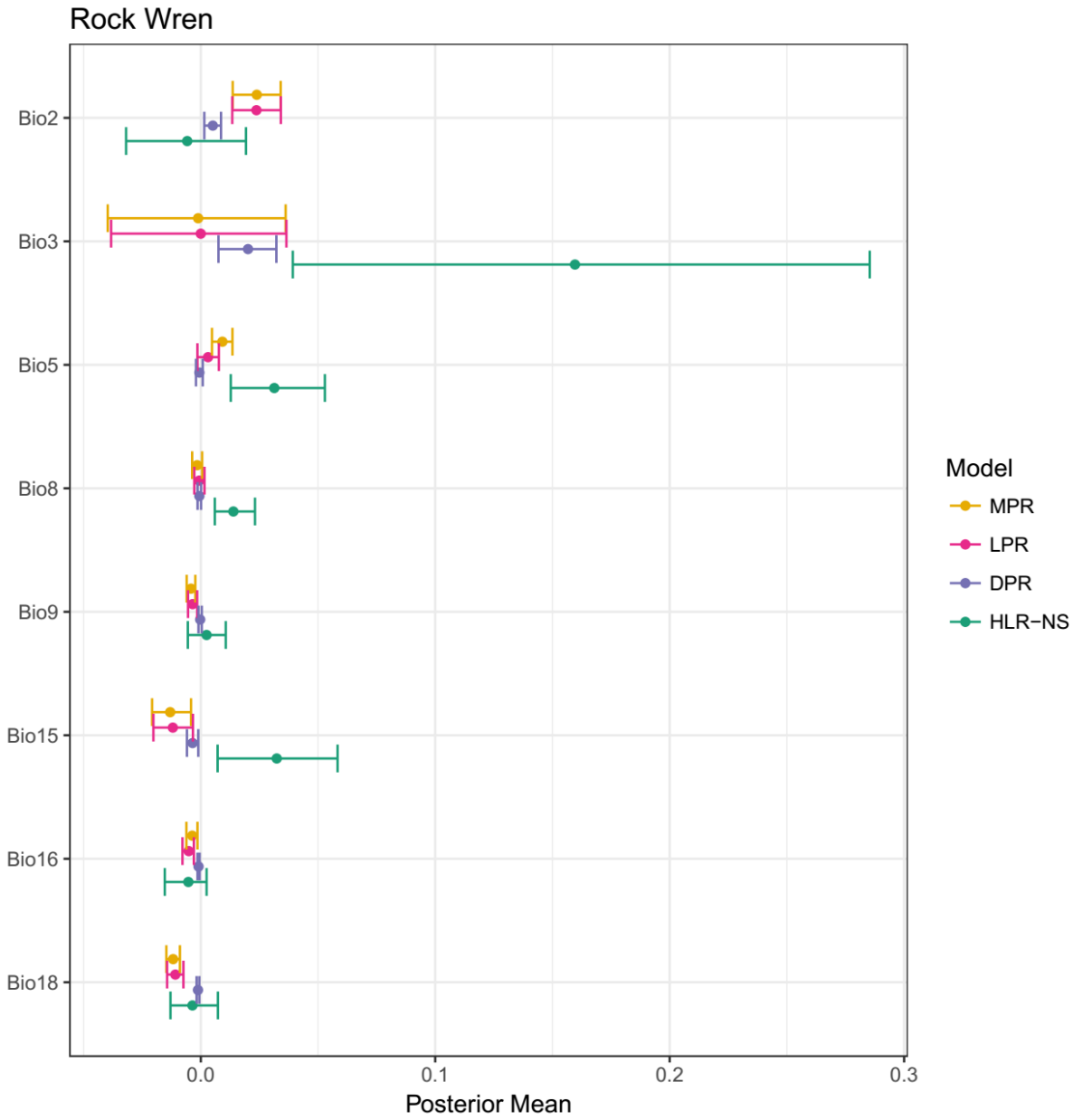


Figure 283: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the rock wren, *Salpinctes obsoletus*.

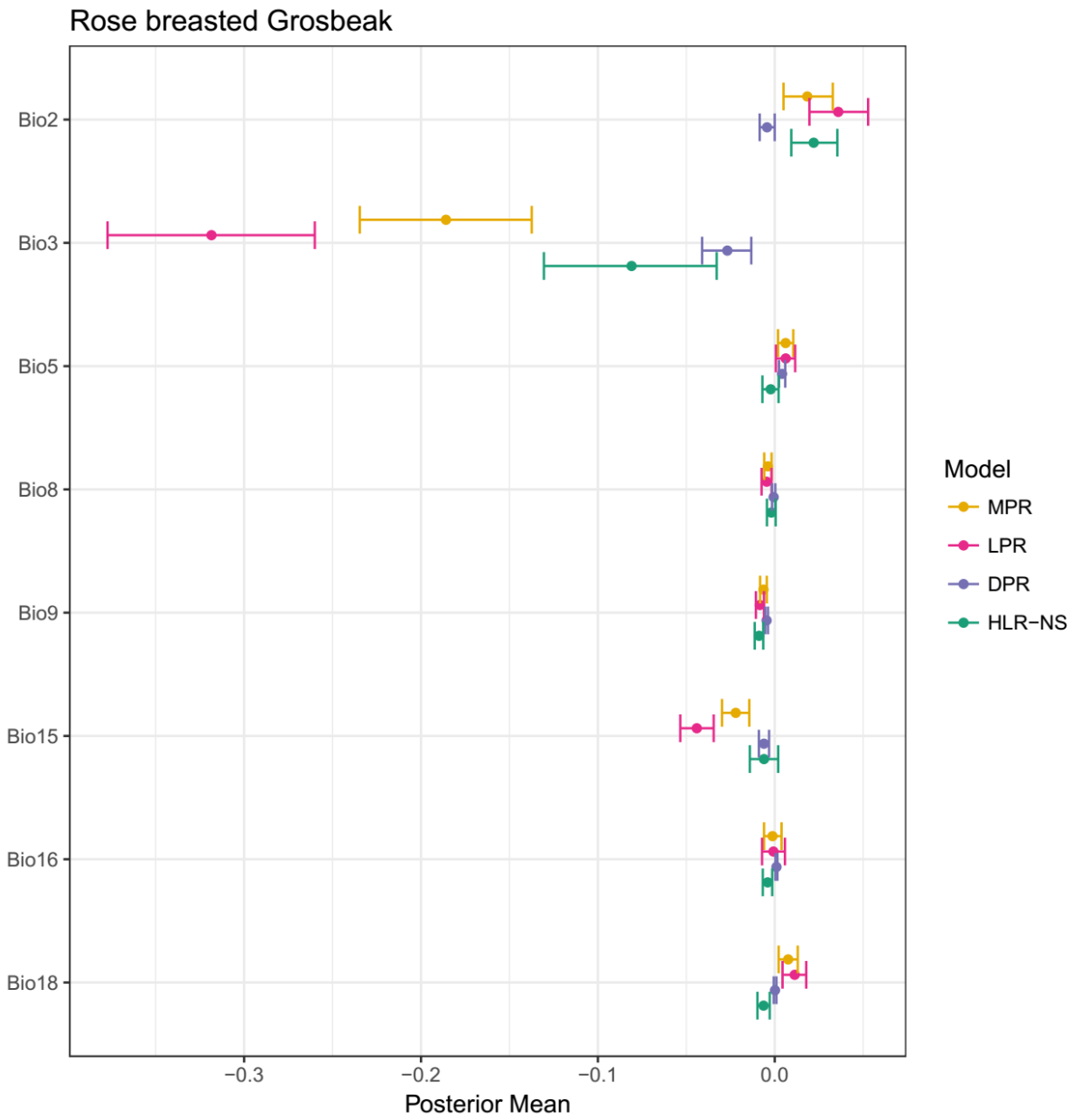


Figure 284: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the rose-breasted grosbeak, *Pheucticus ludovicianus*.

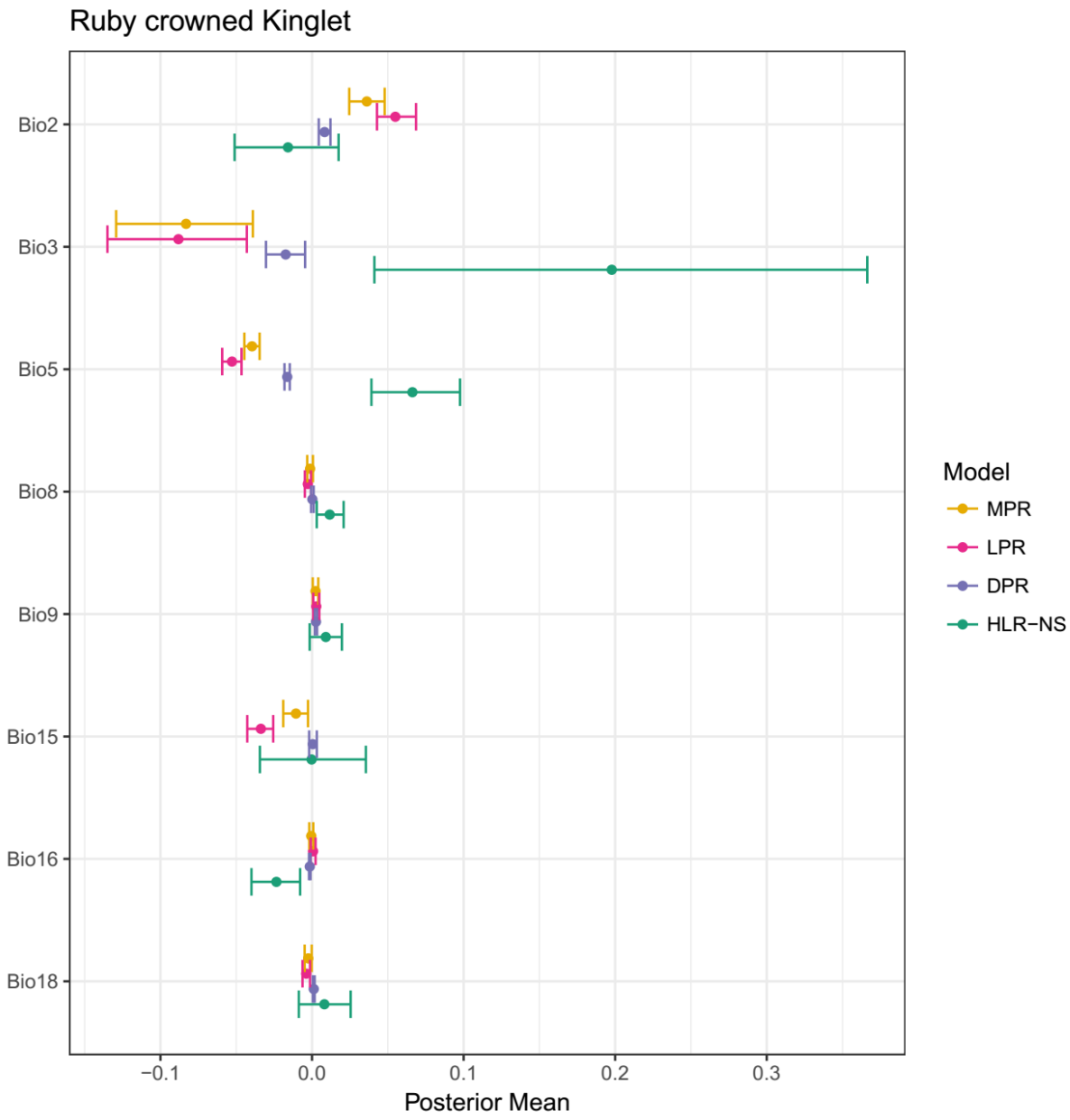


Figure 285: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ruby-crowned kinglet, *Regulus calendula*.

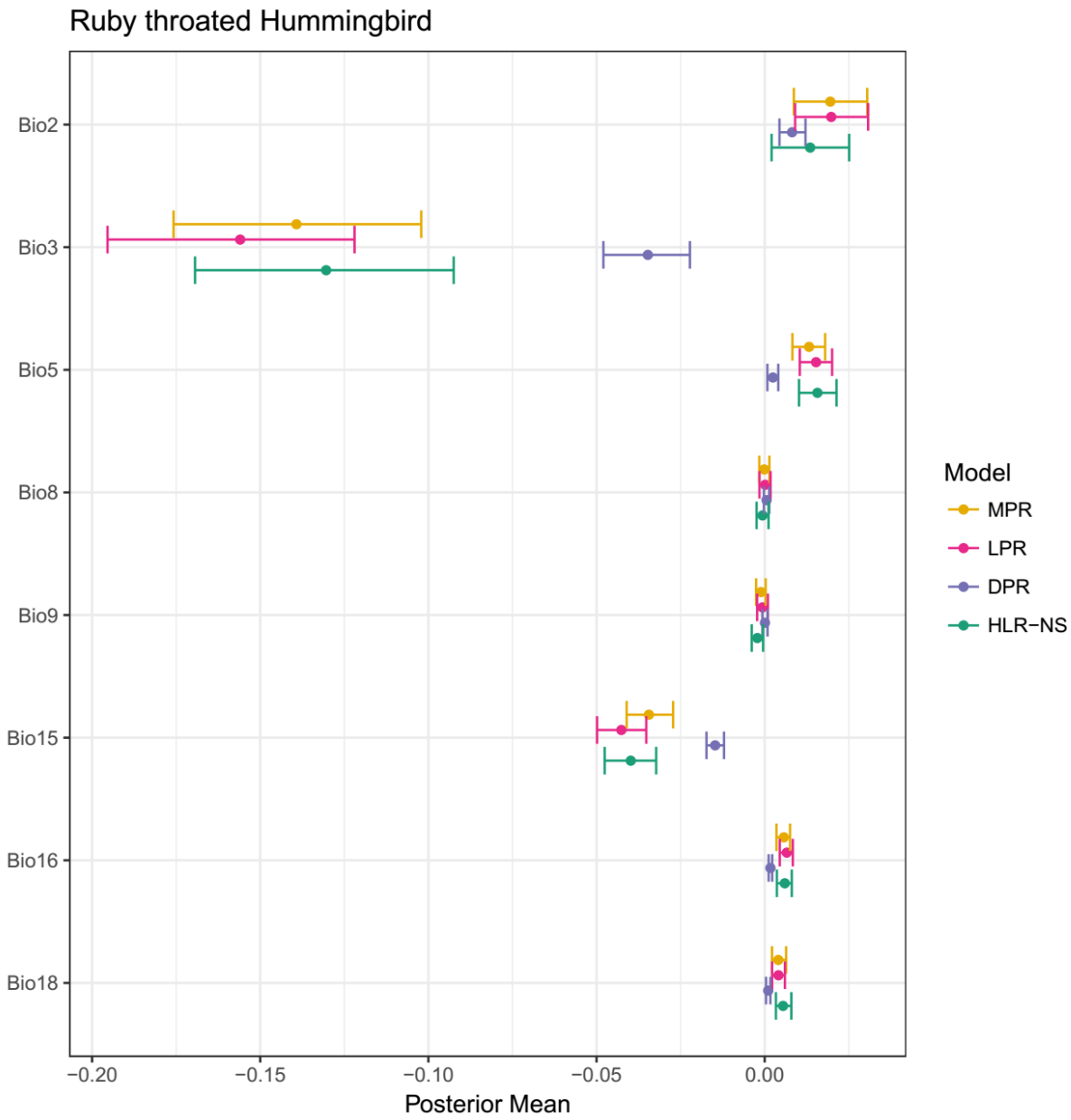


Figure 286: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ruby-throated hummingbird, *Archilochus colubris*.

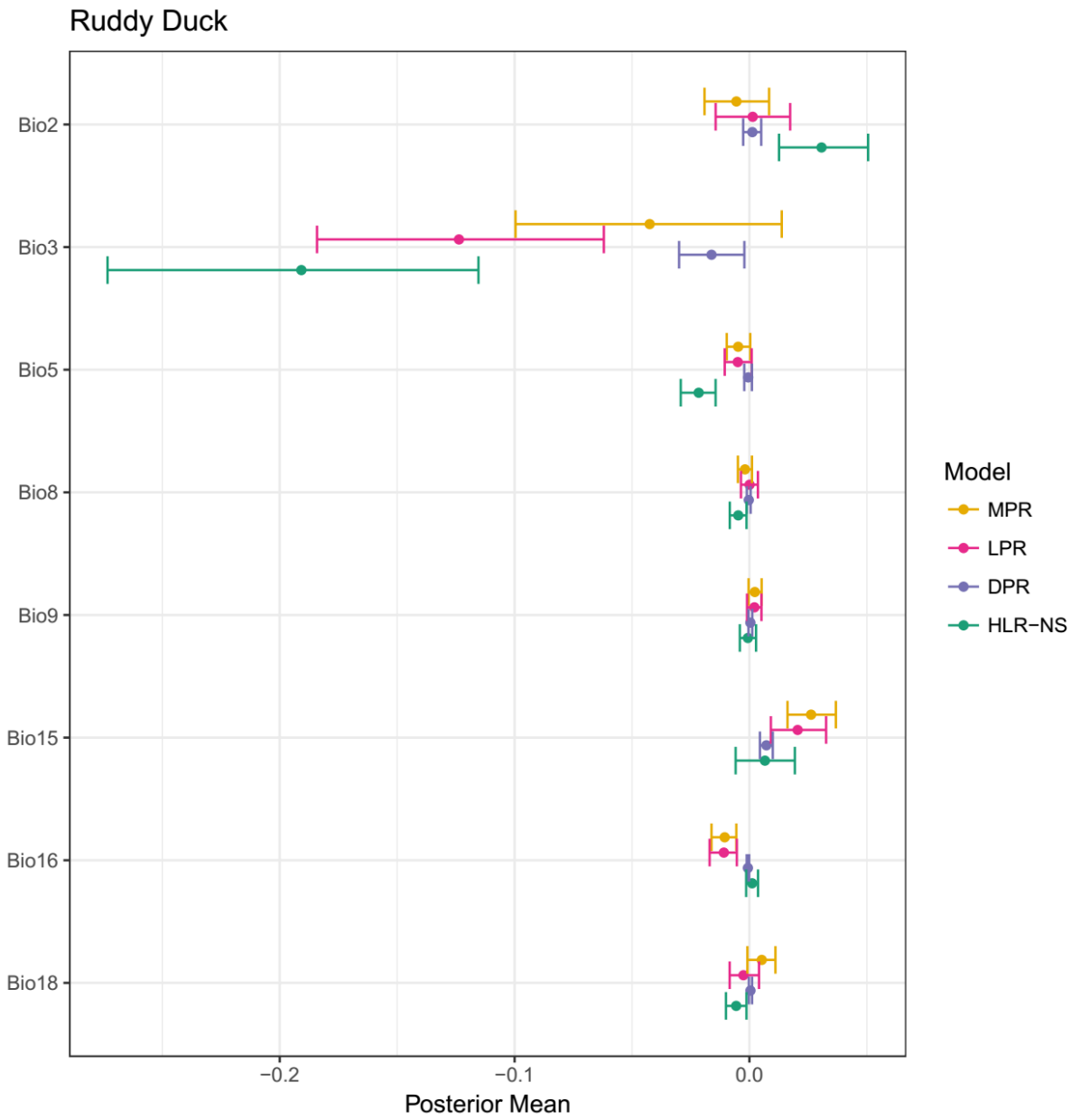


Figure 287: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ruddy duck, *Oxyura jamaicensis*.

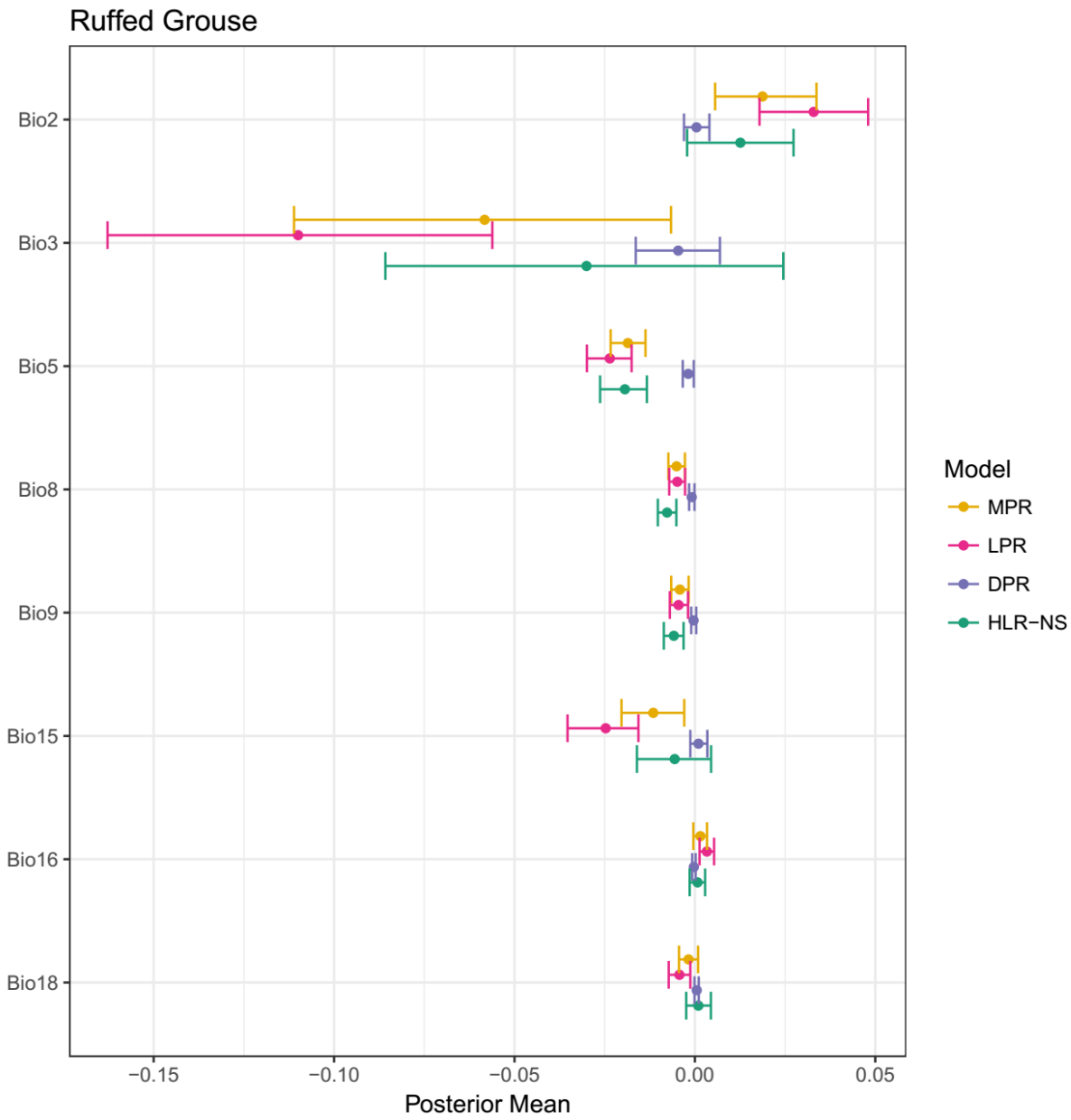


Figure 288: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ruffed grouse, *Bonasa umbellus*.

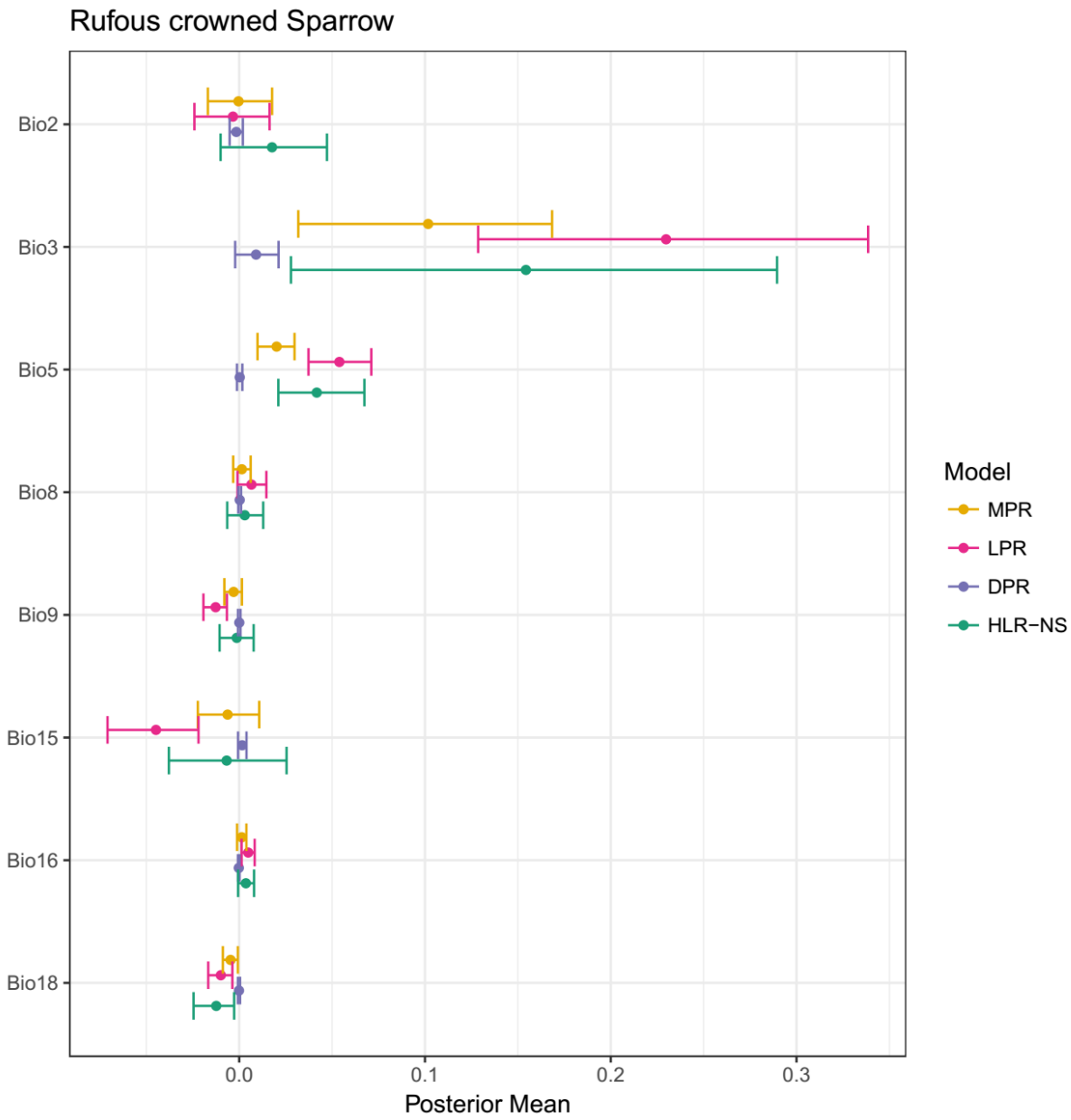


Figure 289: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the rufous-crowned sparrow, *Aimophila ruficeps*.

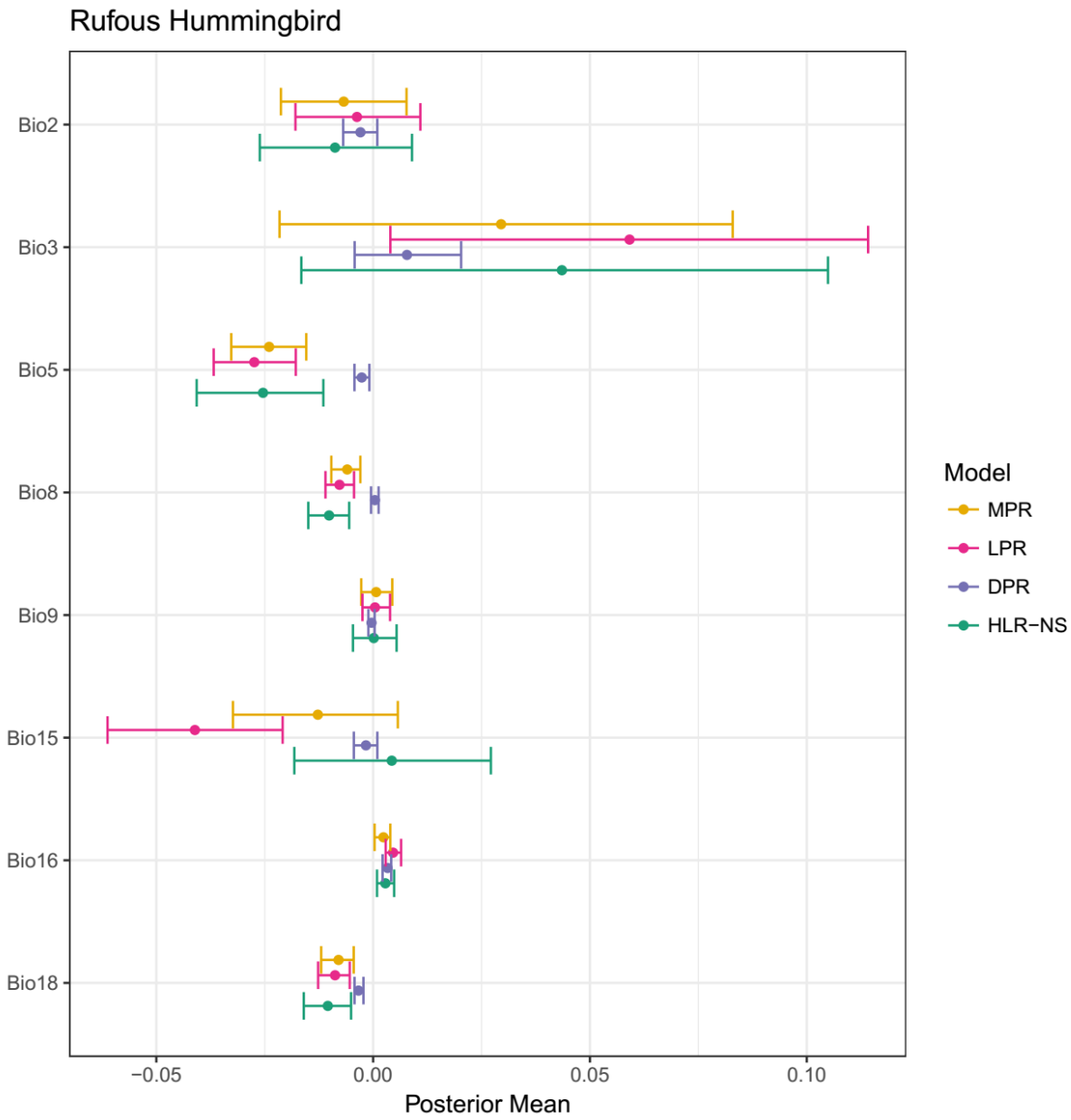


Figure 290: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the rufous hummingbird, *Selasphorus rufus*.

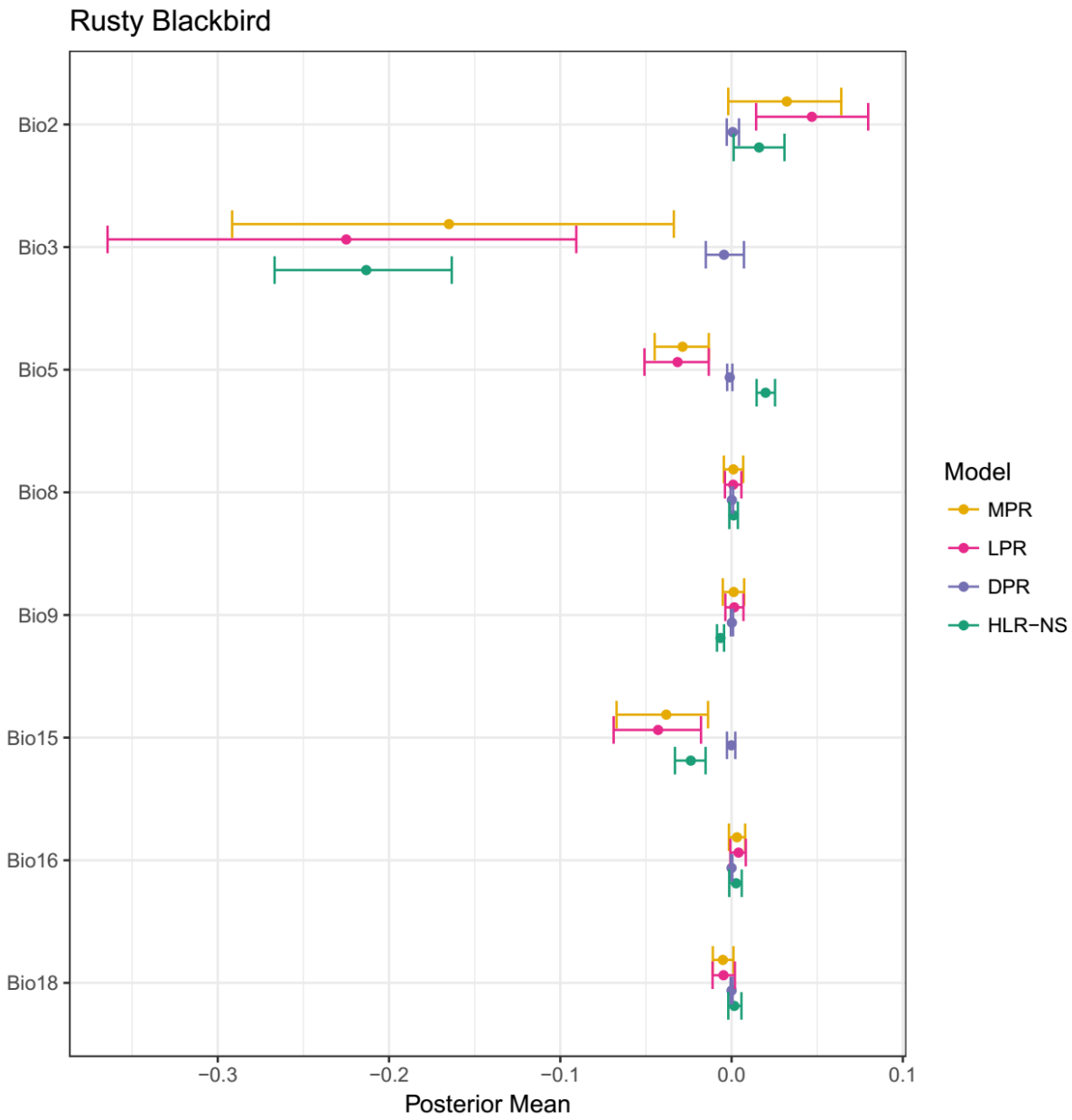


Figure 291: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the rusty blackbird, *Euphagus carolinus*.

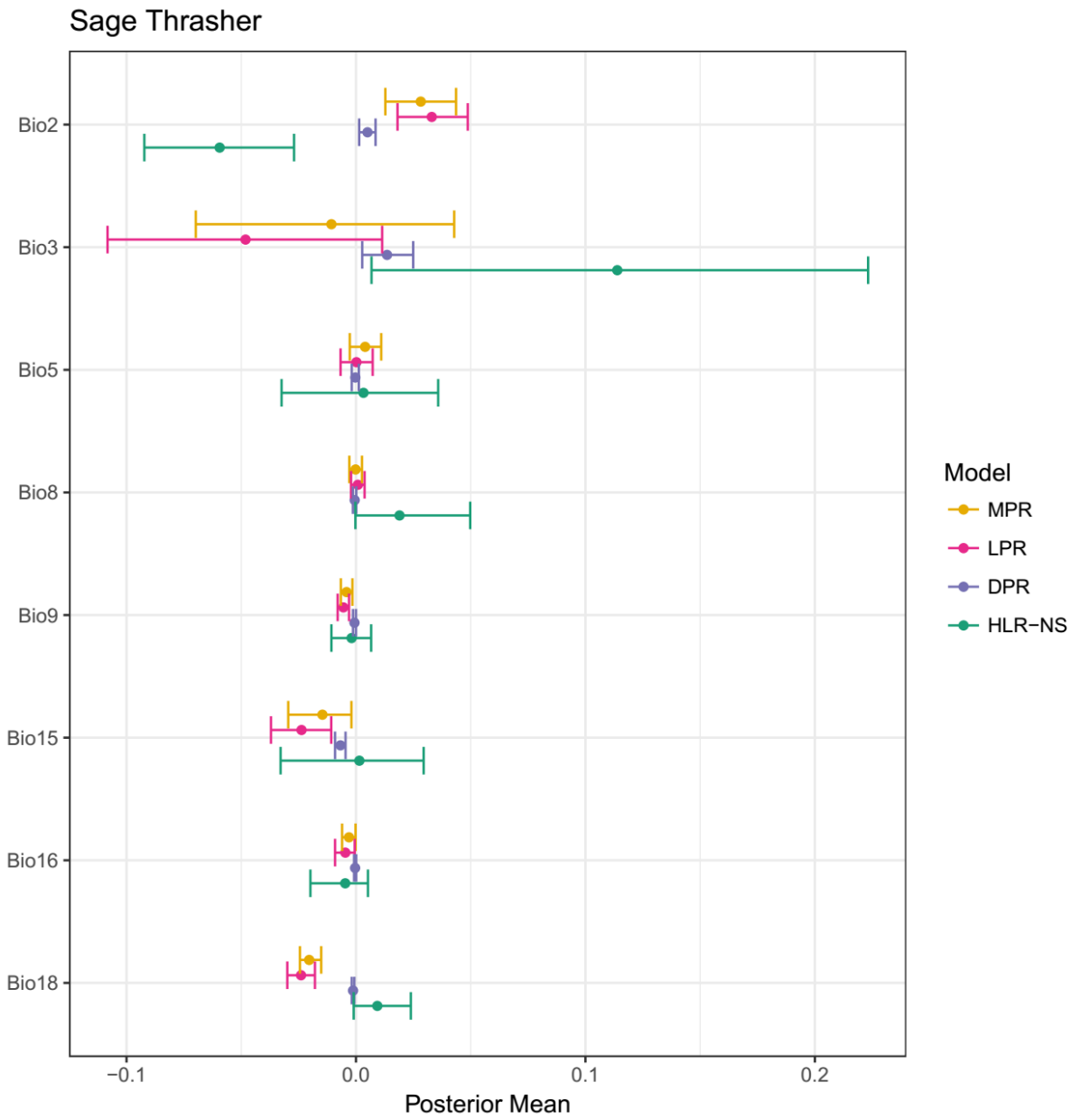


Figure 292: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sage thrasher, *Oreoscoptes montanus*.

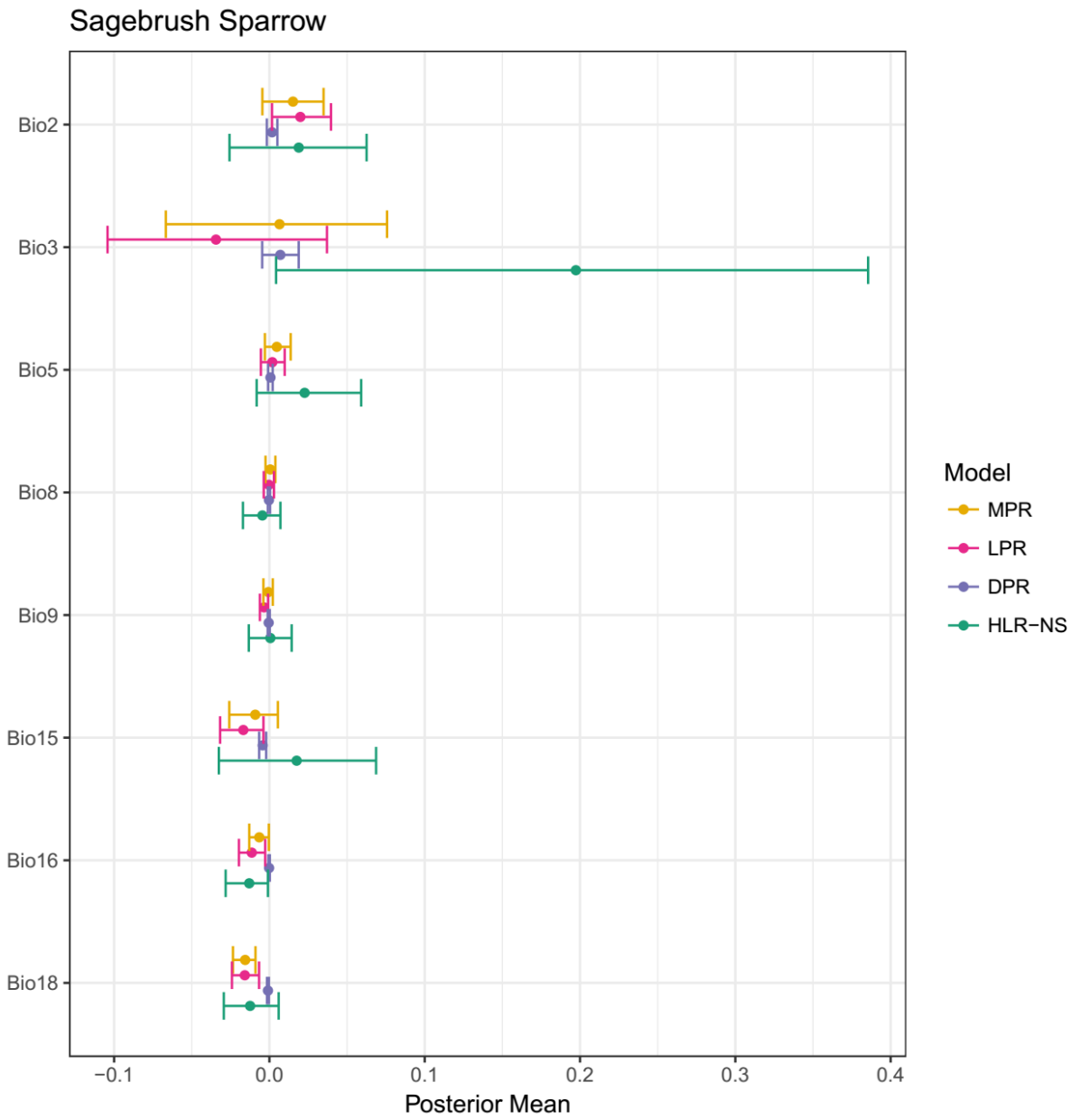


Figure 293: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sagebrush sparrow, *Artemisiospiza nevadensis*.

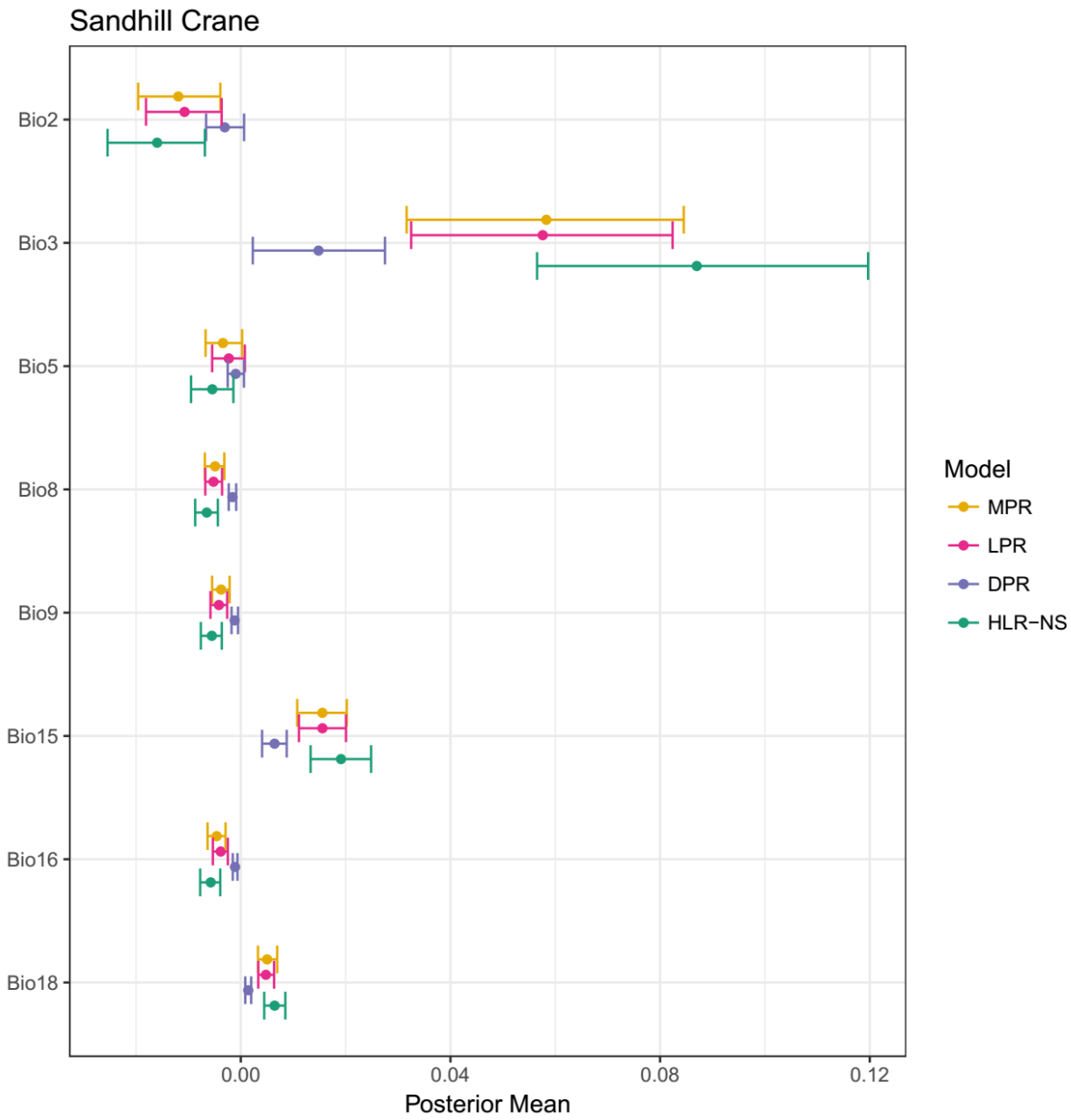


Figure 294: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sandhill crane, *Grus canadensis*.

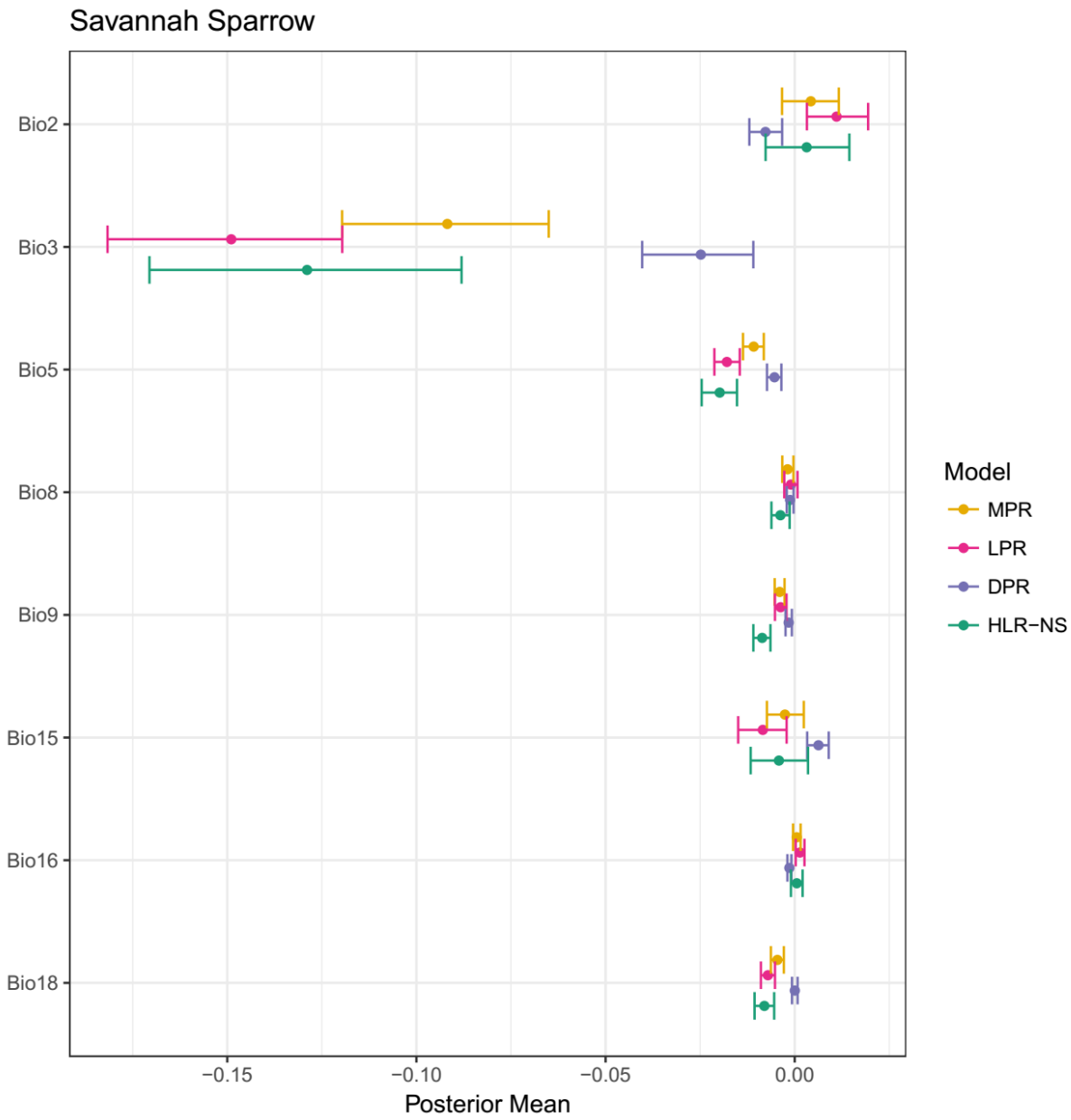


Figure 295: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the savannah sparrow, *Passerculus sandwichensis*.

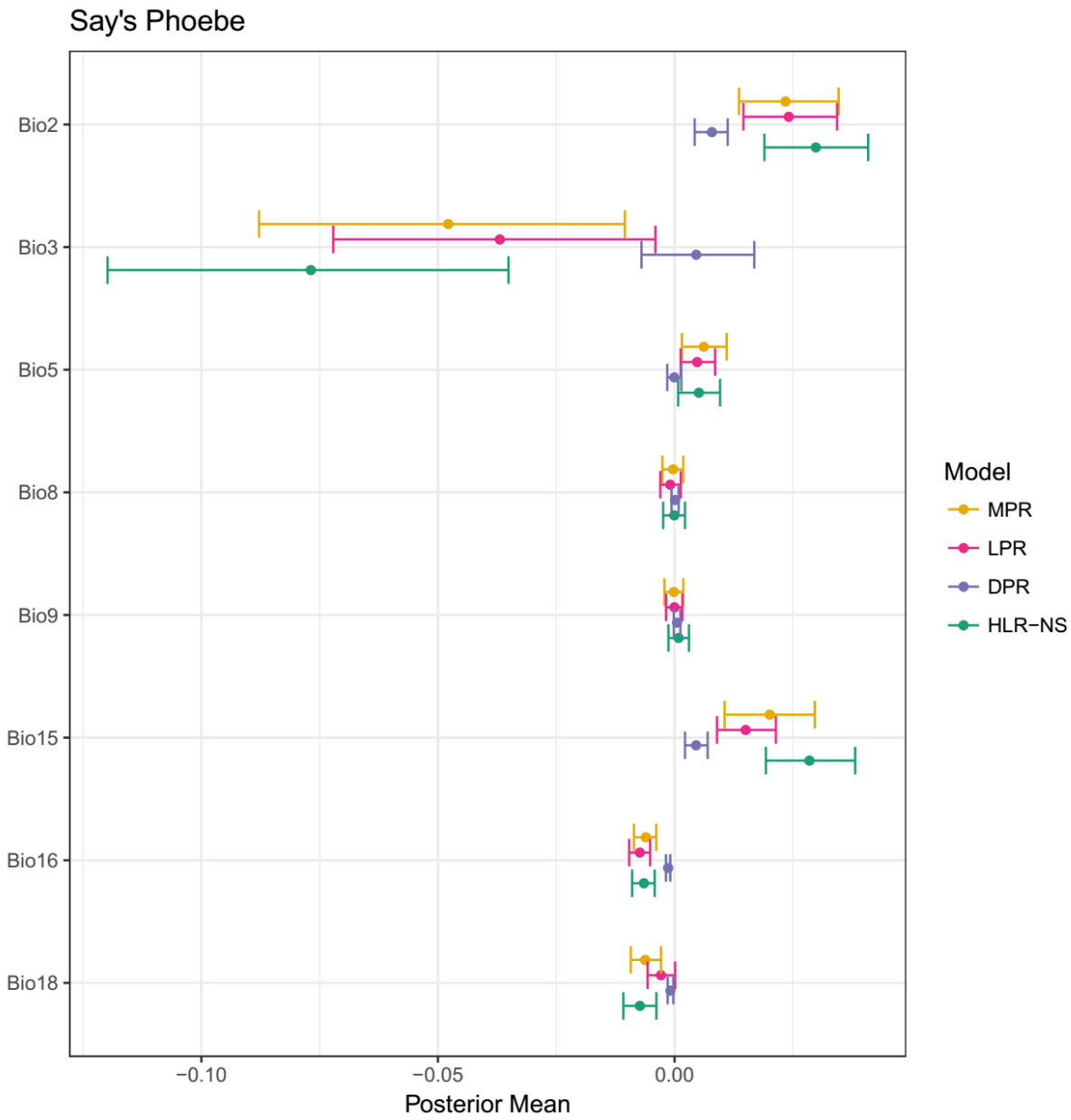


Figure 296: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Say's phoebe, *Sayornis saya*.

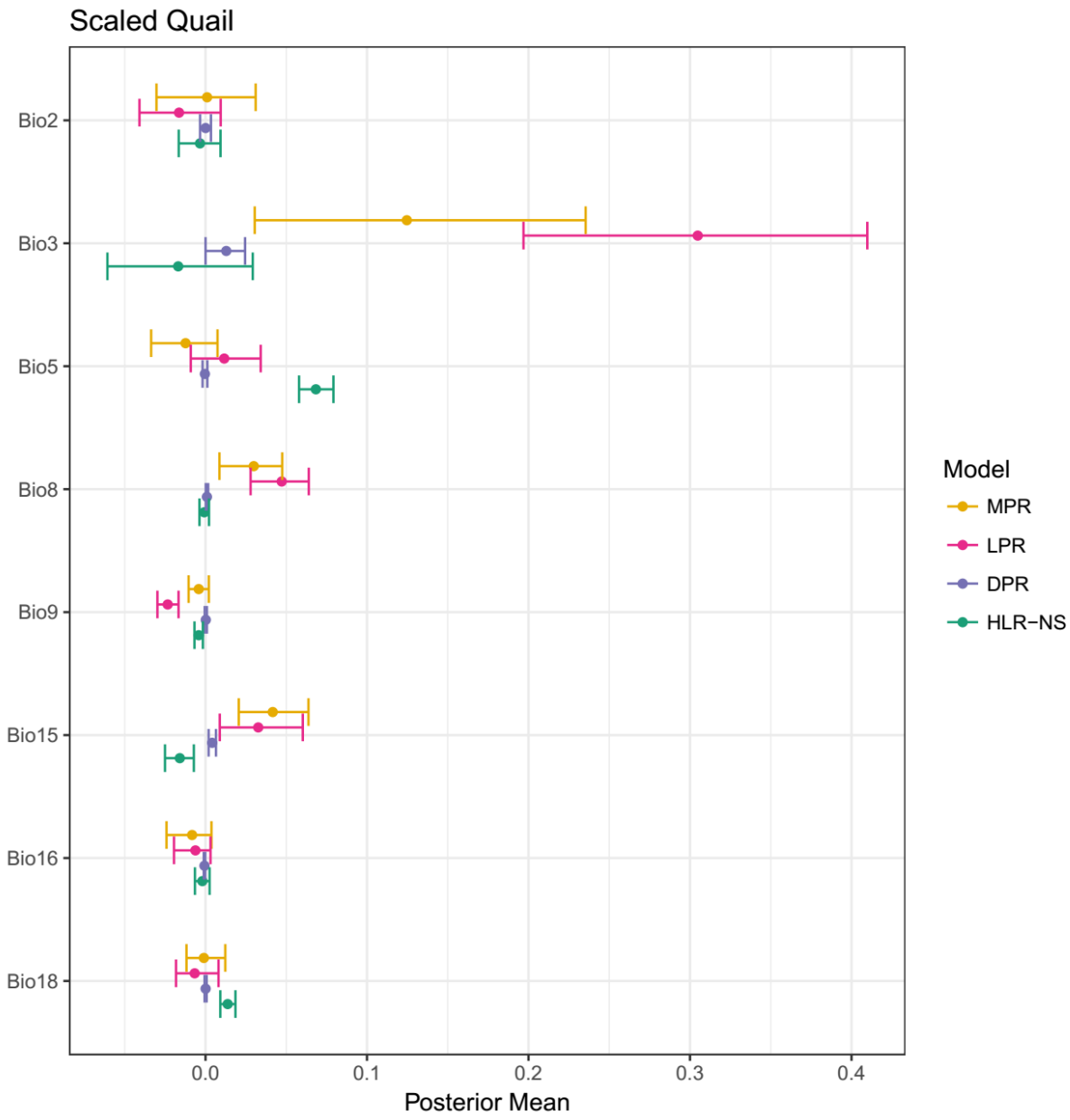


Figure 297: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the scaled quail, *Callipepla squamata*.

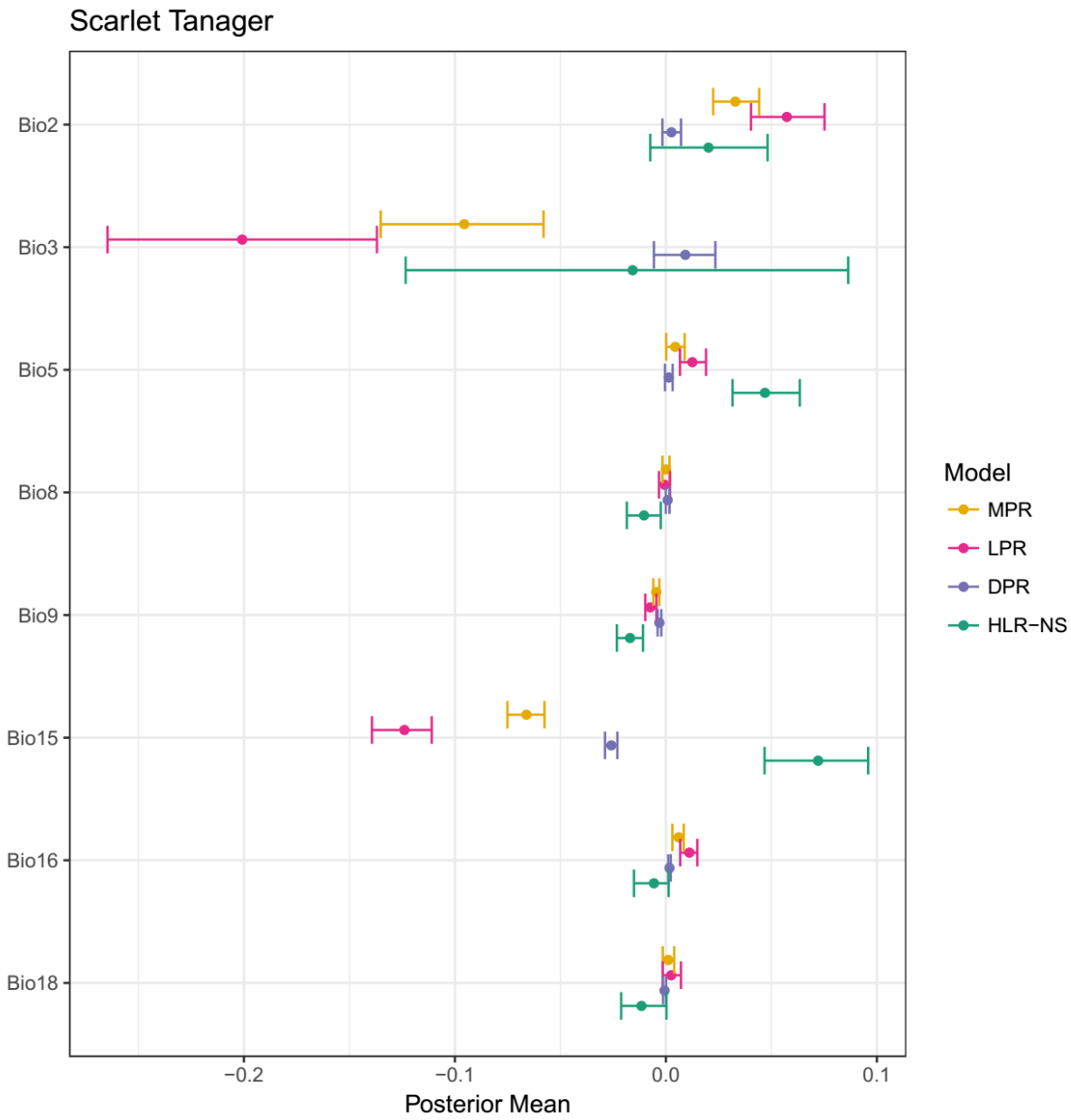


Figure 298: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the scarlet tanager, *Piranga olivacea*.

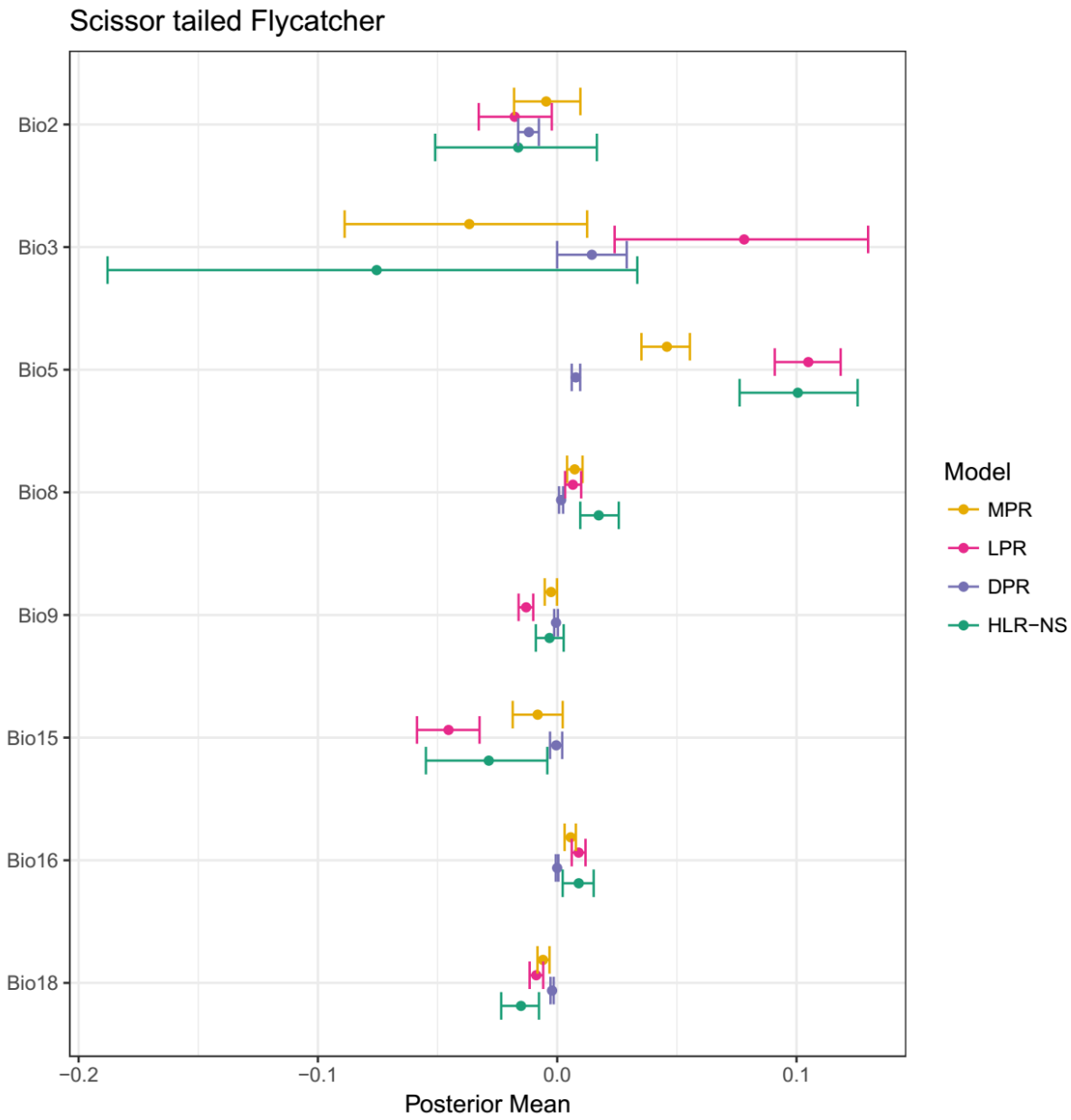


Figure 299: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the scissor-tailed flycatcher, *Tyrannus forficatus*.

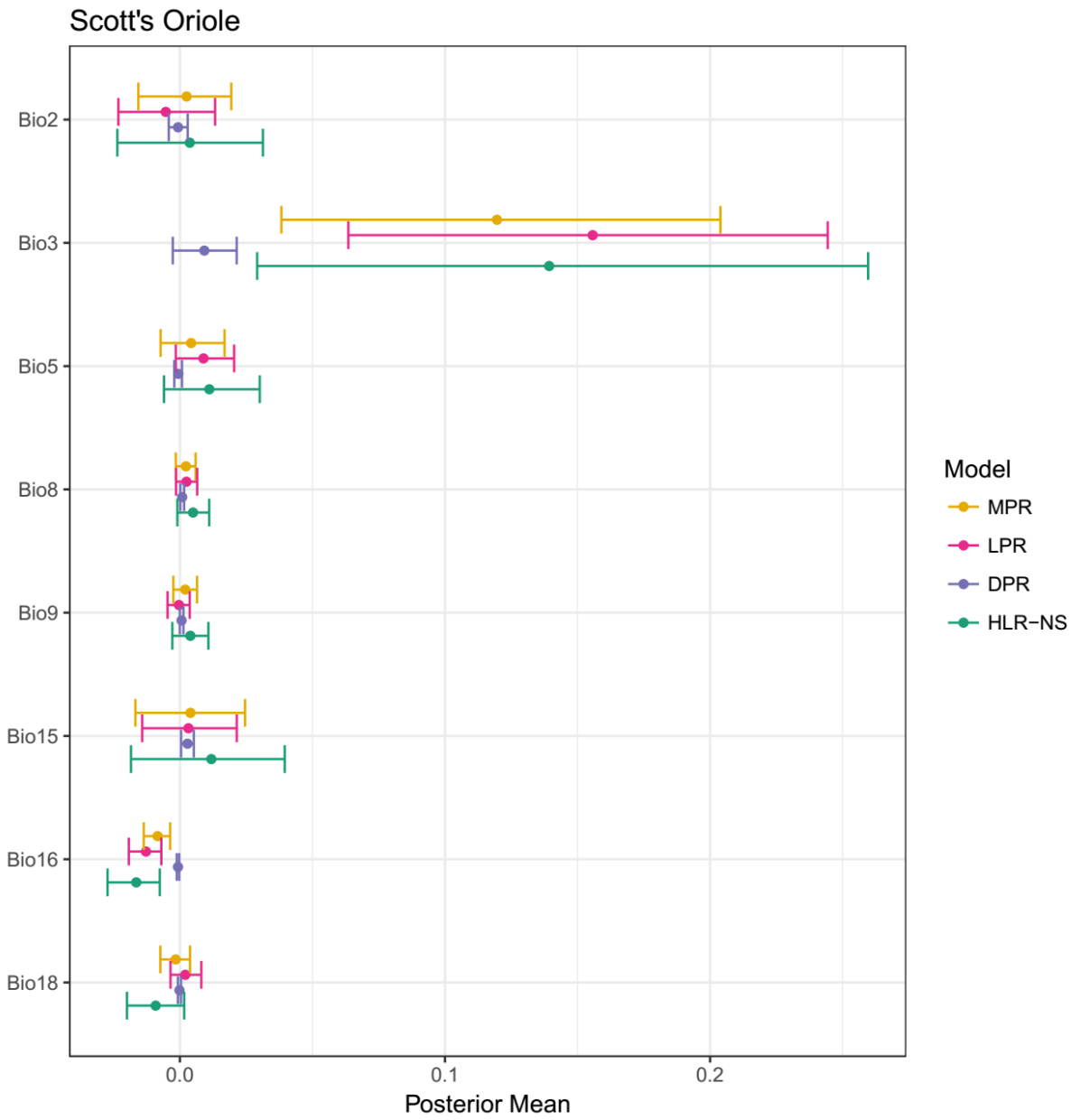


Figure 300: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Scott's oriole, *Icterus parisorum*.

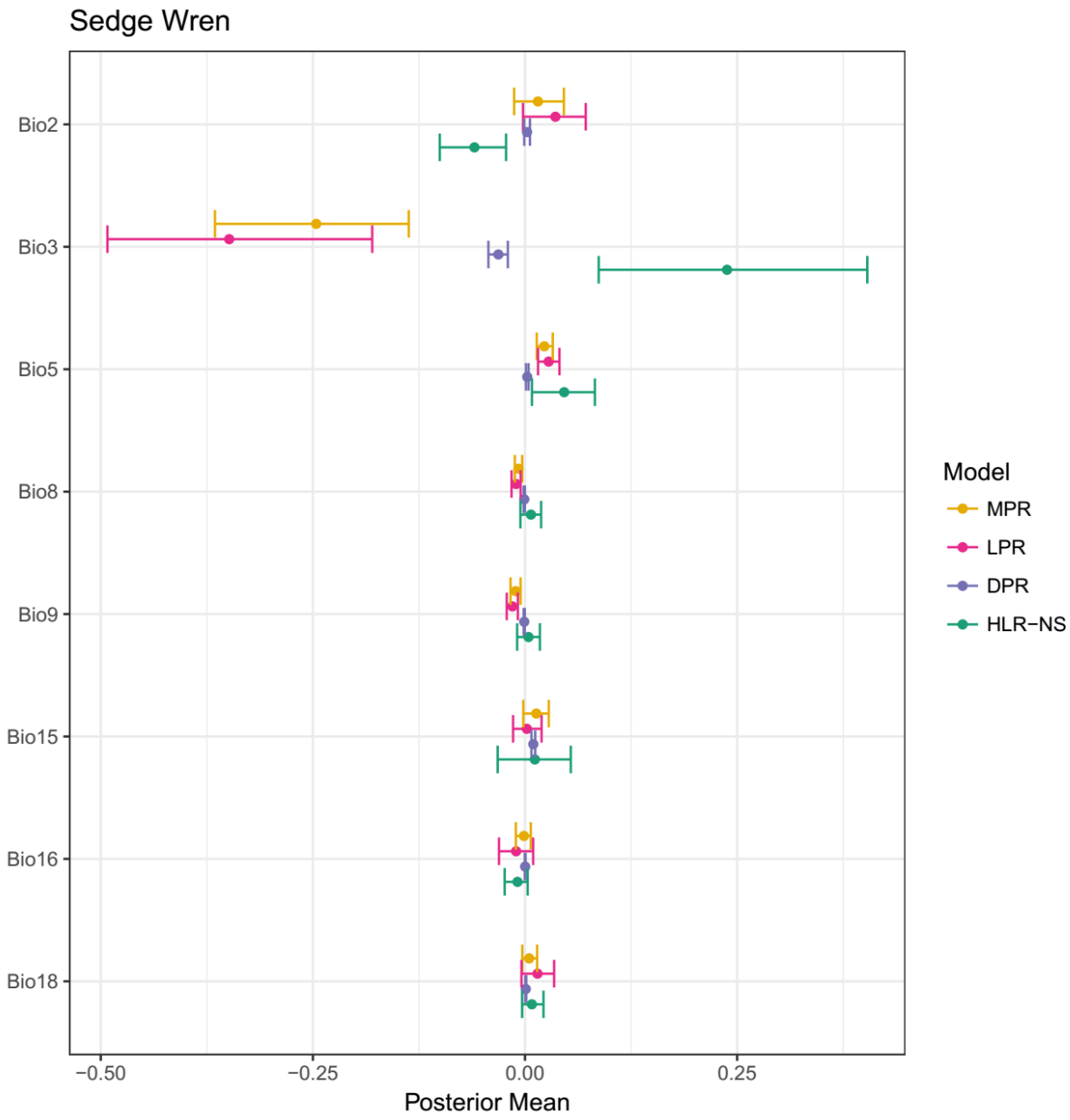


Figure 301: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sedge wren, *Cistothorus platensis*.

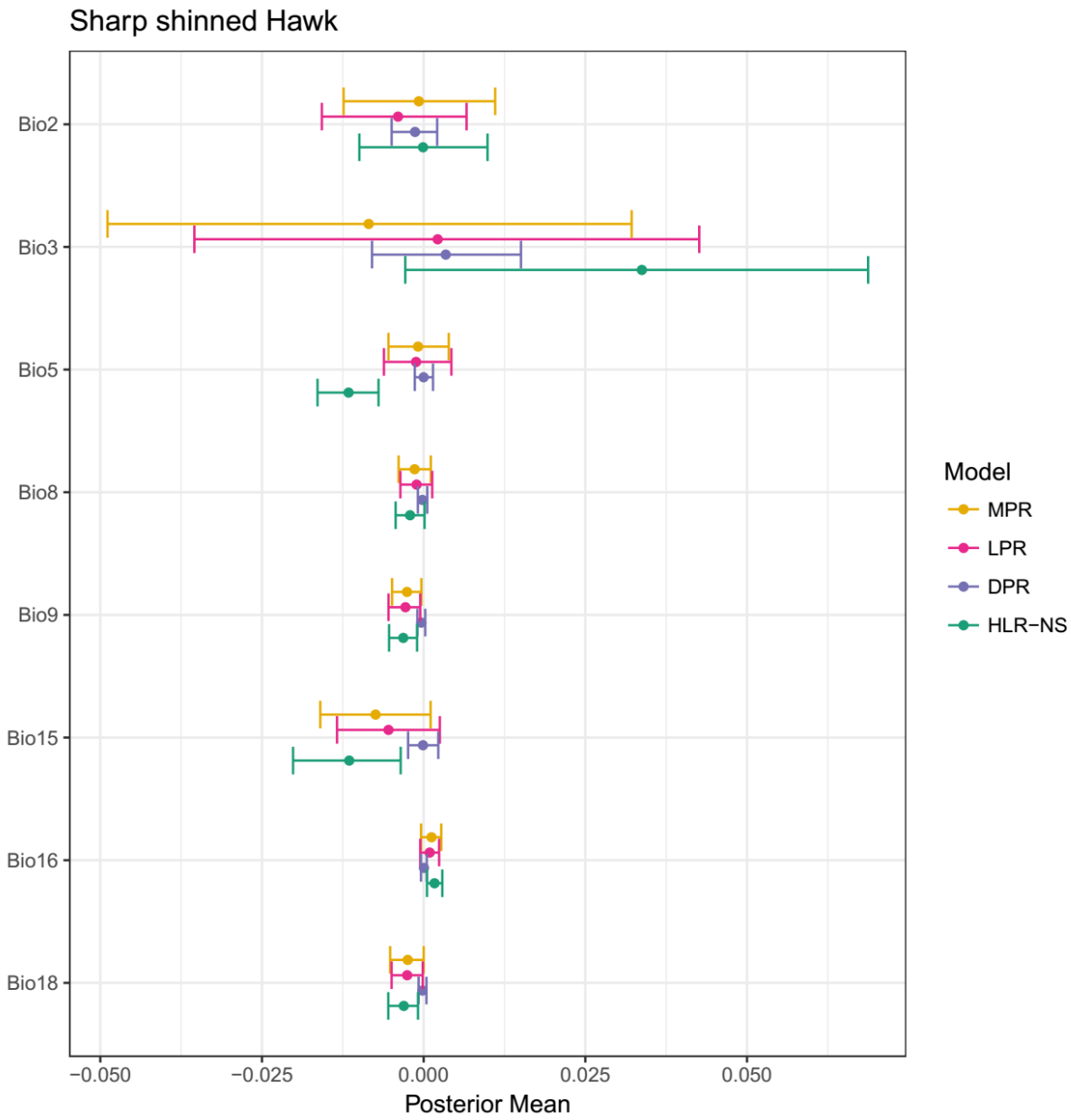


Figure 302: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sharp-shinned hawk, *Accipiter striatus*.

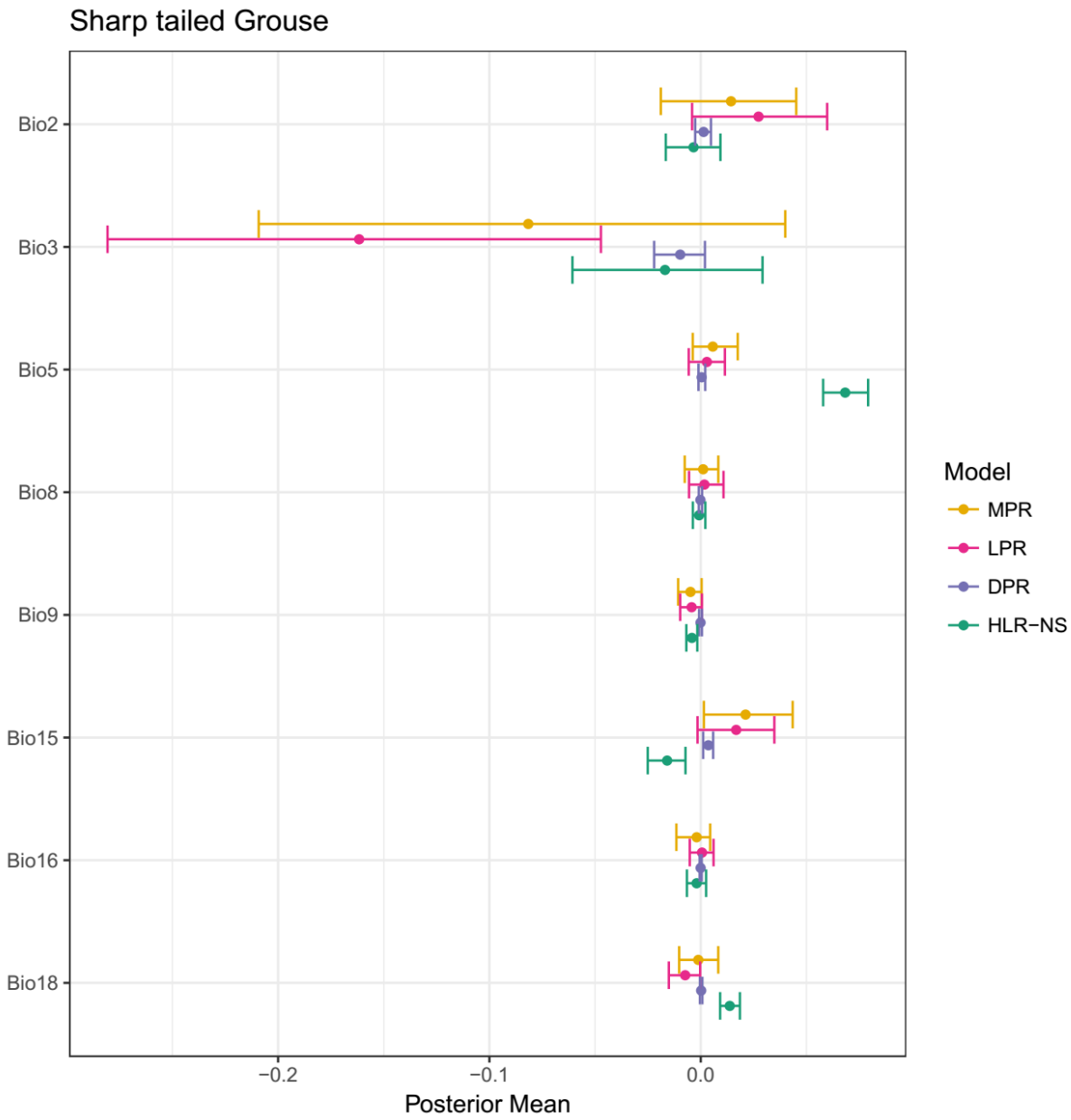


Figure 303: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sharp-tailed grouse, *Tympanuchus phasianellus*.

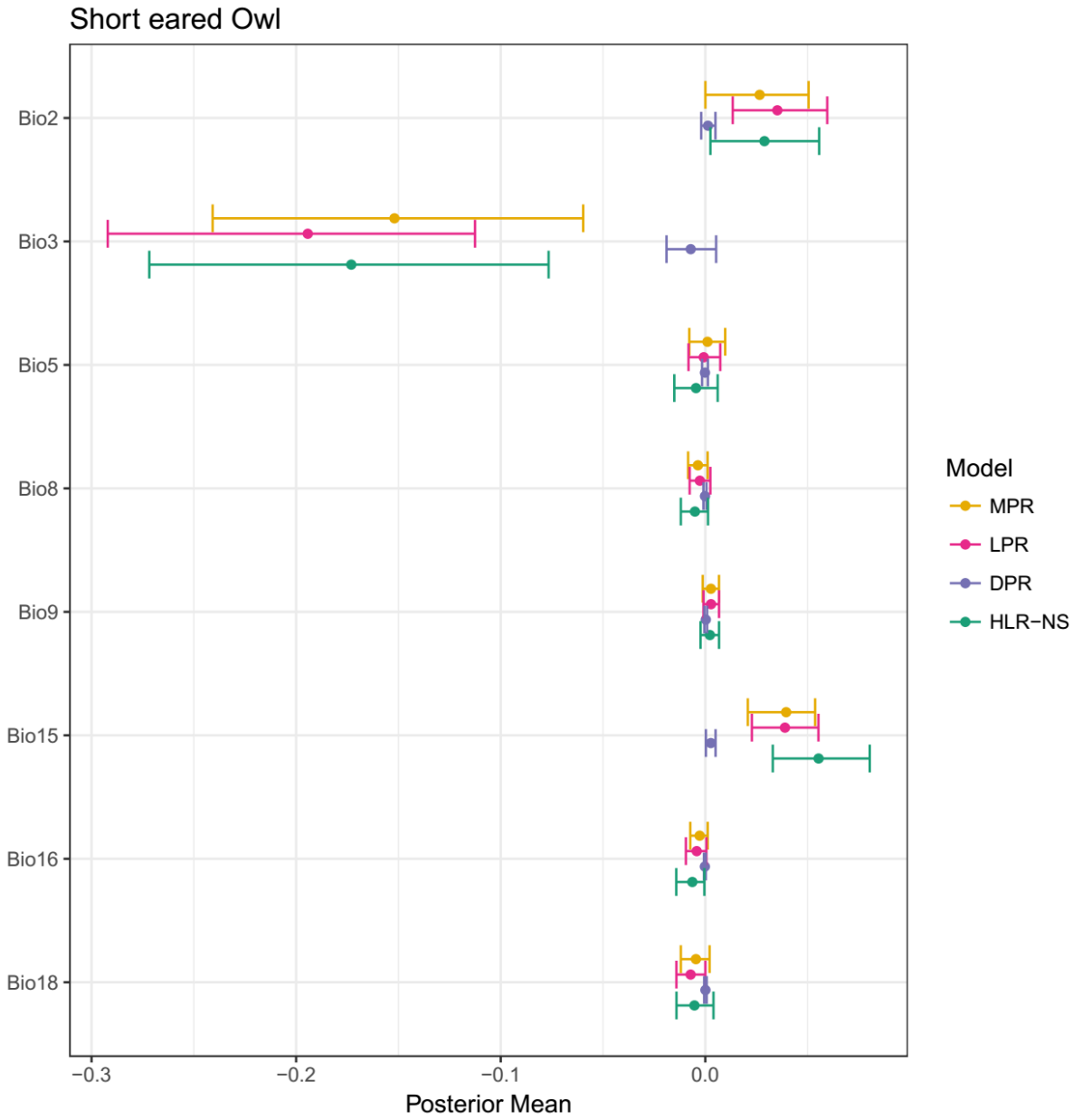


Figure 304: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the short-eared owl, *Asio flammeus*.

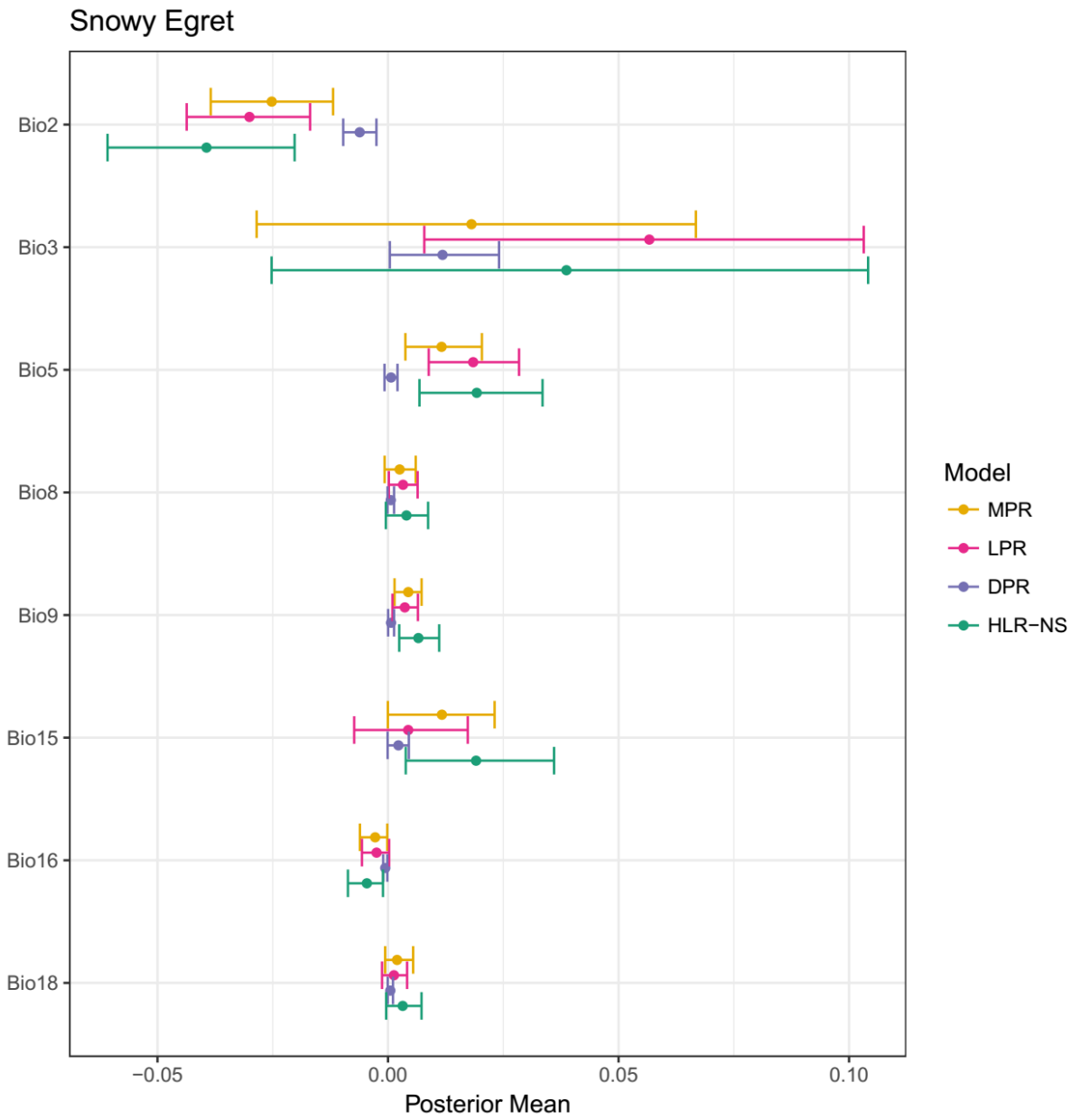


Figure 305: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the snowy egret, *Egretta thula*.

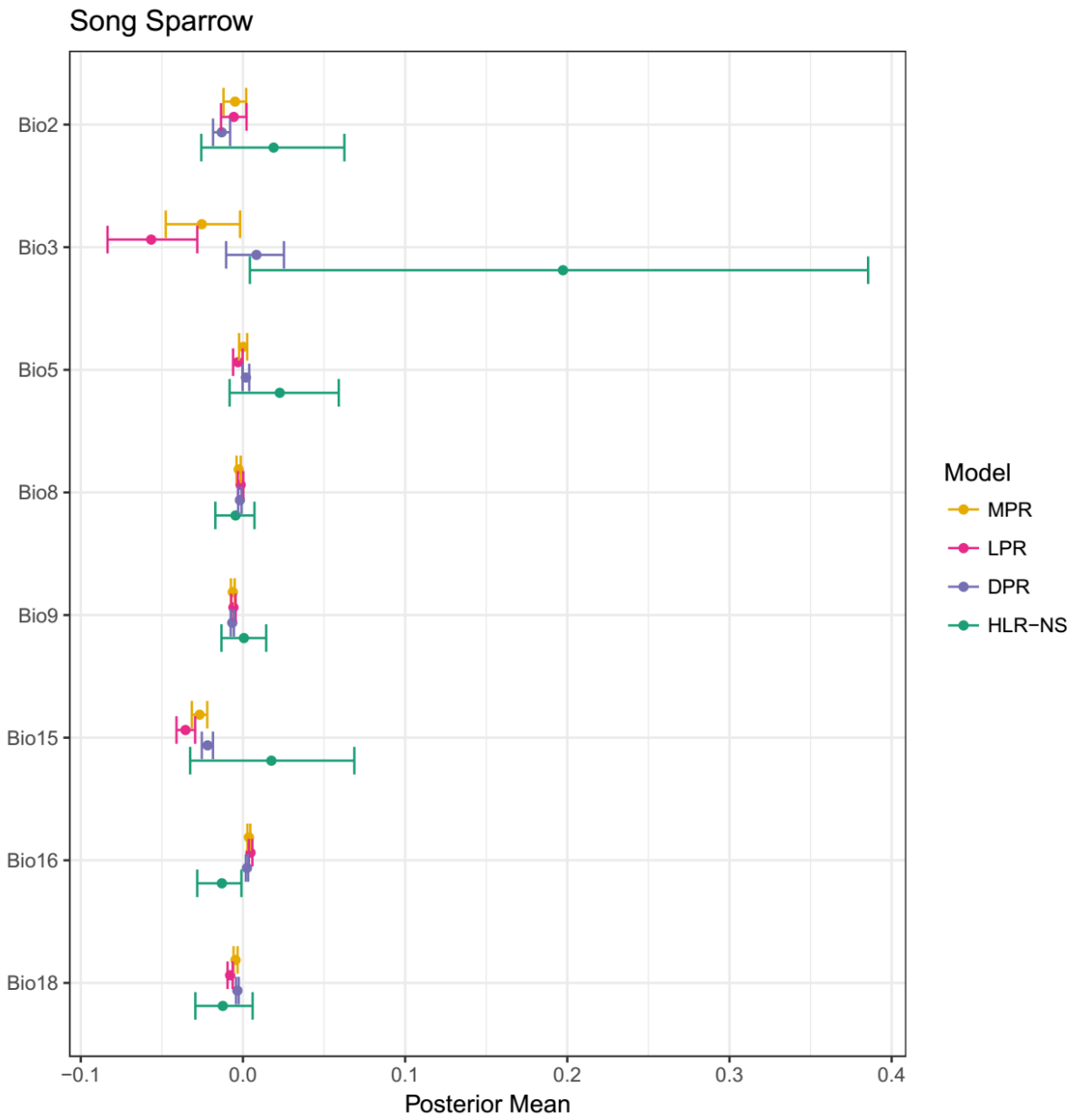


Figure 306: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the song sparrow, *Melospiza melodia*.

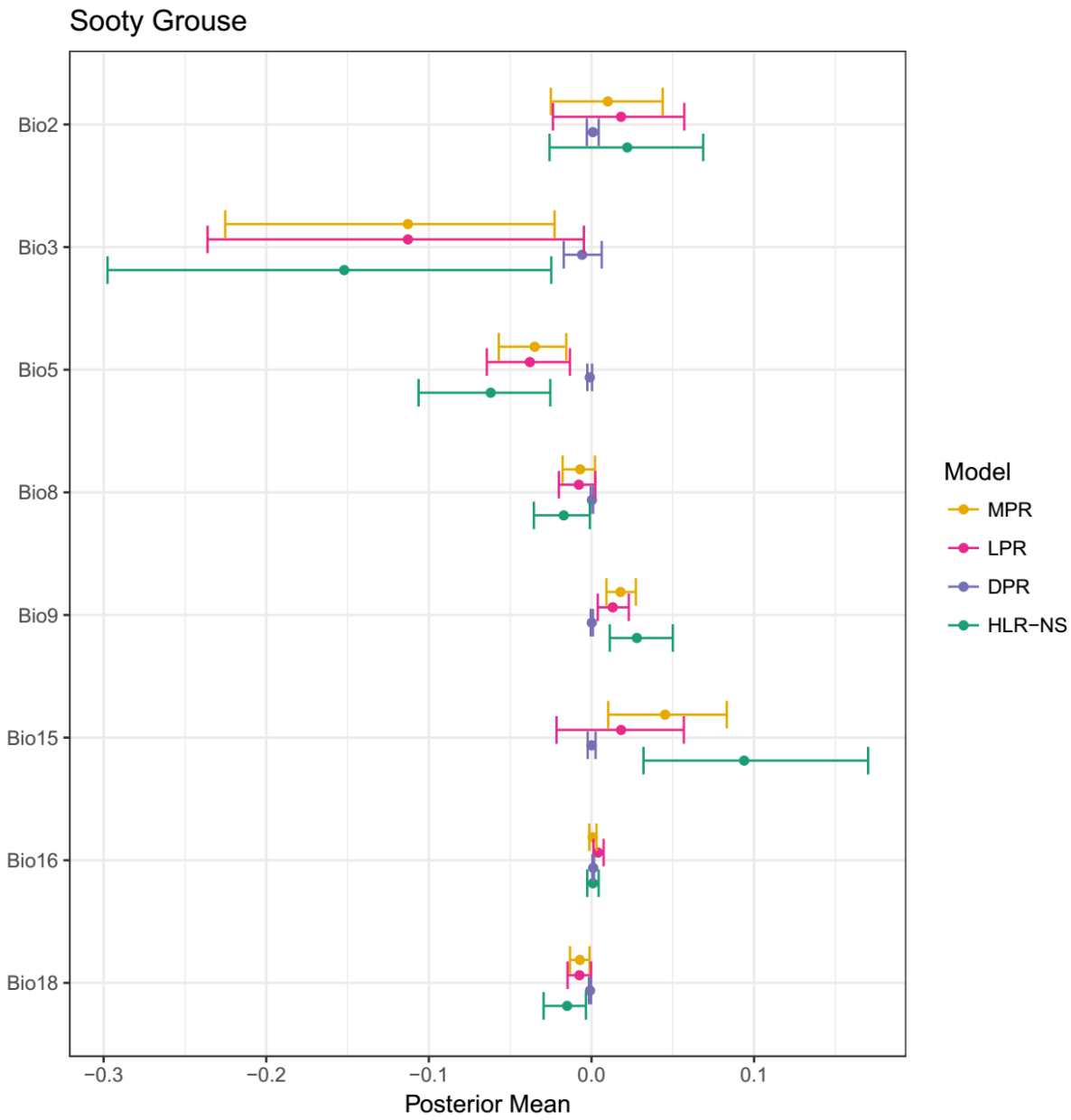


Figure 307: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sooty grouse, *Dendragapus fuliginosus*.

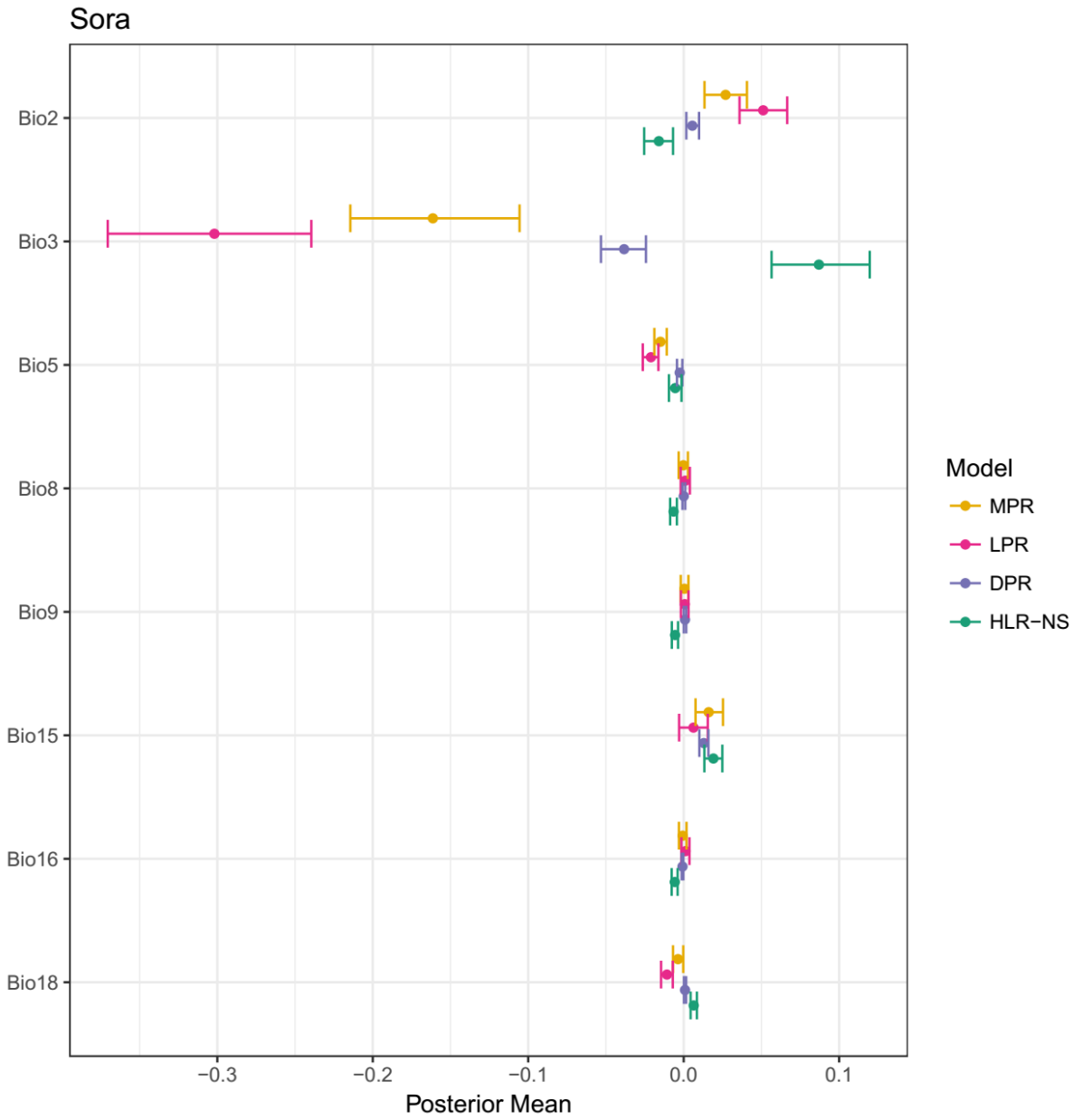


Figure 308: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sora, *Porzana carolina*.

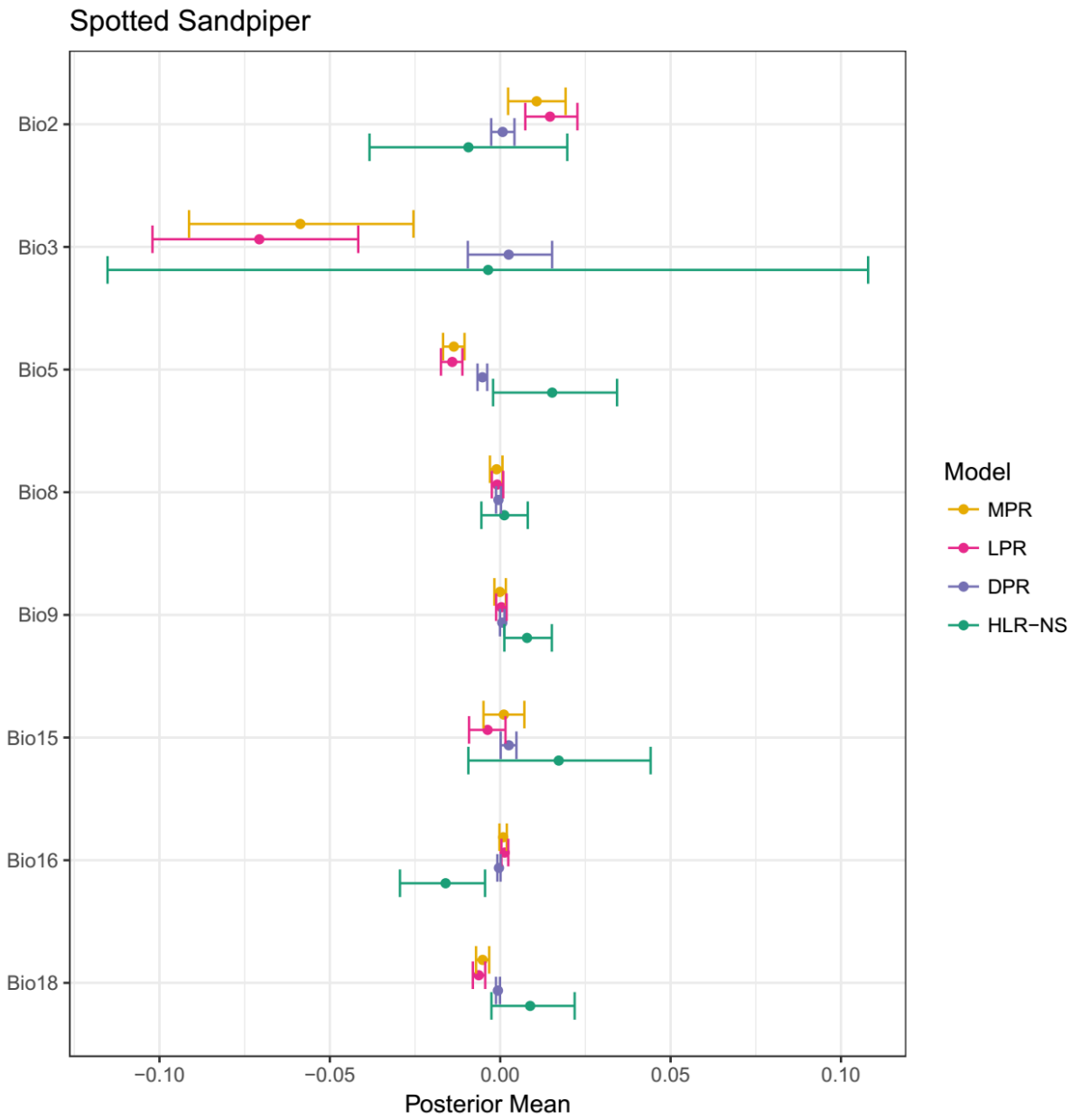


Figure 309: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the spotted sandpiper, *Actitis macularius*.

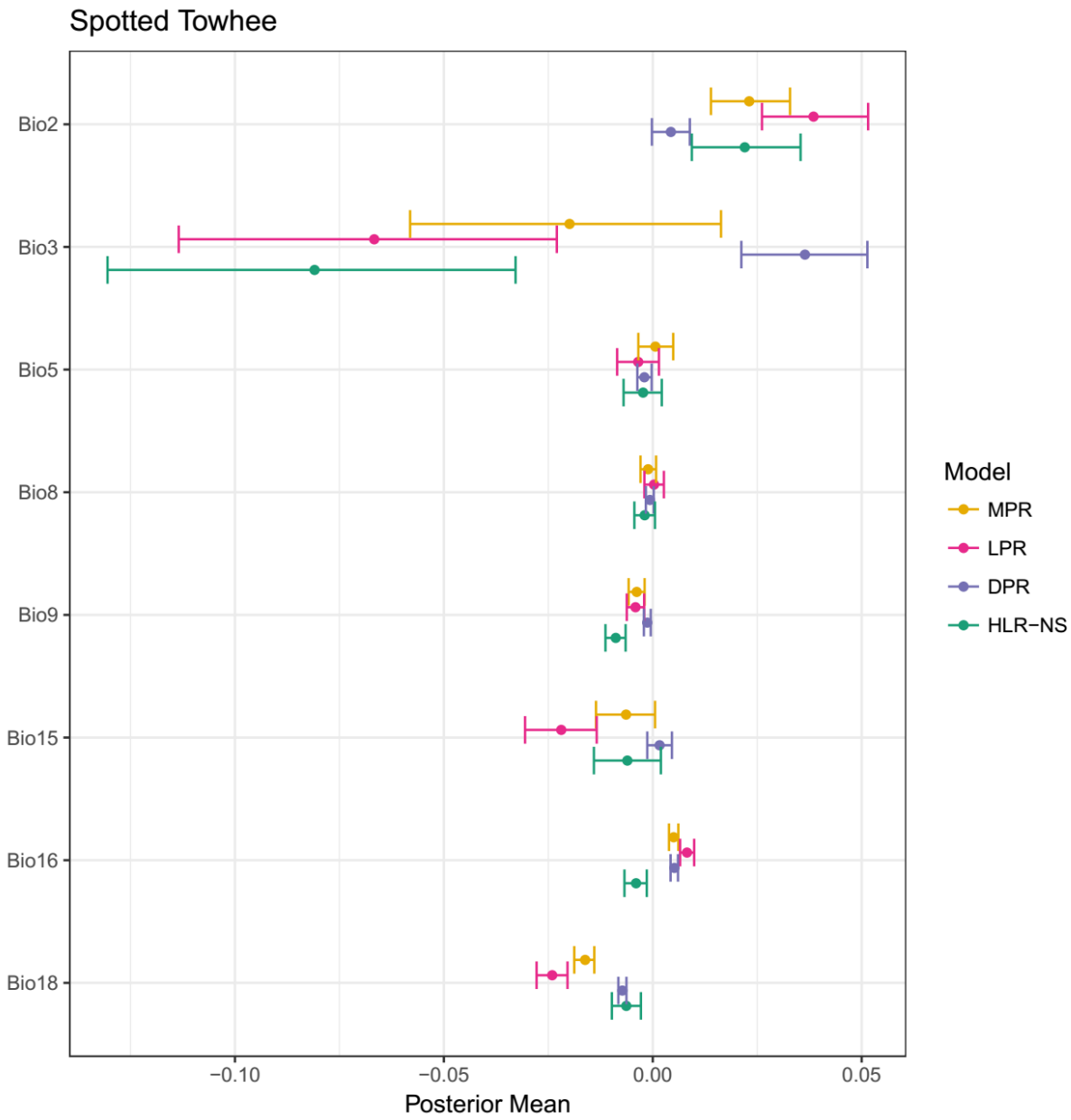


Figure 310: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the spotted towhee, *Pipilo maculatus*.

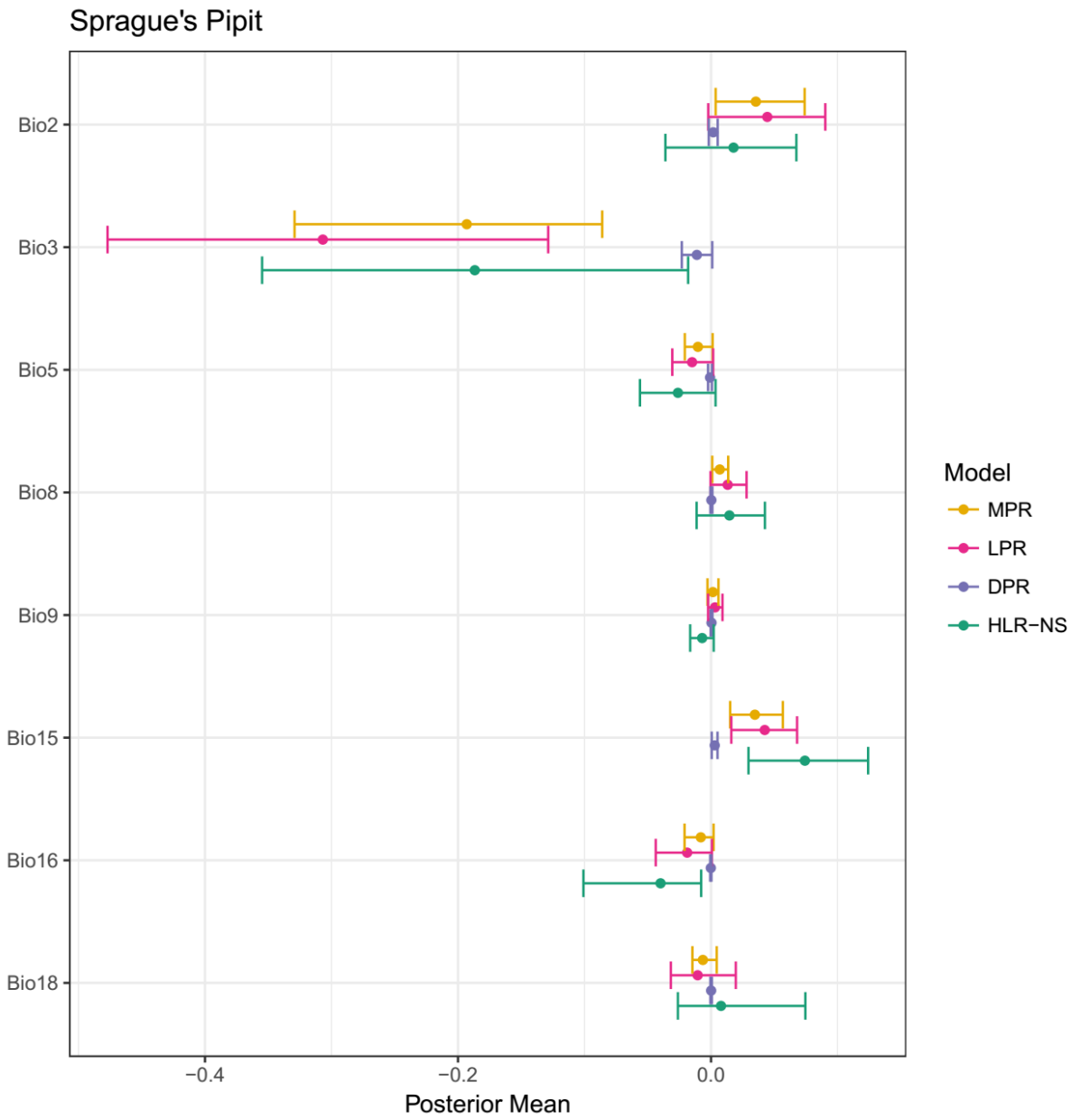


Figure 311: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Sprague's pipit, *Anthus spragueii*.

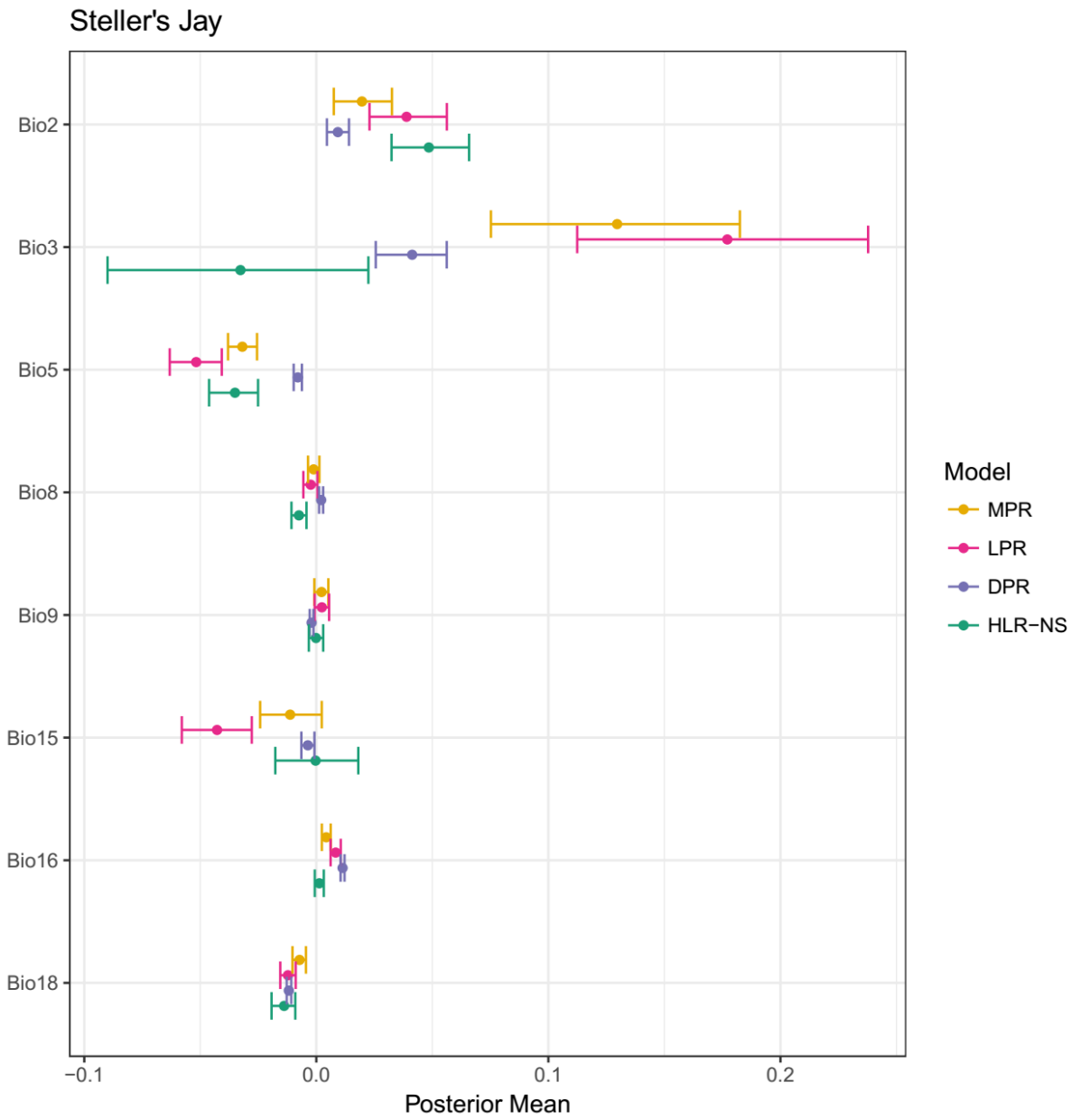


Figure 312: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Stellar's jay, *Cyanocitta stelleri*.

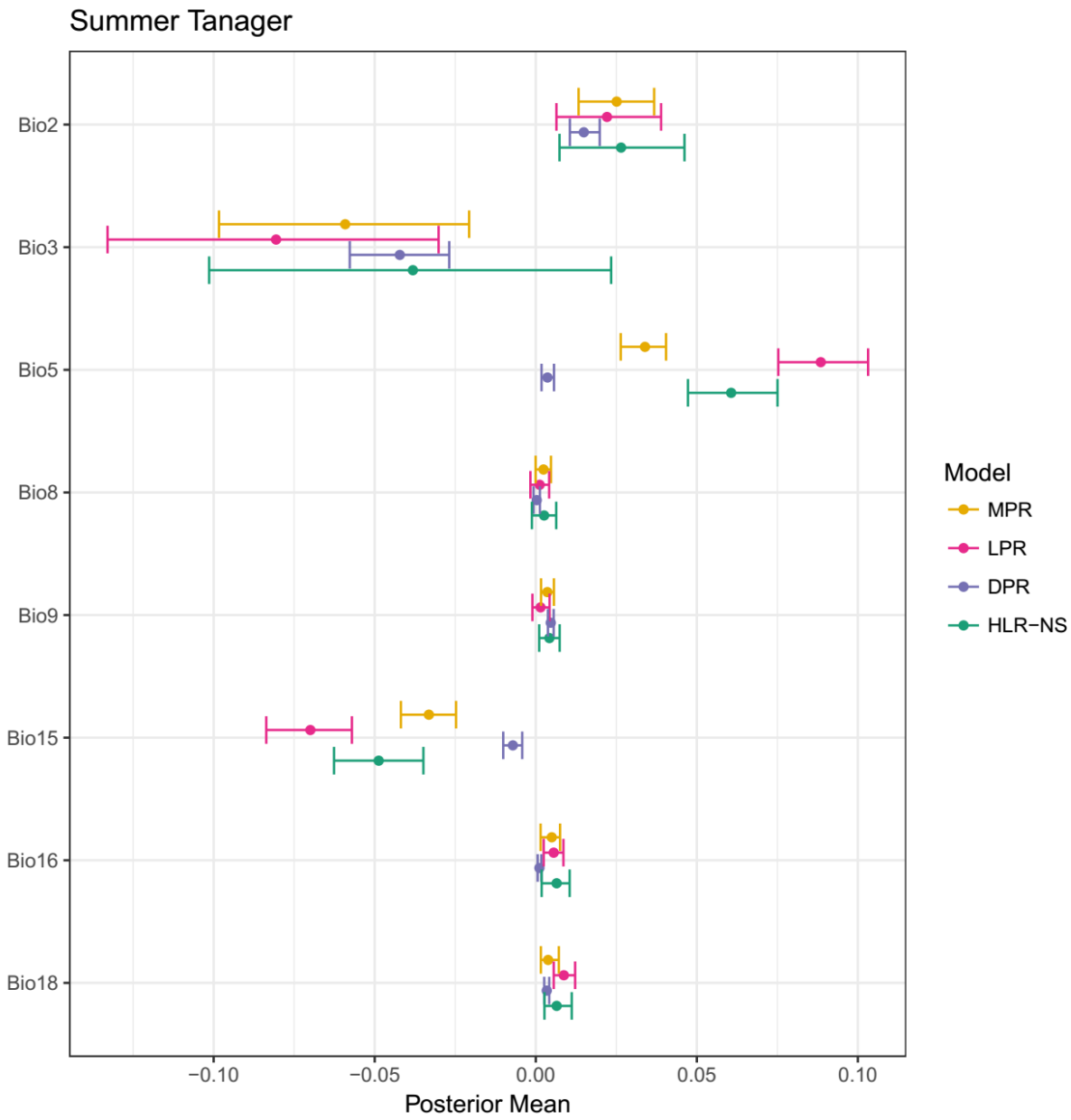


Figure 313: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the summer tanager, *Piranga rubra*.

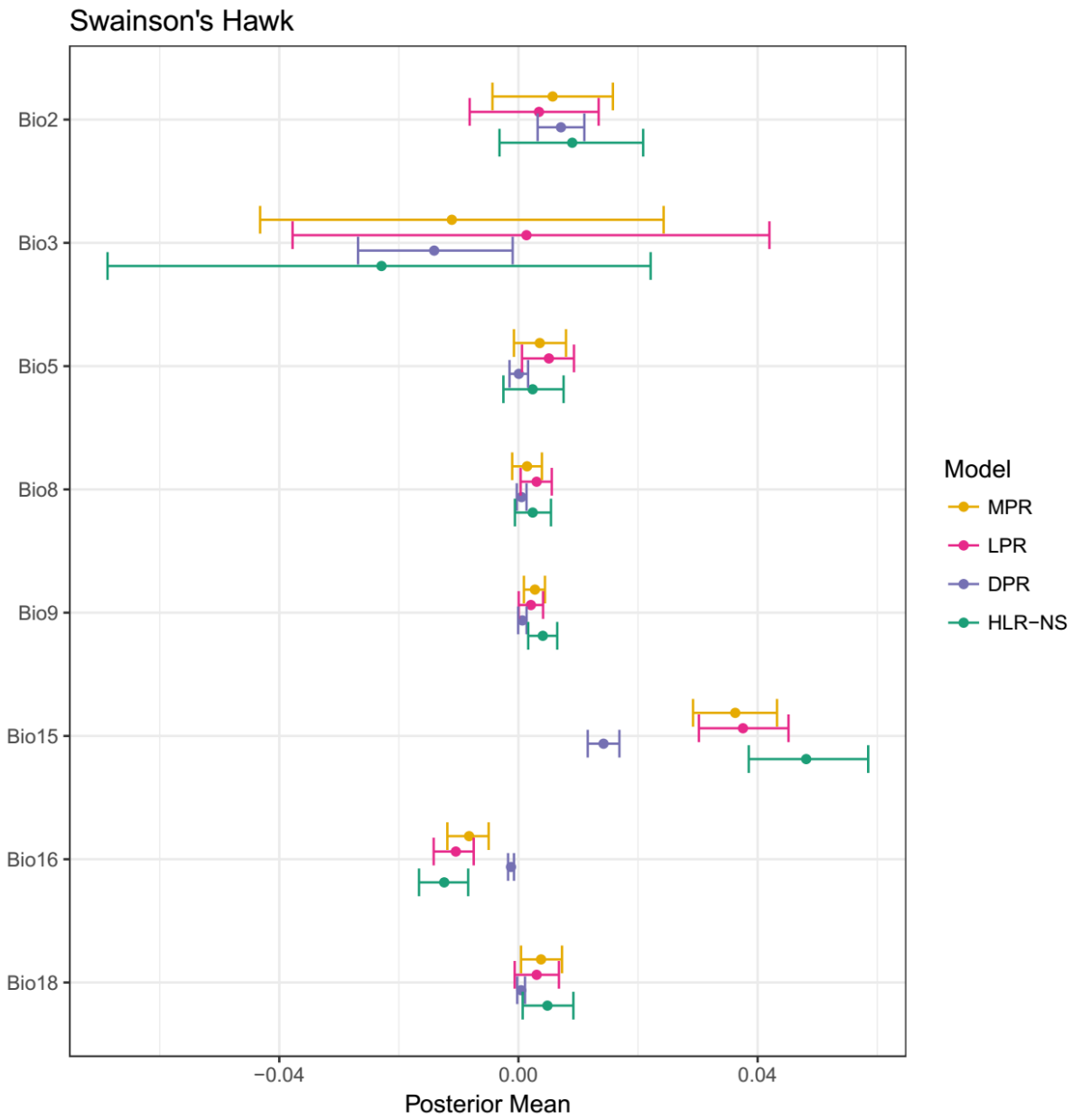


Figure 314: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Swainson's hawk, *Buteo swainsoni*.

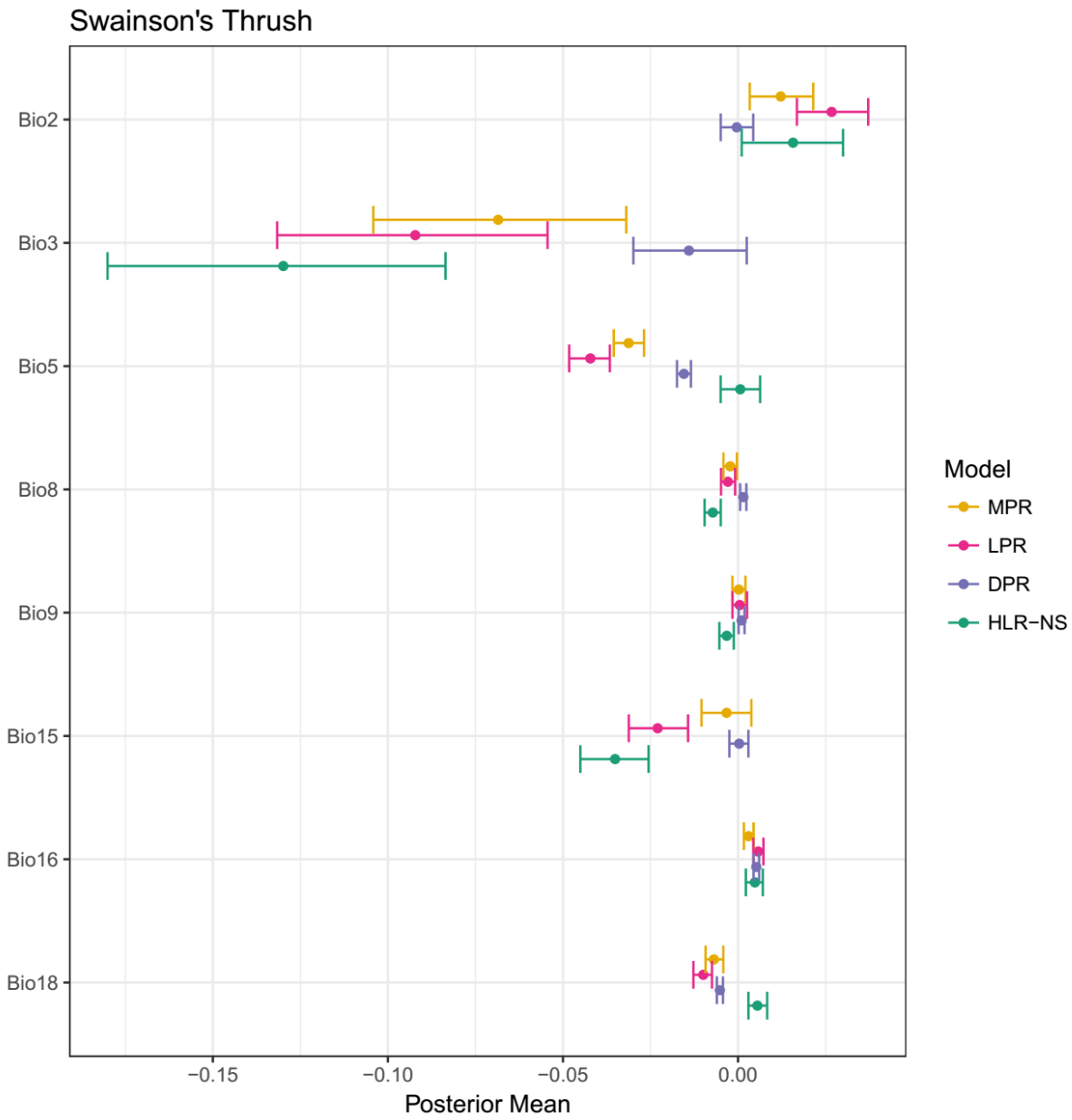


Figure 315: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Swainson's thrush, *Catharus ustulatus*.

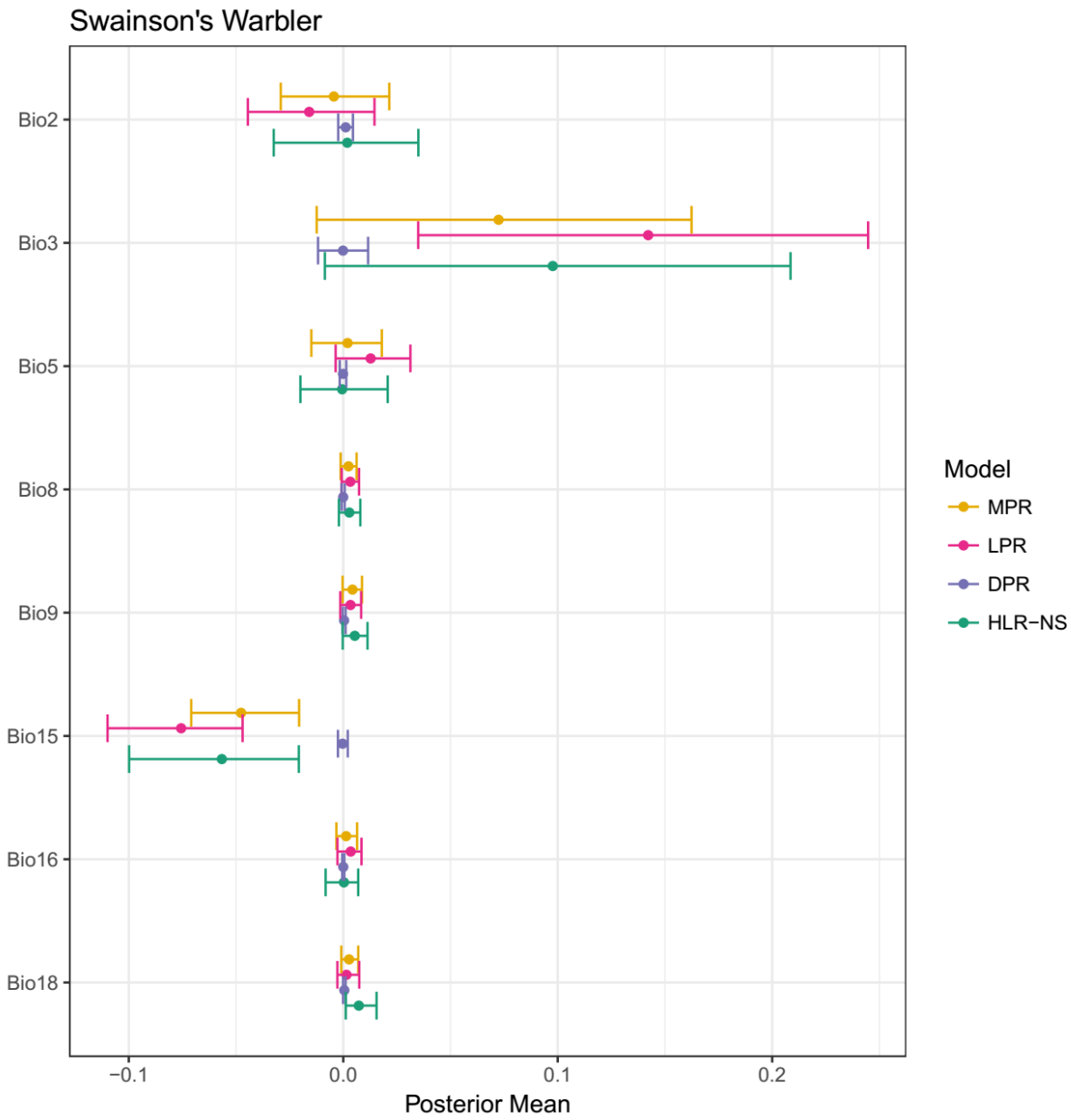


Figure 316: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Swainson's warbler, *Limnothlypis swainsonii*.

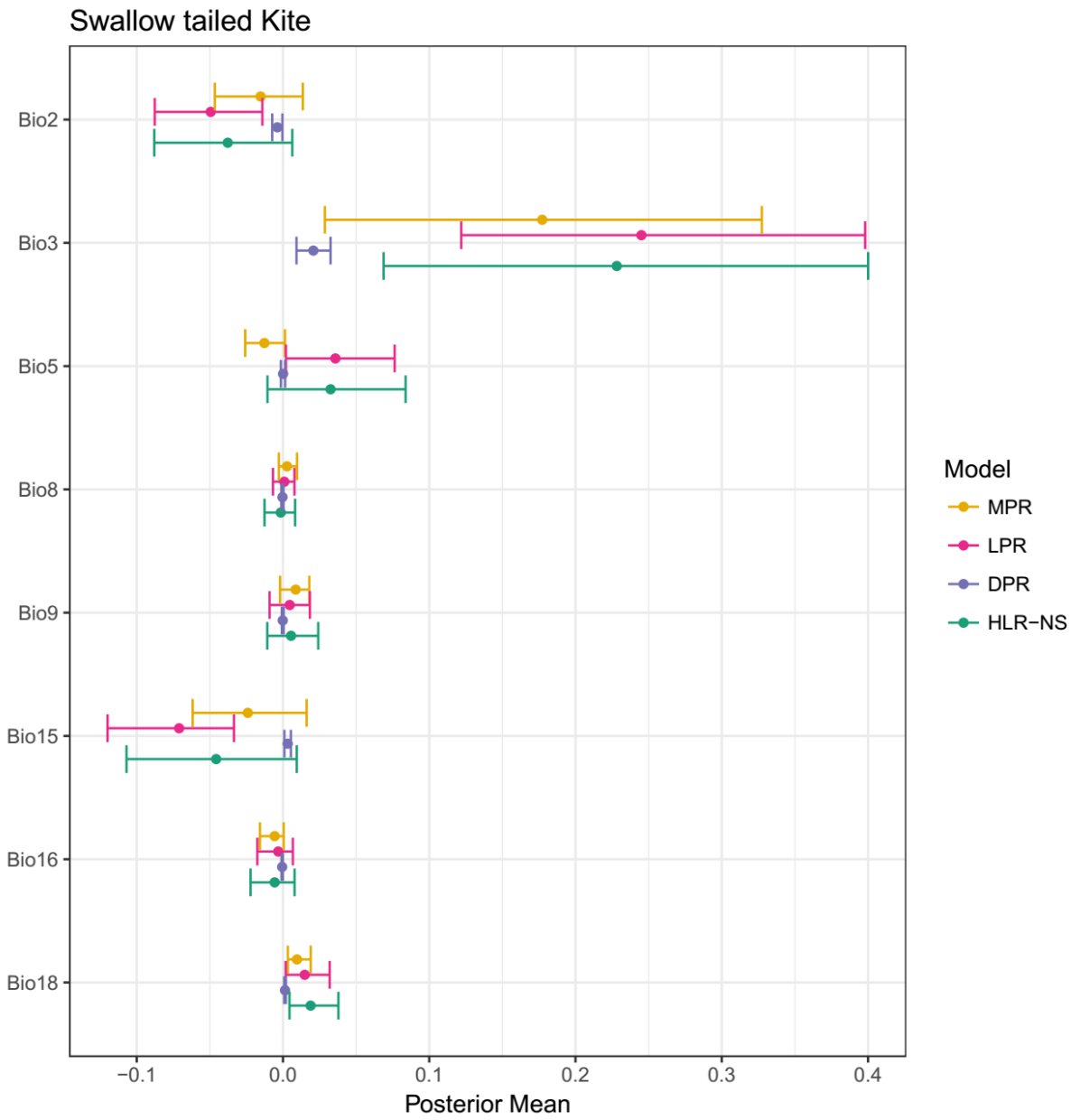


Figure 317: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the swallow-tailed kite, *Elanoides forficatus*.

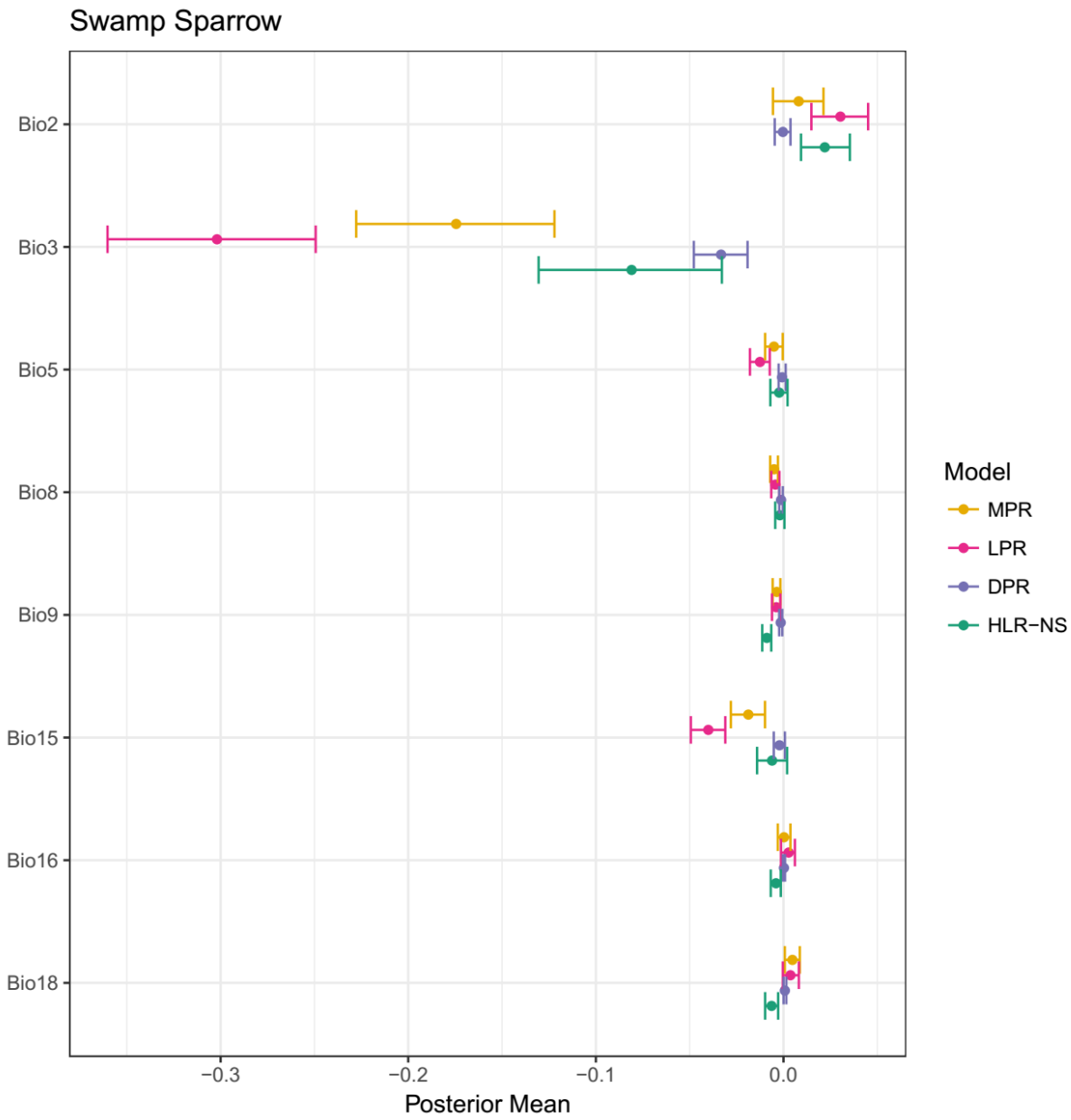


Figure 318: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the swamp sparrow, *Melospiza georgiana*.

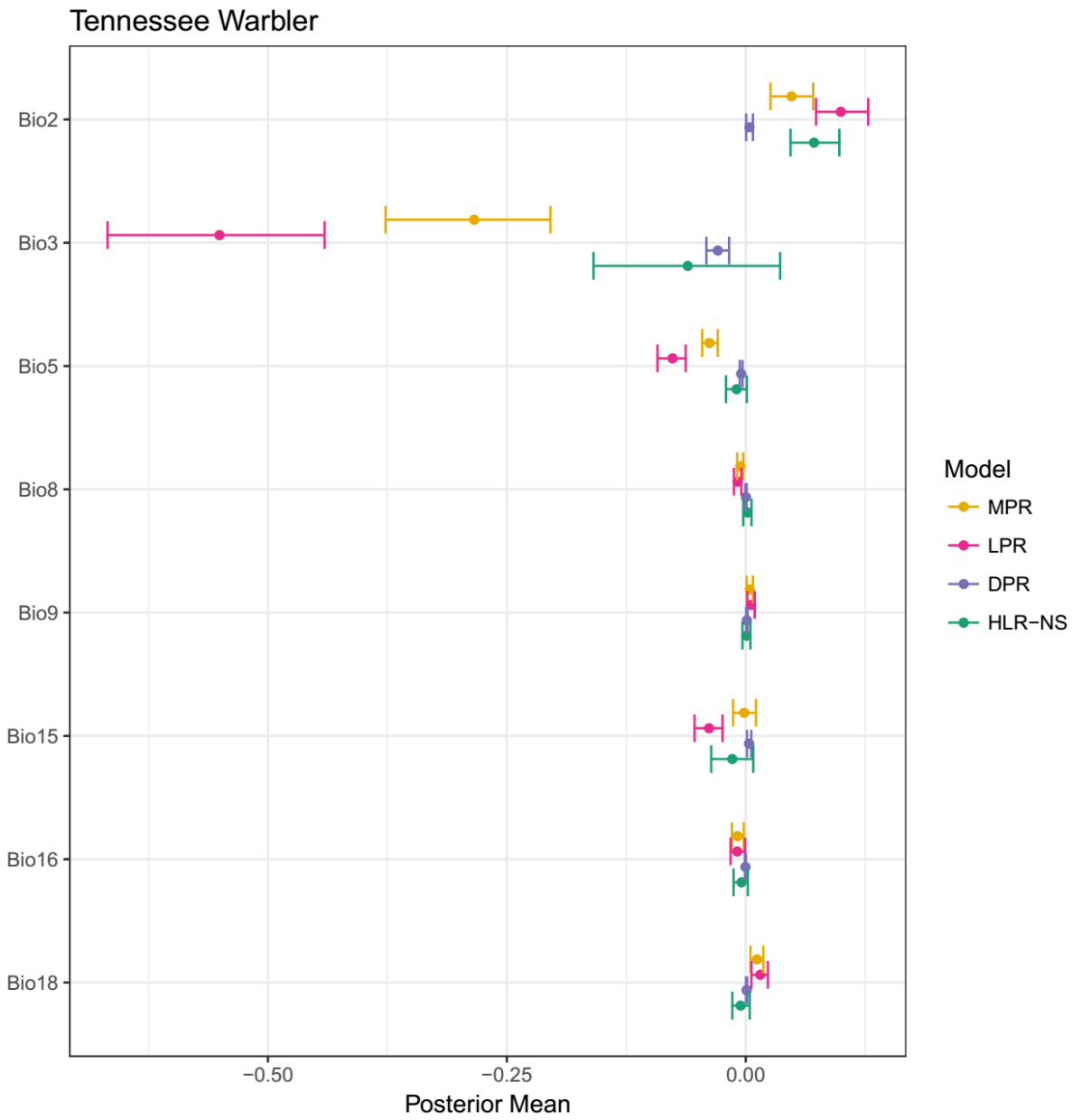


Figure 319: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Tennessee warbler, *Leiothlypis peregrina*.

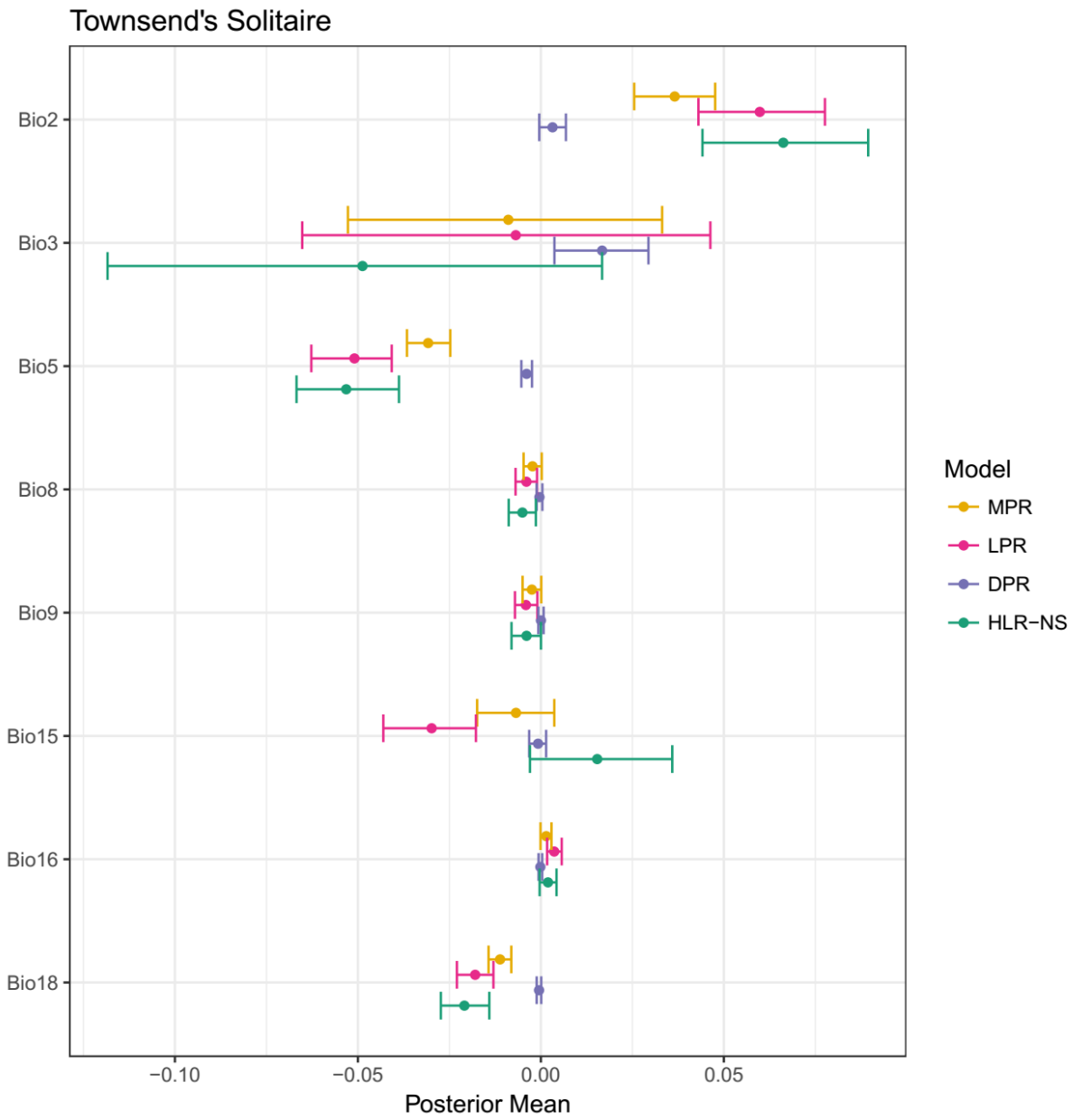


Figure 320: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Townsend's solitaire, *Myadestes townsendi*.

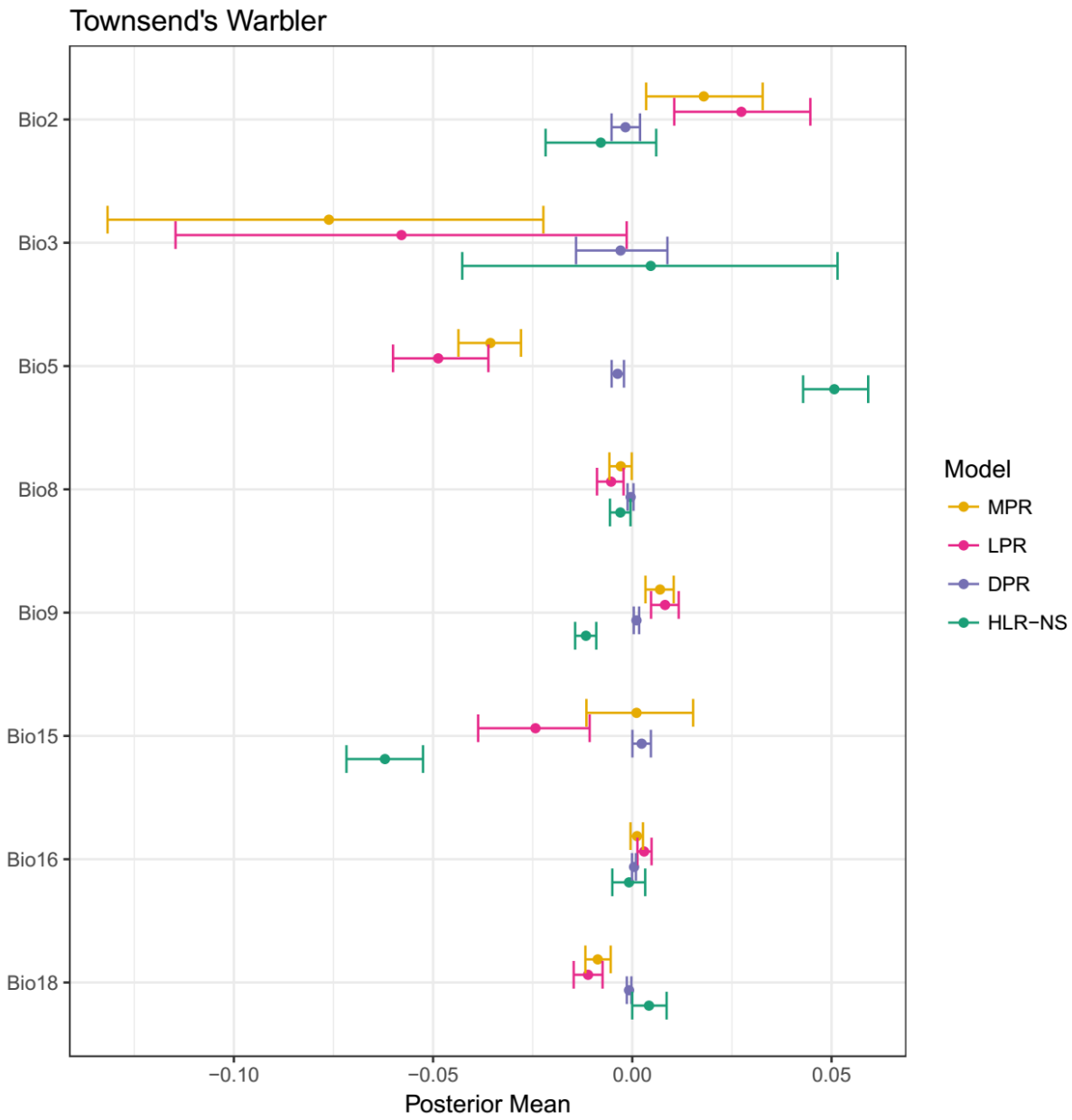


Figure 321: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Townsend's warbler, *Setophaga townsendi*.

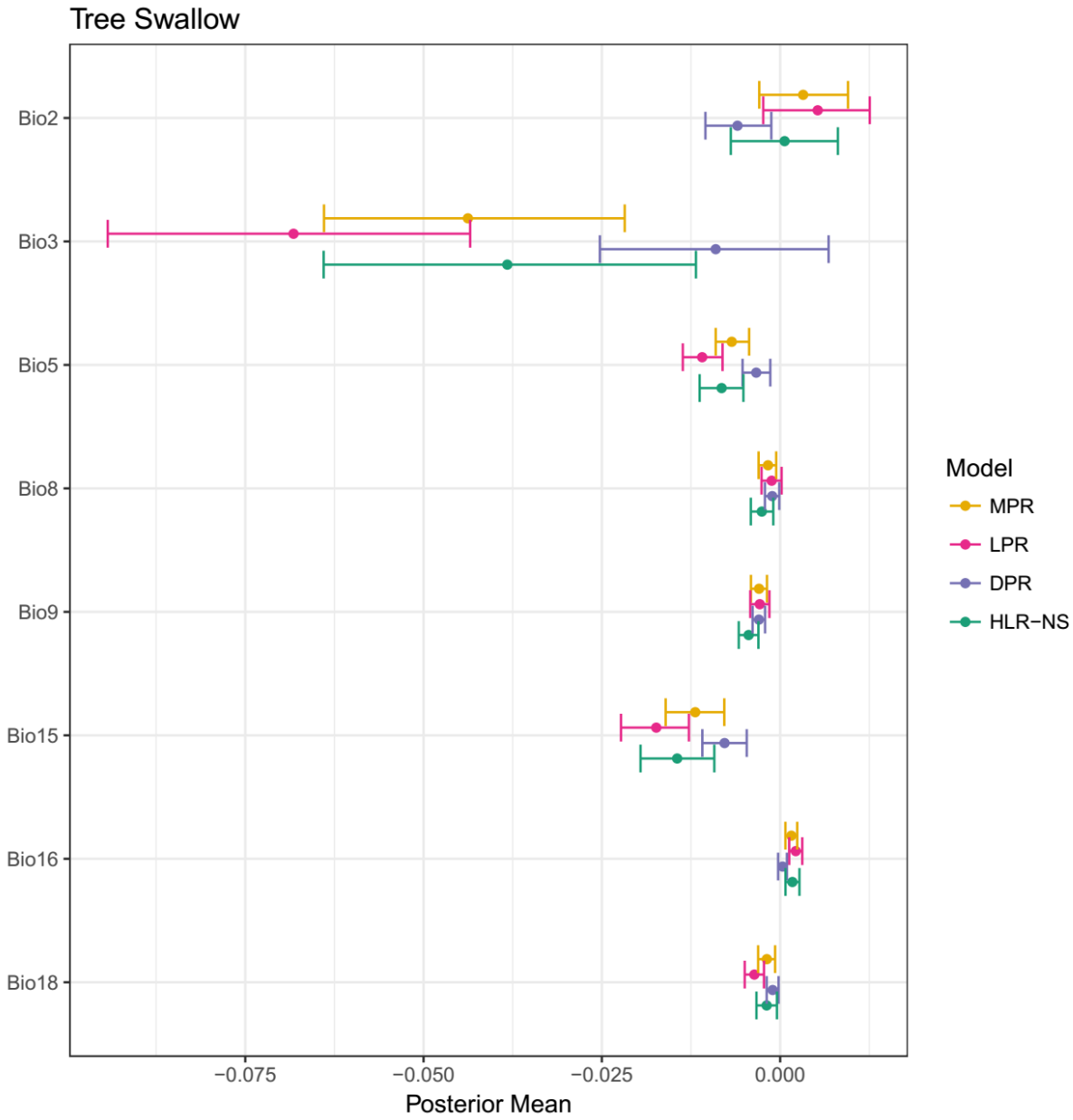


Figure 322: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the tree swallow, *Tachycineta bicolor*.

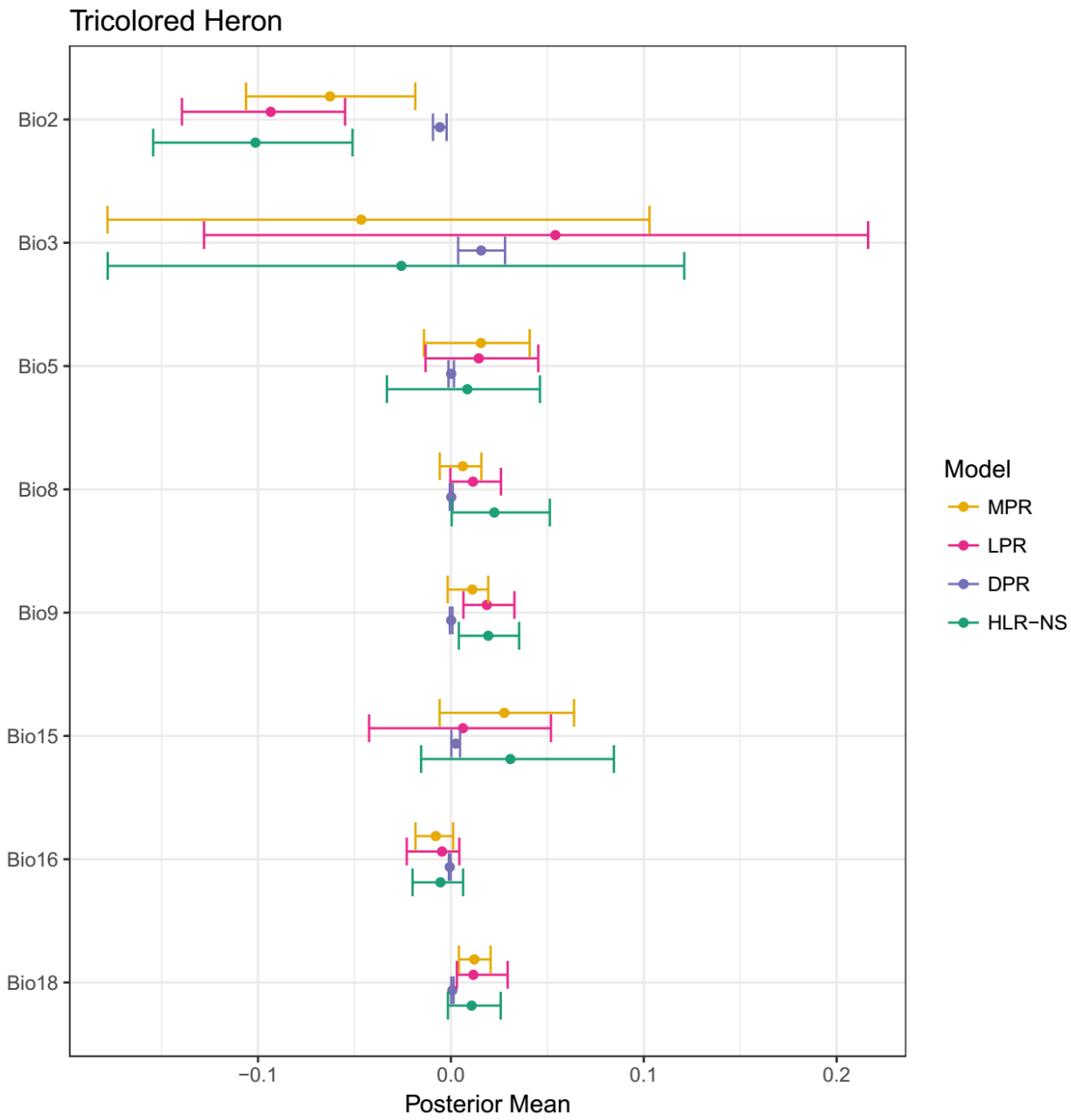


Figure 323: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the tricoloured heron, *Egretta tricolor*.

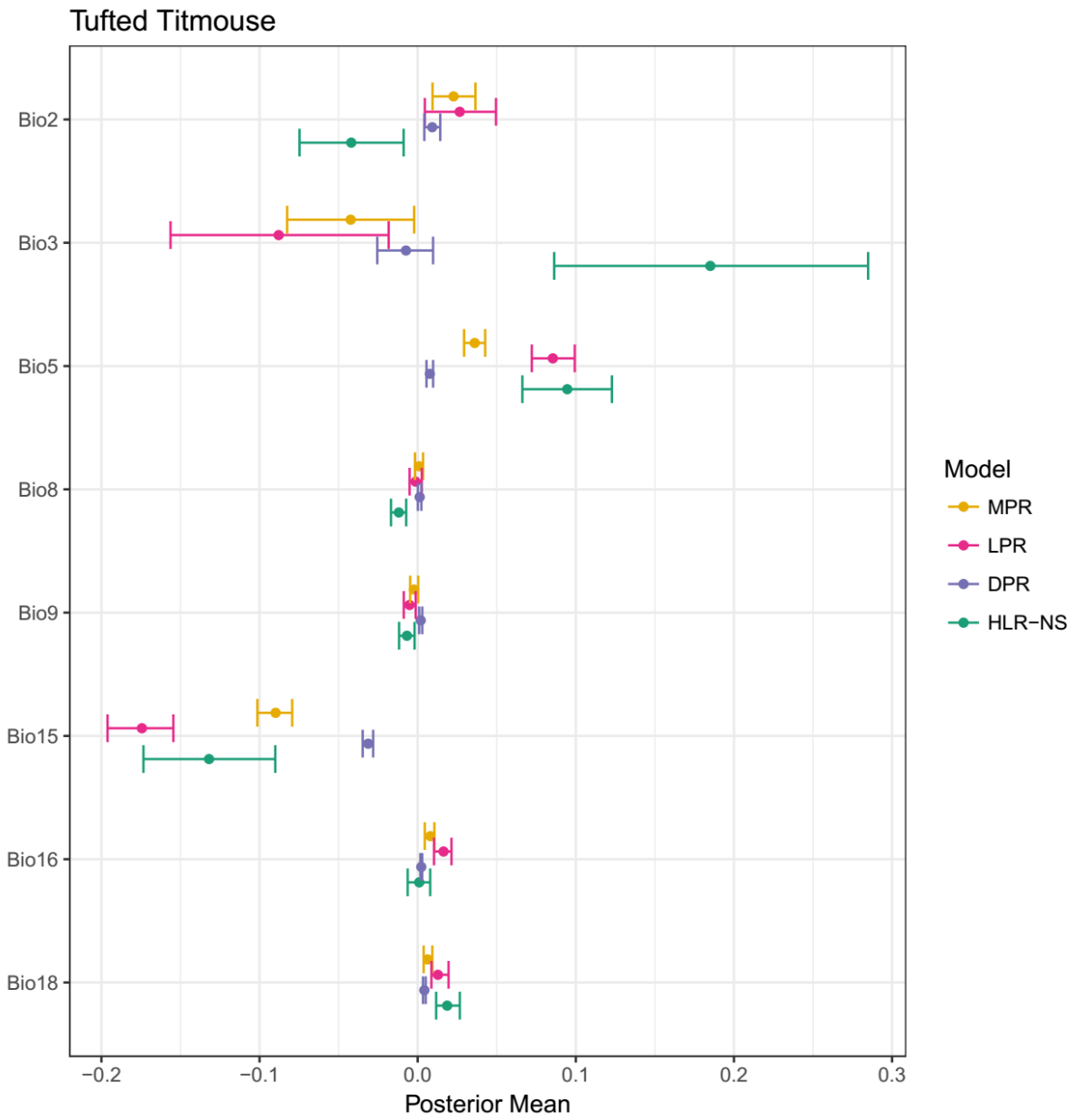


Figure 324: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the tufted titmouse, *Baeolophus bicolor*.

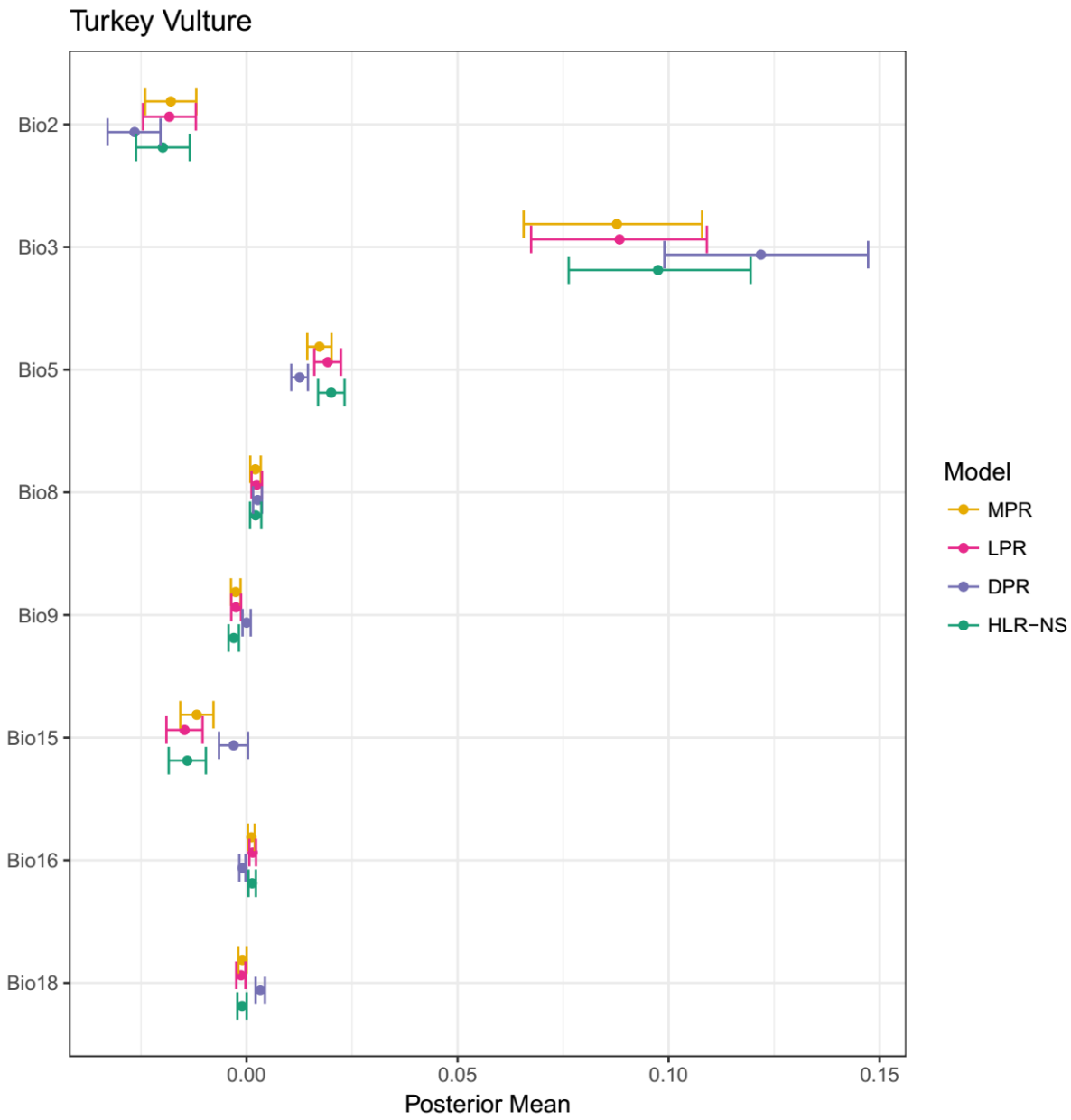


Figure 325: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the turkey vulture, *Cathartes aura*.

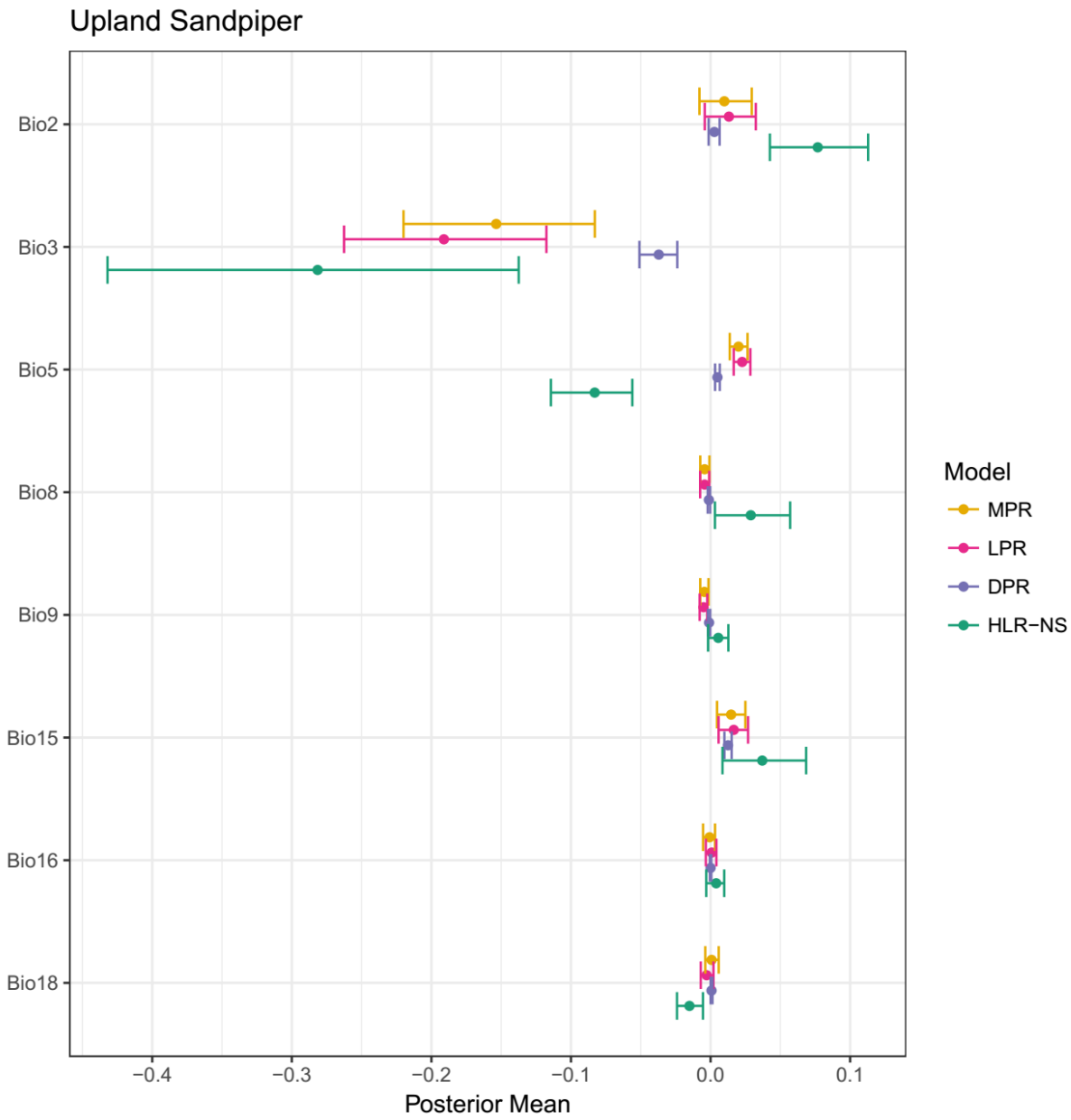


Figure 326: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the upland sandpiper, *Bartramia longicauda*.

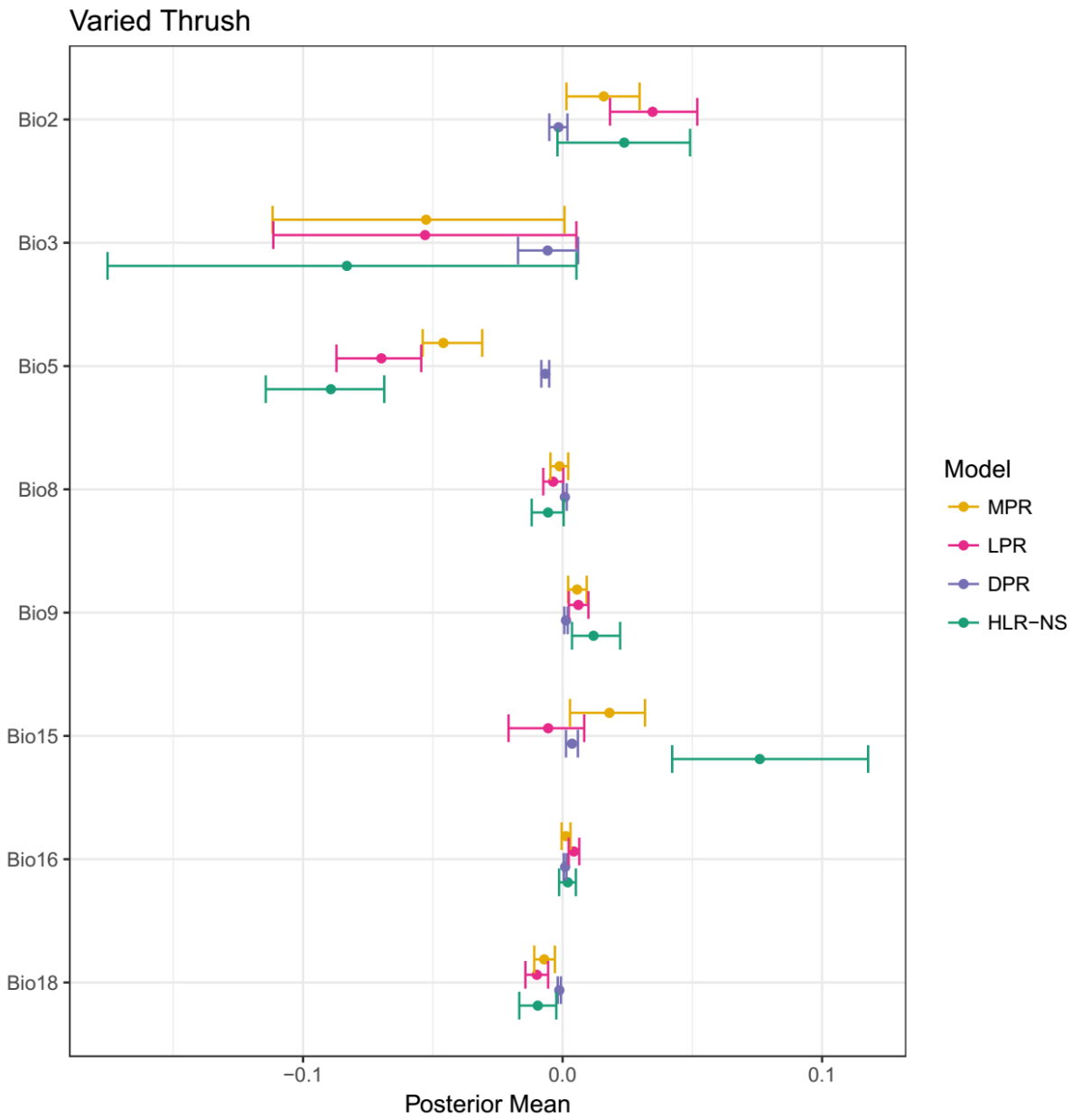


Figure 327: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the varied thrush, *Ixoreus naevius*.

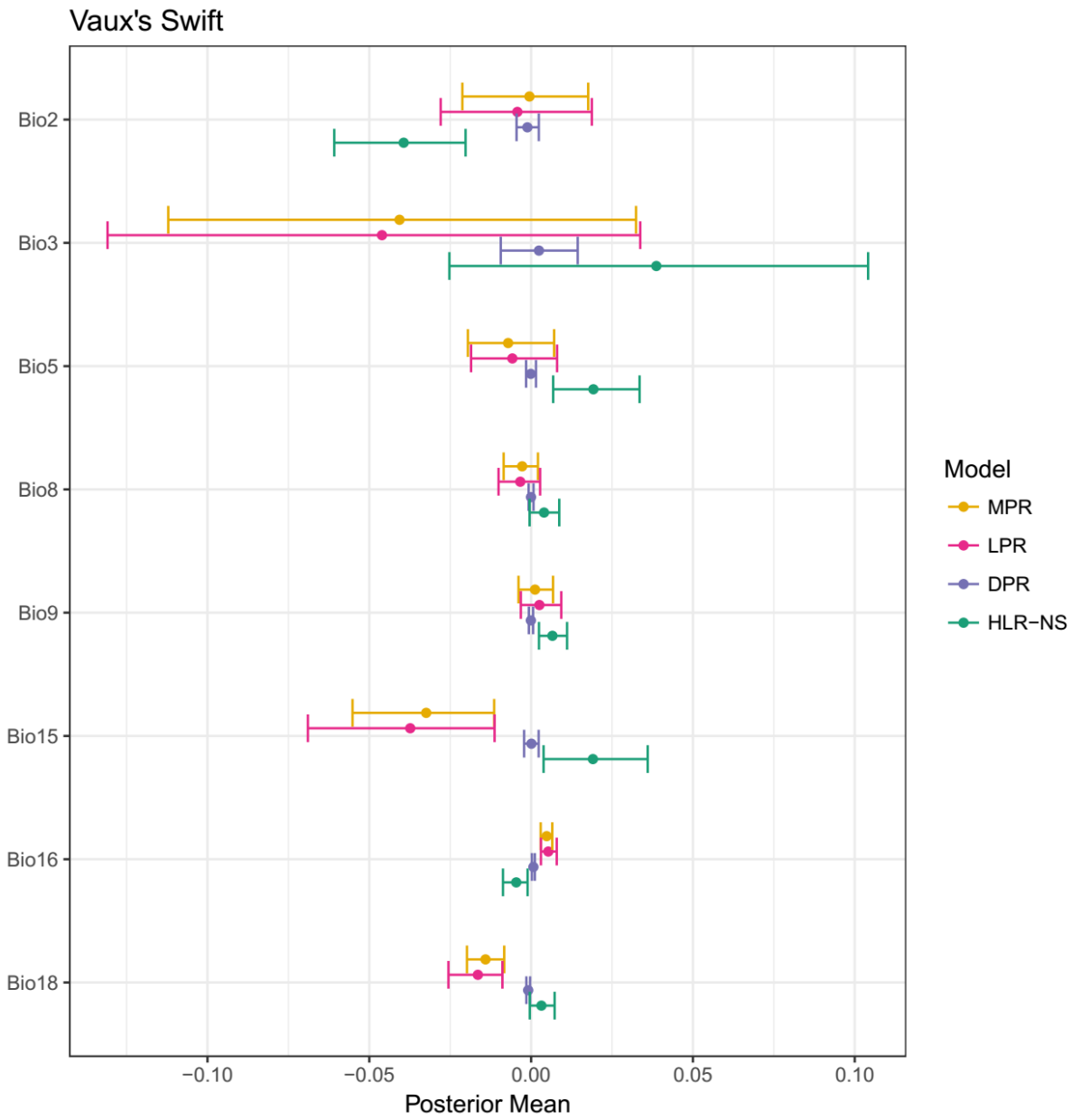


Figure 328: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Vaux's swift, *Chaetura vauxi*.

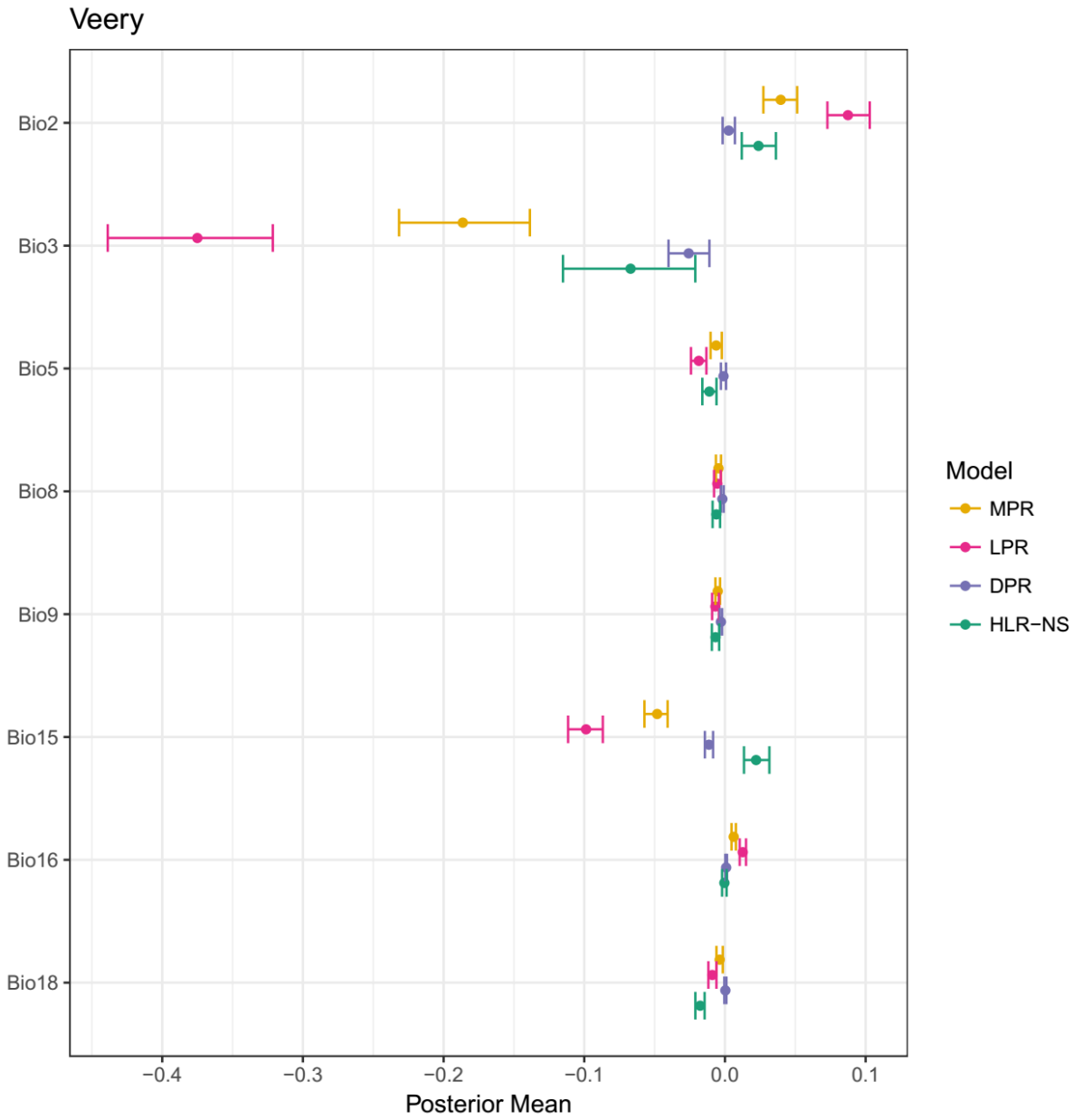


Figure 329: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the veery, *Catharus fuscescens*.

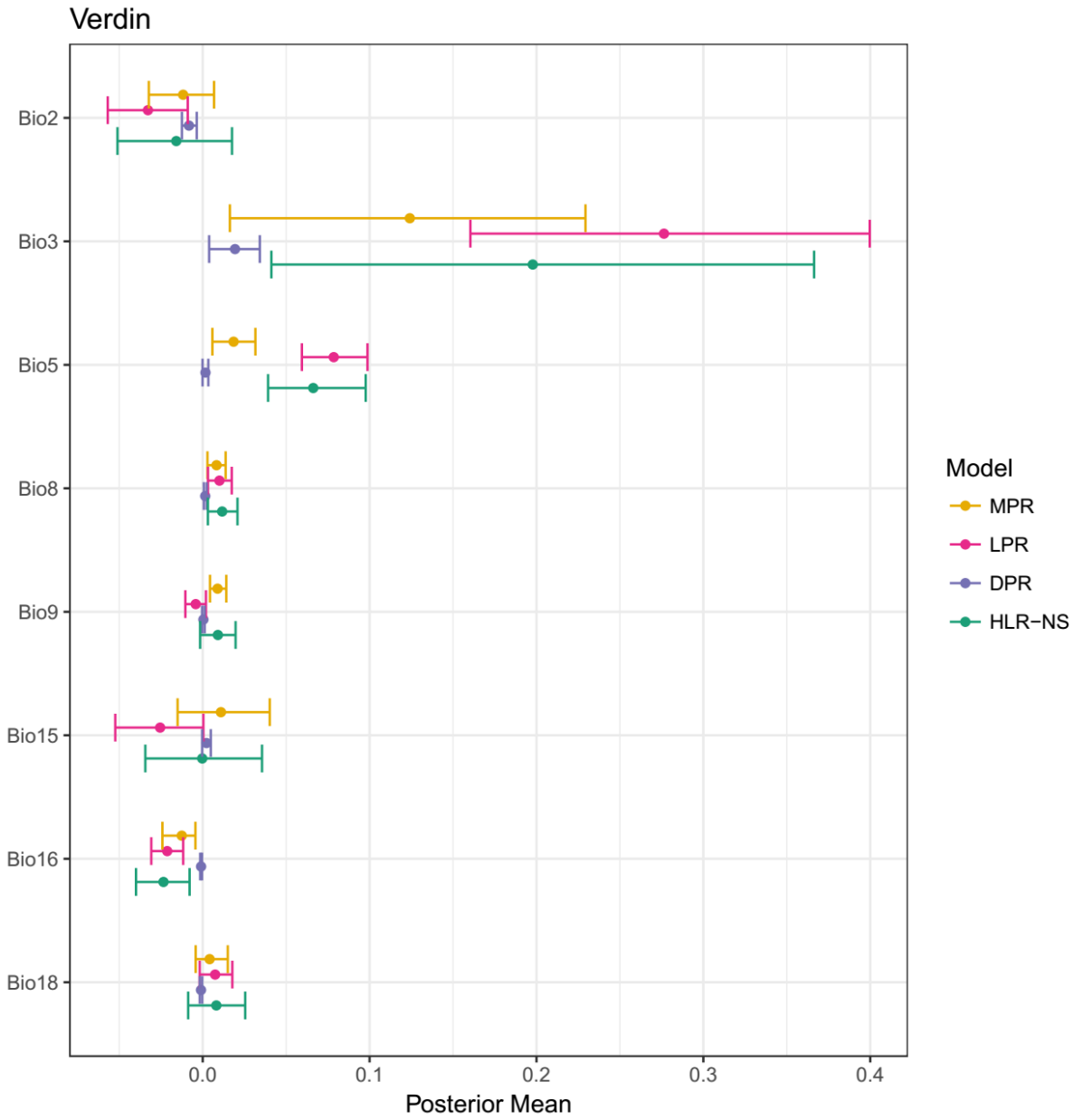


Figure 330: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the verdin, *Auriparus flaviceps*.

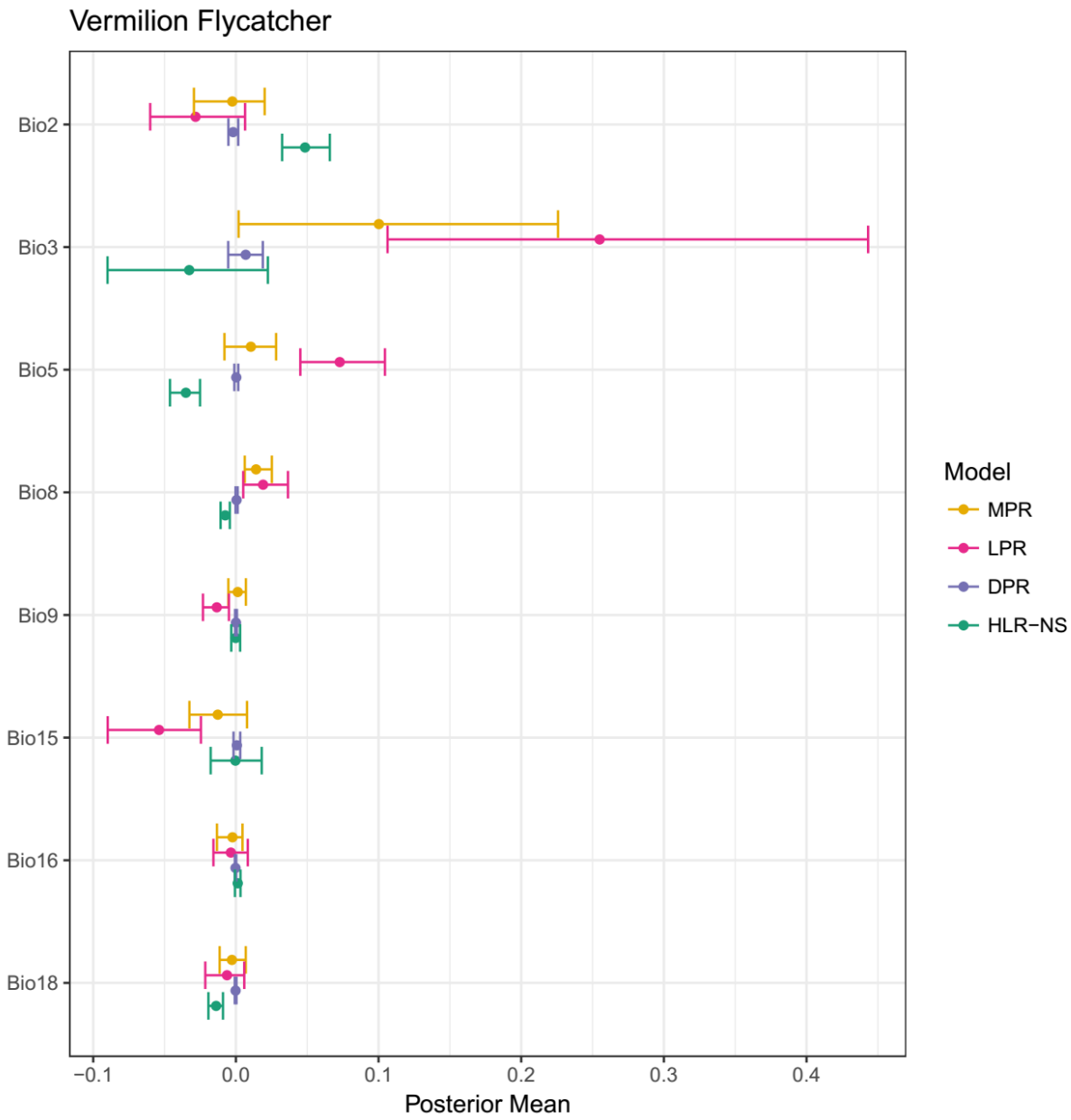


Figure 331: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the vermilion flycatcher, *Pyrocephalus rubinus*.

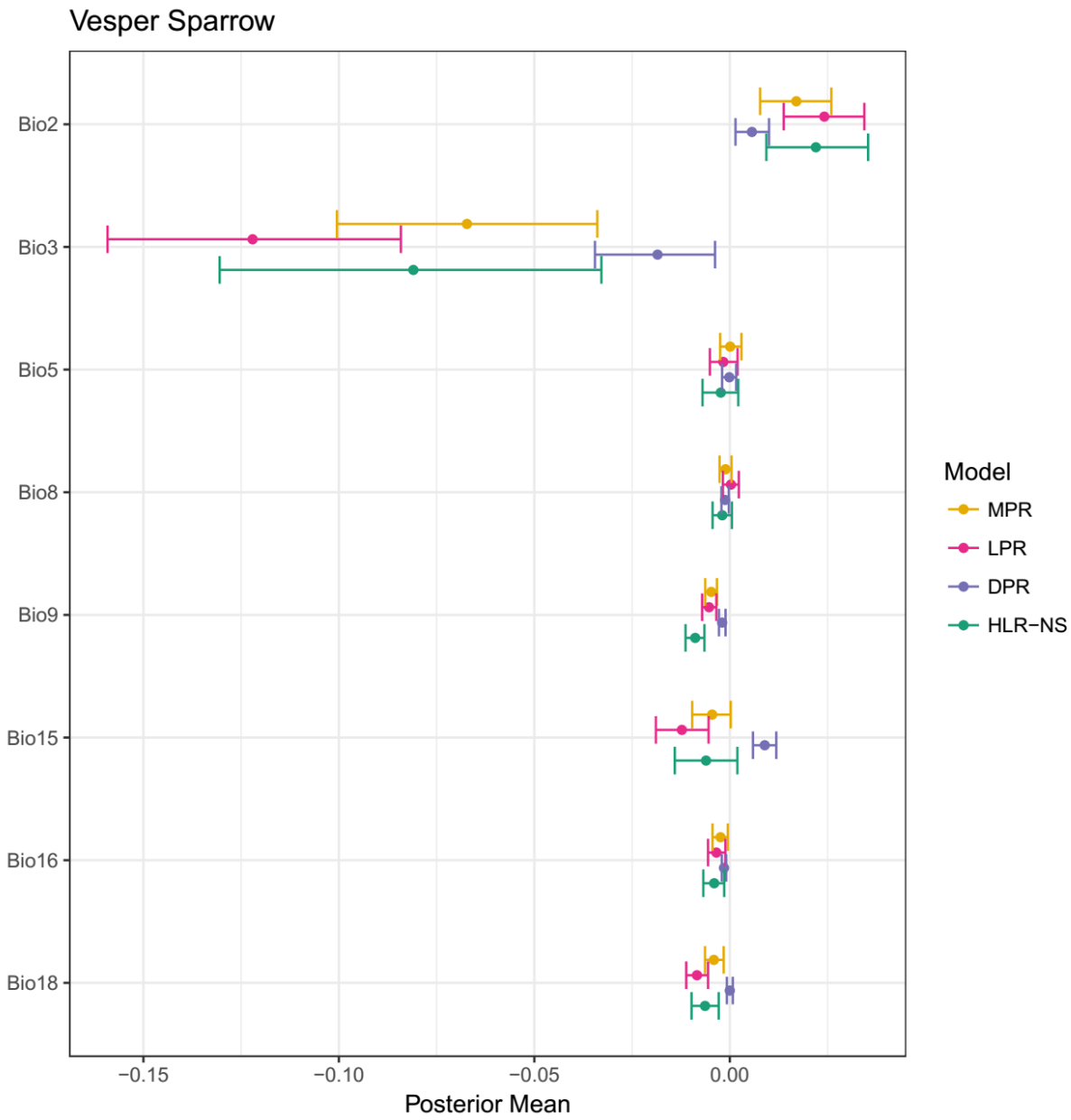


Figure 332: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the vesper sparrow, *Pooecetes gramineus*.

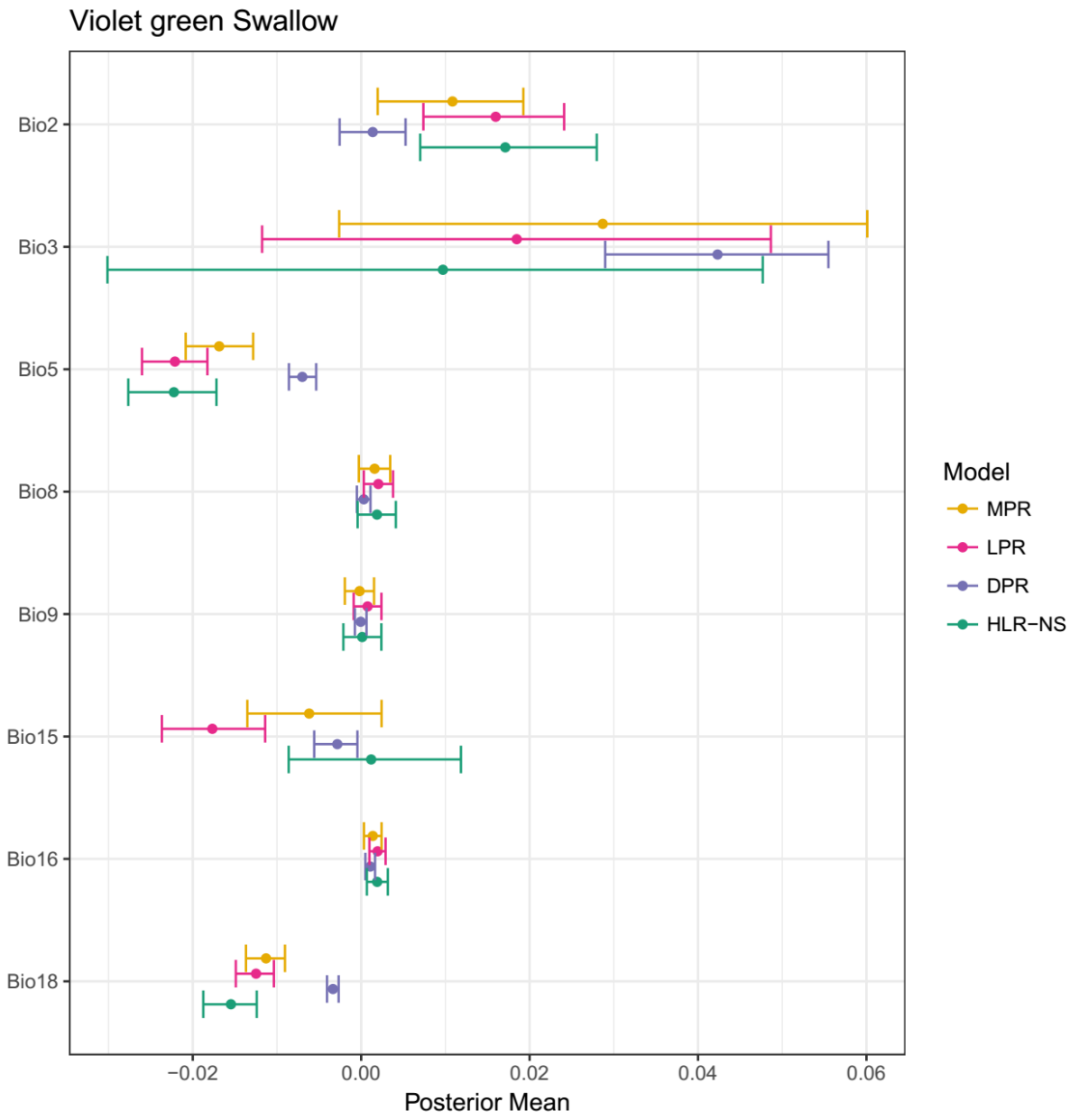


Figure 333: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the violet-green swallow, *Tachycineta thalassina*.

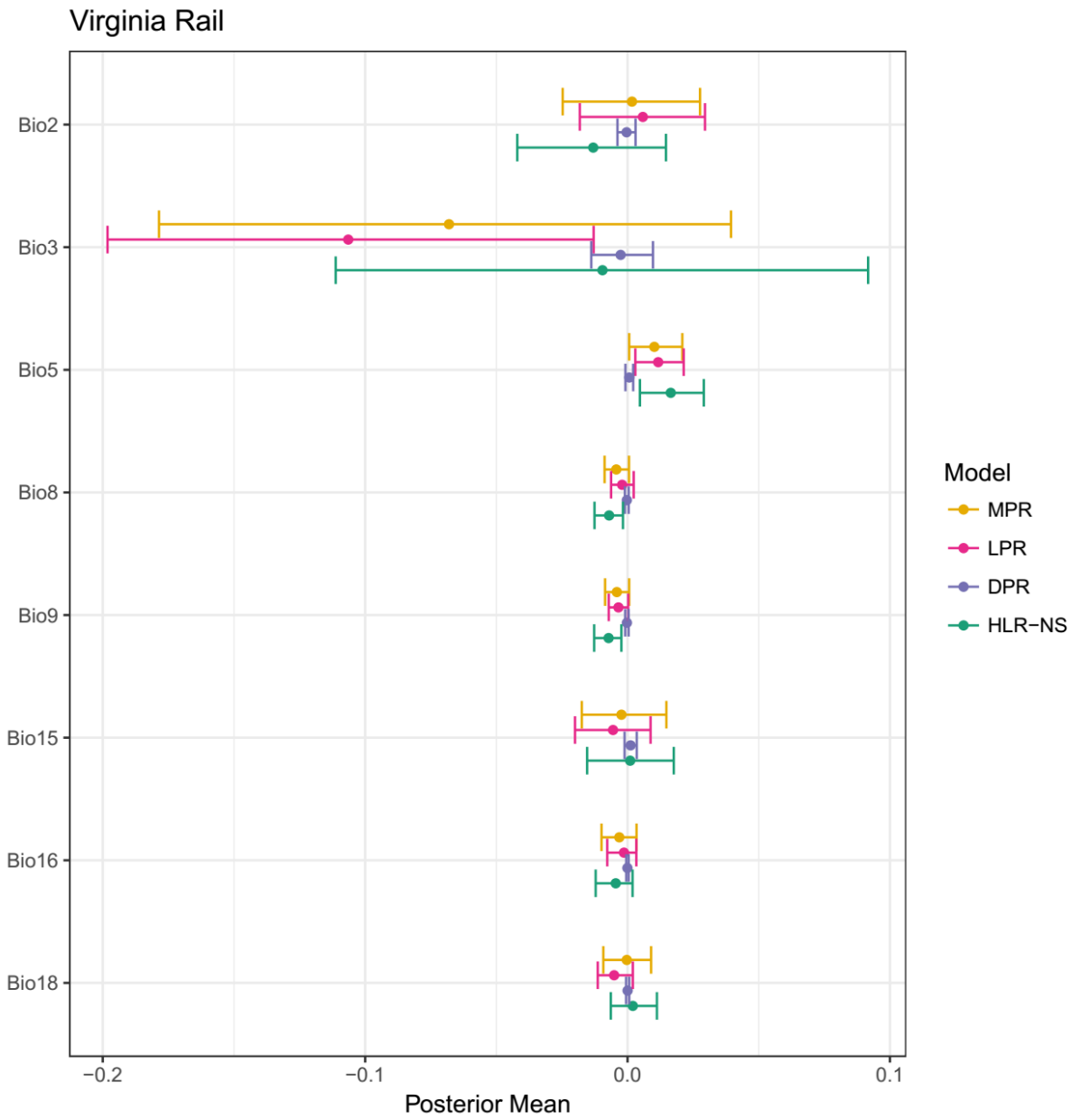


Figure 334: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Virginia rail, *Rallus limicola*.

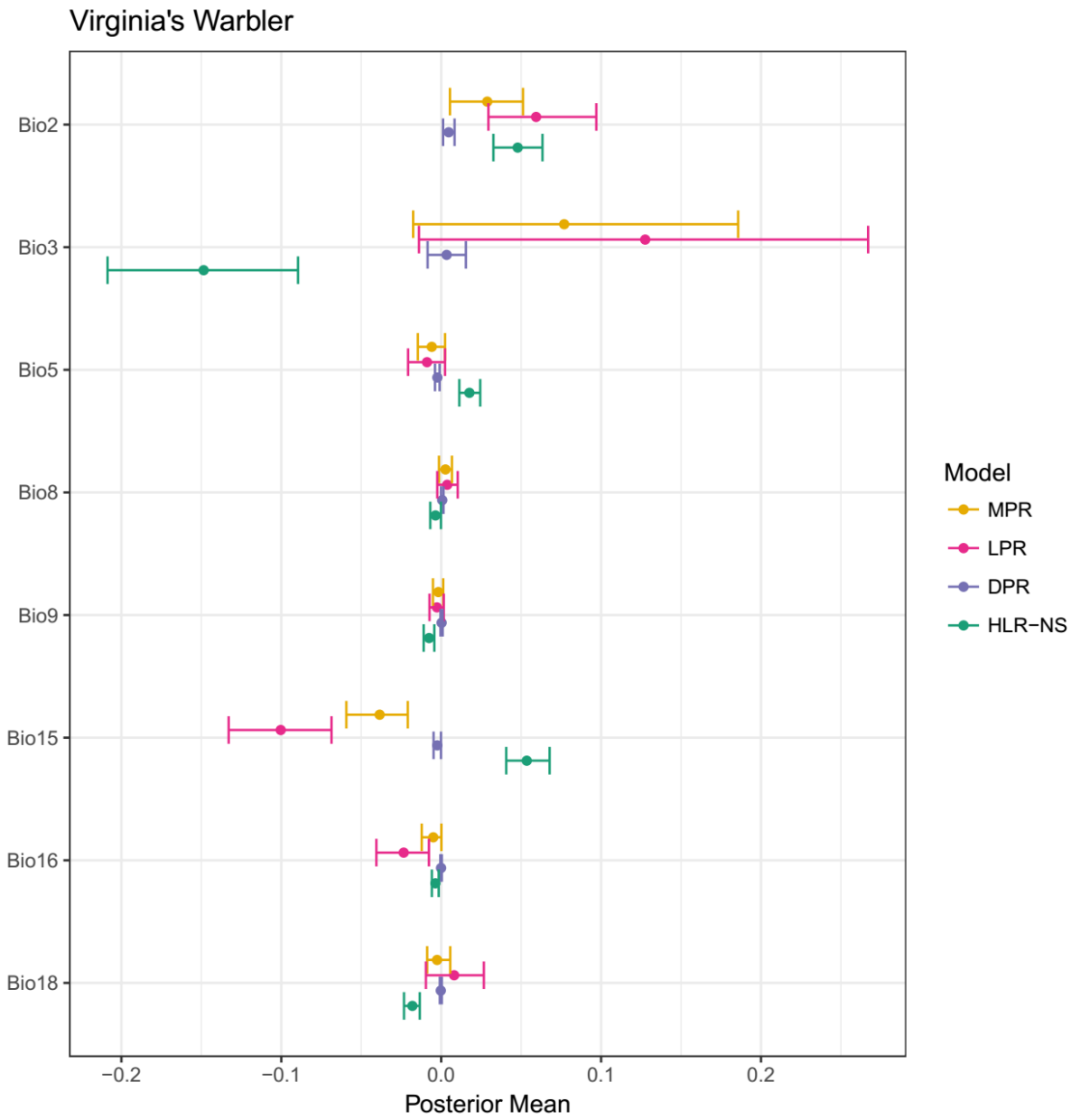


Figure 335: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Virginia's warbler, *Leiothlypis virginiae*.

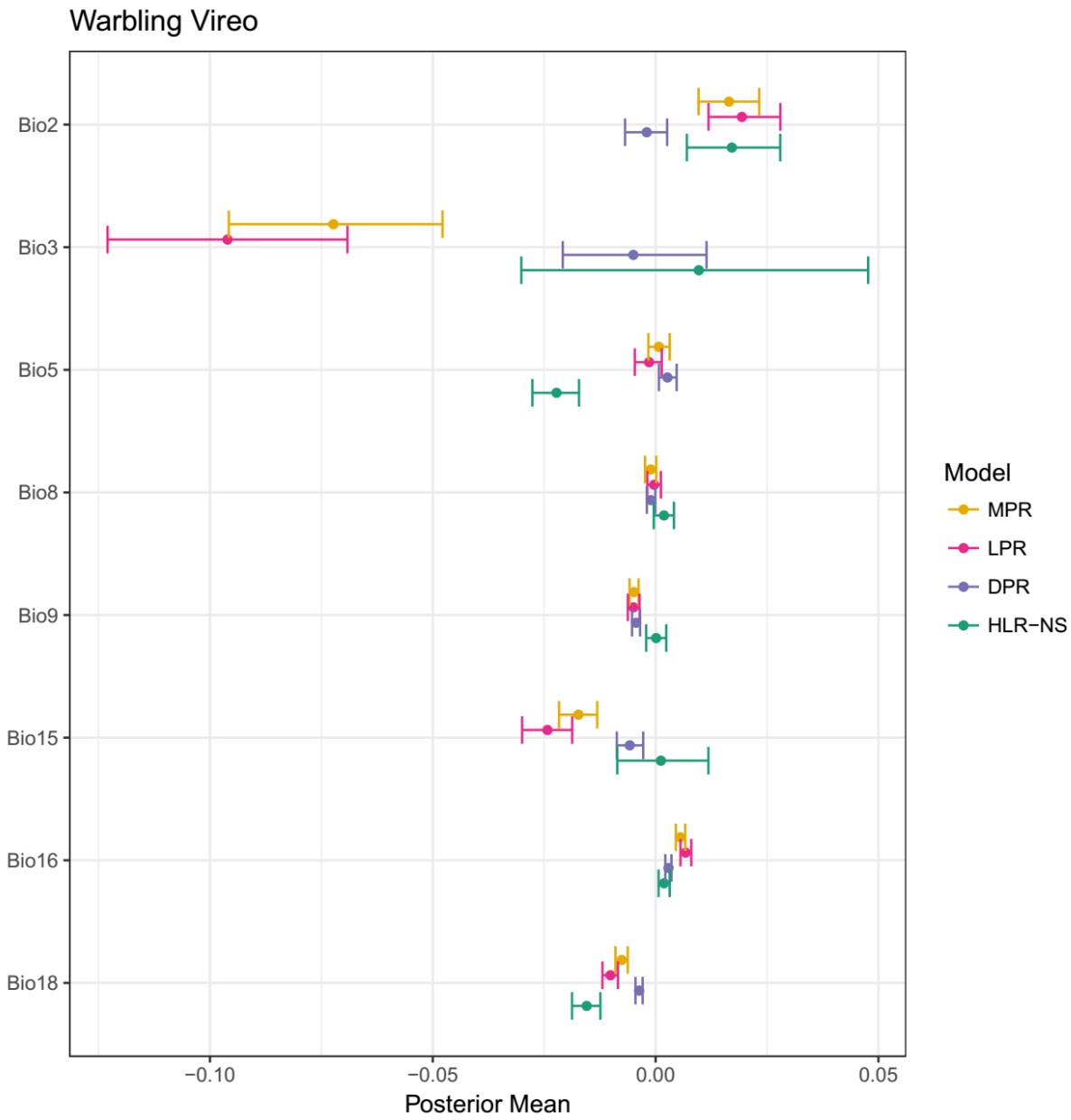


Figure 336: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the warbling vireo, *Vireo gilvus*.

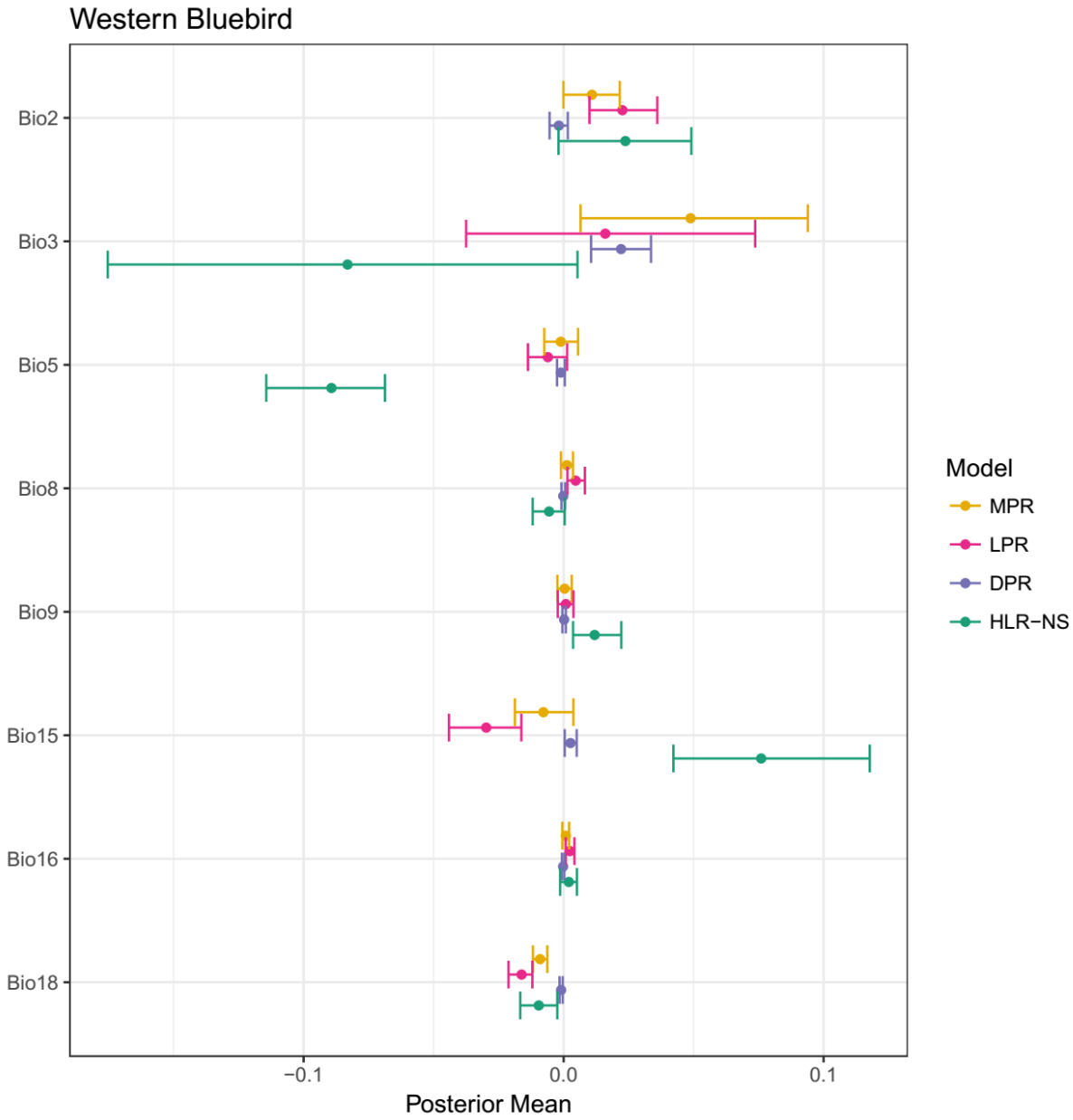


Figure 337: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the western bluebird, *Sialia mexicana*.

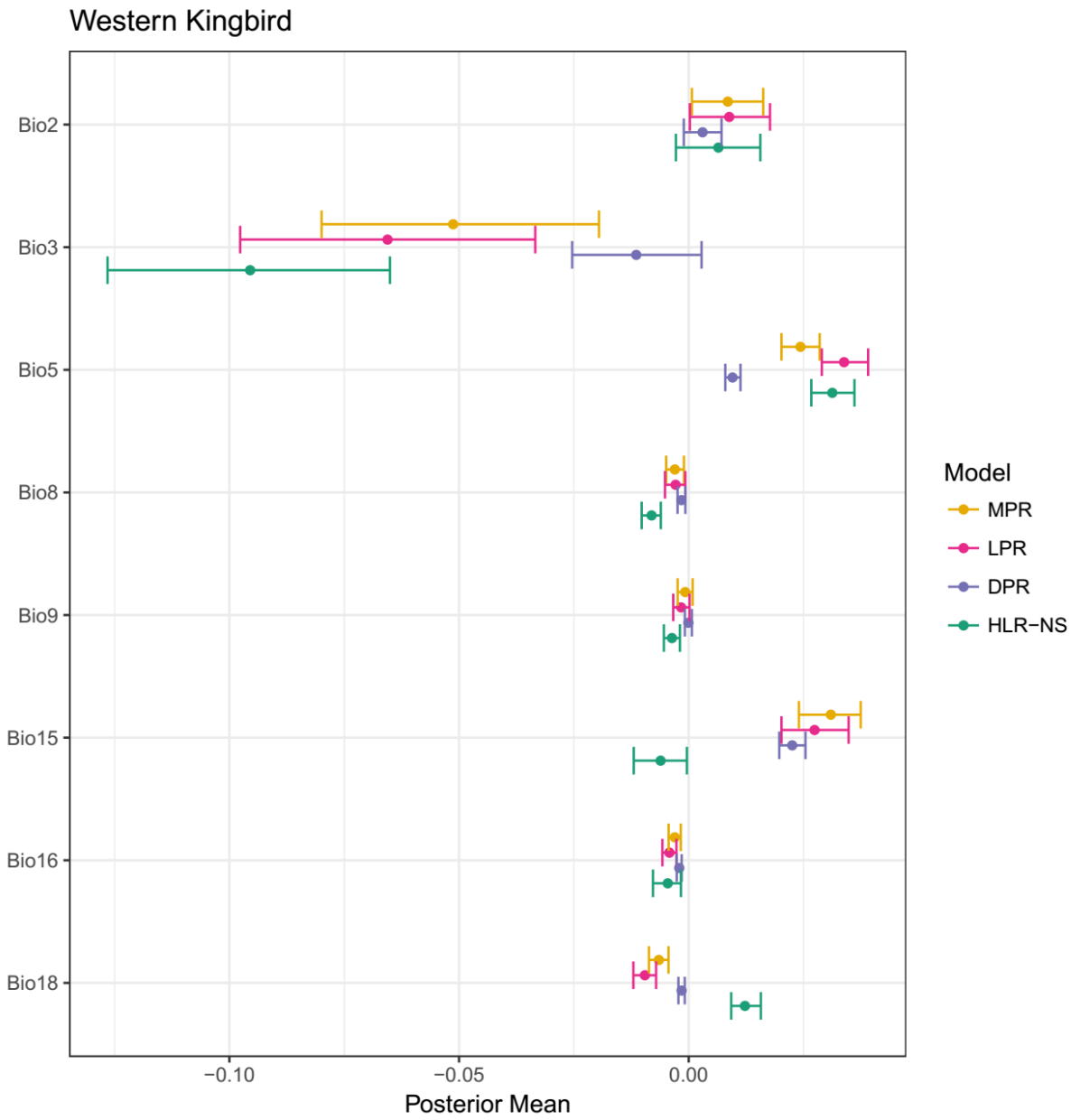


Figure 338: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the western kingbird, *Tyrannus verticalis*.

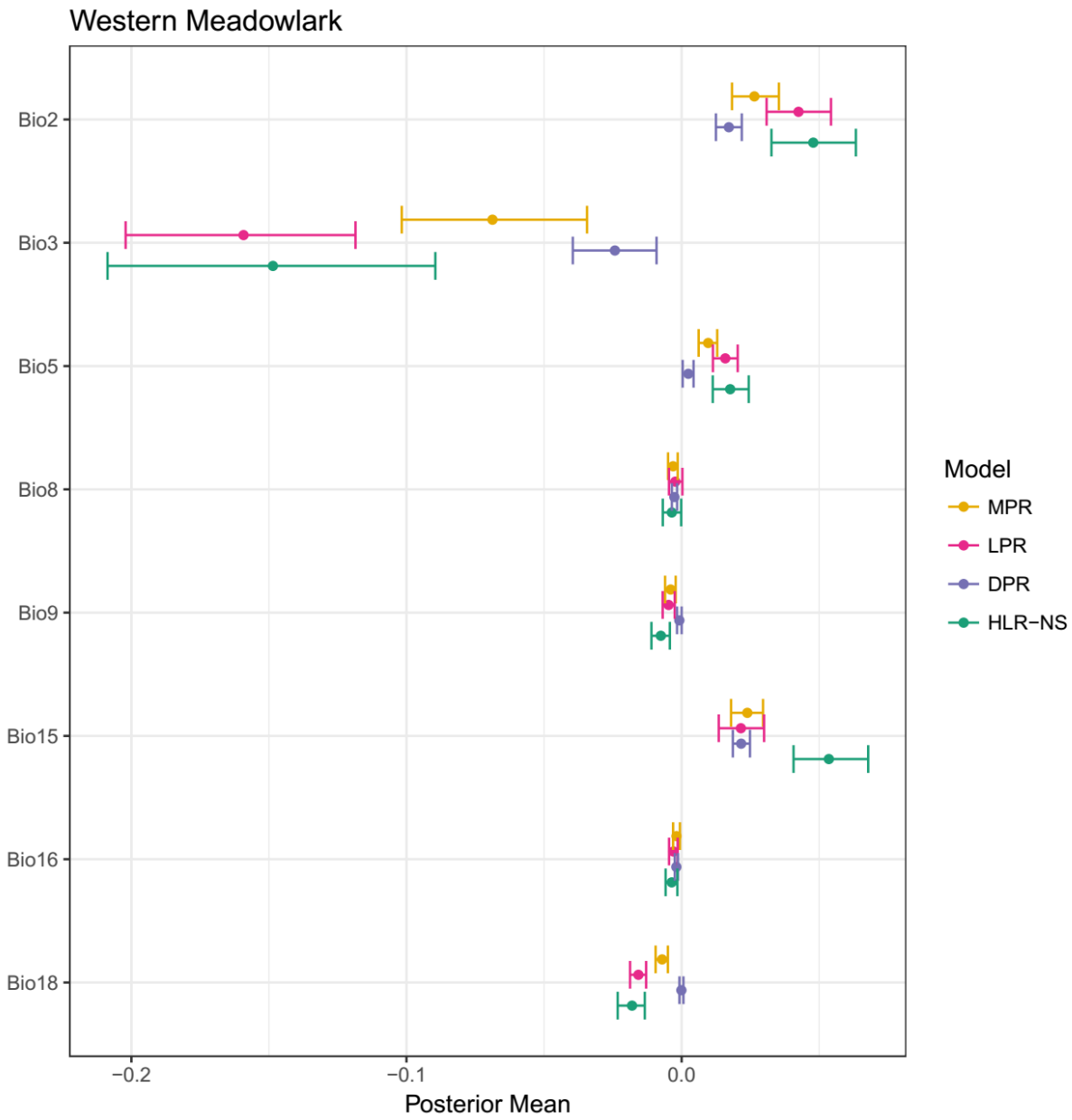


Figure 339: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the western meadowlark, *Sturnella neglecta*.

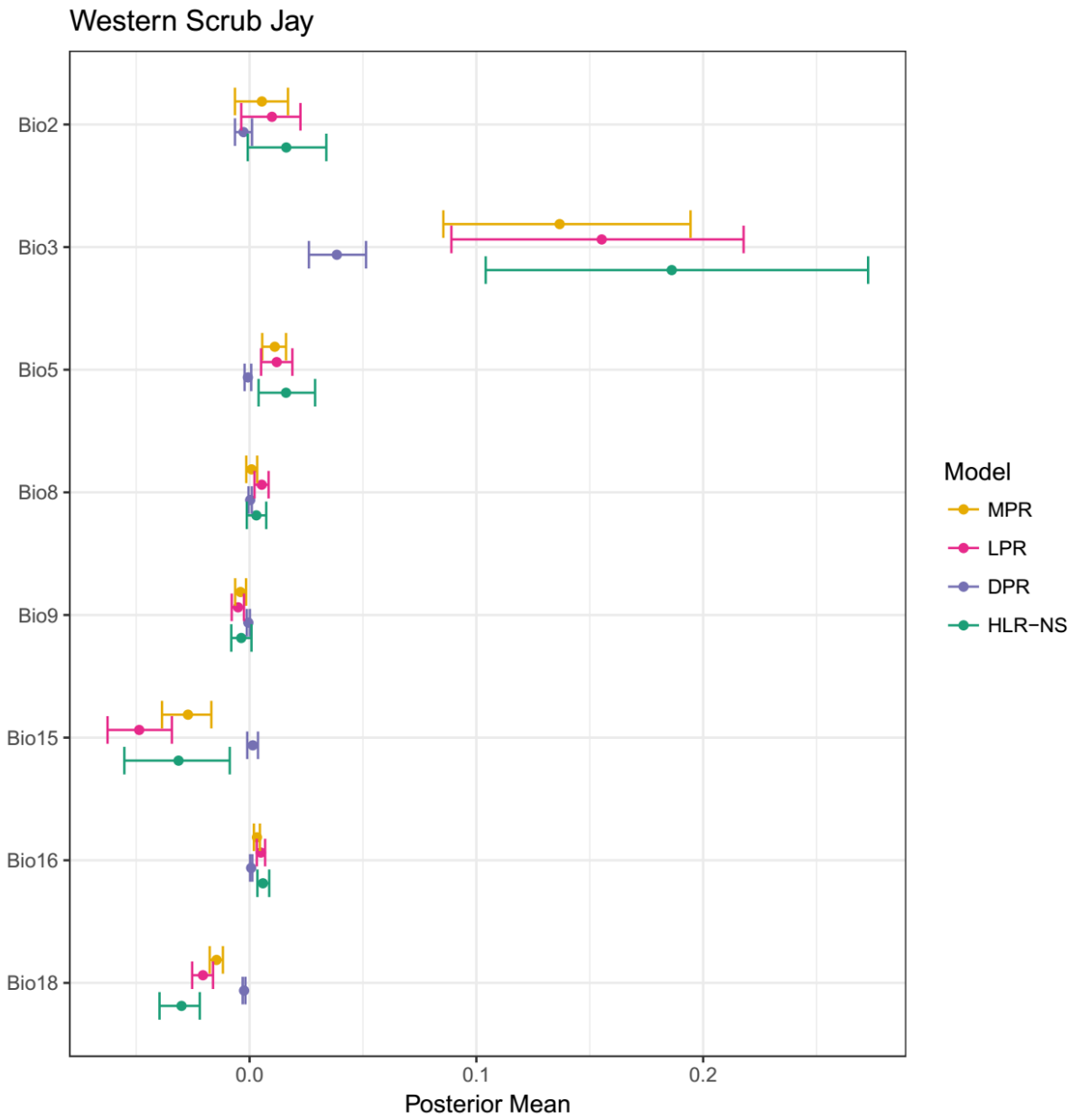


Figure 340: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the western scrub jay, *Aphelocoma californica*.

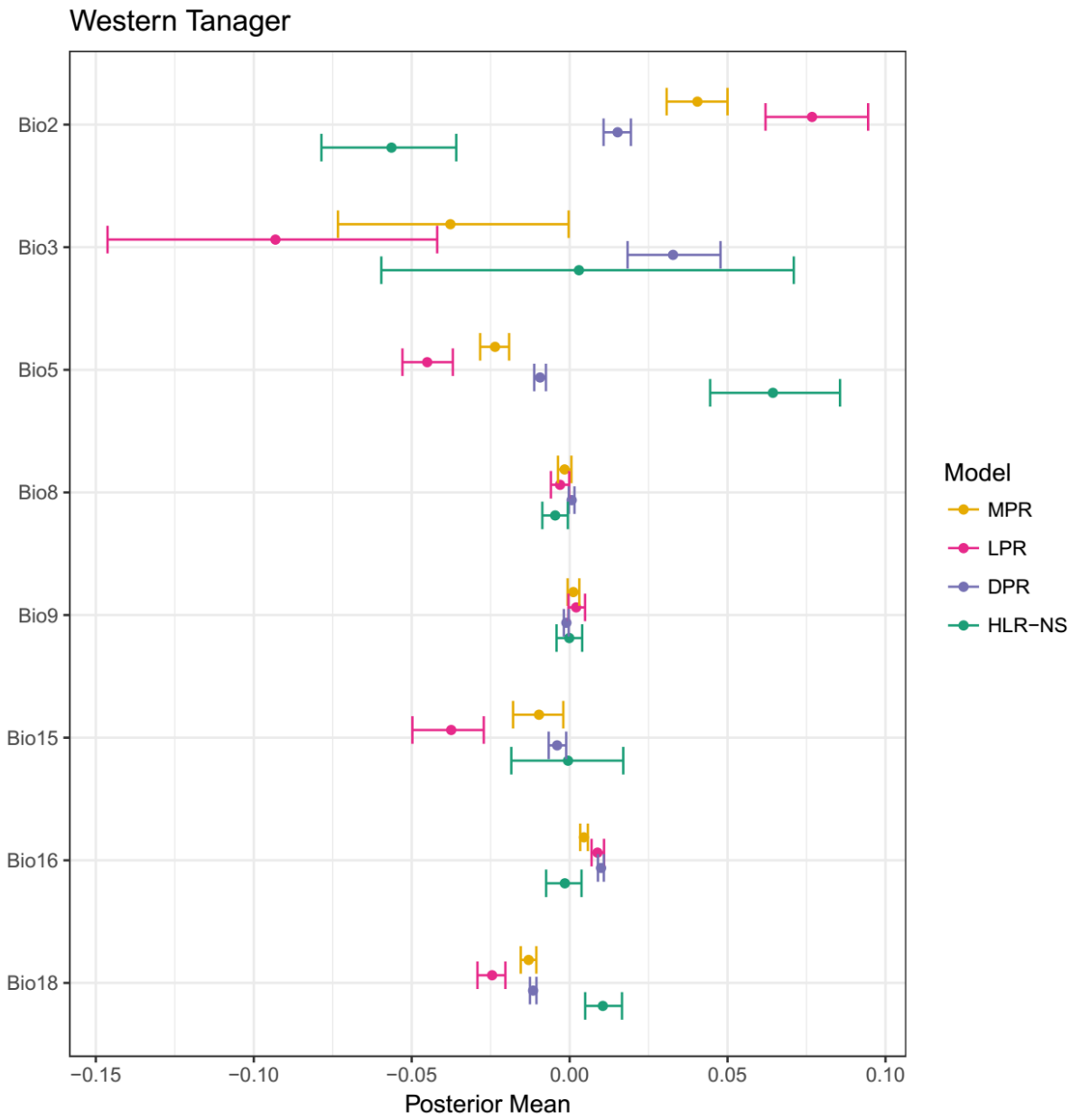


Figure 341: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the western tanager, *Piranga ludoviciana*.

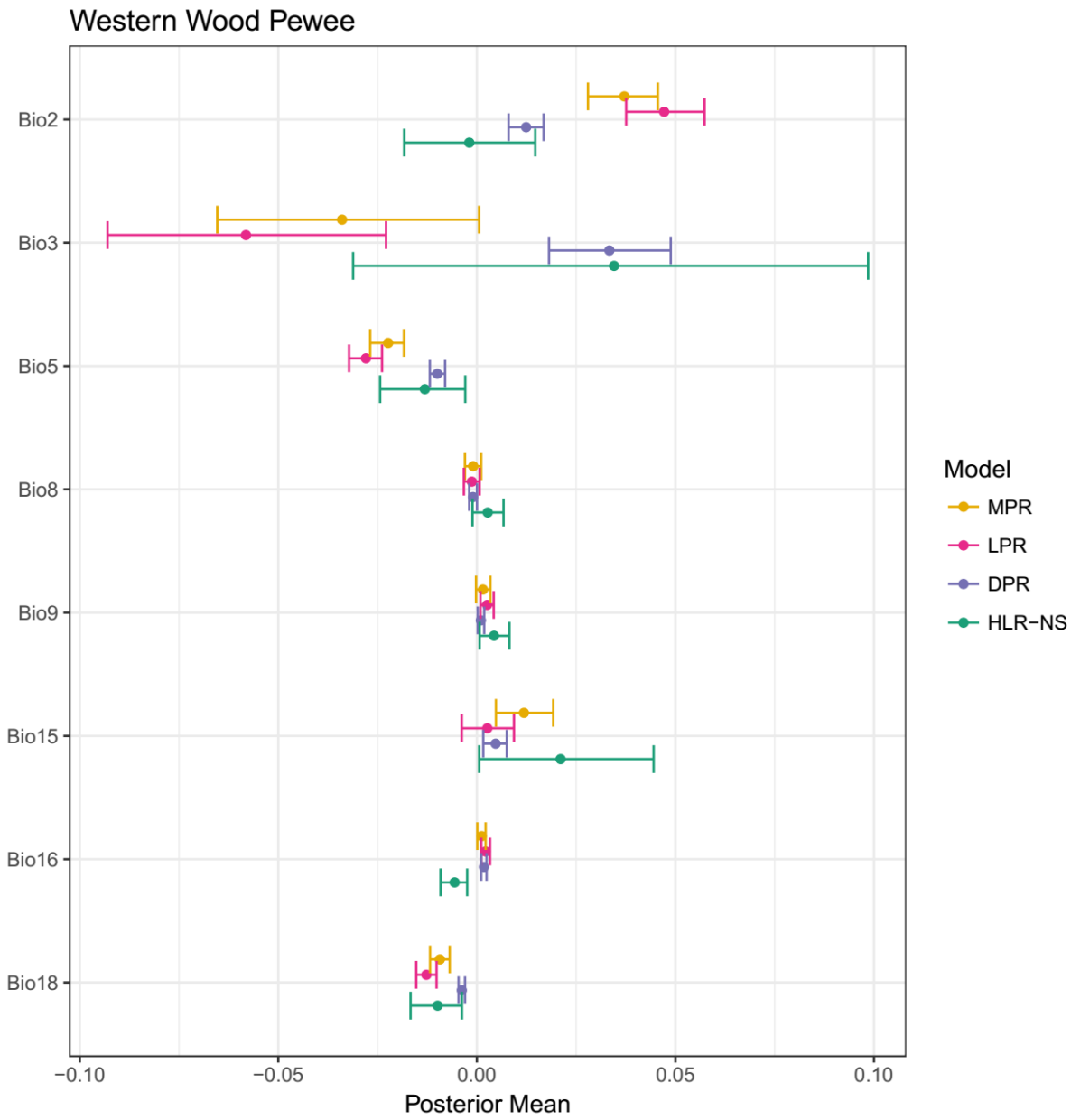


Figure 342: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the western wood pewee, *Contopus sordidulus*.

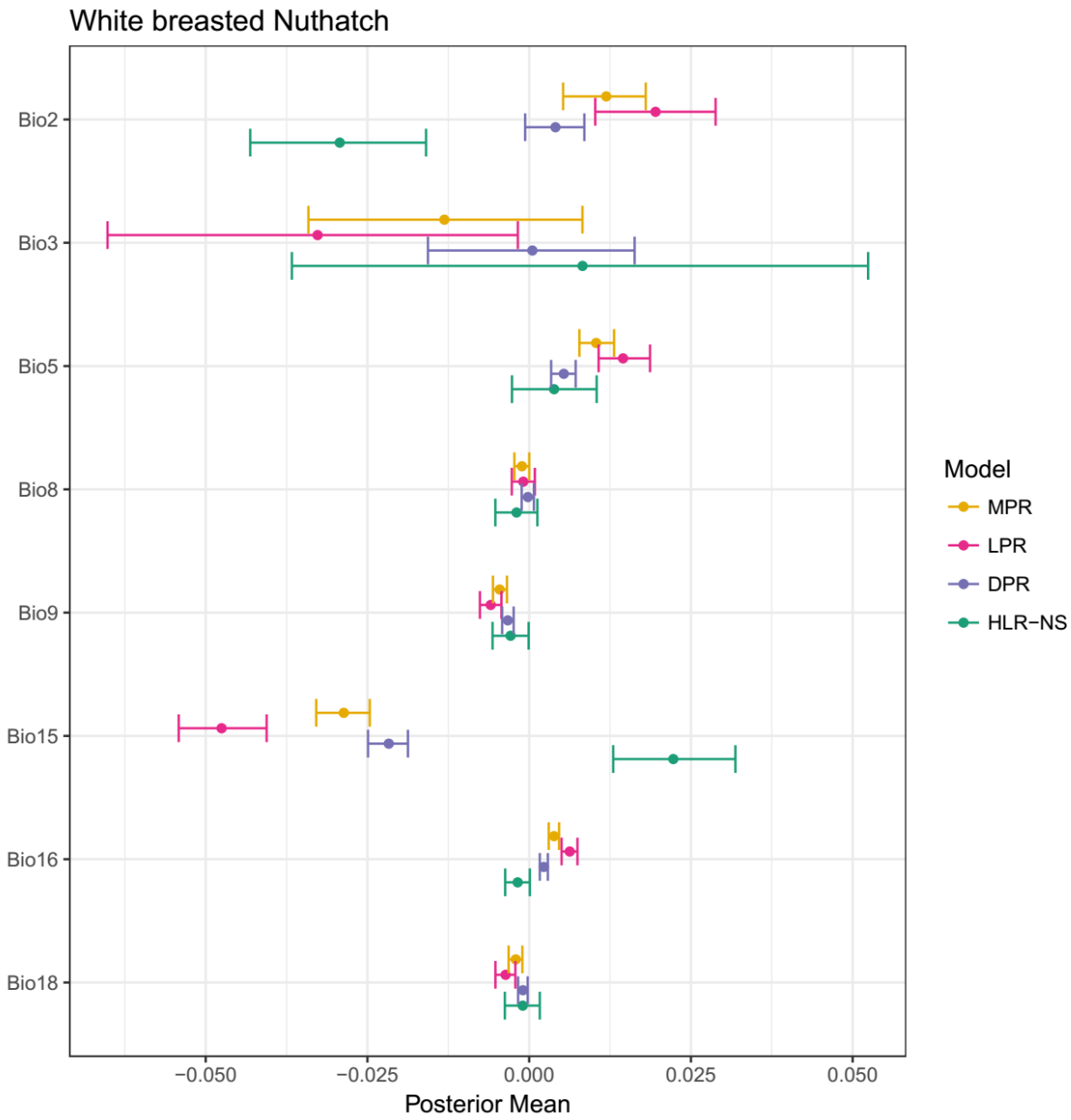


Figure 343: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-breasted nuthatch, *Sitta carolinensis*.

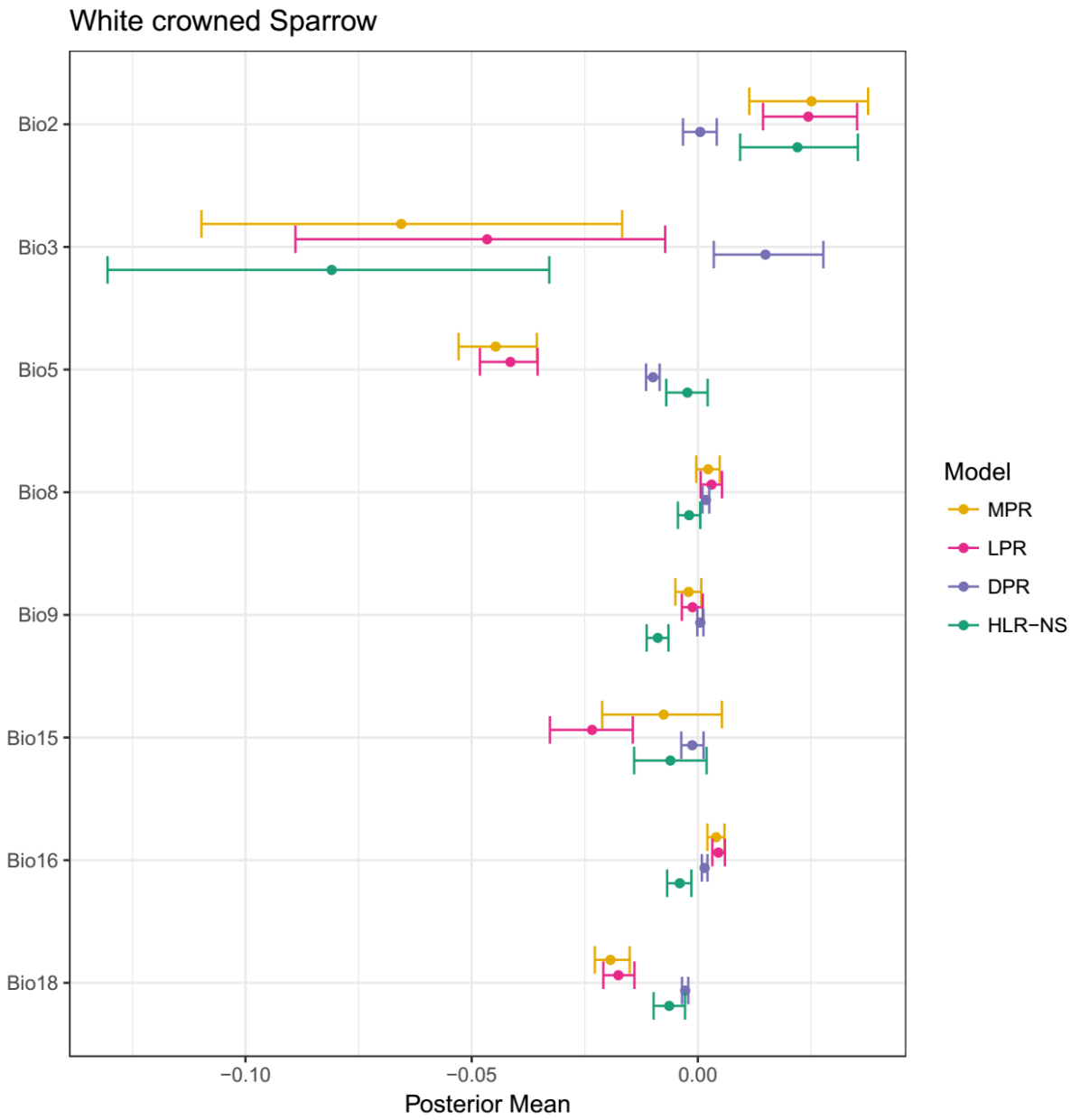


Figure 344: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-crowned sparrow, *Zonotrichia leucophrys*.

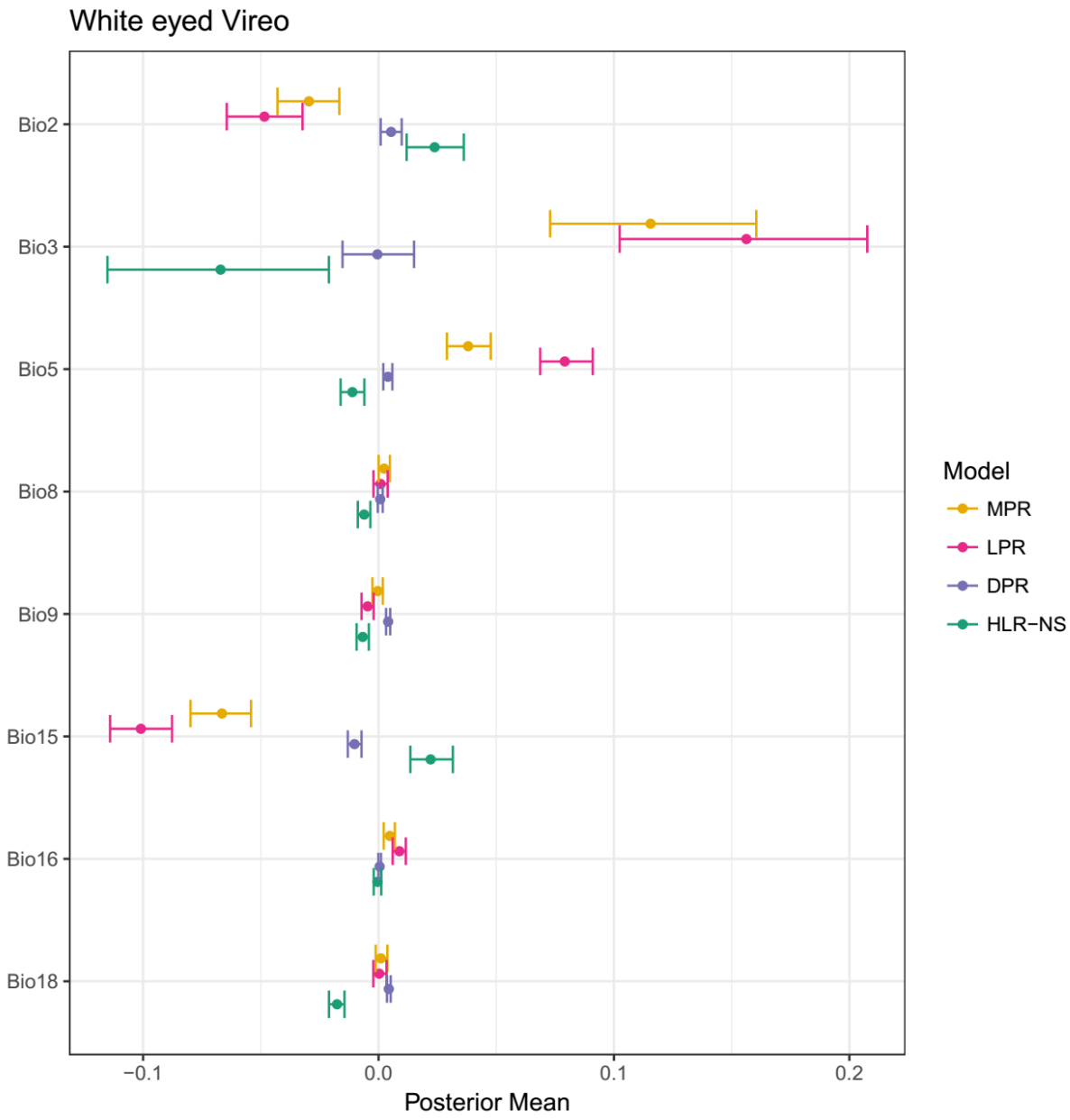


Figure 345: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-eyed vireo, *Vireo griseus*.

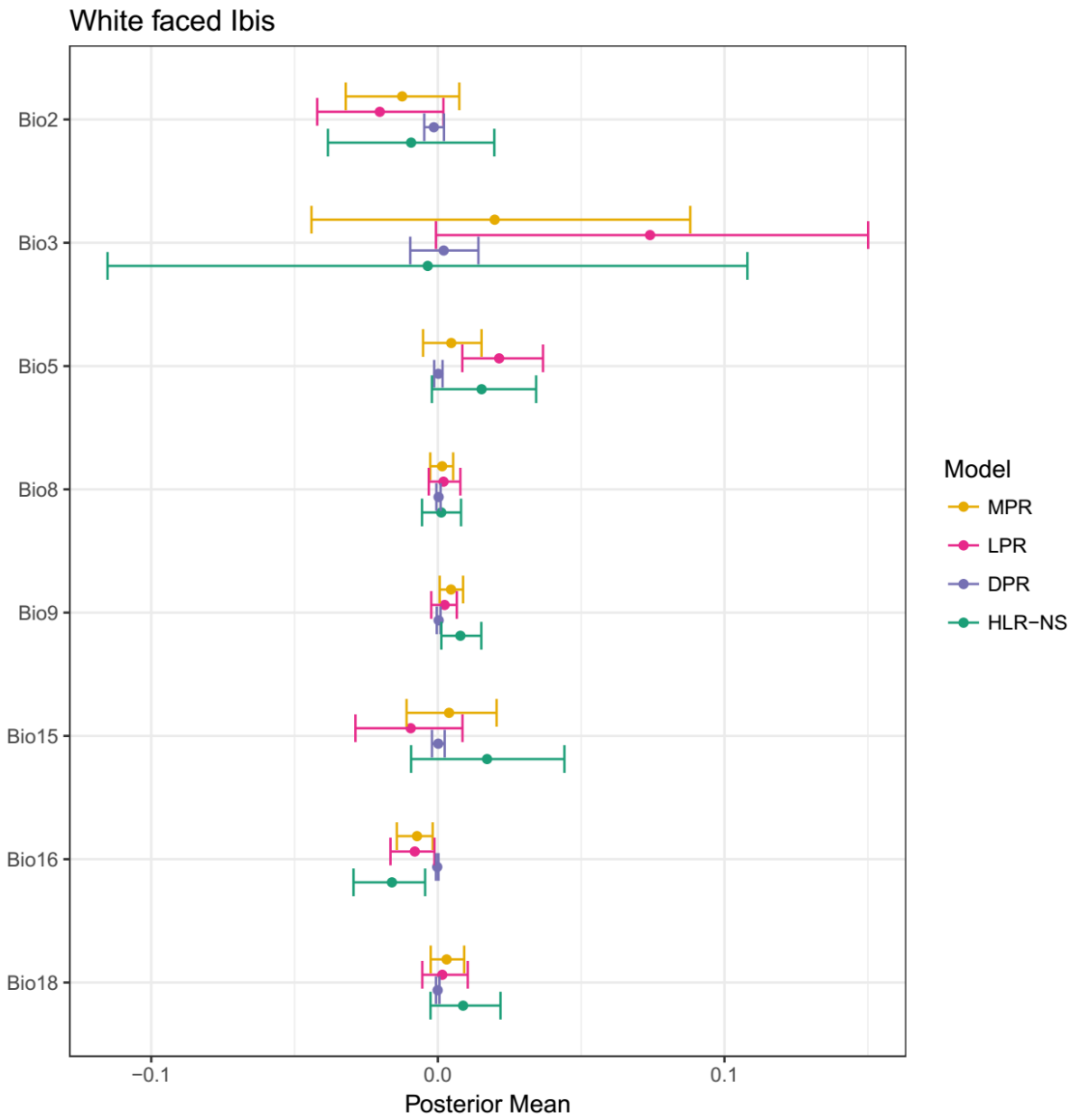


Figure 346: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-faced ibis, *Plegadis chihi*.

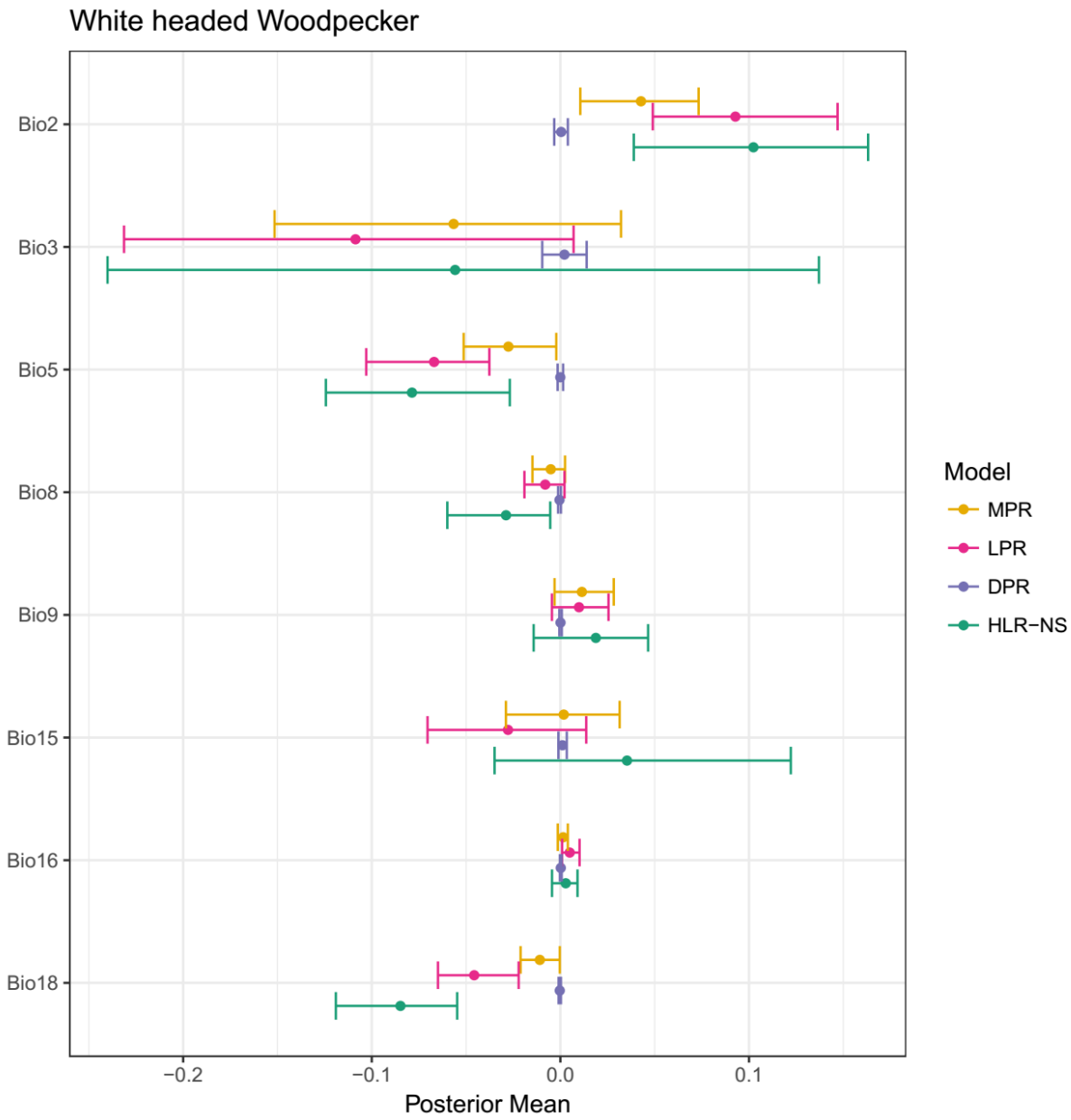


Figure 347: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-headed woodpecker, *Picoides albolarvatus*.

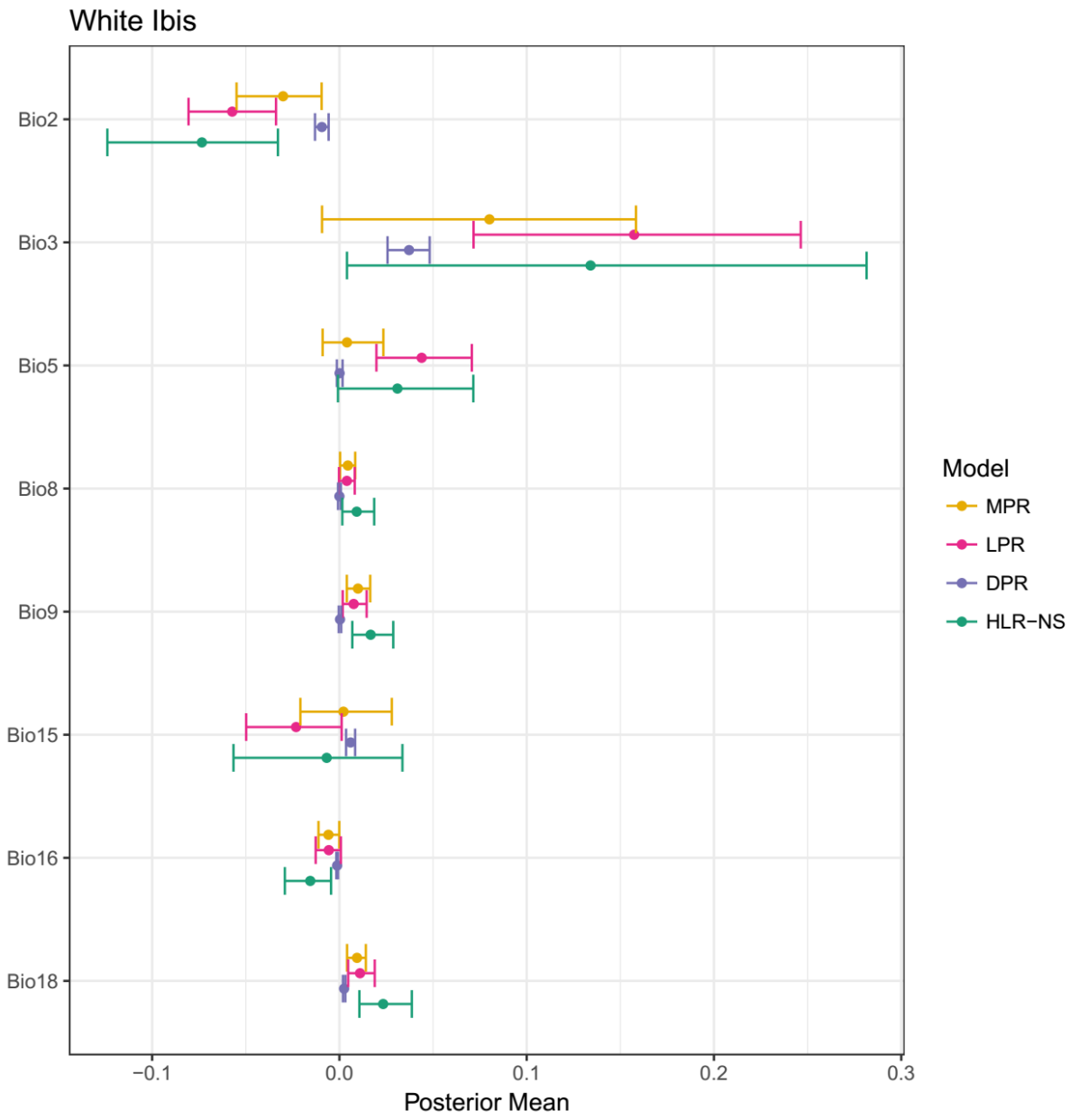


Figure 348: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white ibis, *Eudocimus albus*.

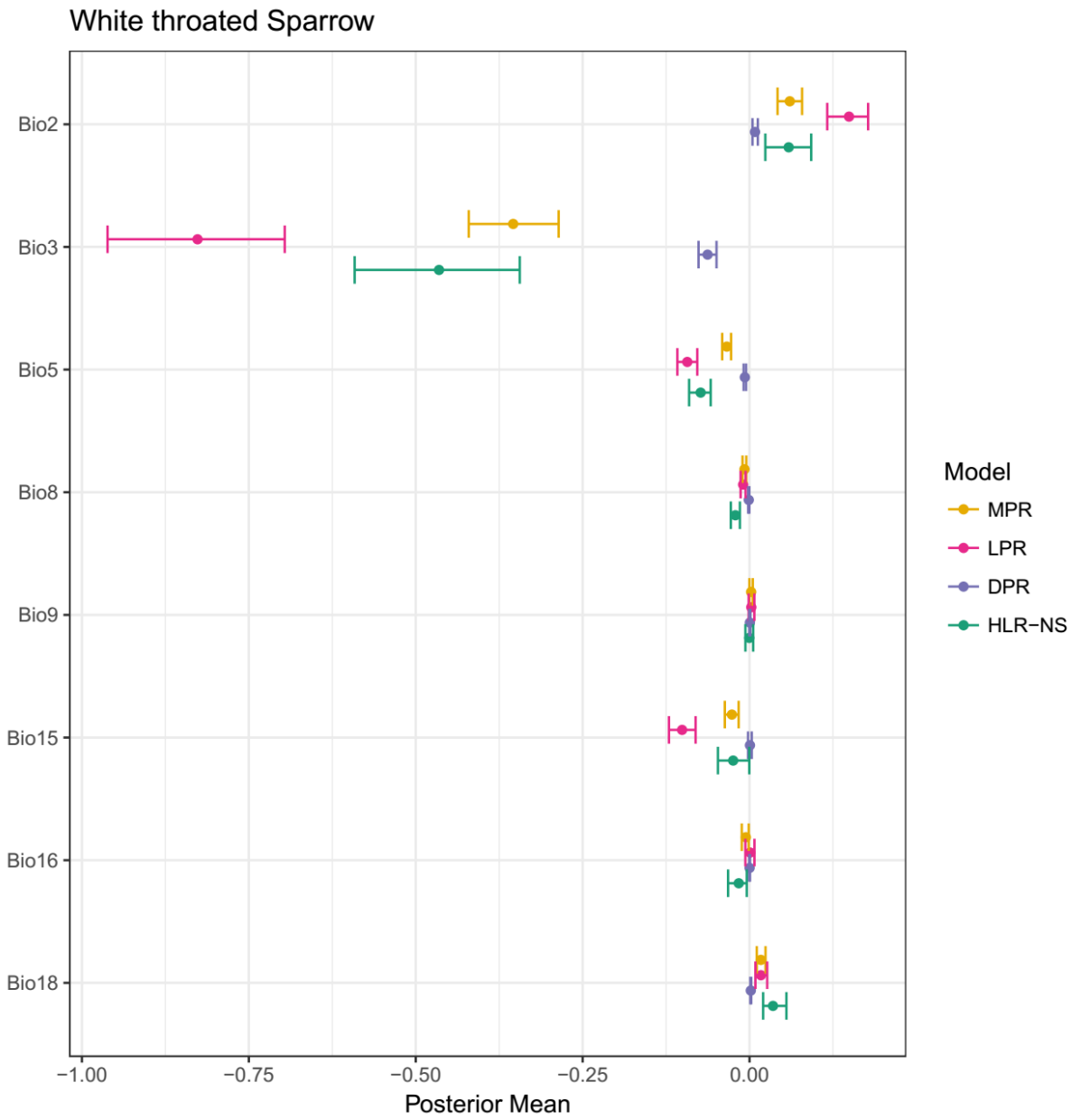


Figure 349: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-throated sparrow, *Zonotrichia albicollis*.

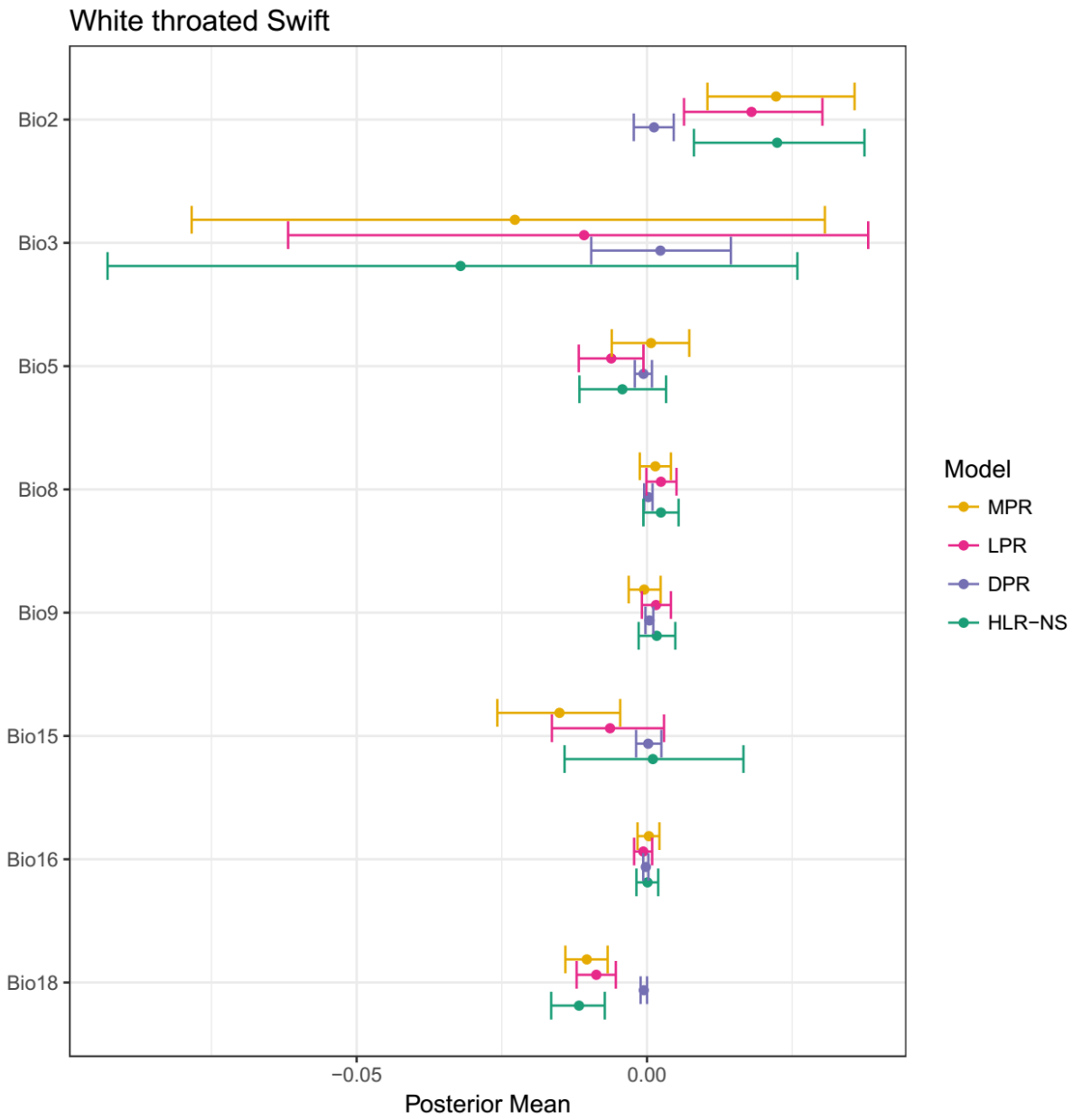


Figure 350: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-throated swift, *Aeronautes saxatalis*.

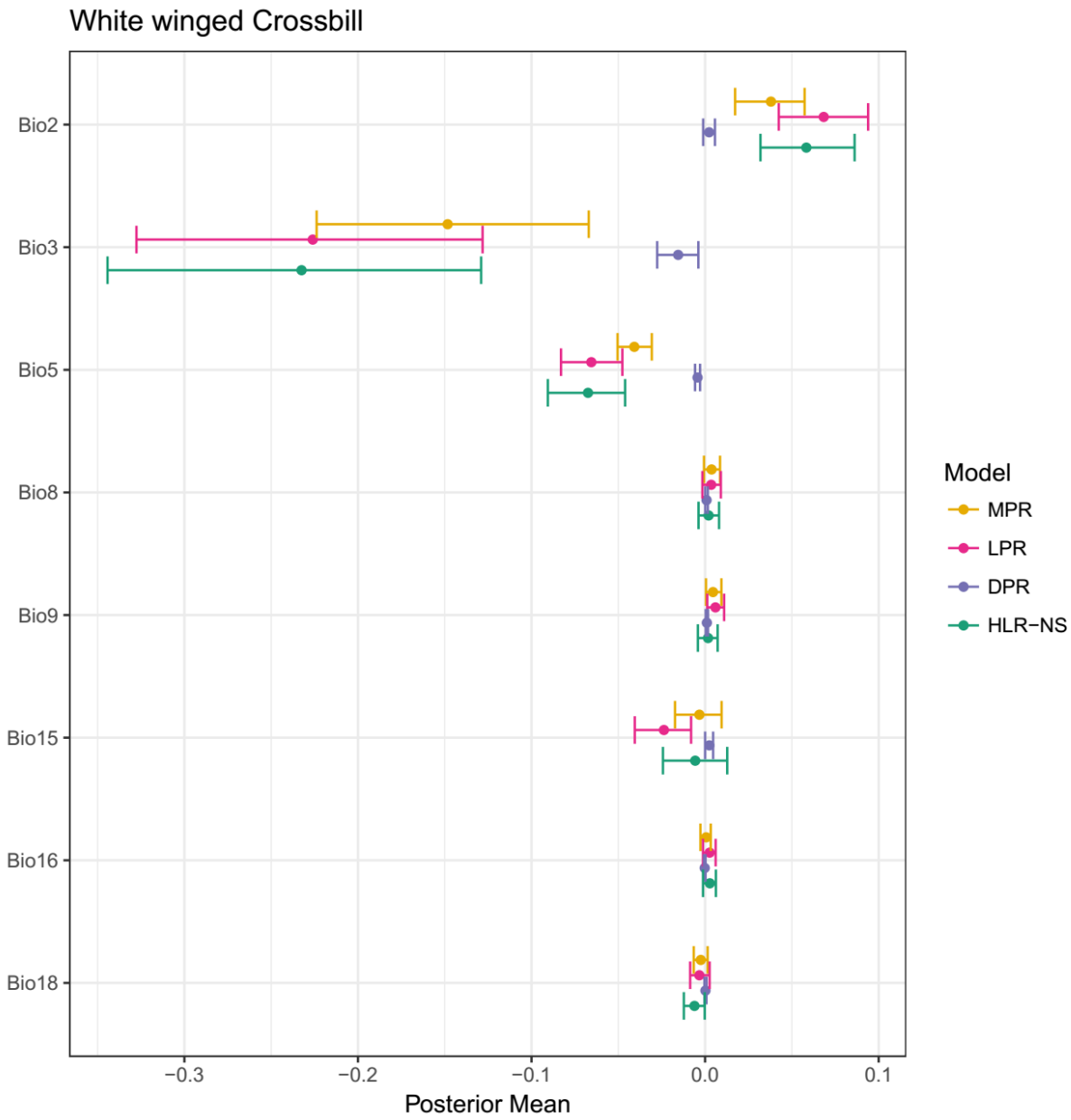


Figure 351: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-winged crossbill, *Loxia leucoptera*.

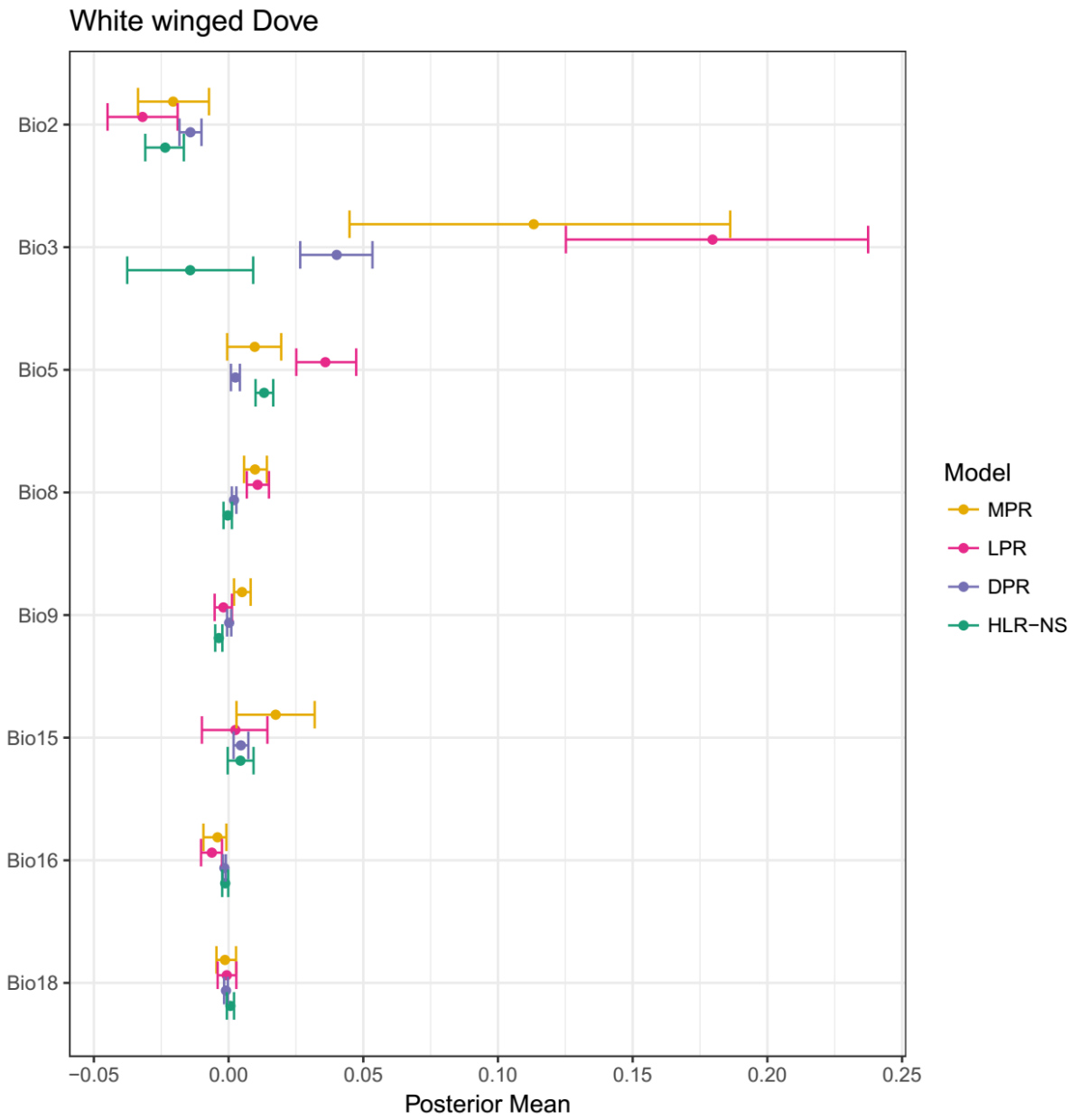


Figure 352: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the white-winged dove, *Zenaida asiatica*.

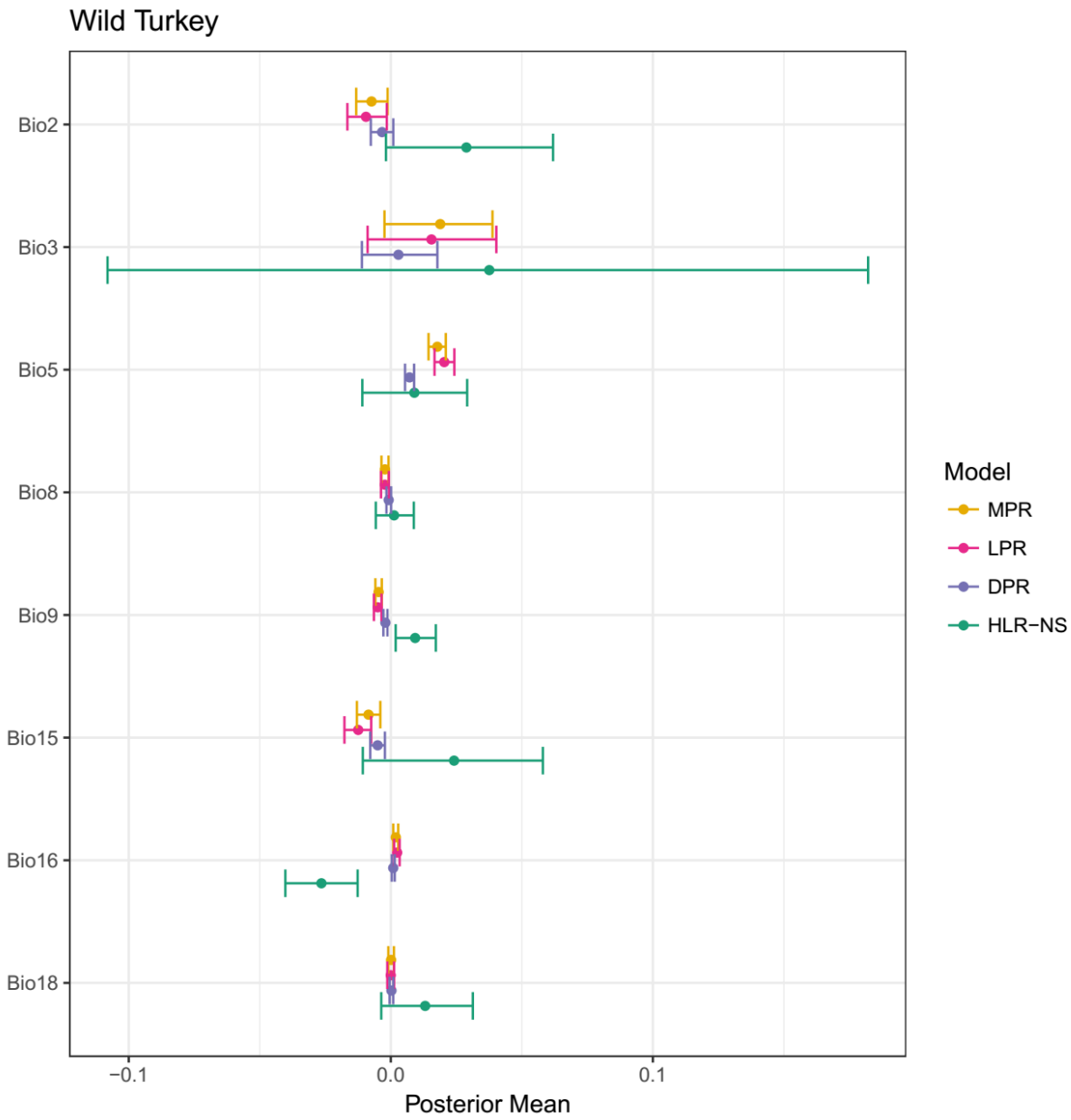


Figure 353: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the wild turkey, *Meleagris gallopavo*.

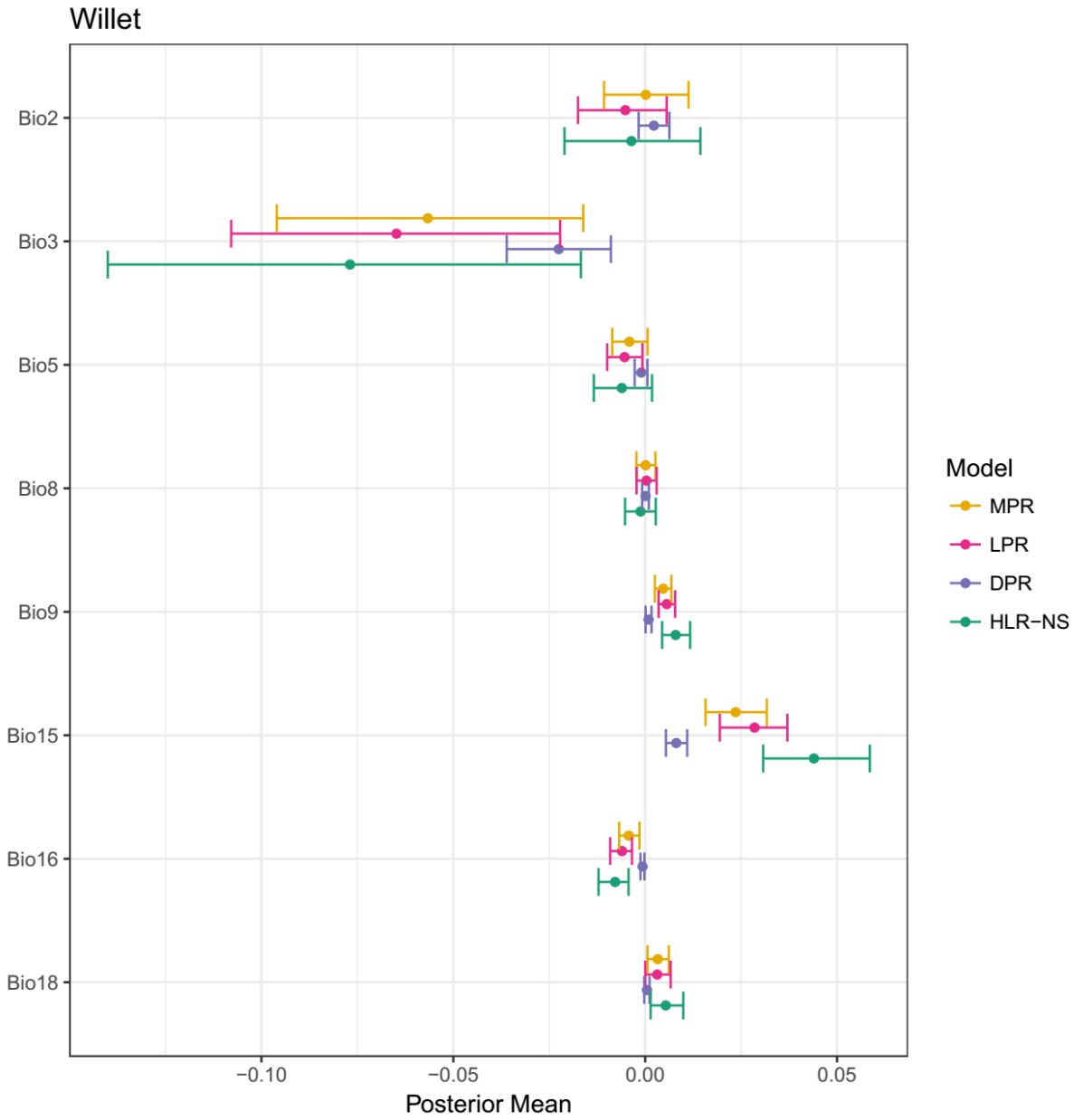


Figure 354: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the willet, *Tringa semipalmata*.

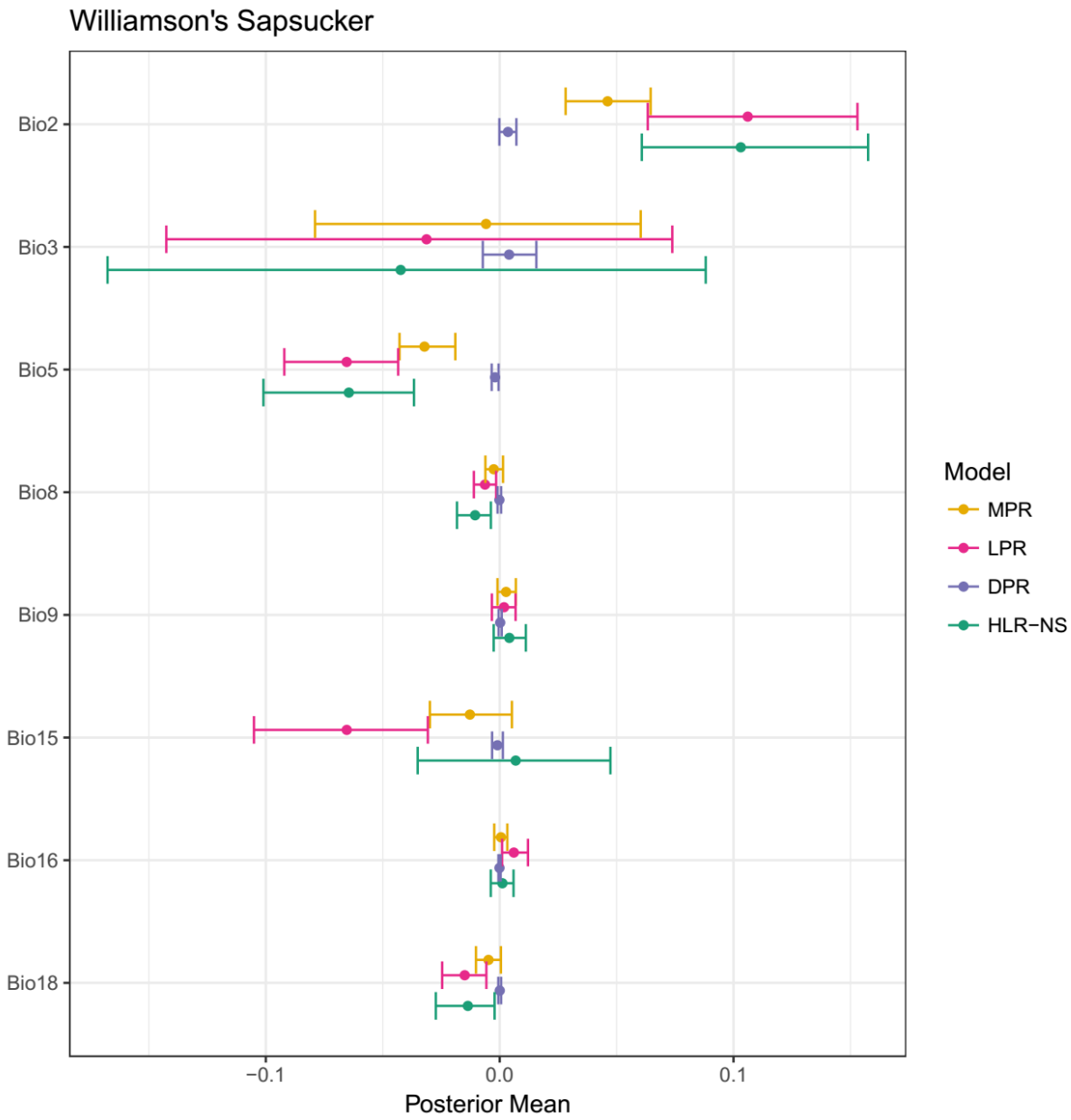


Figure 355: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Williamson's sapsucker, *Sphyrapicus thyroideus*.

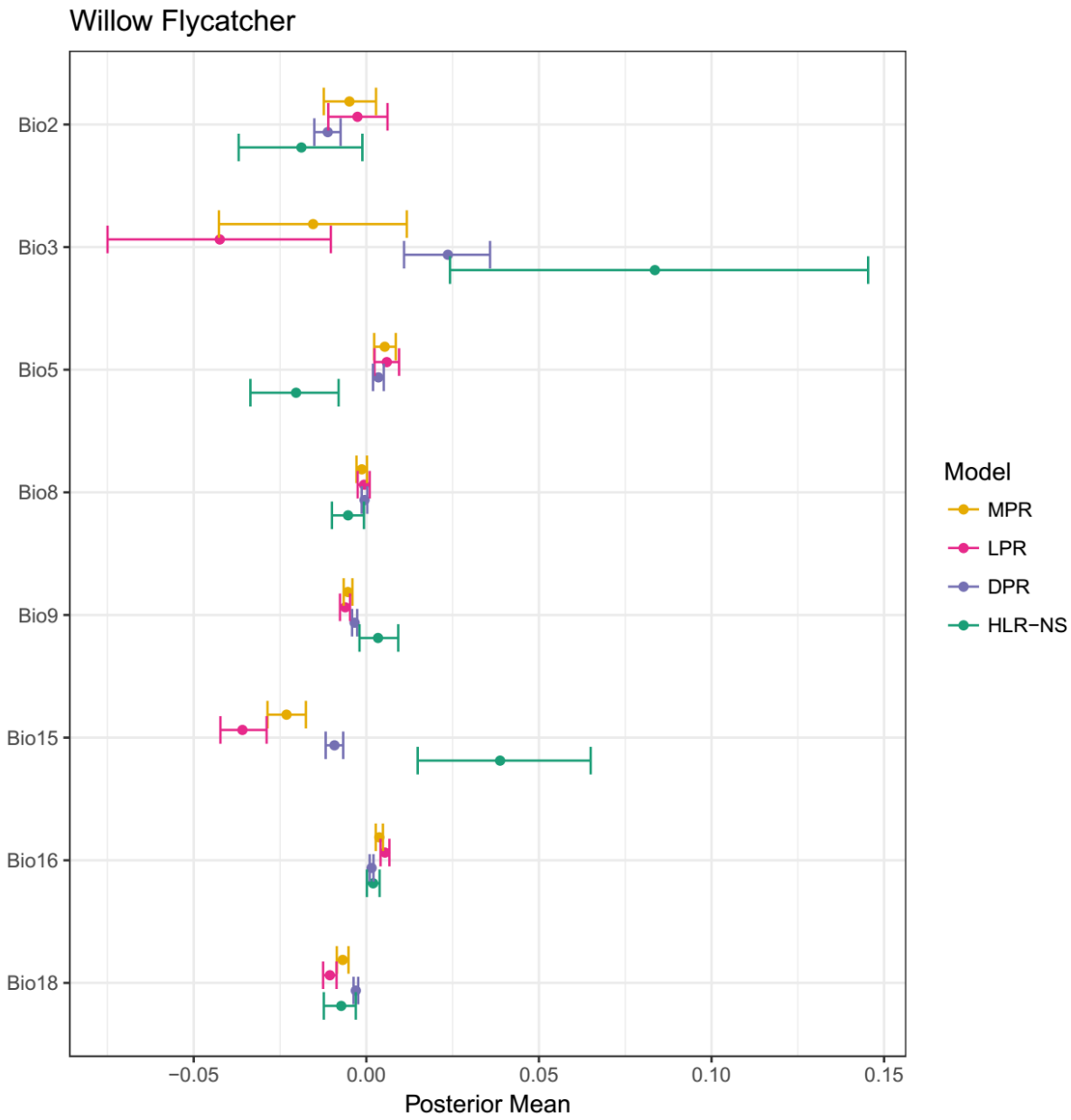


Figure 356: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the willow flycatcher, *Empidonax traillii*.

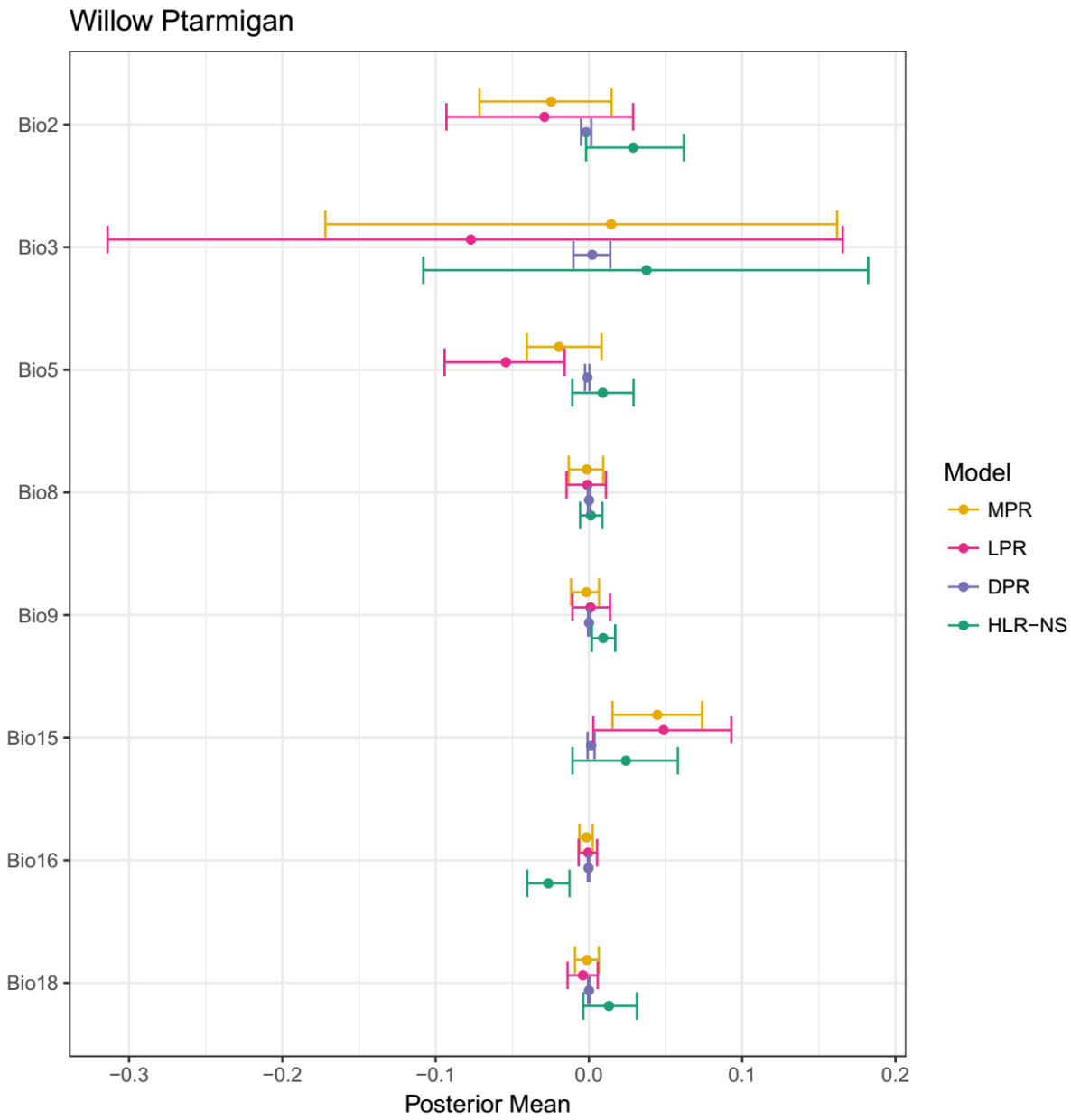


Figure 357: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the willow ptarmigan, *Lagopus lagopus*.

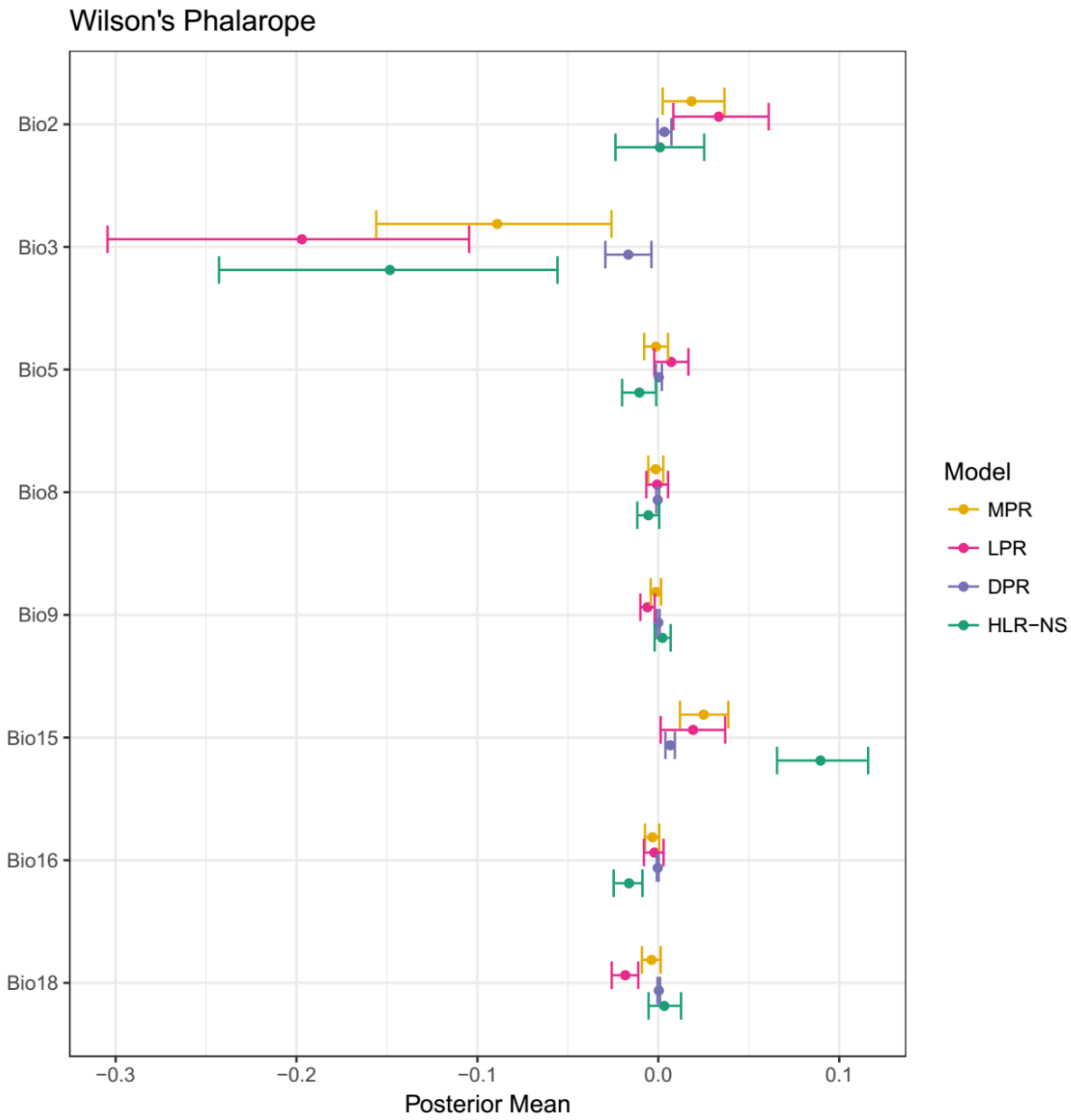


Figure 358: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Wilson's pharalope, *Phalaropus tricolor*.

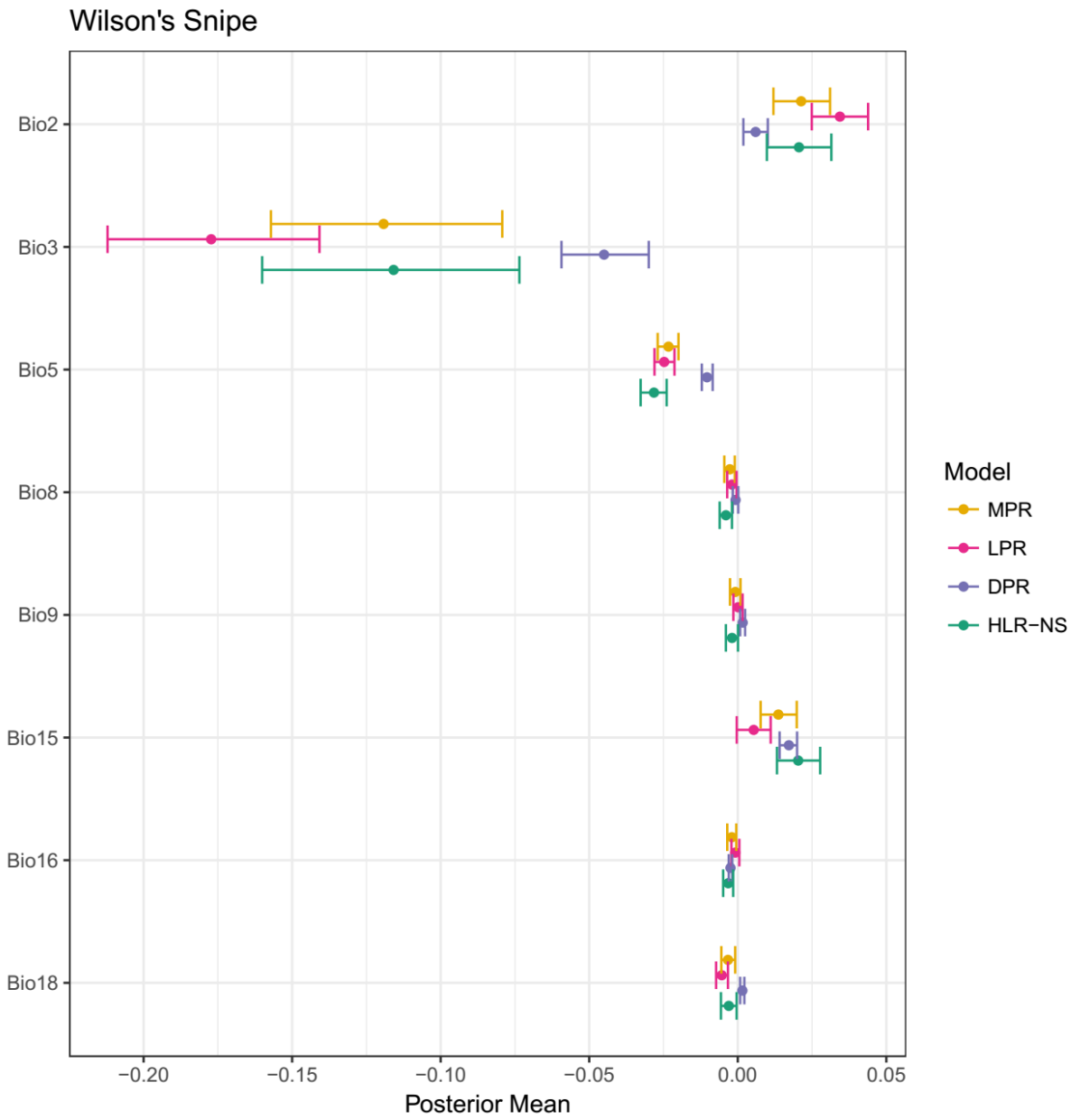


Figure 359: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Wilson's snipe, *Gallinago delicata*.

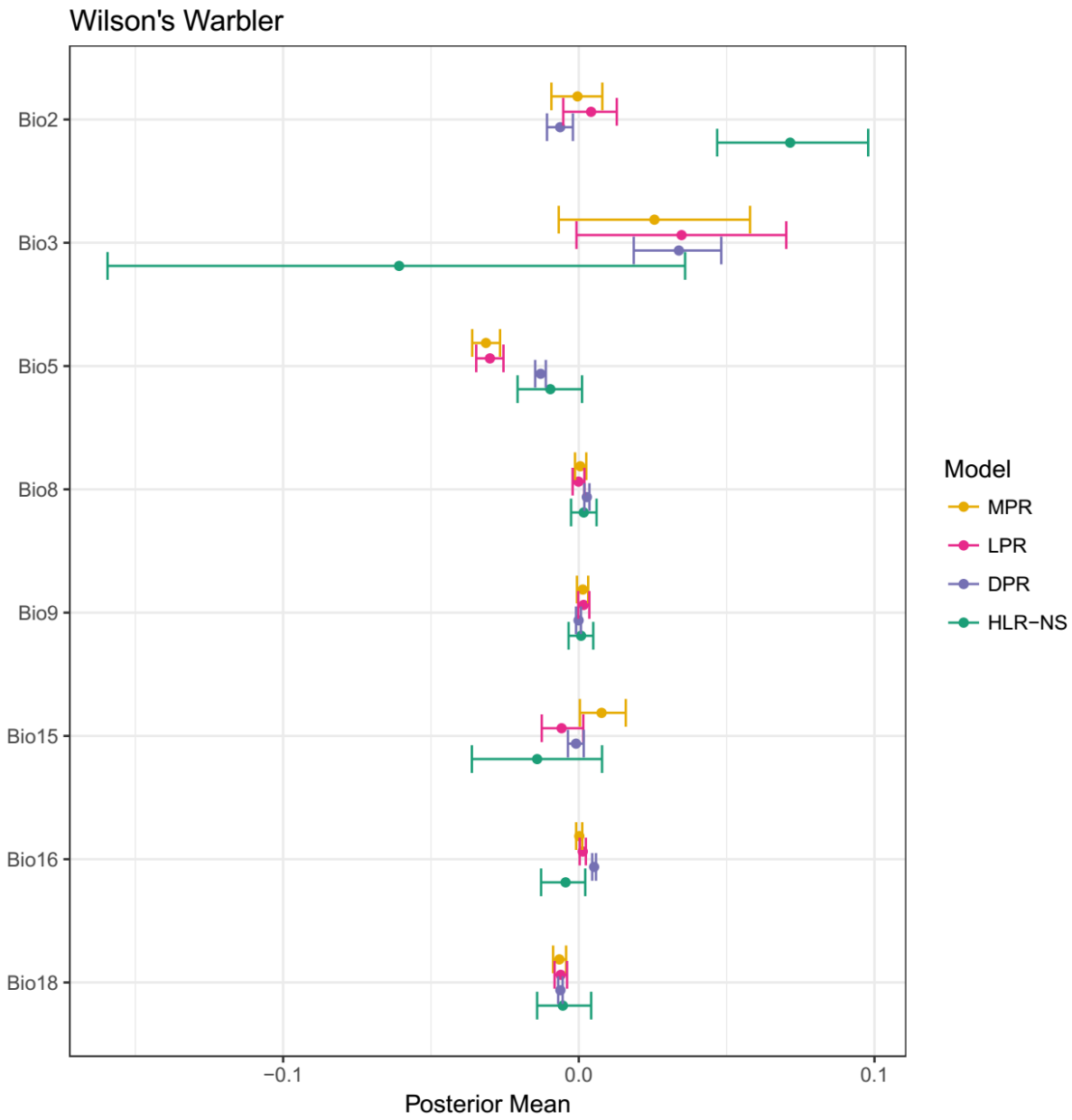


Figure 360: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Wilson's warbler, *Cardellina pusilla*.

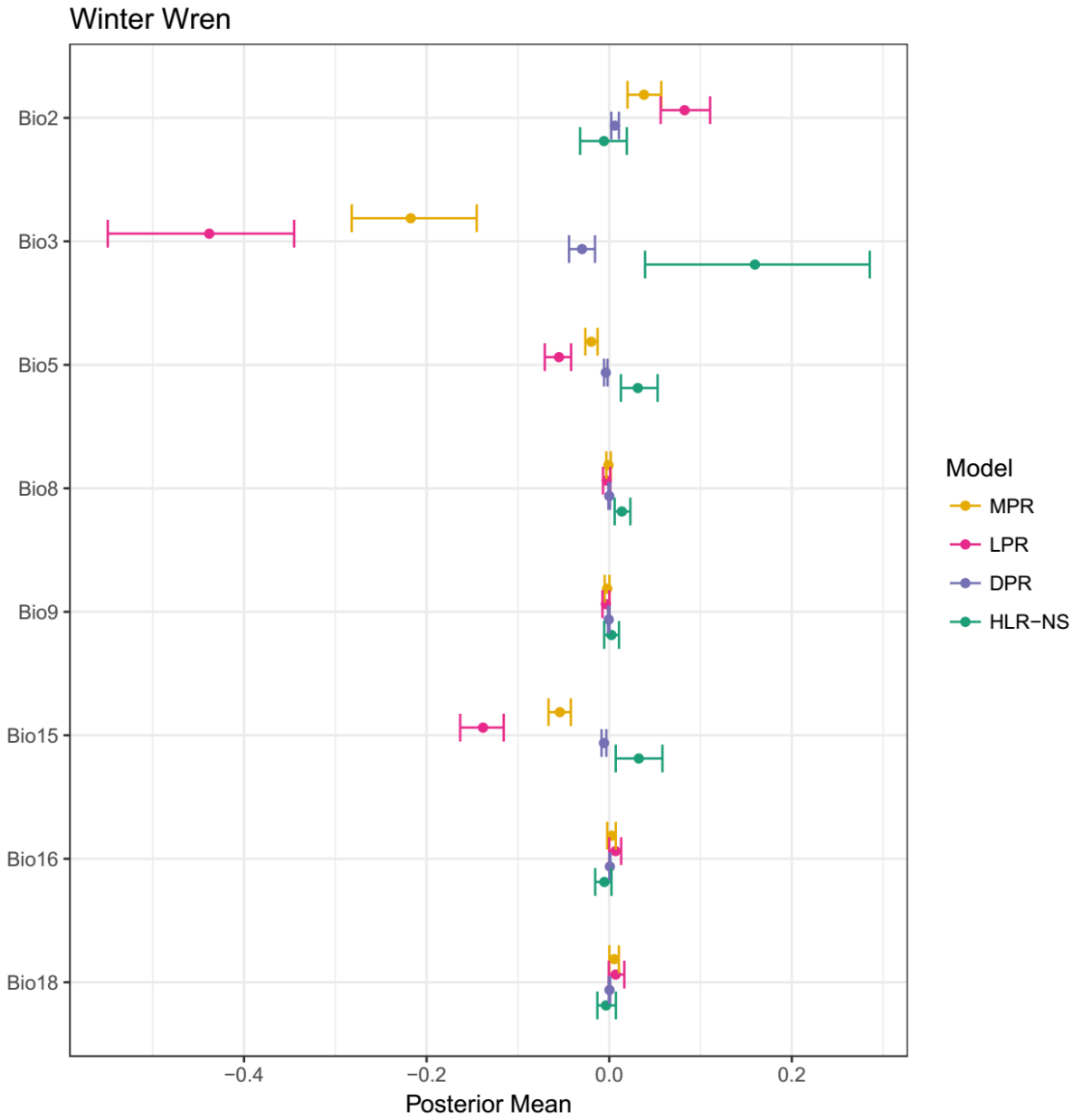


Figure 361: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the winter wren, *Troglodytes hiemalis*.

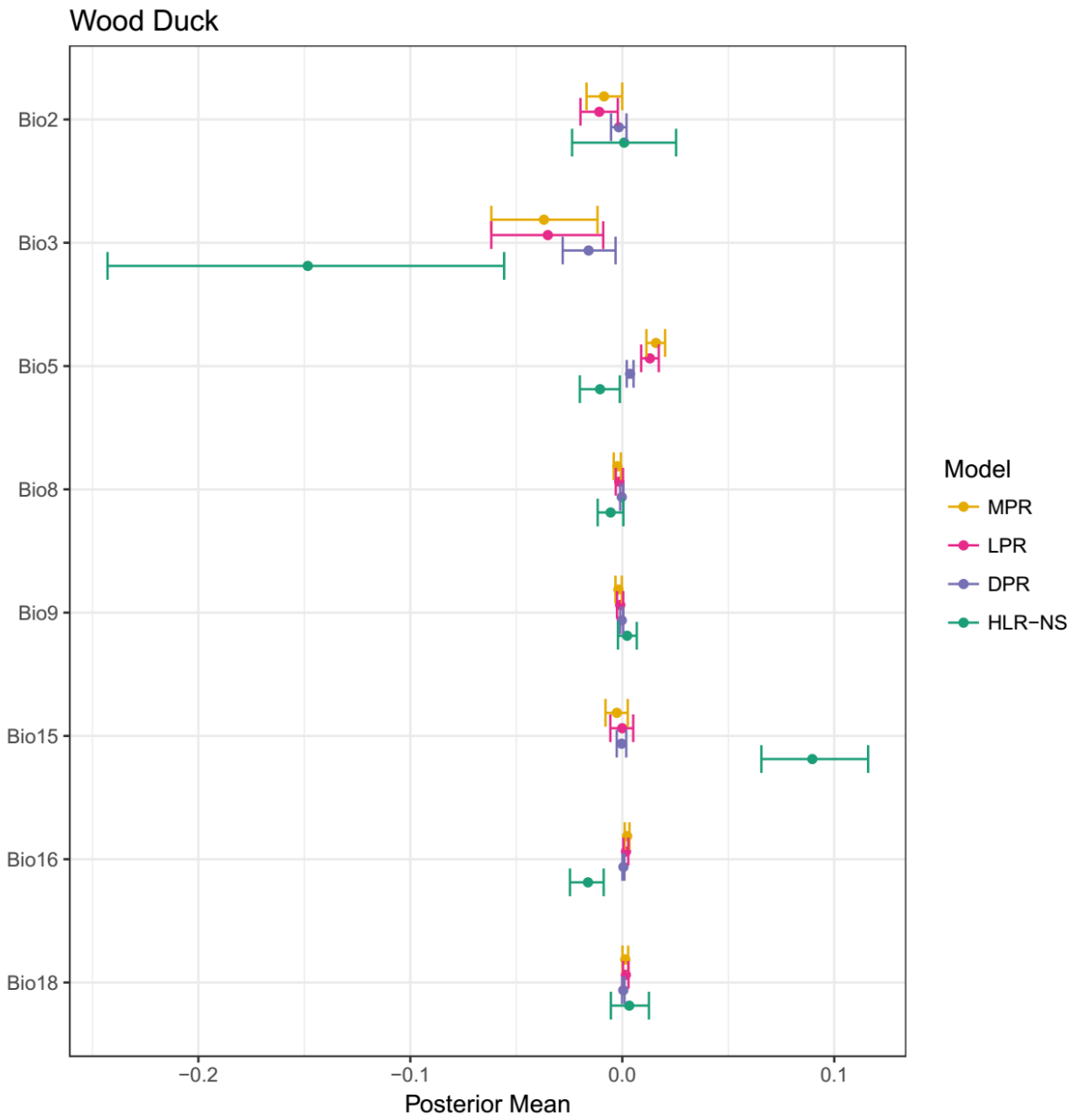


Figure 362: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the wood duck, *Aix sponsa*.

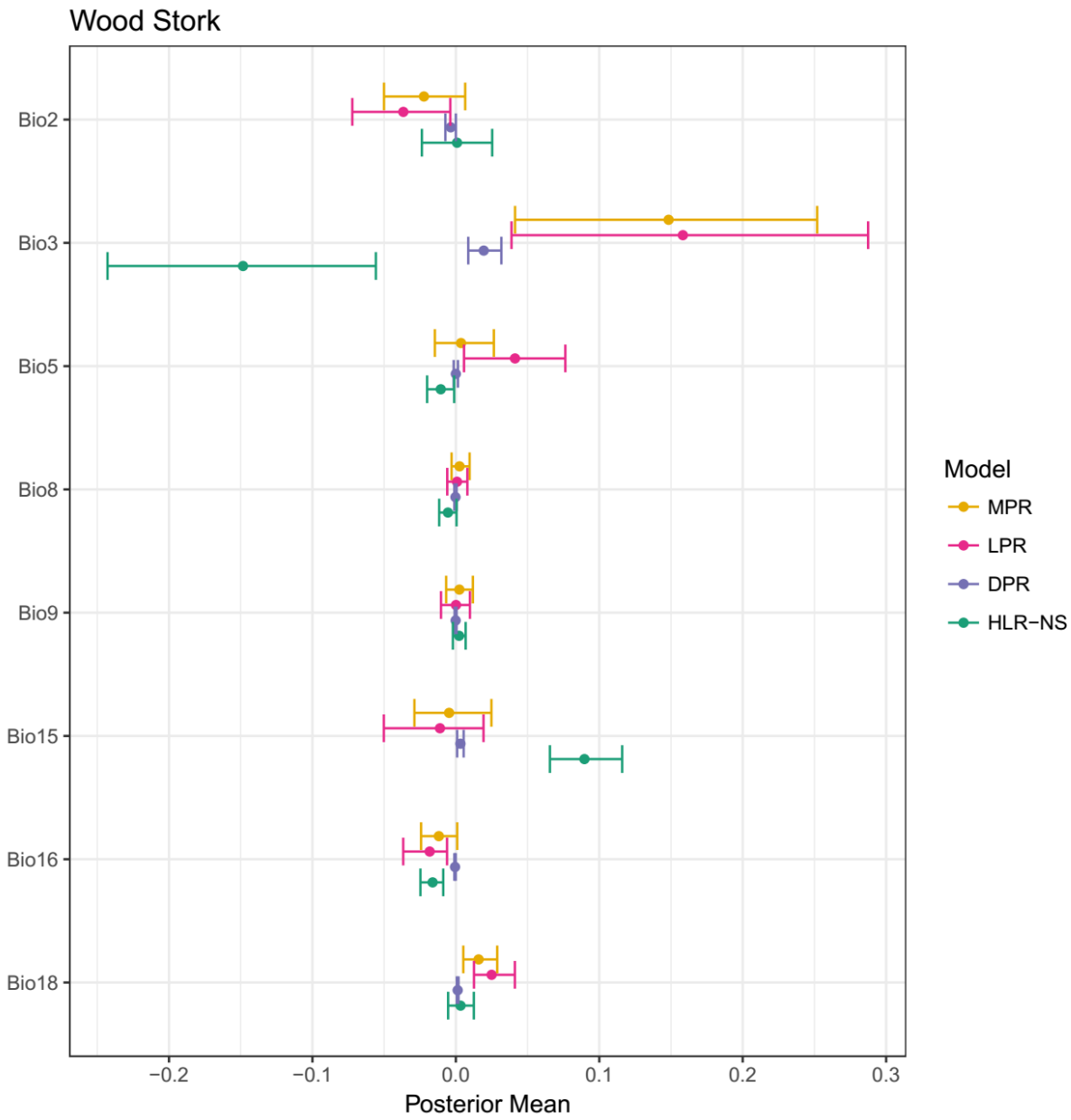


Figure 363: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the wood stork, *Mycteria americana*.

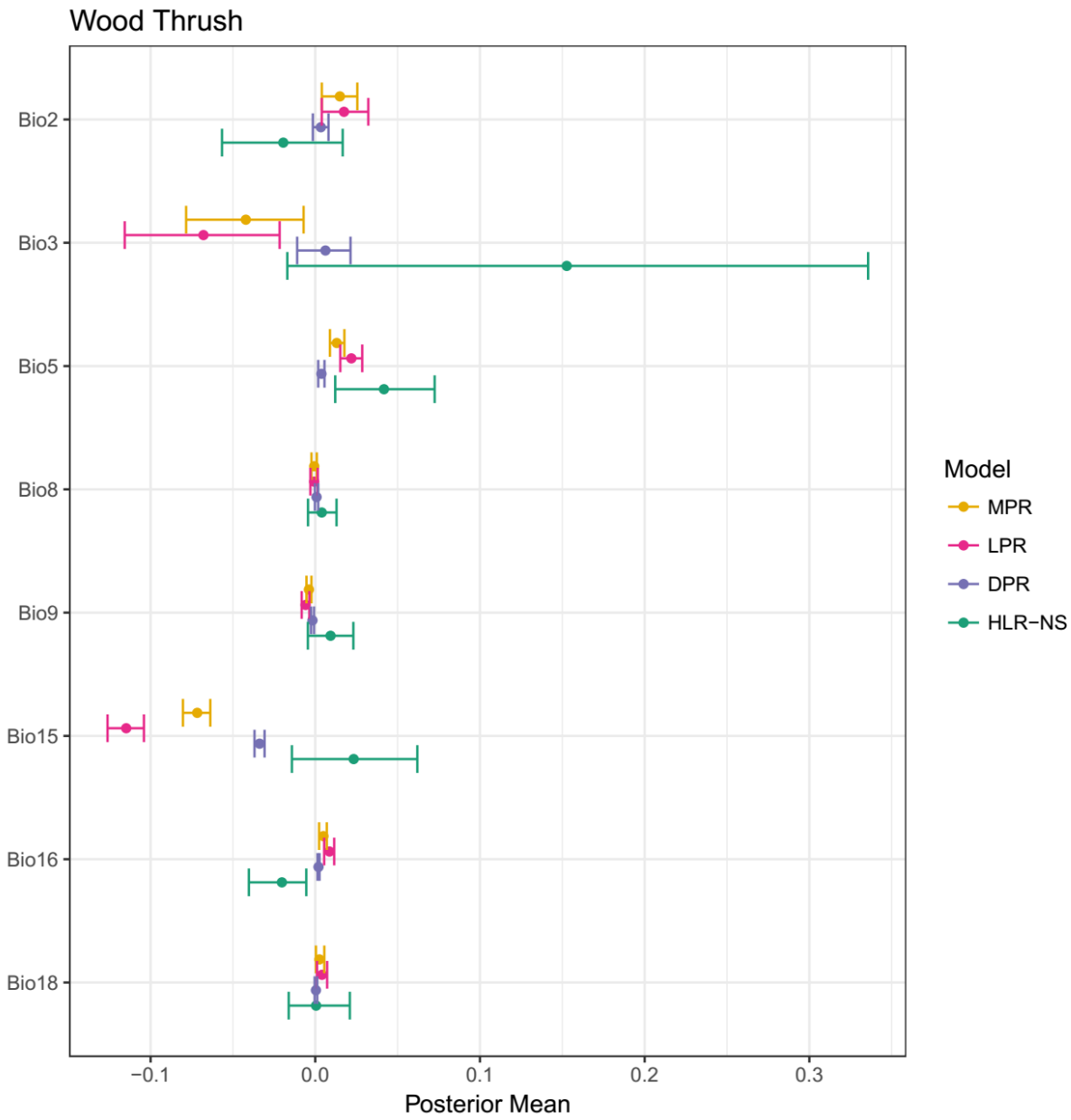


Figure 364: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the wood thrush, *Hylocichla mustelina*.

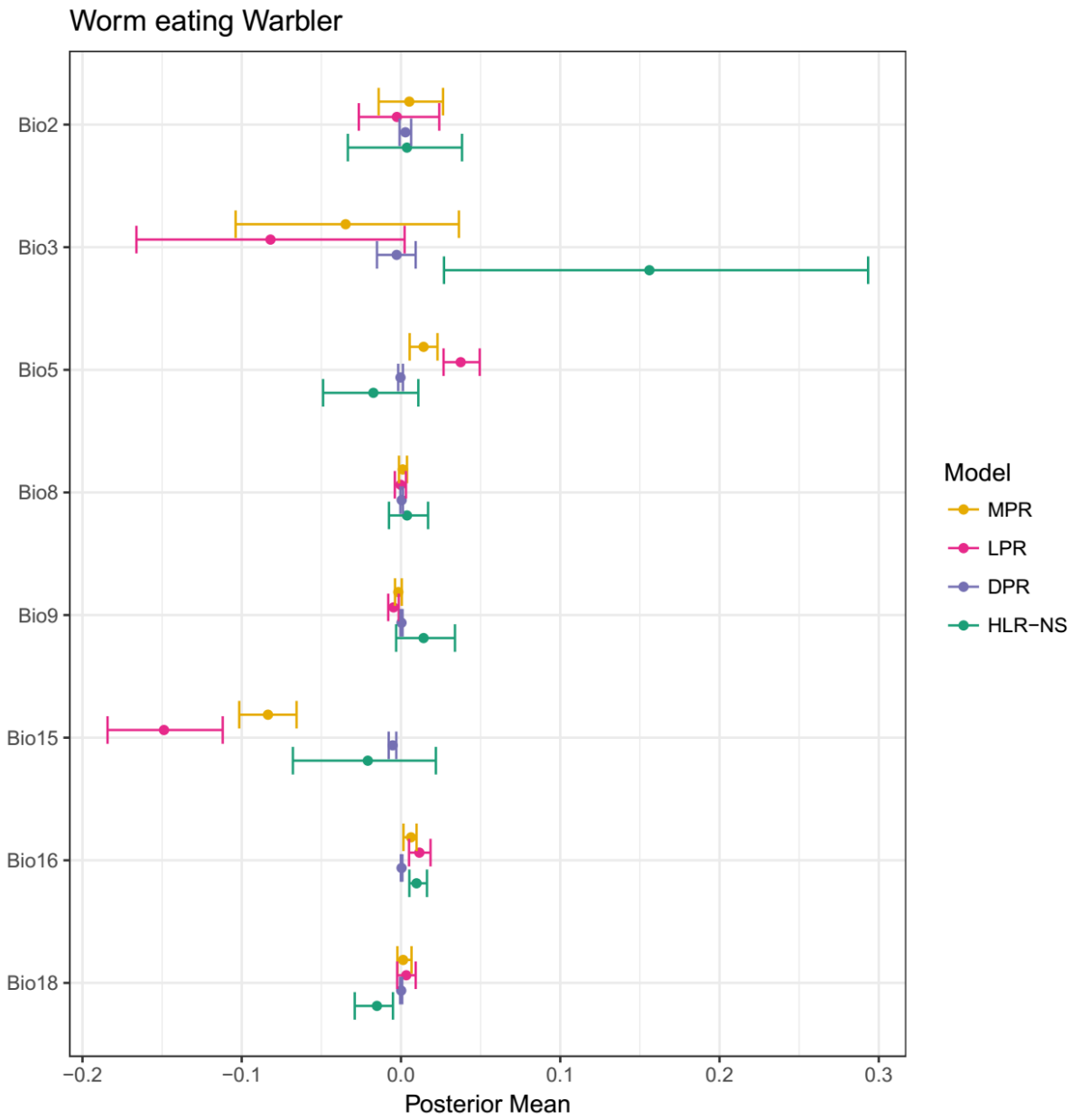


Figure 365: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the worm-eating warbler, *Helmitheros vermivorum*.

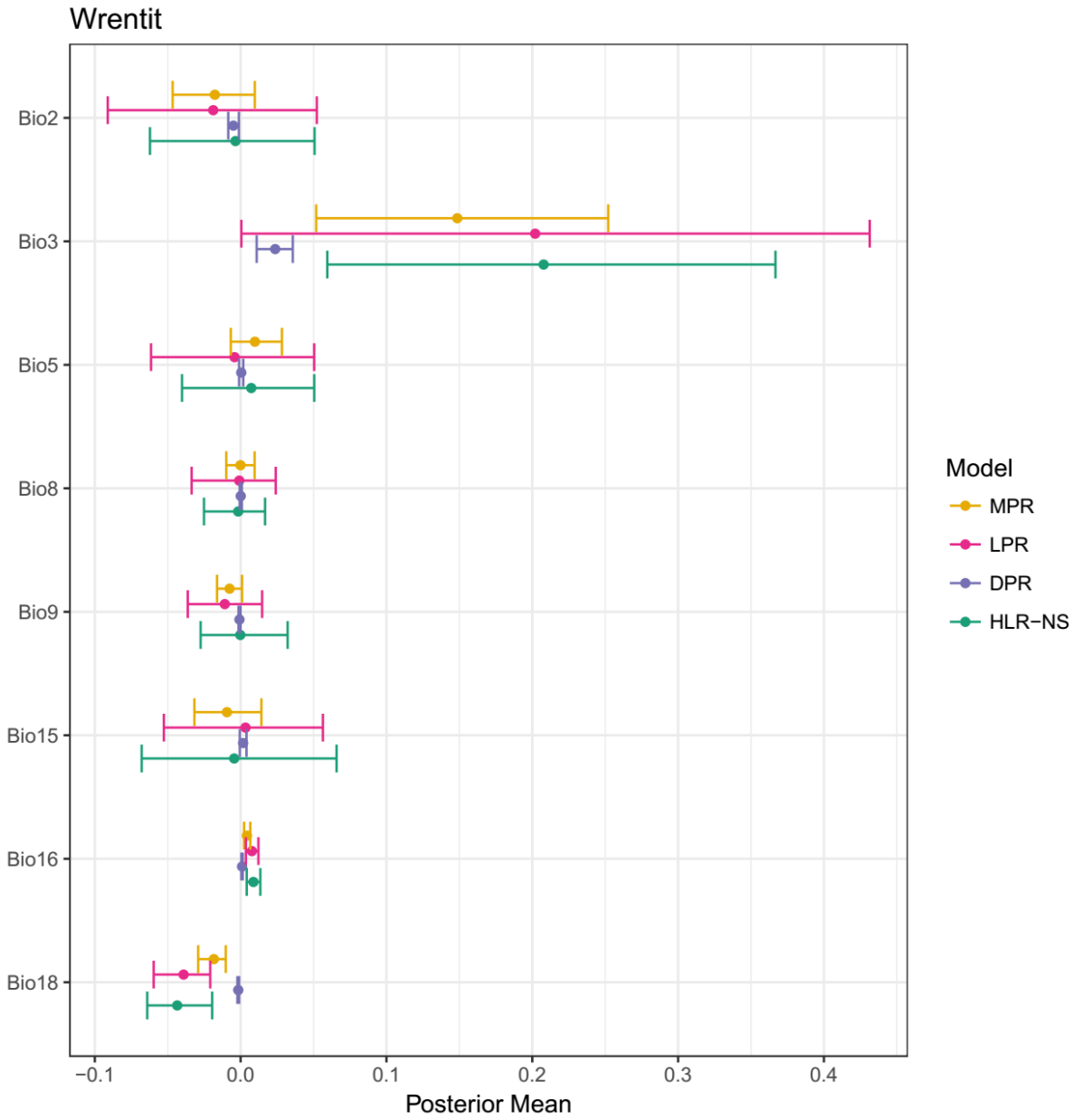


Figure 366: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the wrentit, *Chamaea fasciata*.

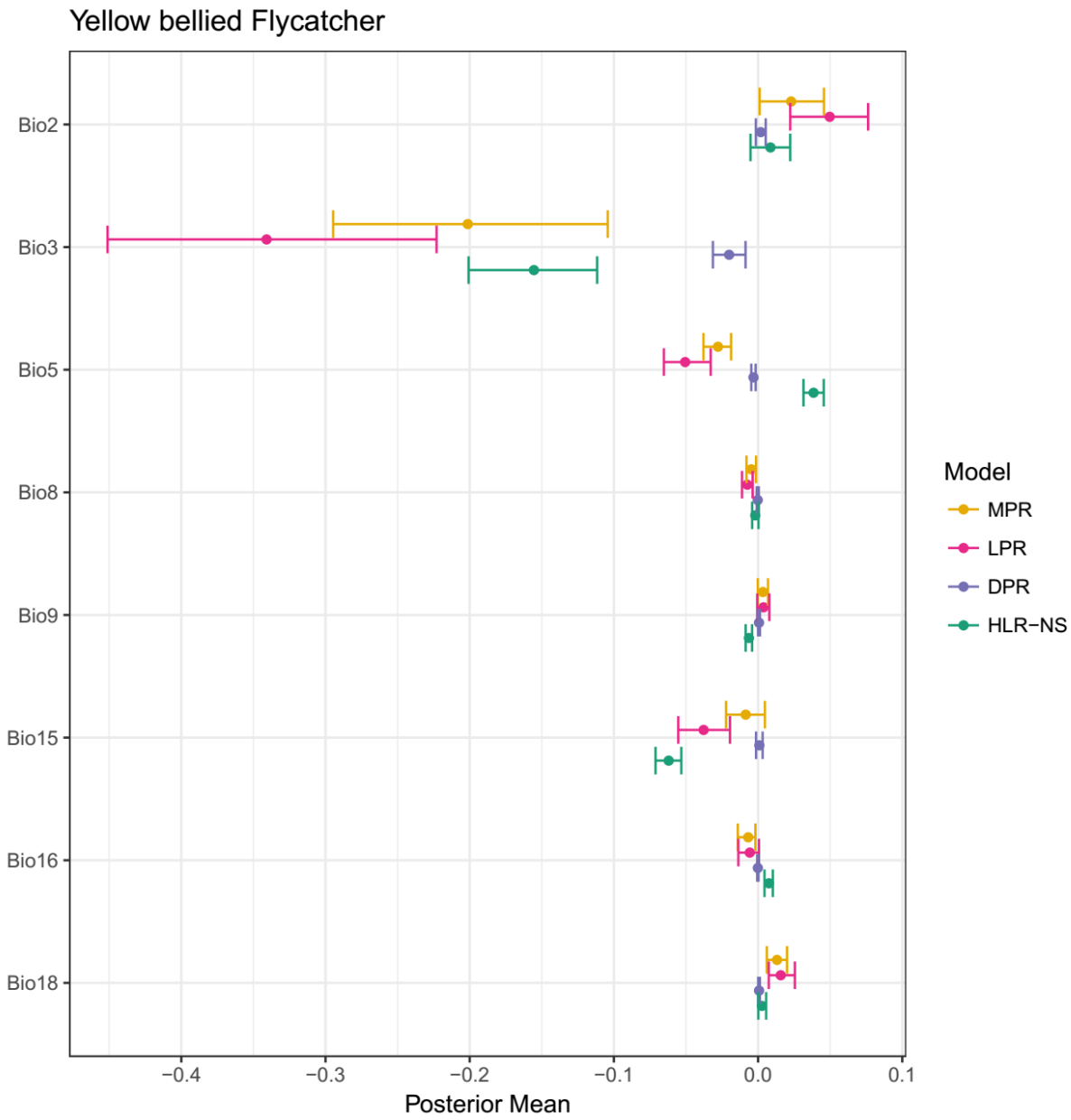


Figure 367: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-bellied flycatcher, *Empidonax flaviventris*.

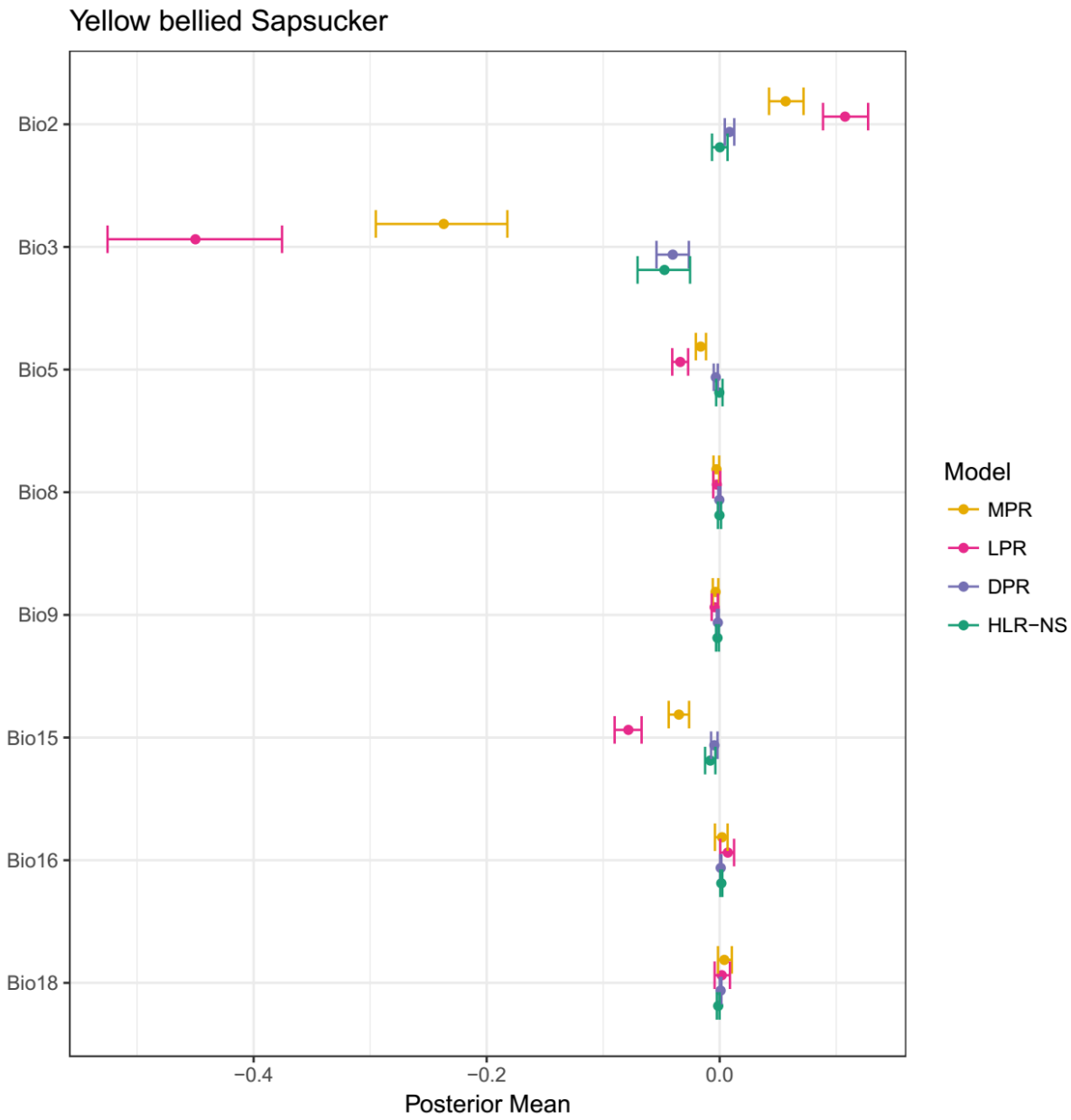


Figure 368: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-bellied sapsucker, *Sphyrapicus varius*.

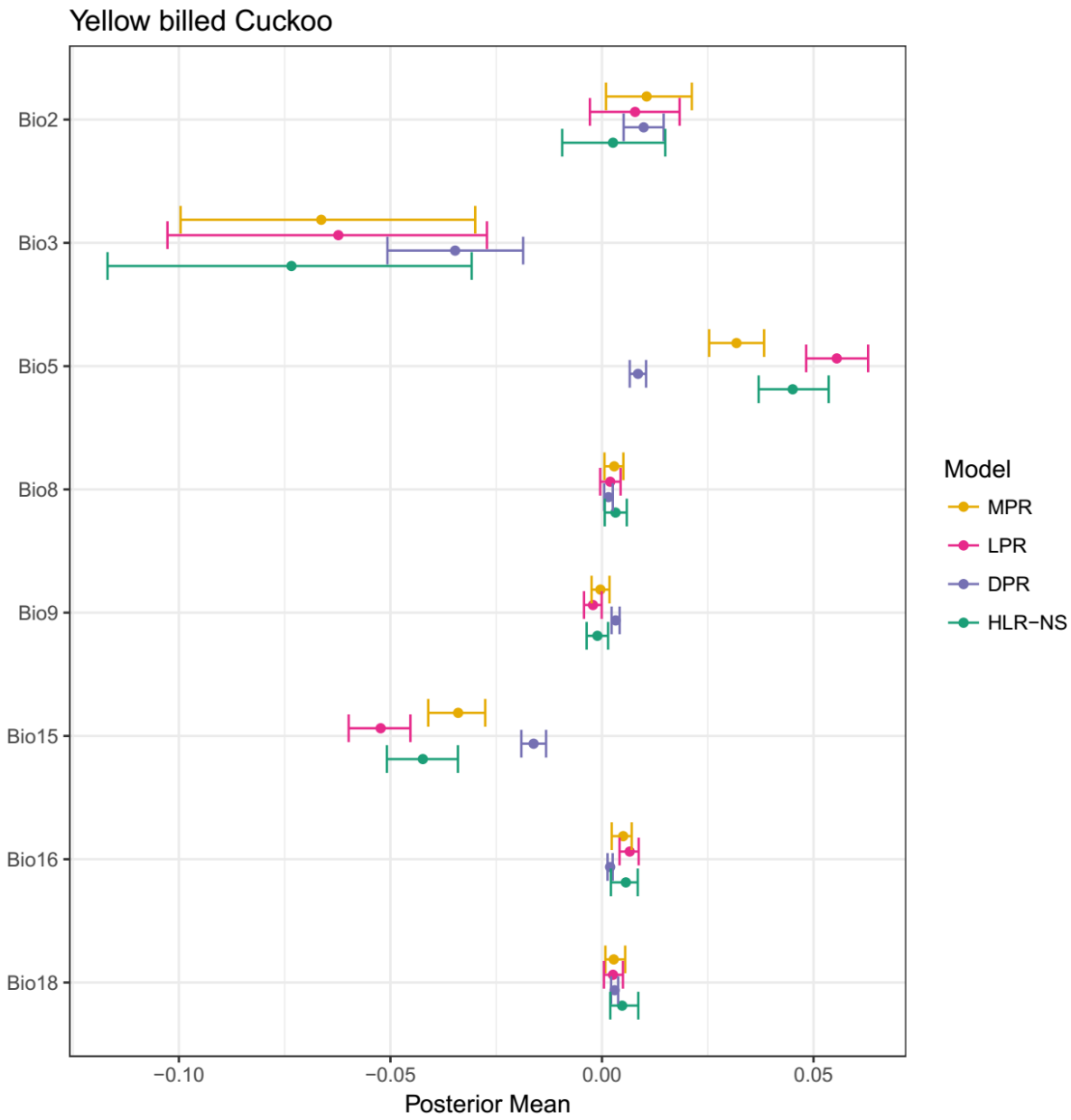


Figure 369: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-billed cuckoo, *Coccyzus americanus*.

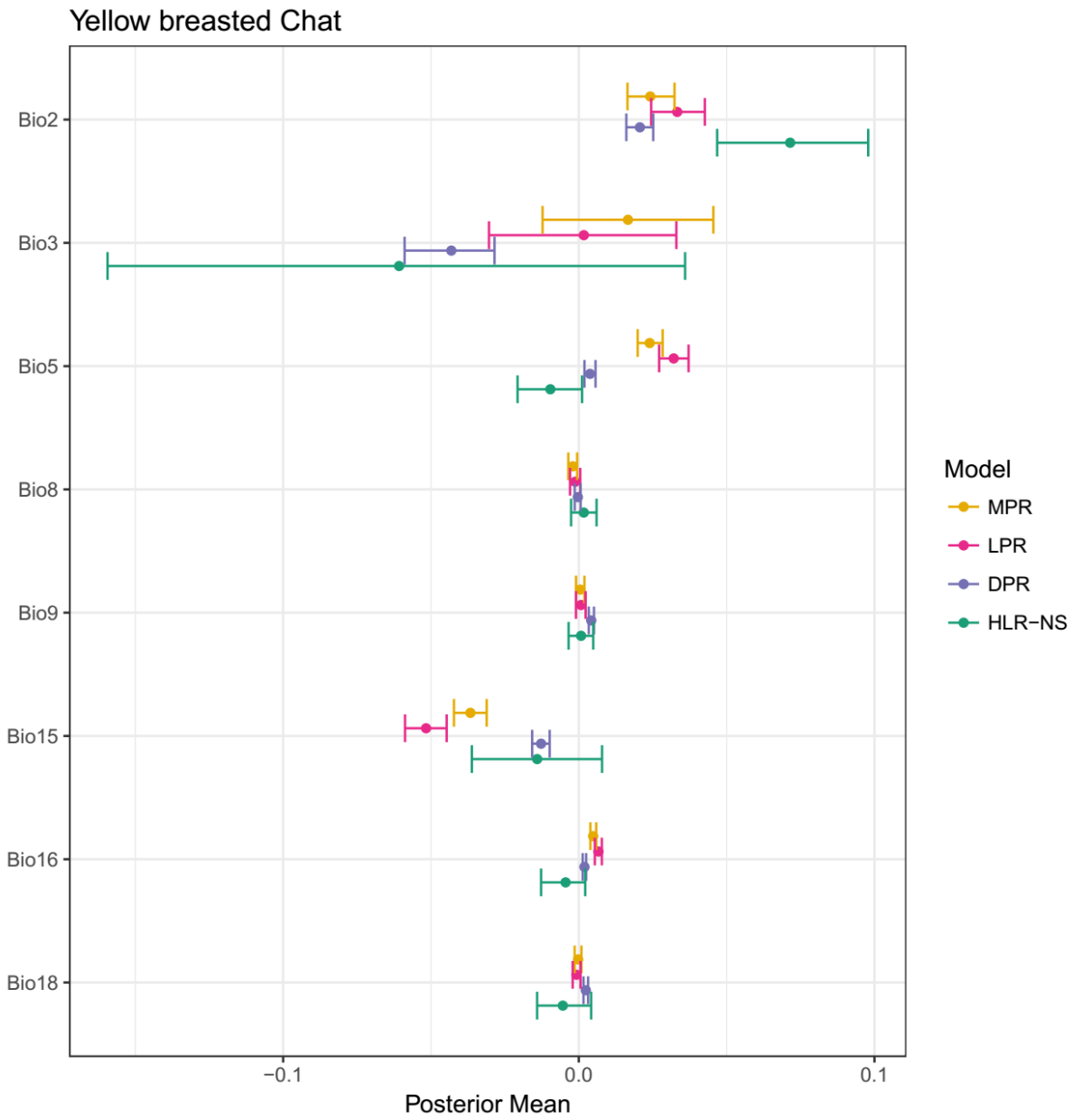


Figure 370: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-breasted chat, *Icteria virens*.

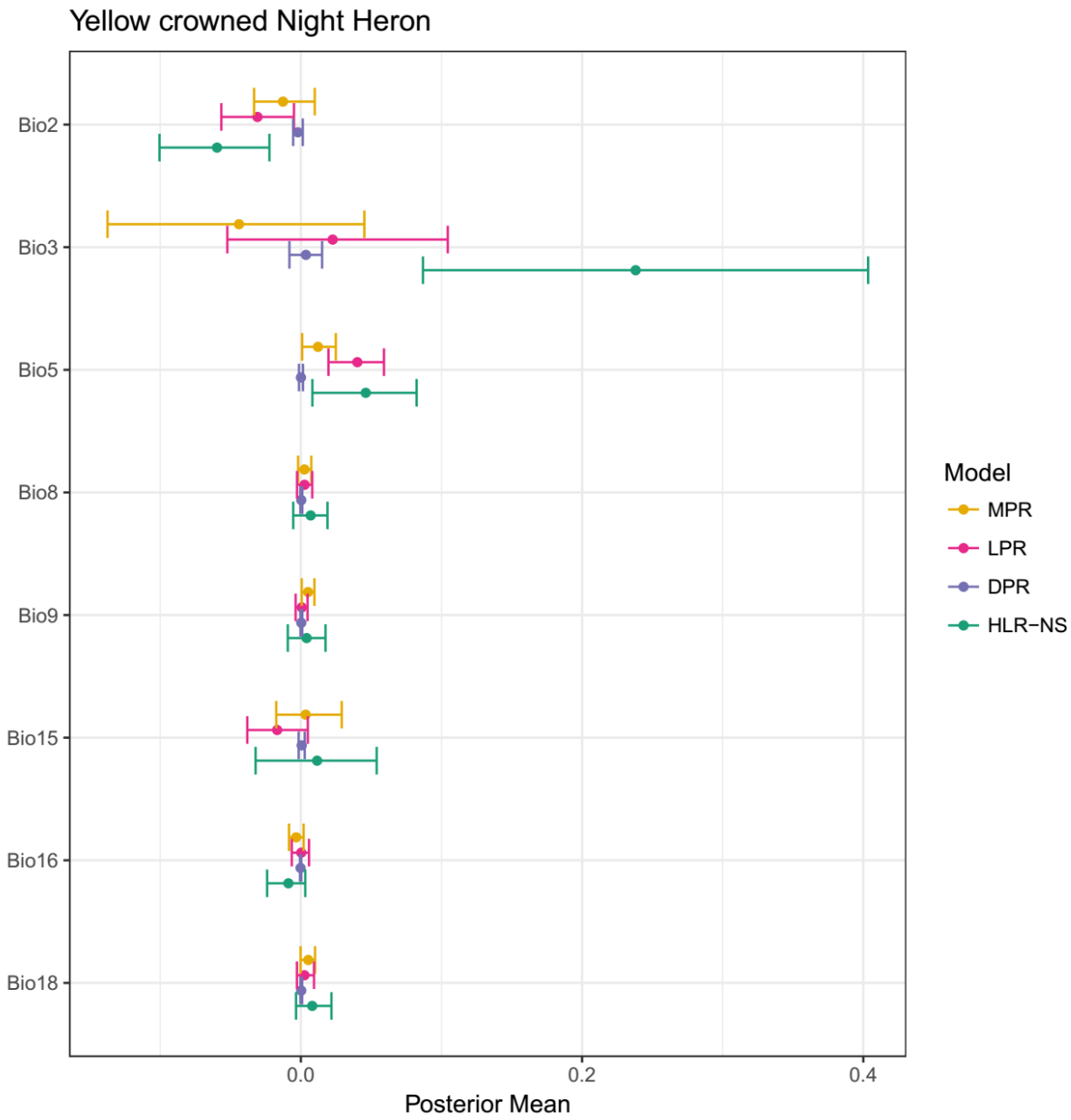


Figure 371: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-crowned night heron, *Nyctanassa violacea*.

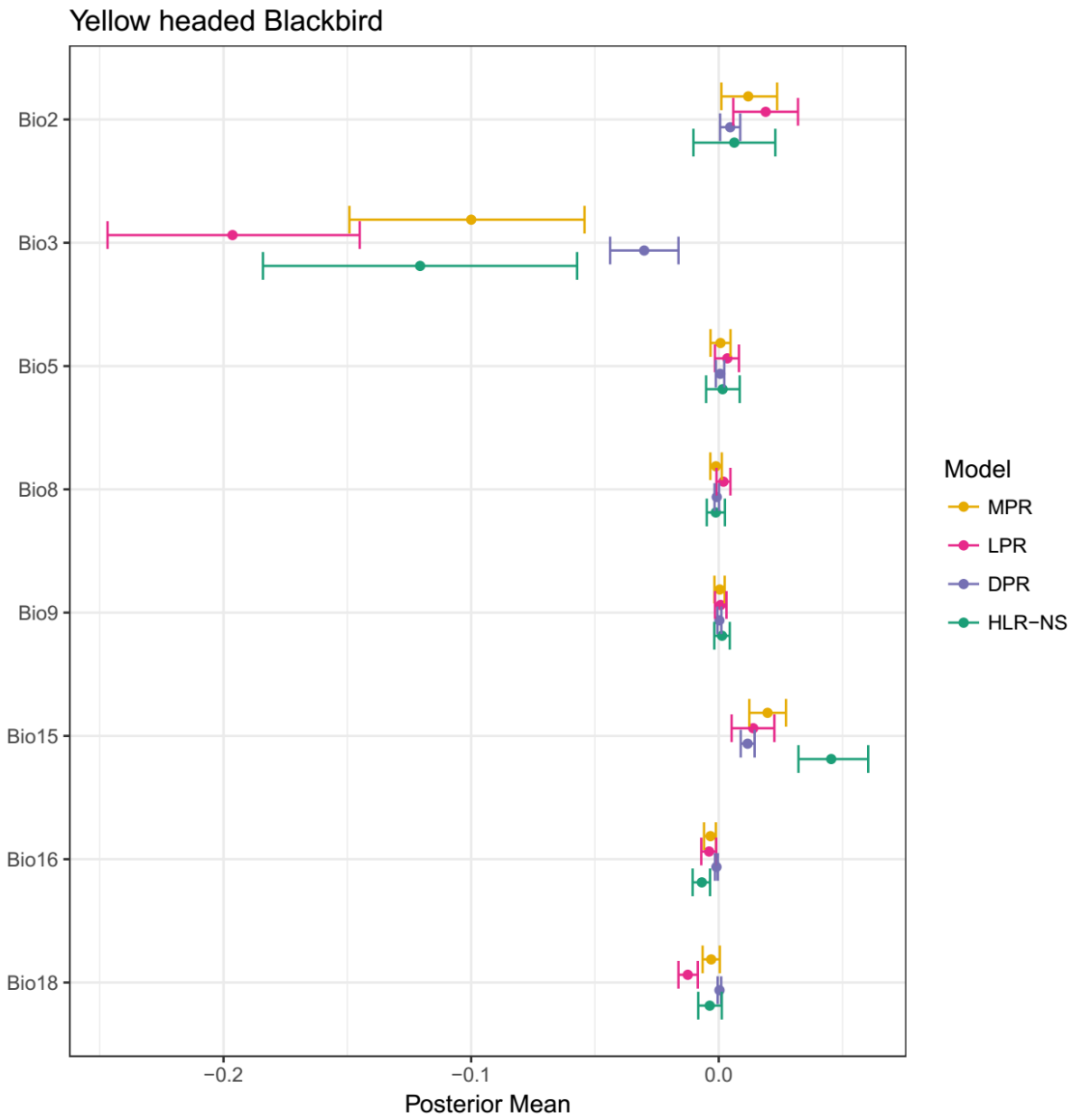


Figure 372: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-headed blackbird, *Xanthocephalus xanthocephalus*.

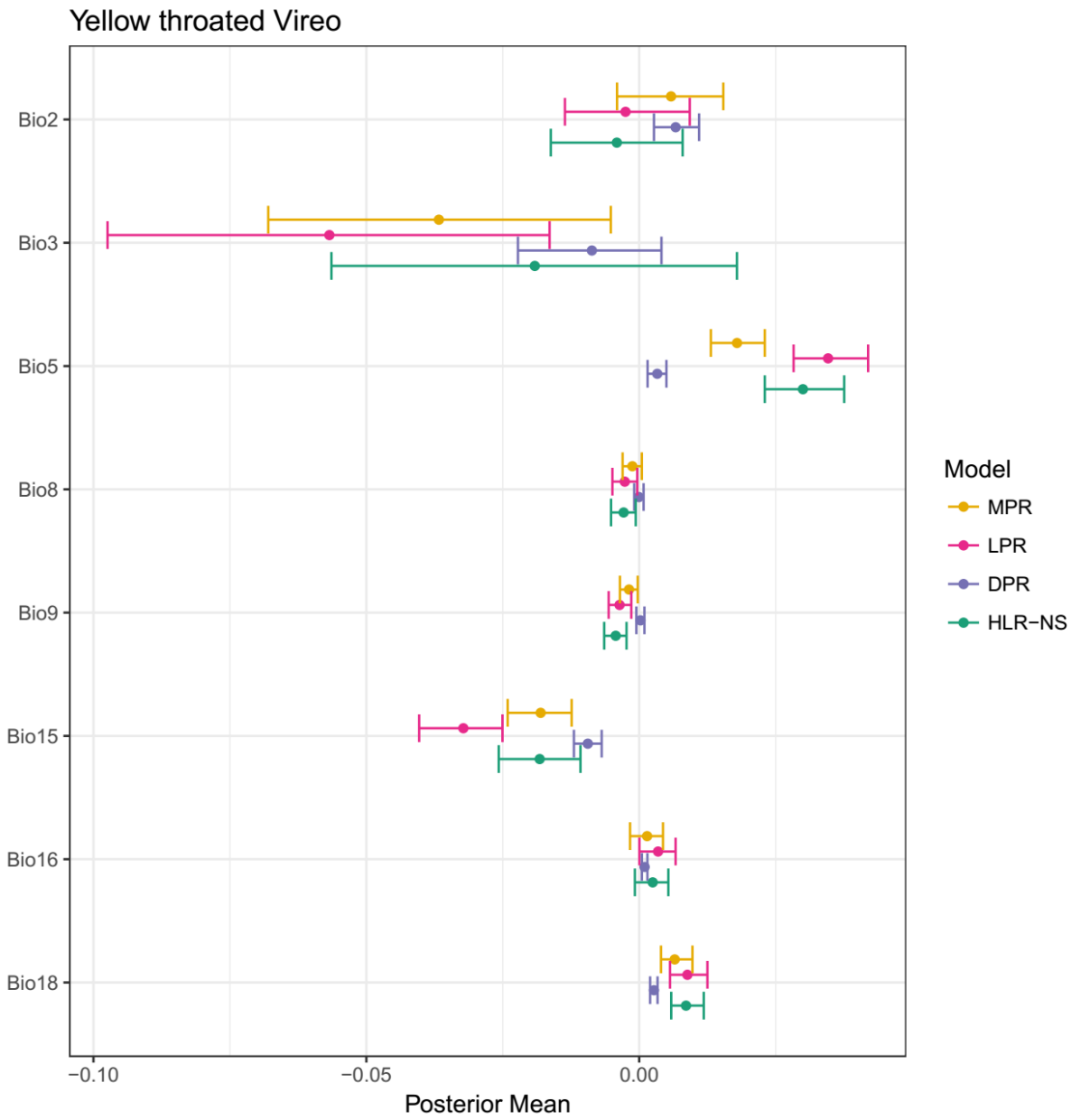


Figure 373: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-throated vireo, *Vireo flavifrons*.

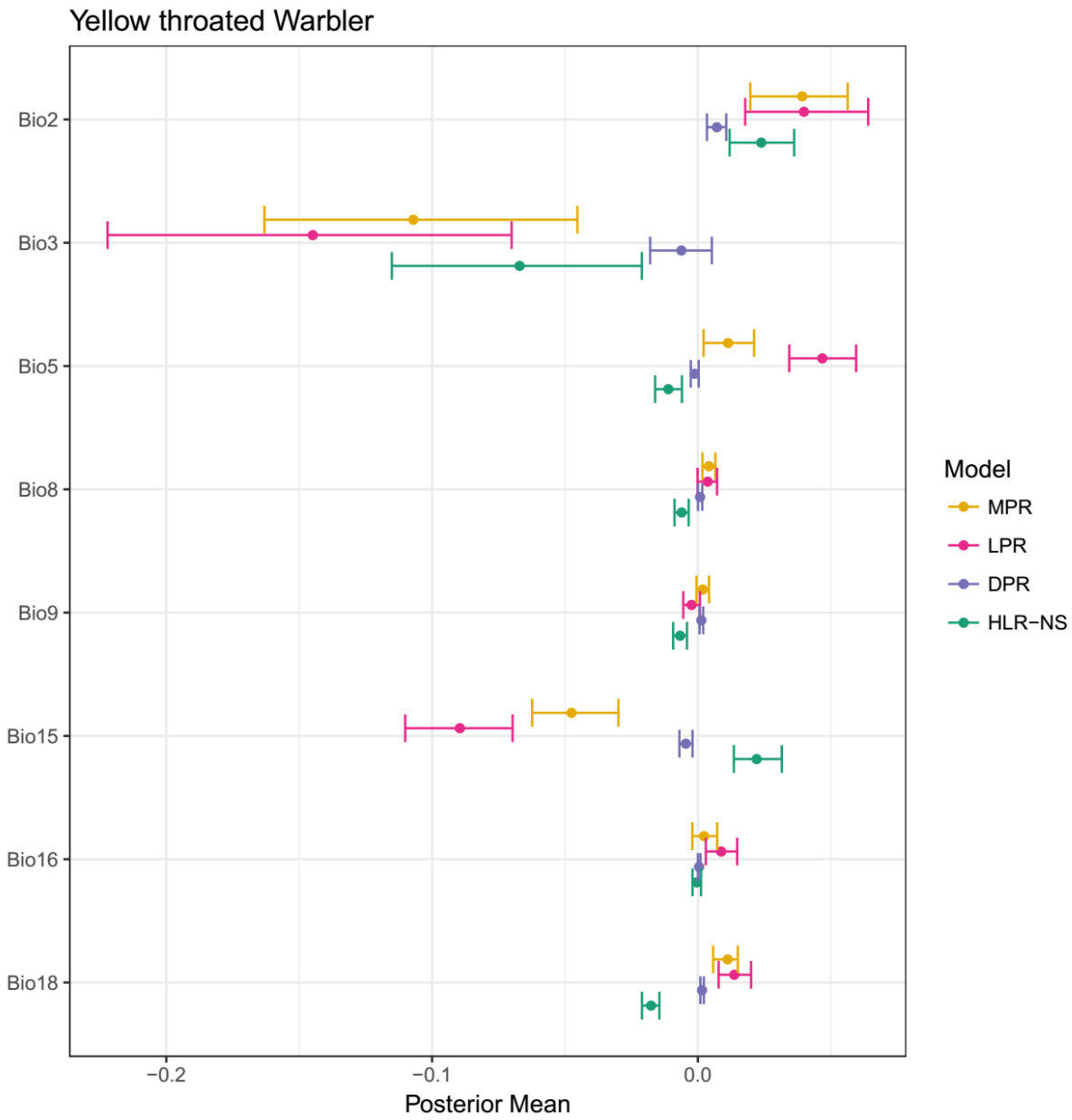


Figure 374: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow-throated warbler, *Setophaga dominica*.

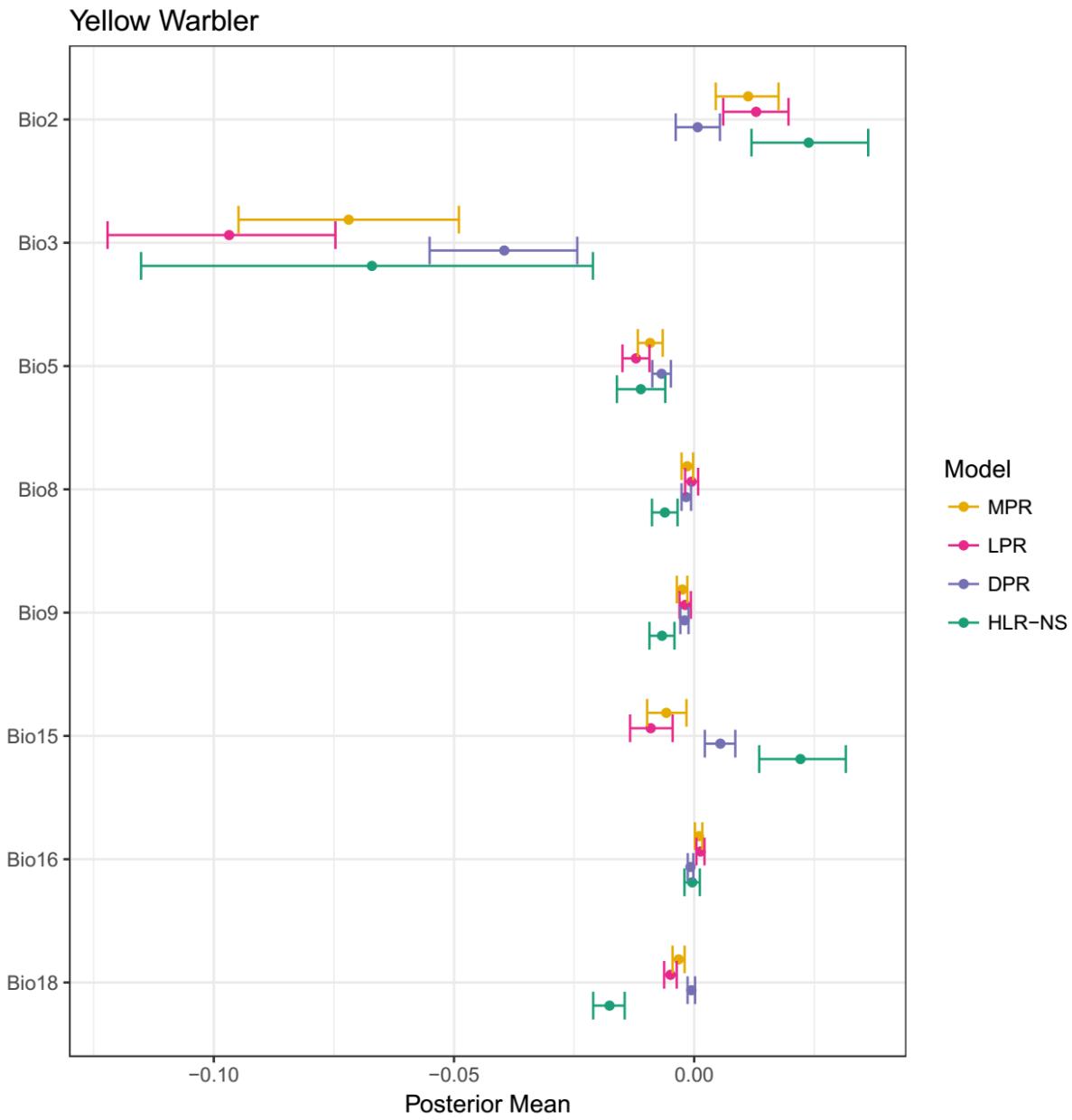


Figure 375: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow warbler, *Setophaga petechia*.

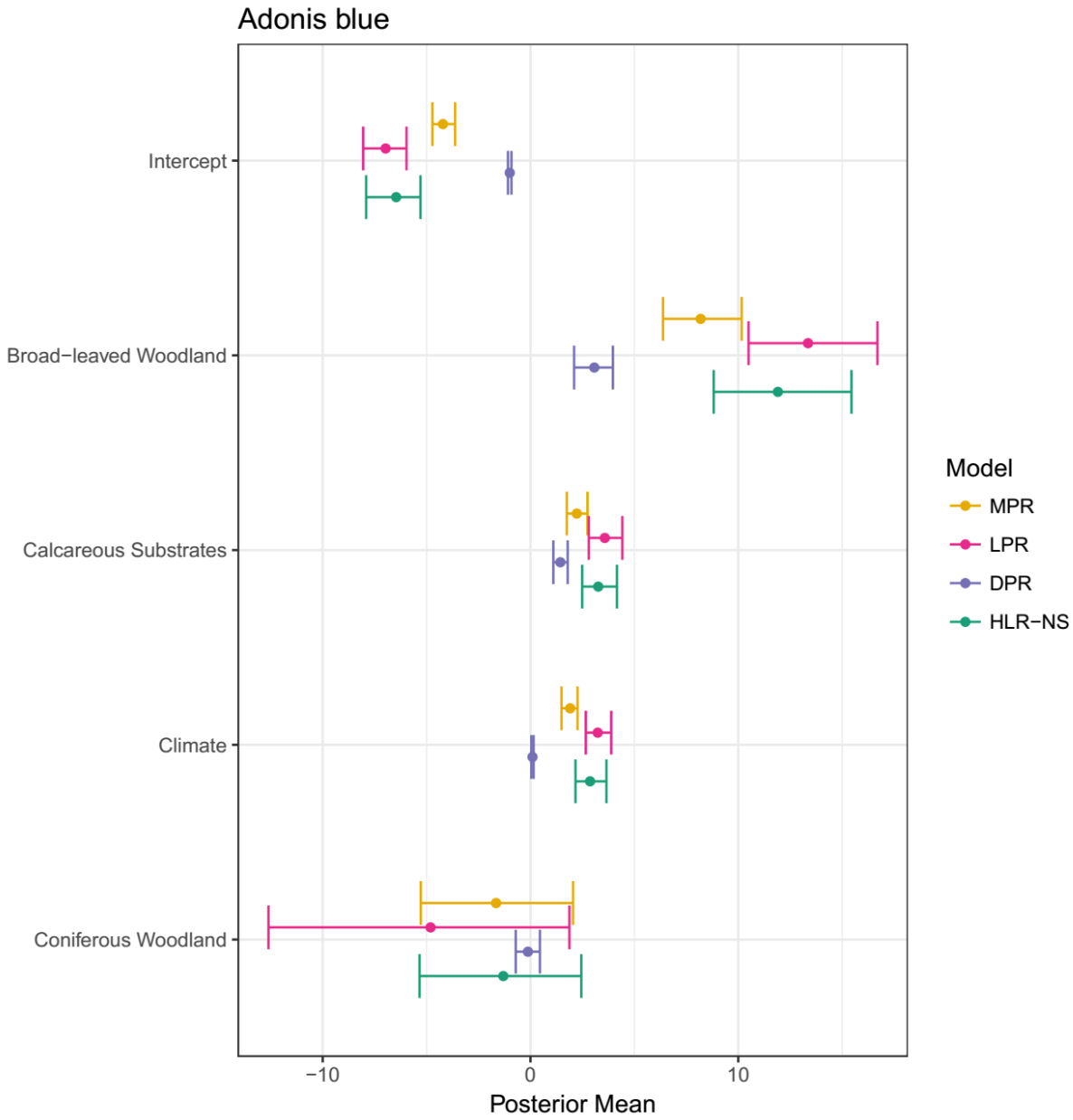


Figure 376: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Adonis blue, *Polyommatus bellargus*.

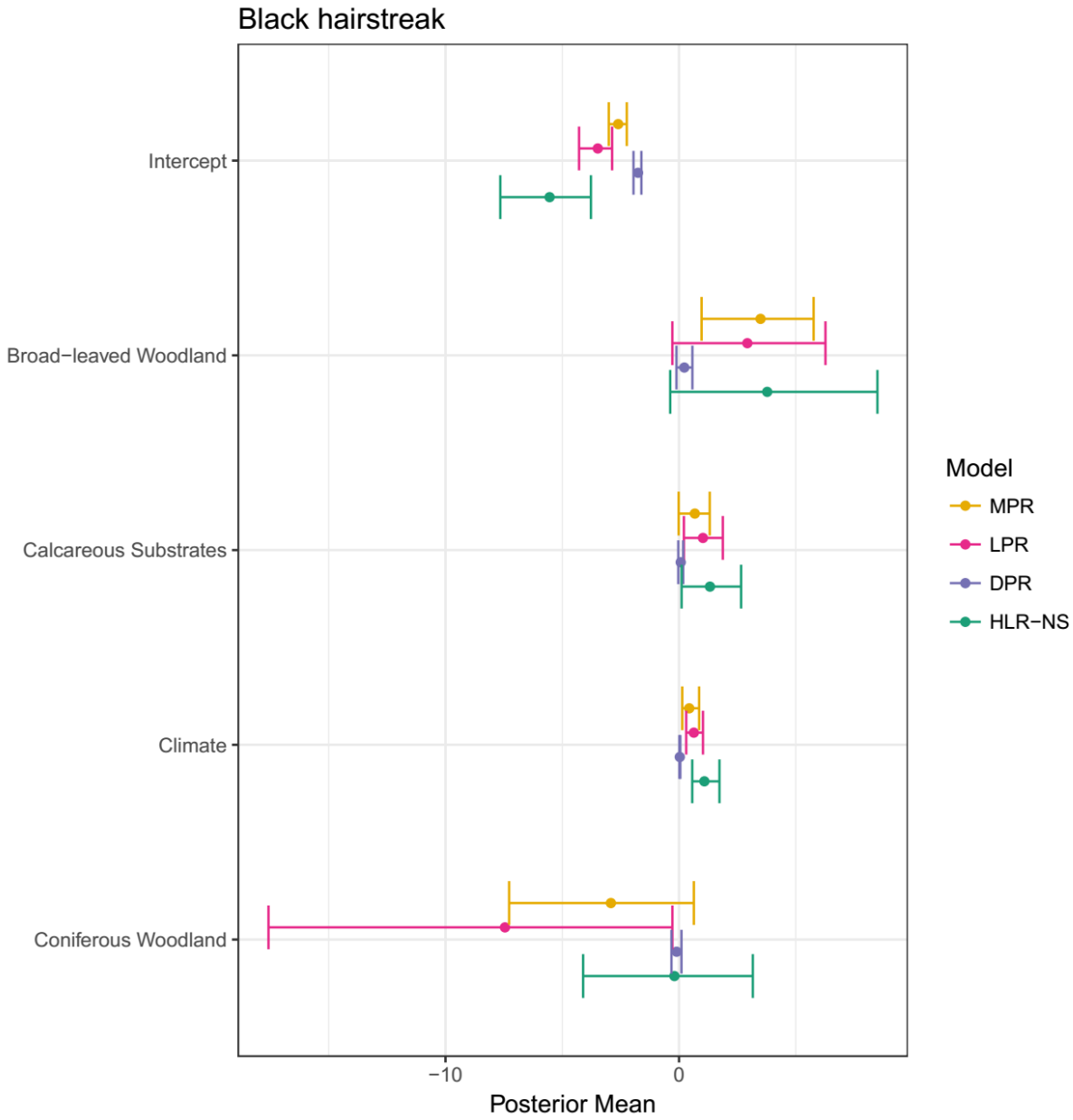


Figure 377: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the black hairstreak, *Satyrium pruni*.

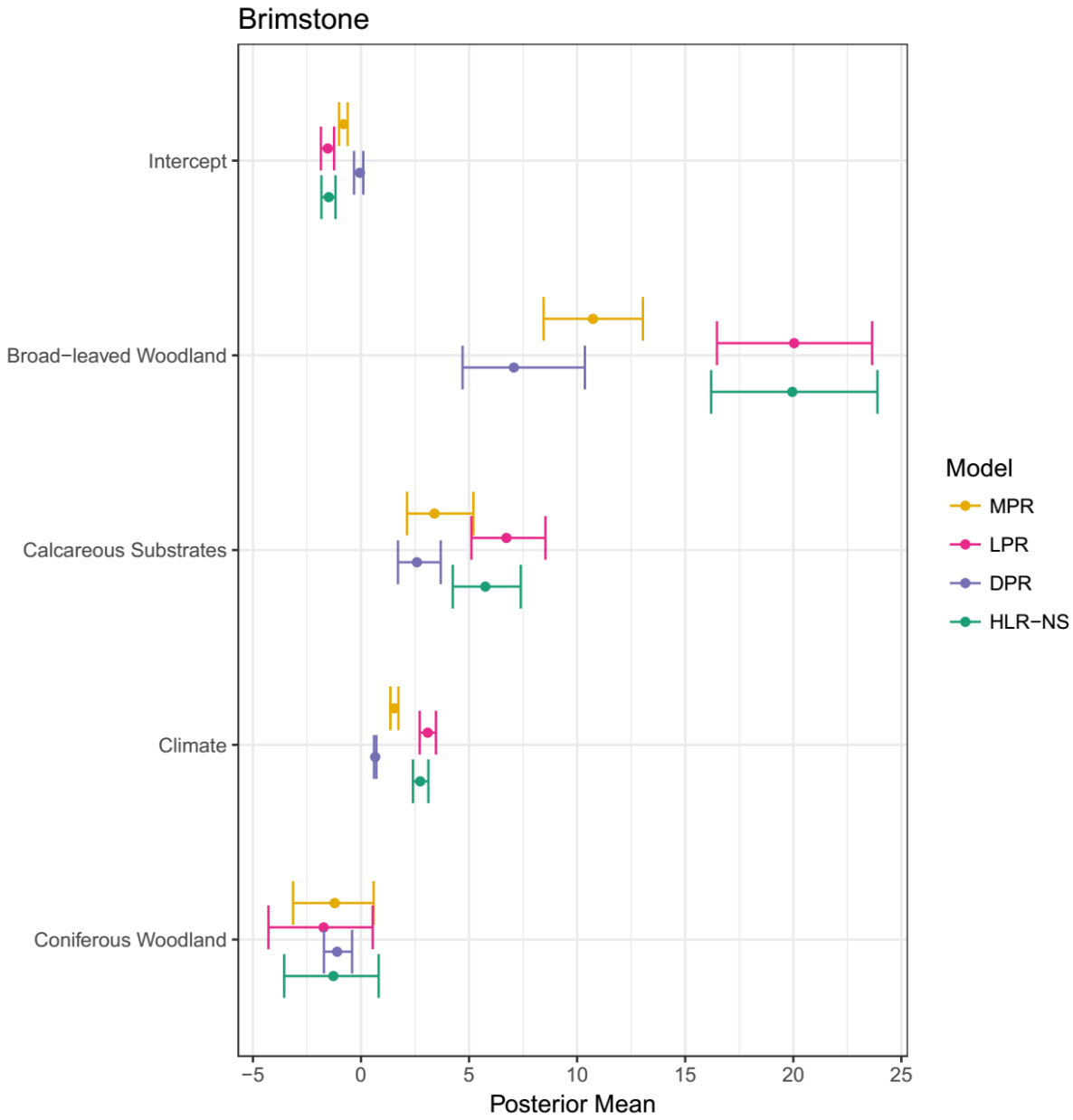


Figure 378: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brimstone, *Gonepteryx rhamni*.

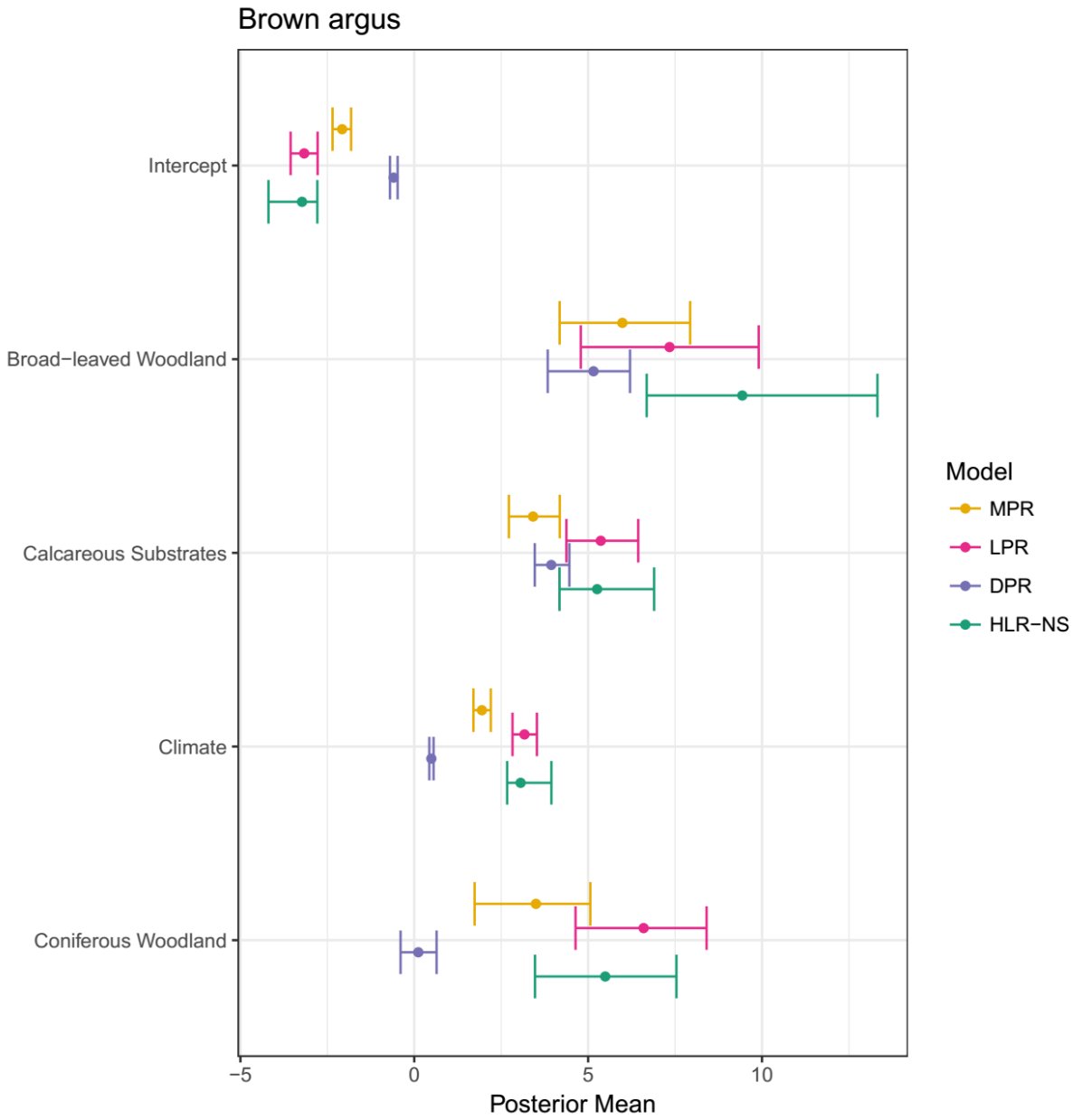


Figure 379: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown argus, *Aricia agestis*.

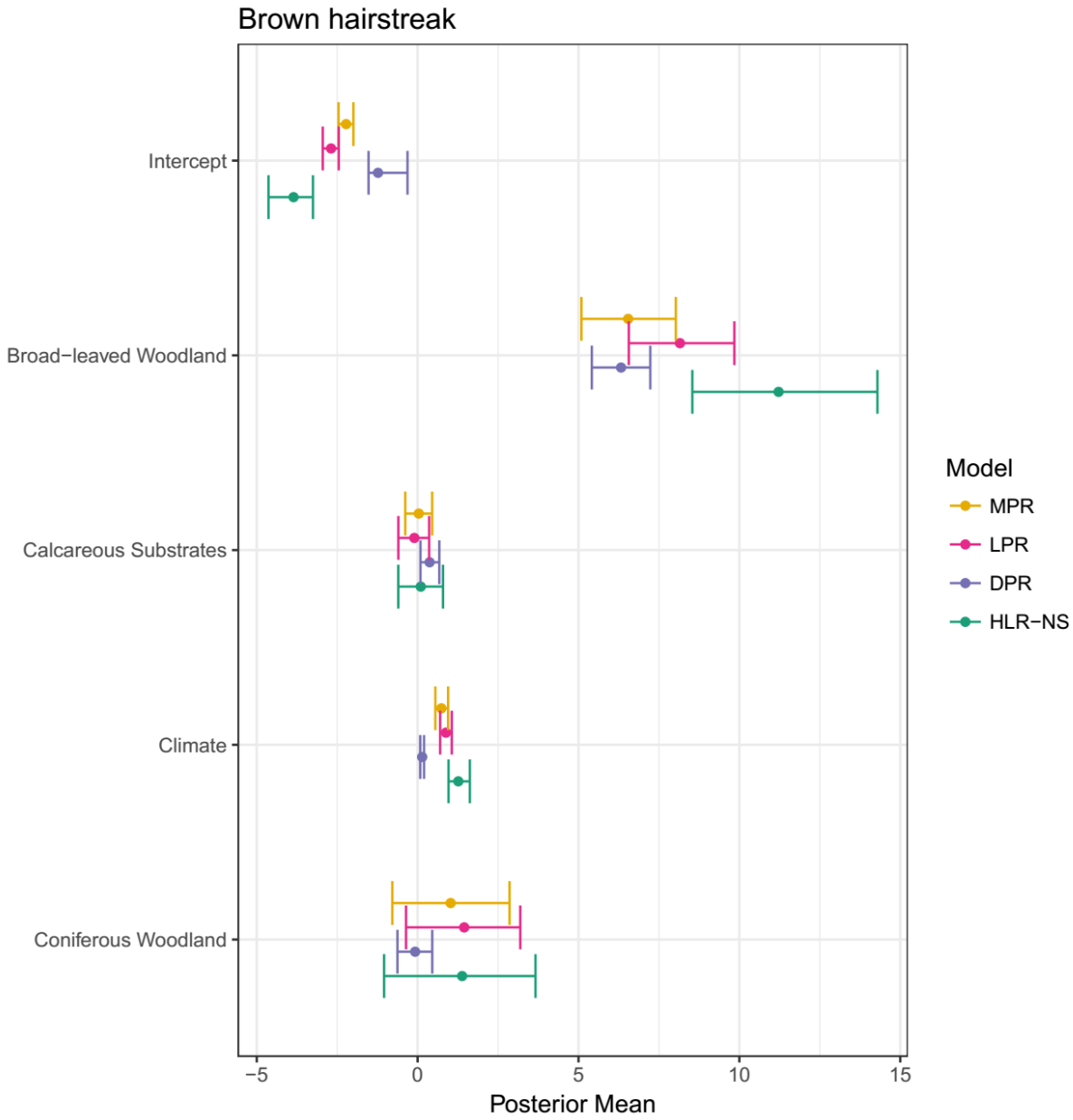


Figure 380: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown-hairstreak, *Thecla betulae*.

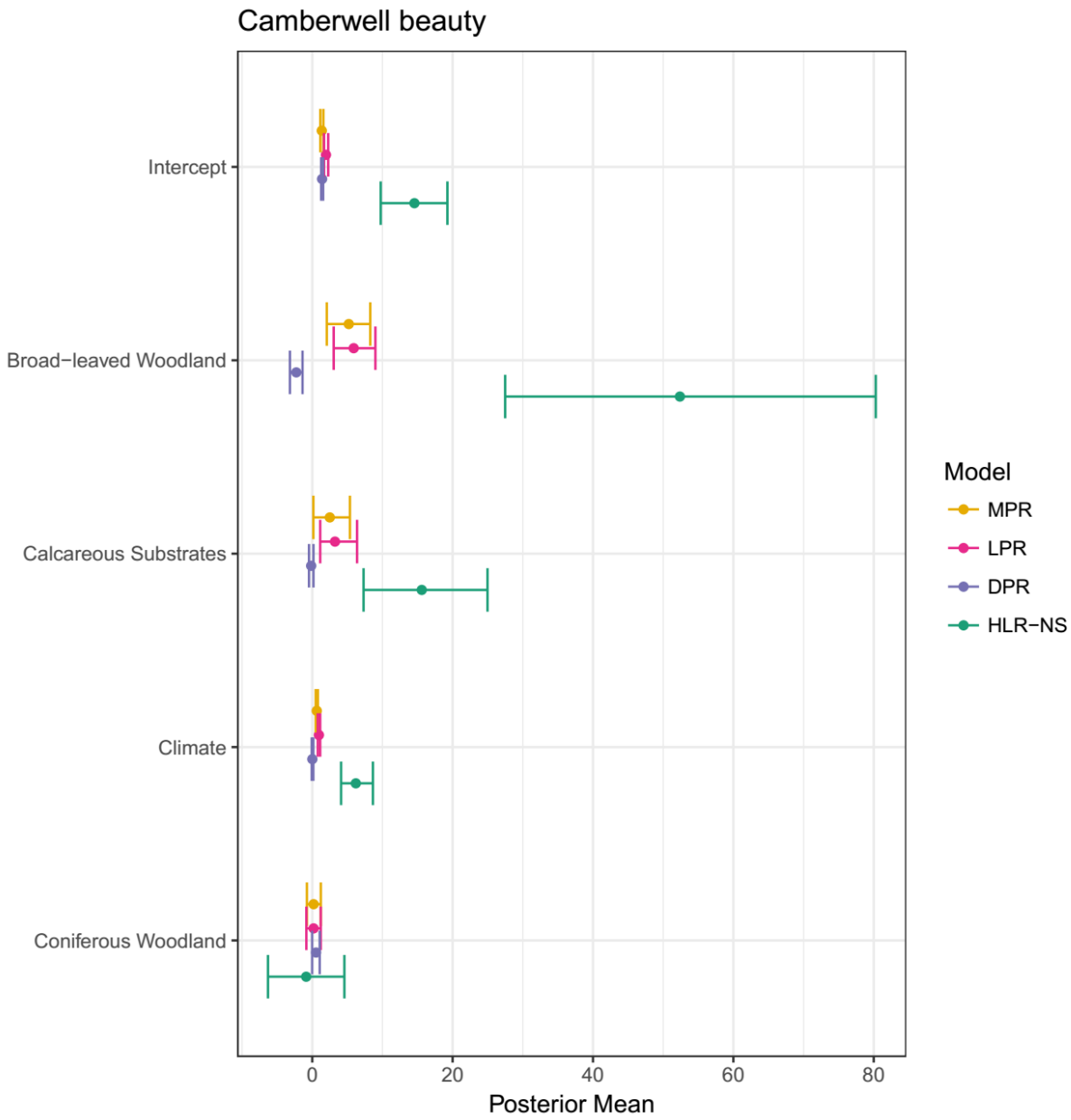


Figure 381: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Camberwell beauty, *Nymphalis antiopa*.

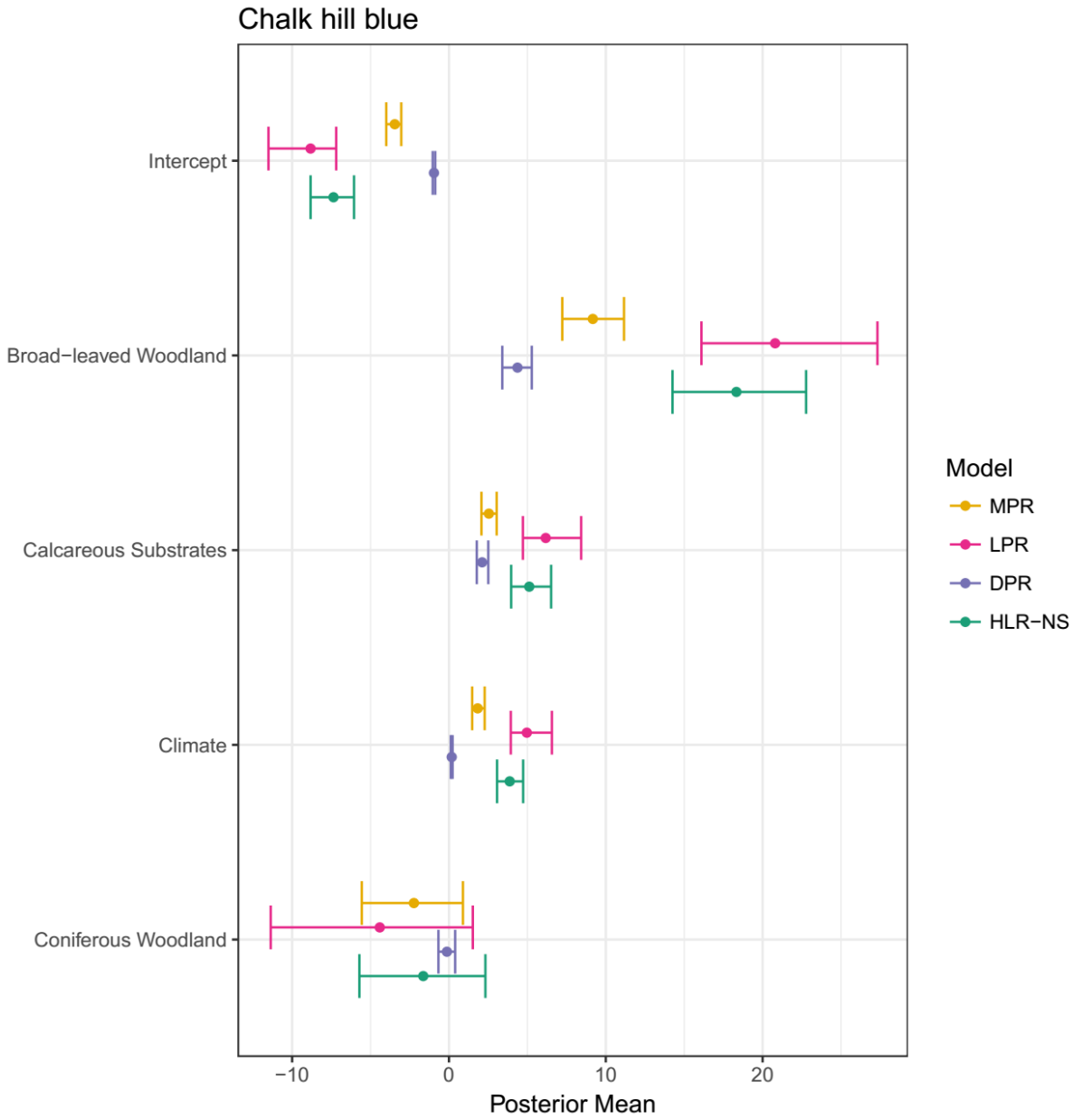


Figure 382: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the chalkhill bue, *Polyommatus coridon*.

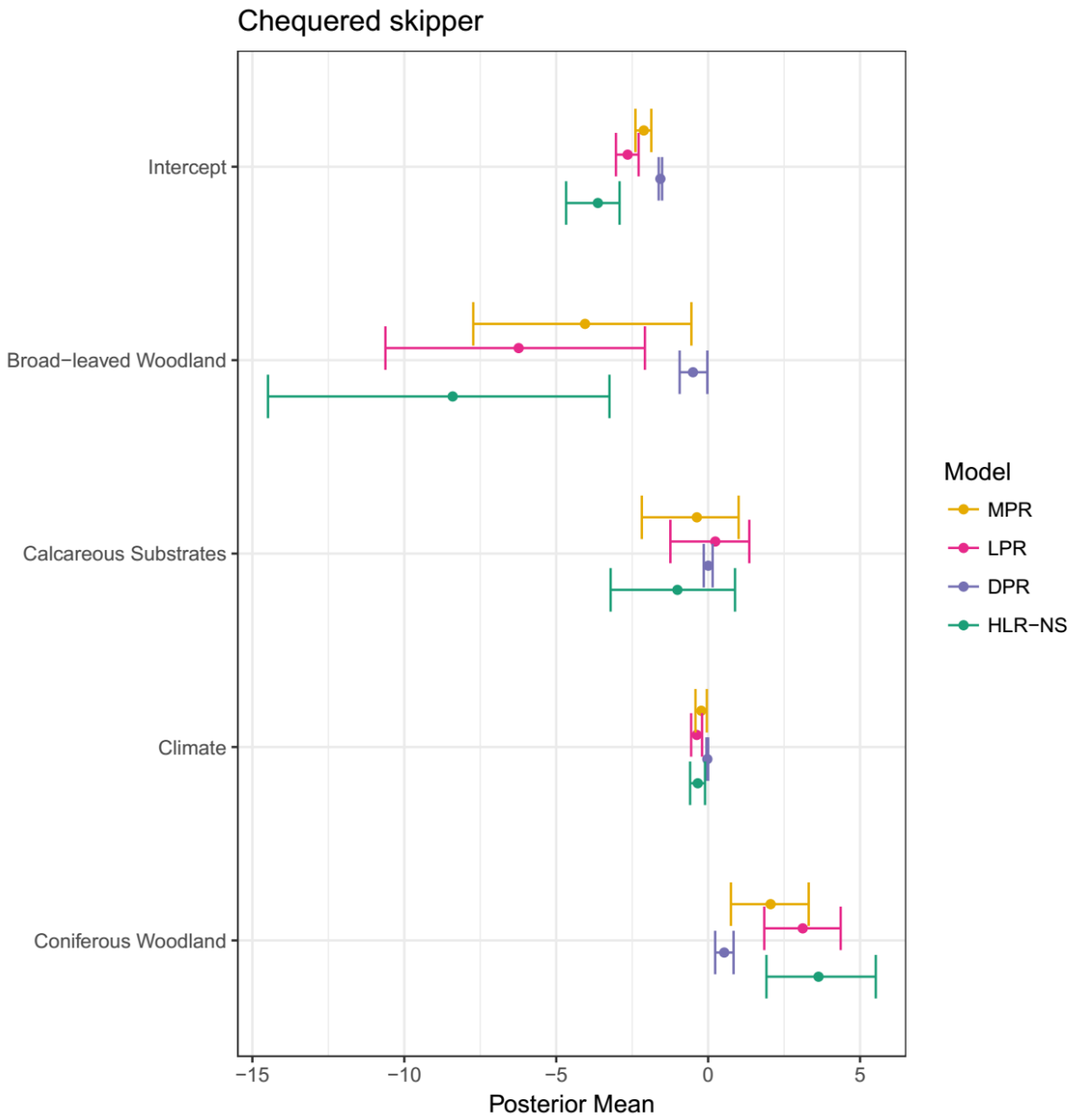


Figure 383: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Chequered skipper, *Carterocephalus palaemon*.

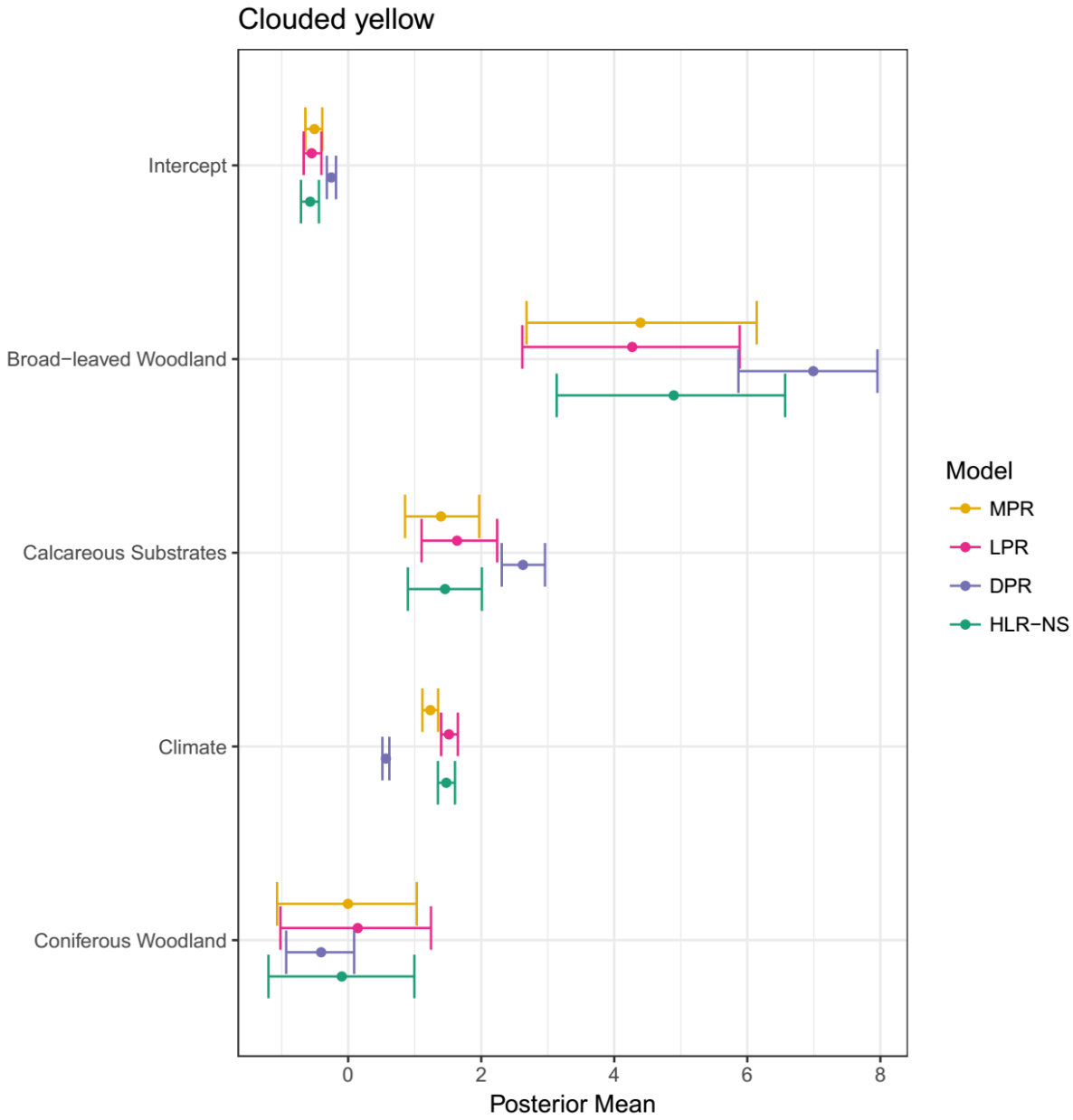


Figure 384: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the clouded yellow, *Colias croceus*.

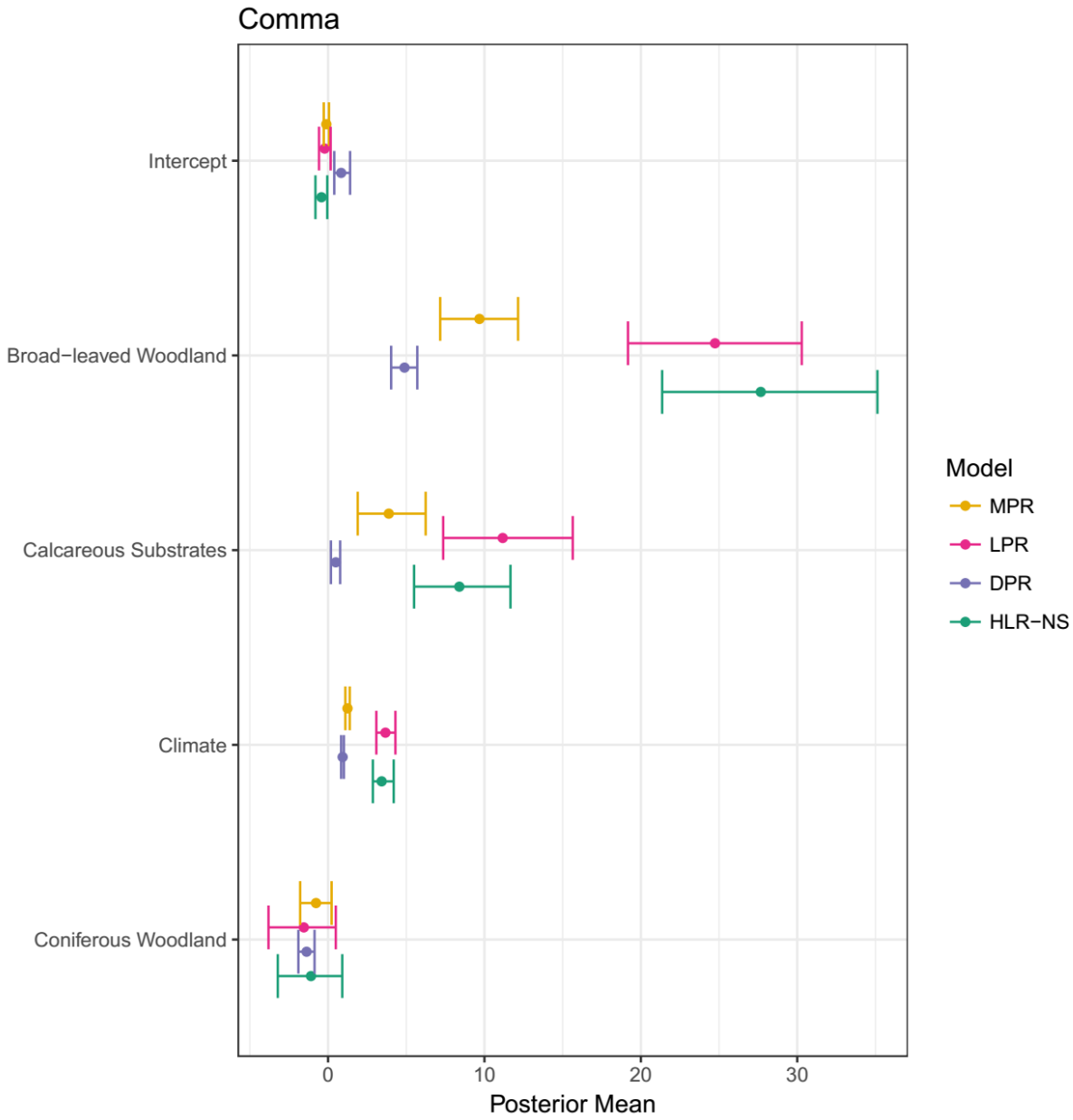


Figure 385: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the comma, *Polygonia c-album*.

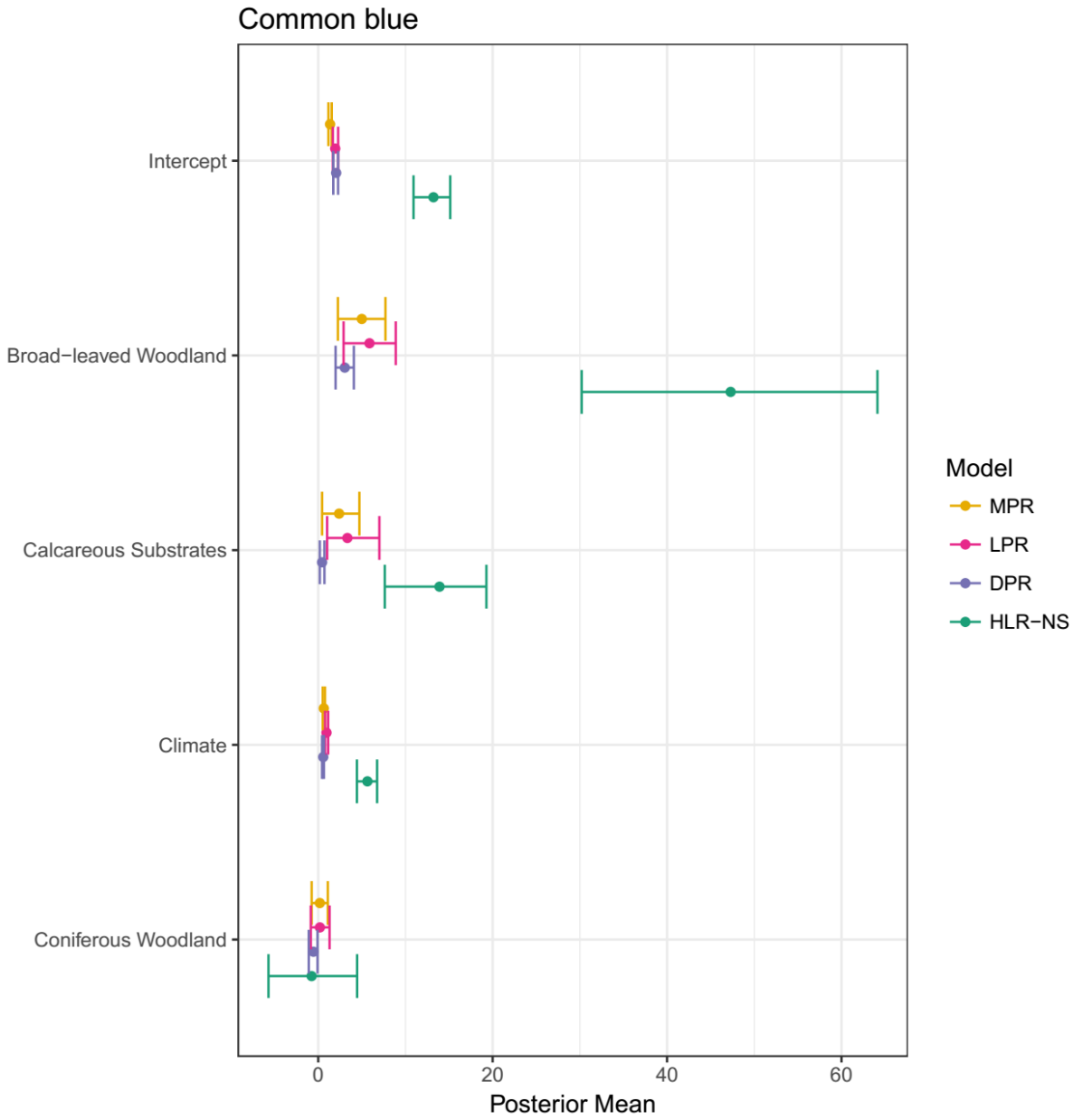


Figure 386: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common blue, *Polyommatus icarus*.

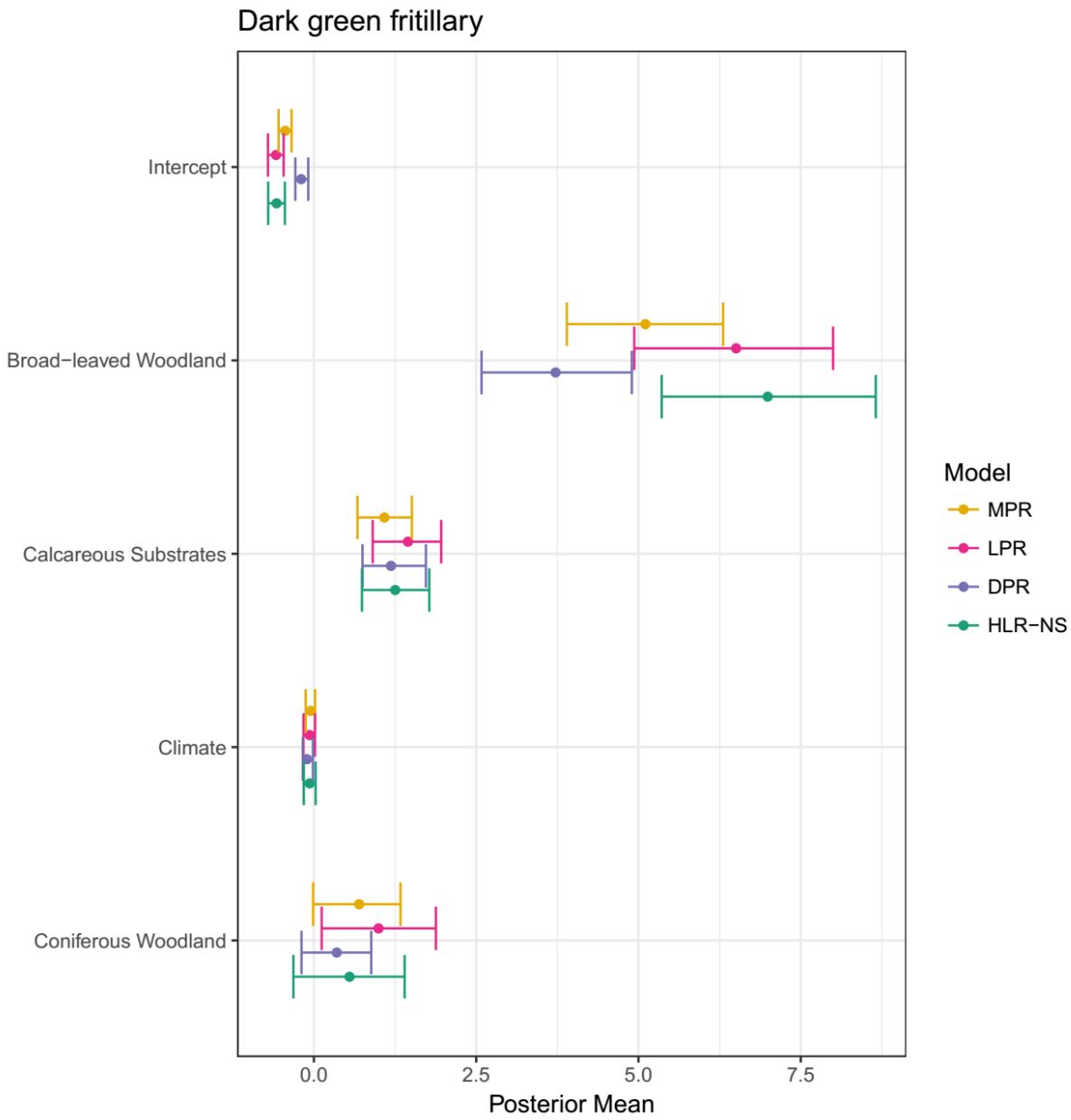


Figure 387: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the dark green fritillary, *Argynnis aglaja*.

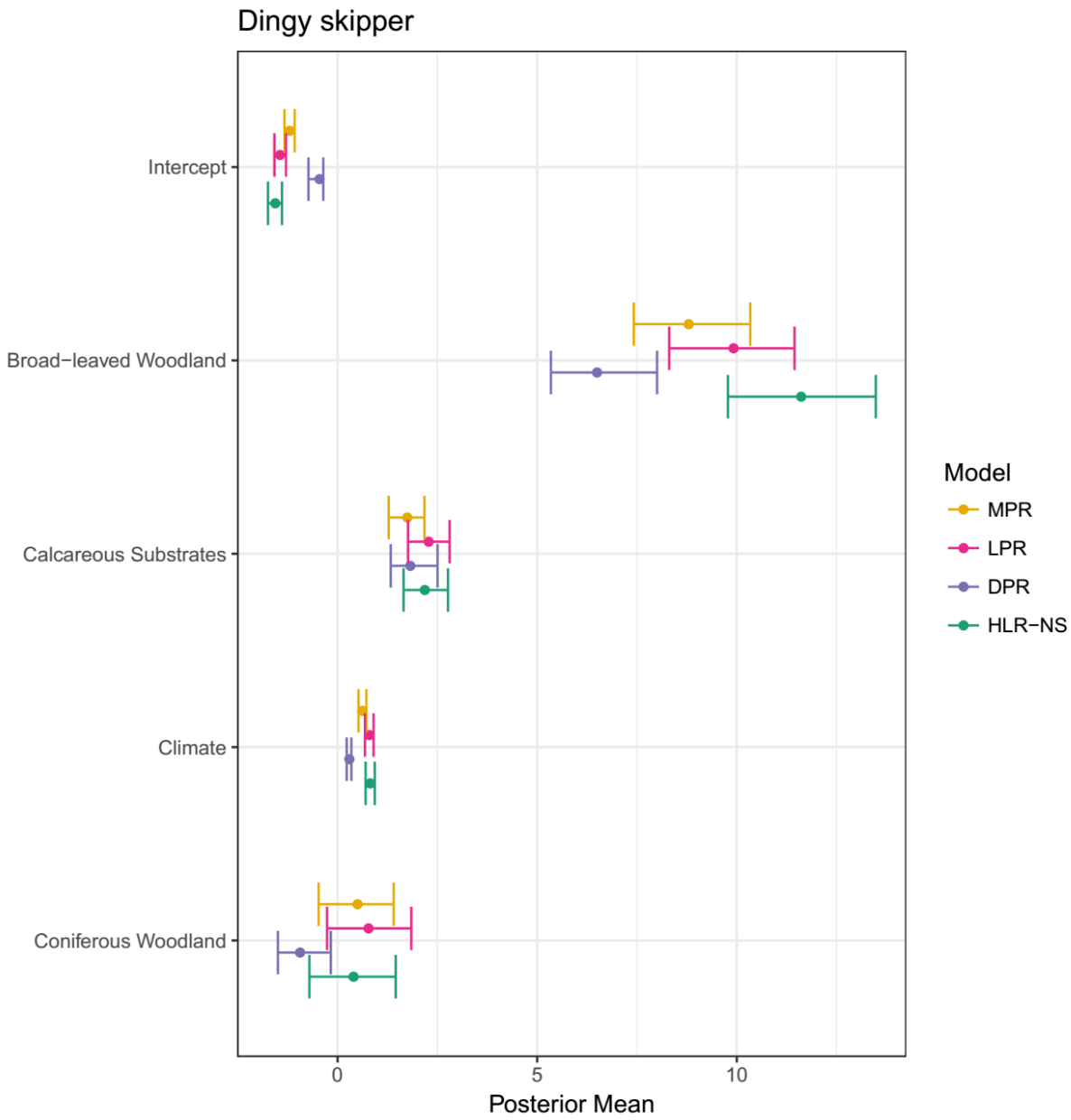


Figure 388: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the dingy skipper, *Erynnis tages*.

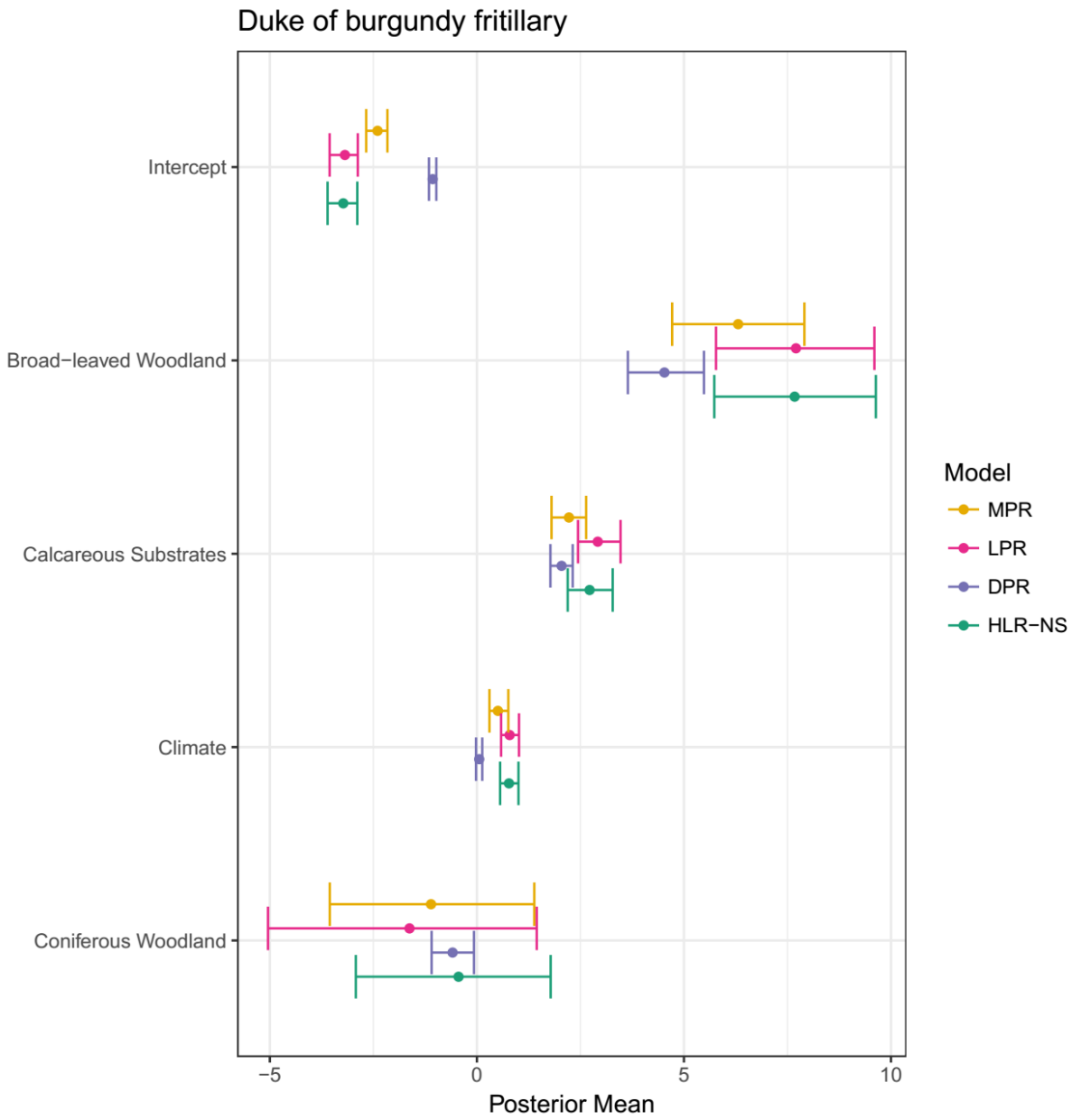


Figure 389: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Duke of Burgundy fritillary, *Hamearis lucina*.

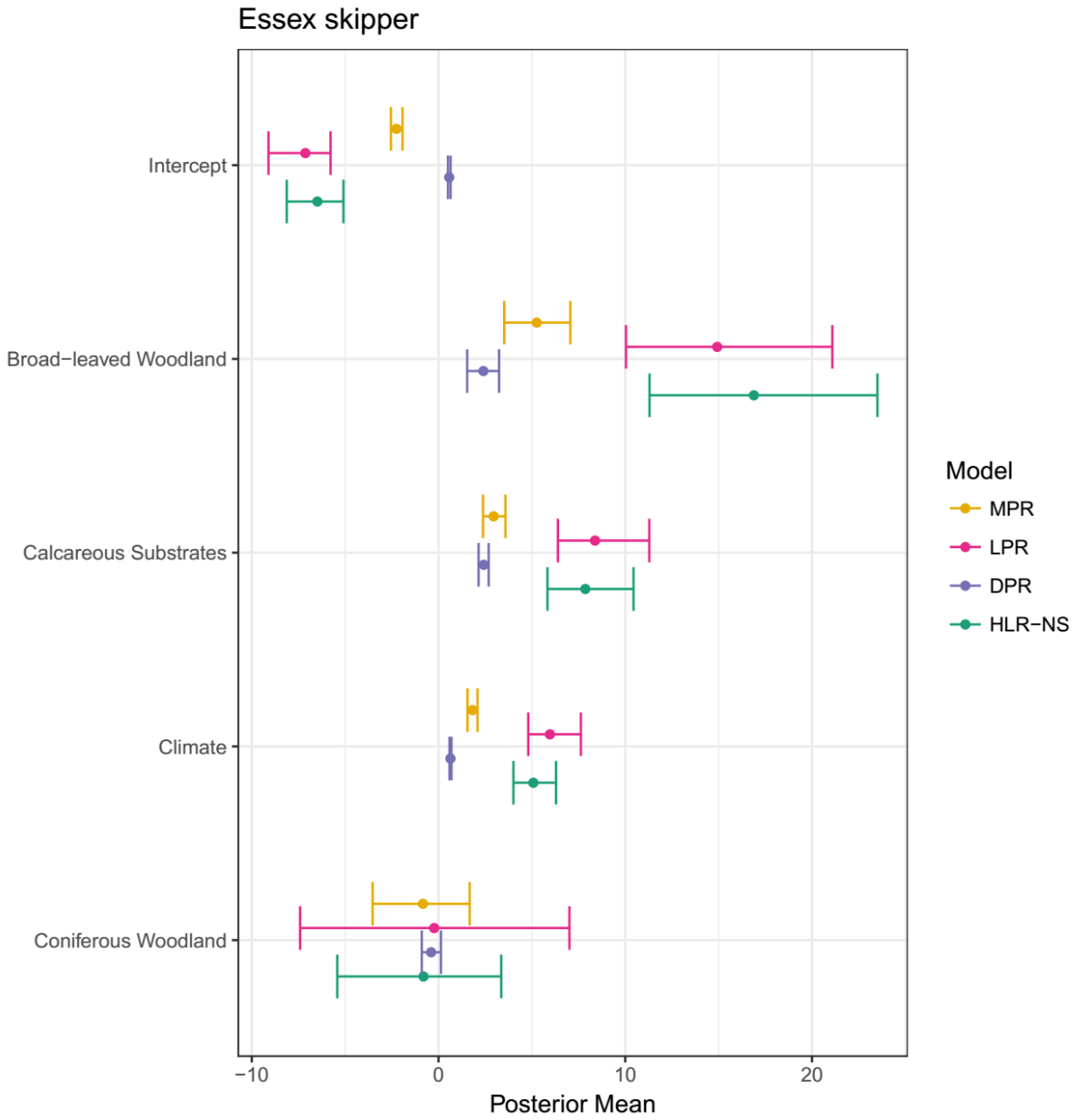


Figure 390: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Essex skipper, *Thymelicus lineola*.

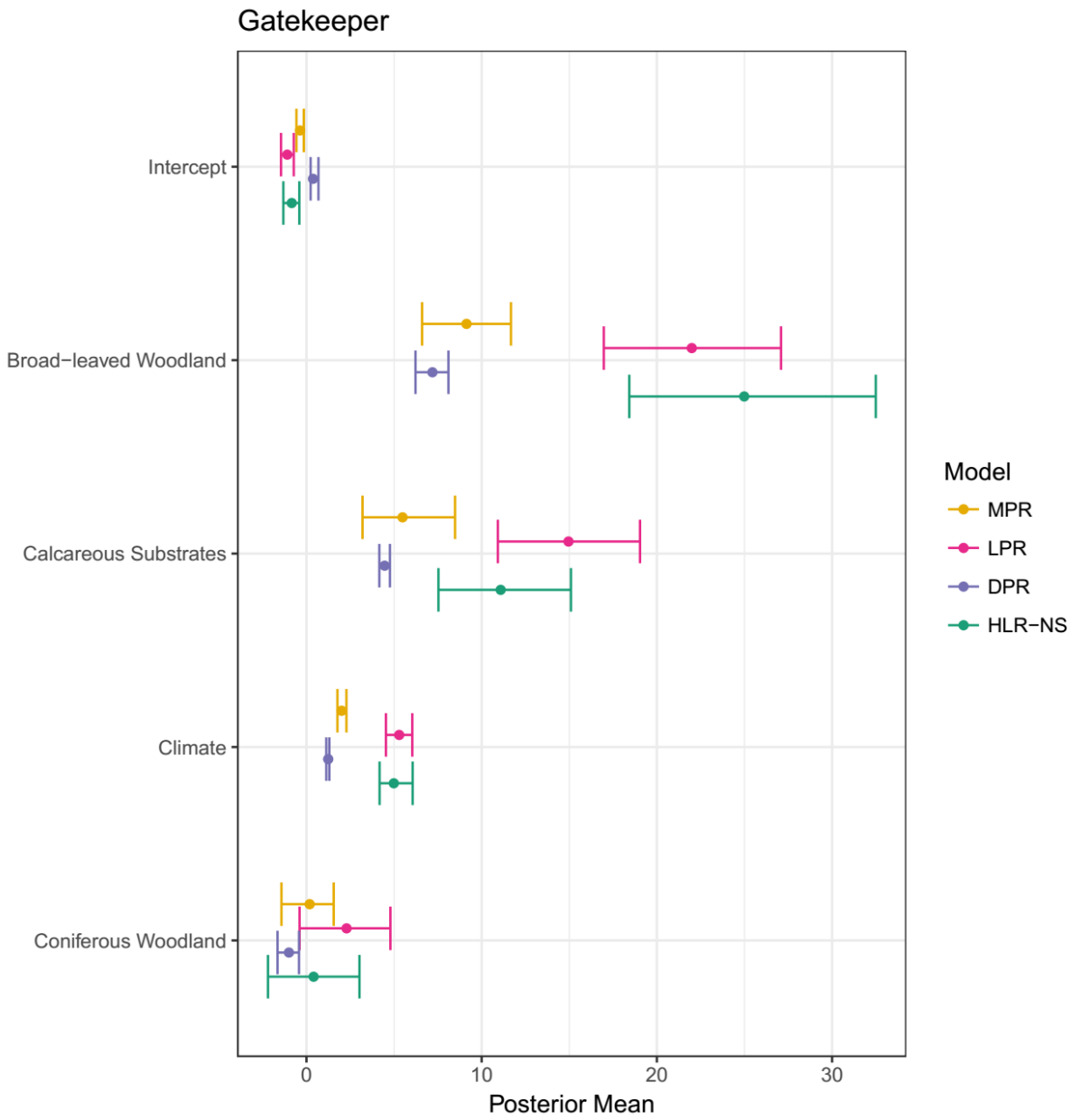


Figure 391: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the gatekeeper, *Pyronia tithonus*.

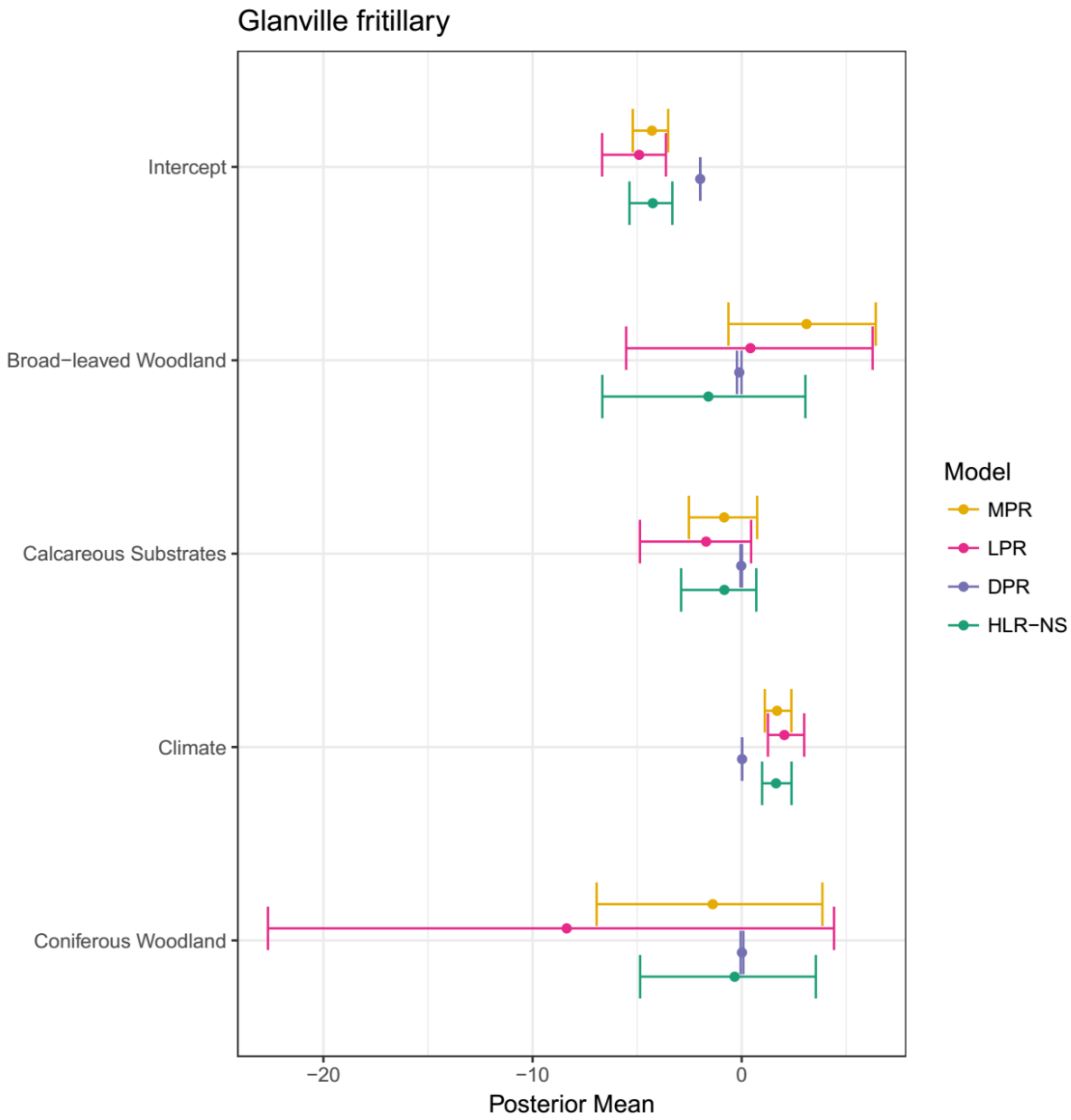


Figure 392: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Glanville fritillary, *Melitaea cinxia*.

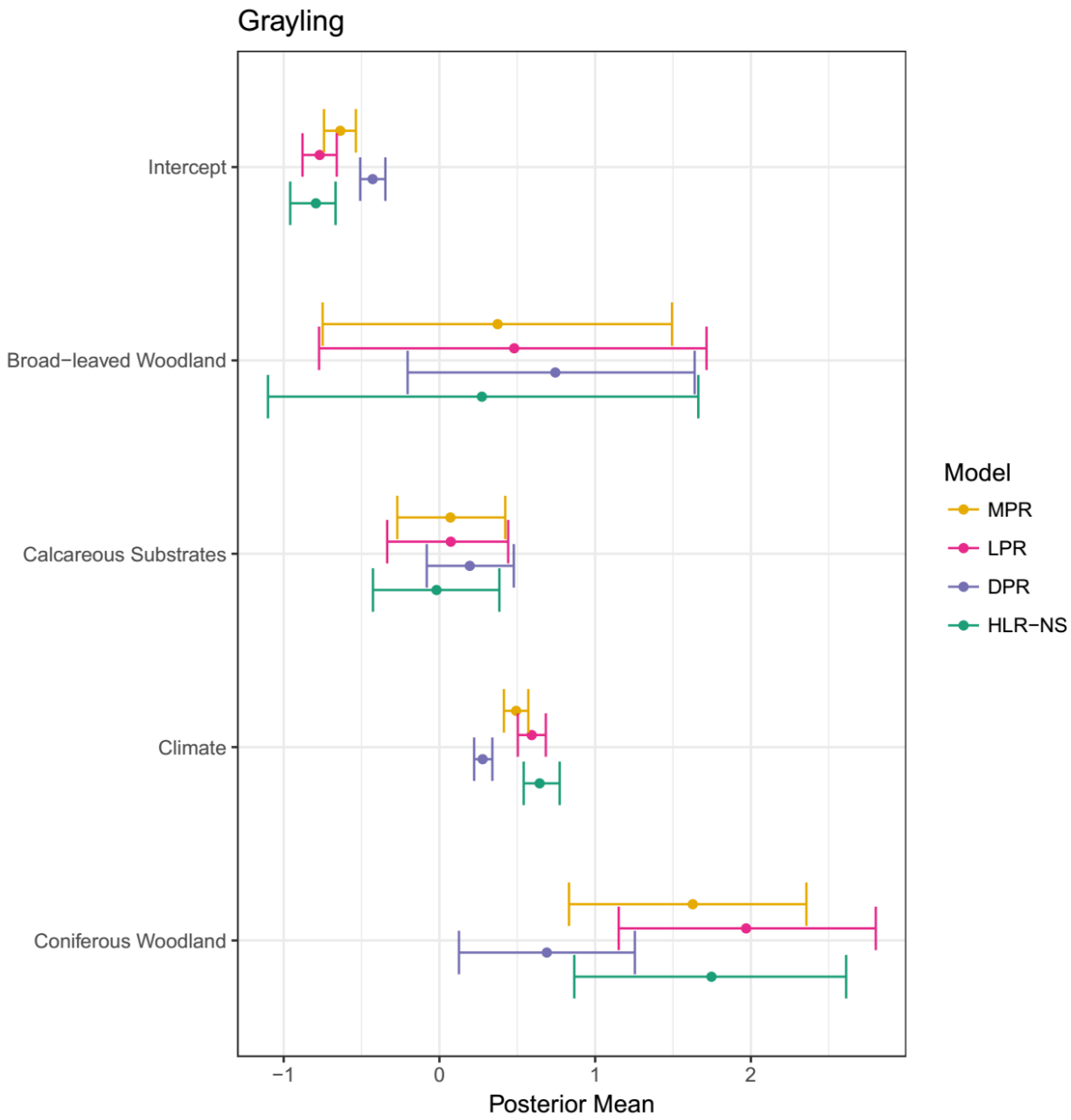


Figure 393: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the grayling, *Hipparchia semele*.

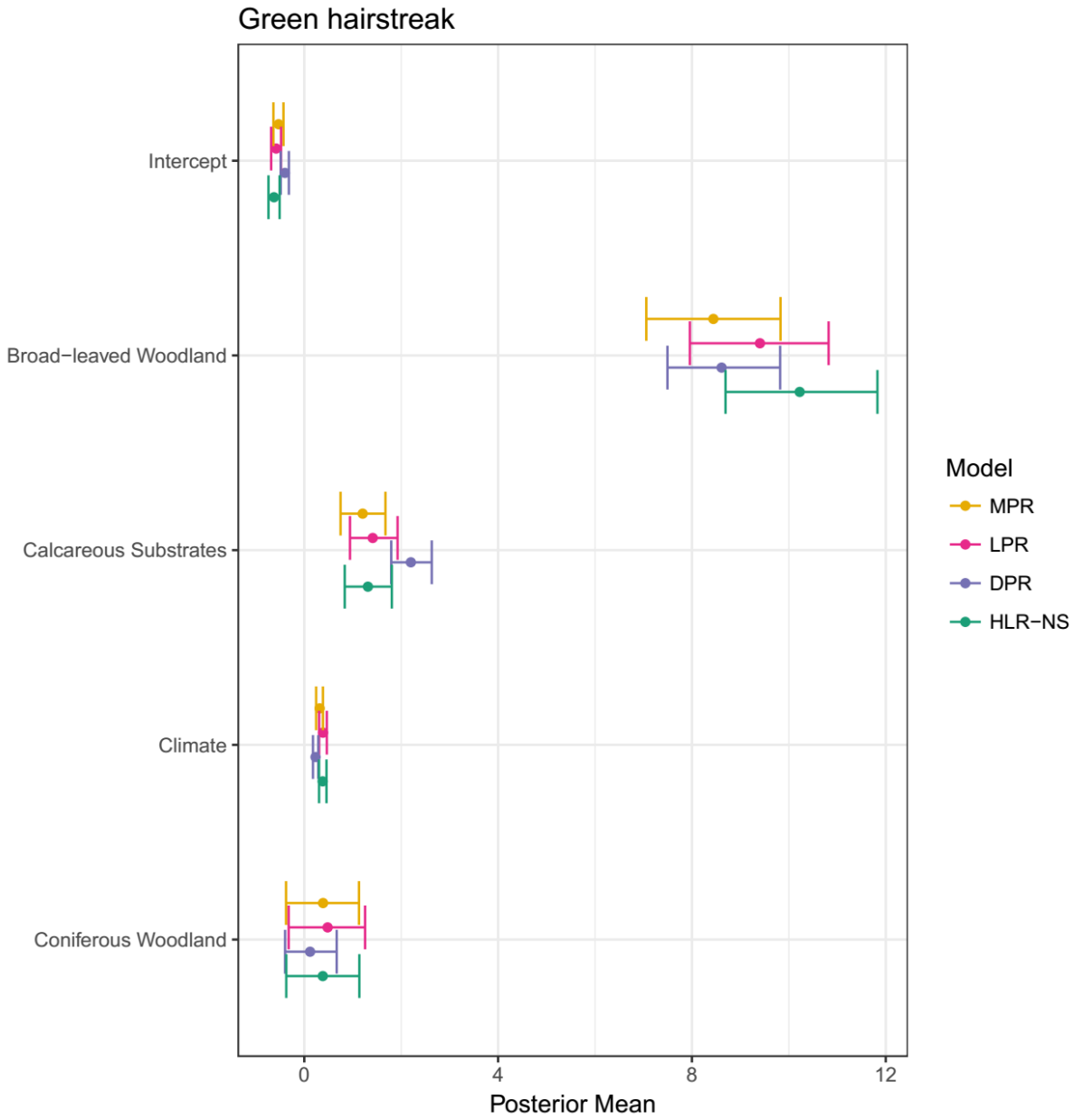


Figure 394: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the green hairstreak, *Callophrys rubi*.

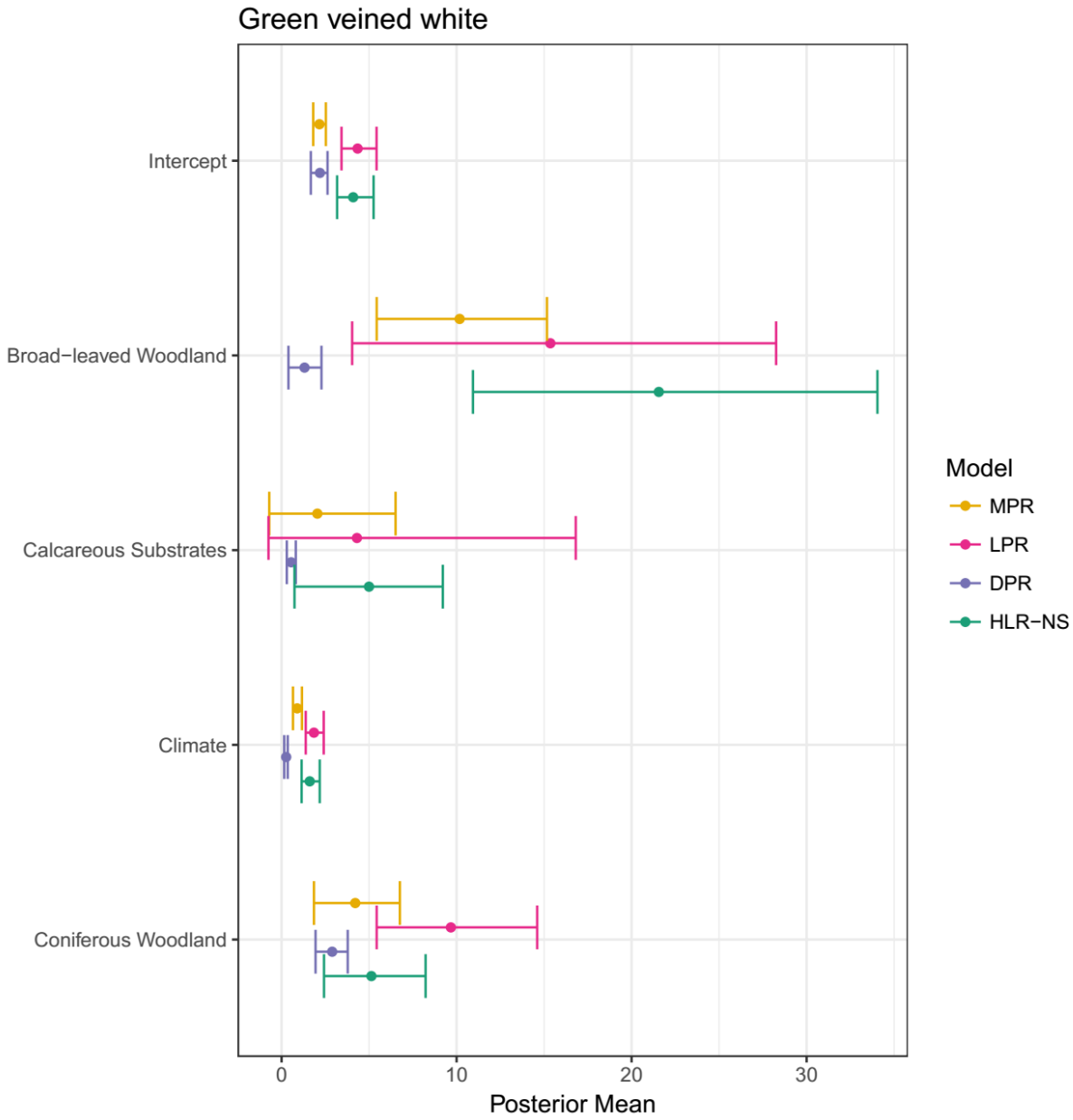


Figure 395: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the green-veined white, *Pieris napi*.

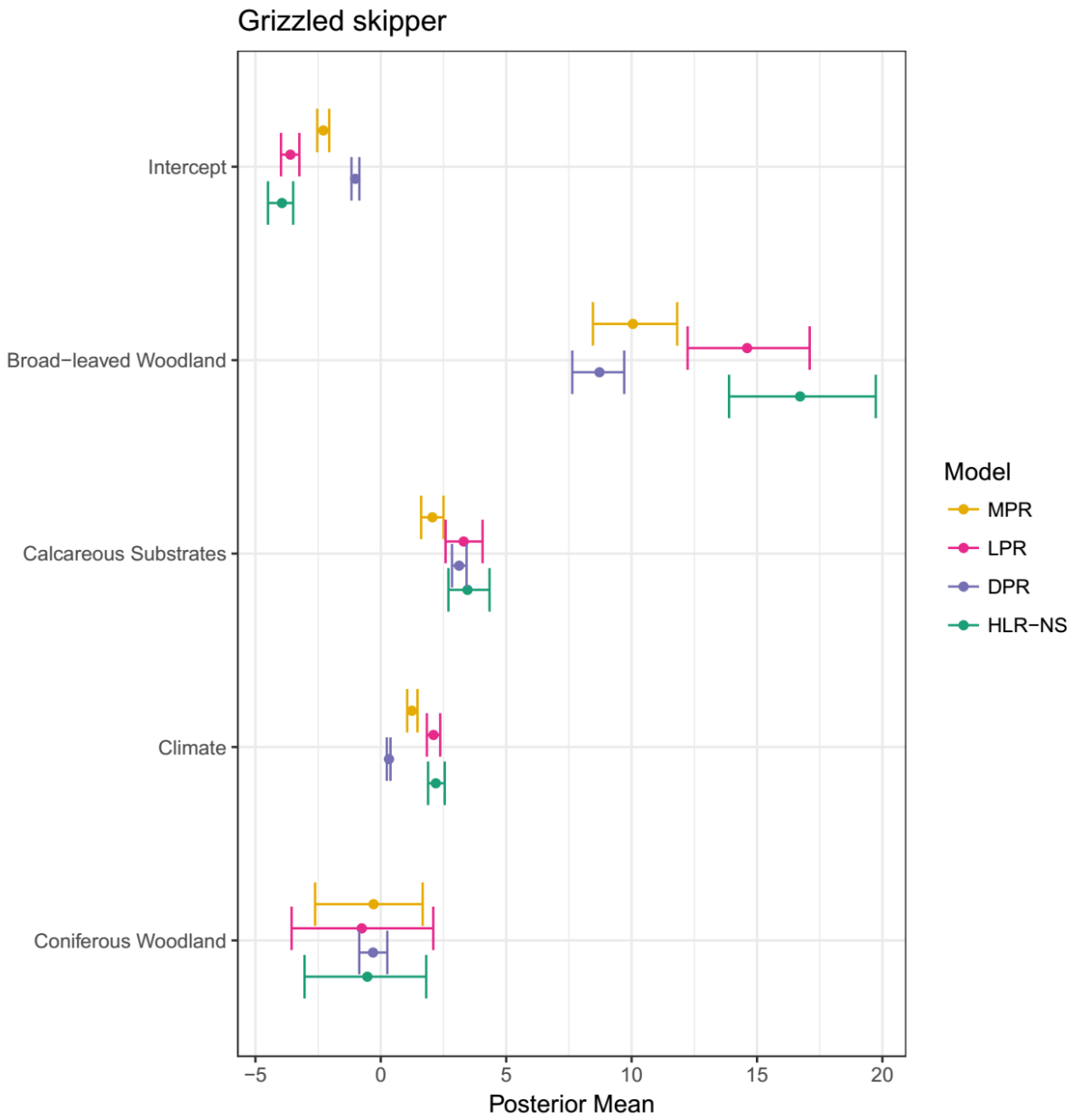


Figure 396: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the grizzled skipper, *Pyrgus malvae*.

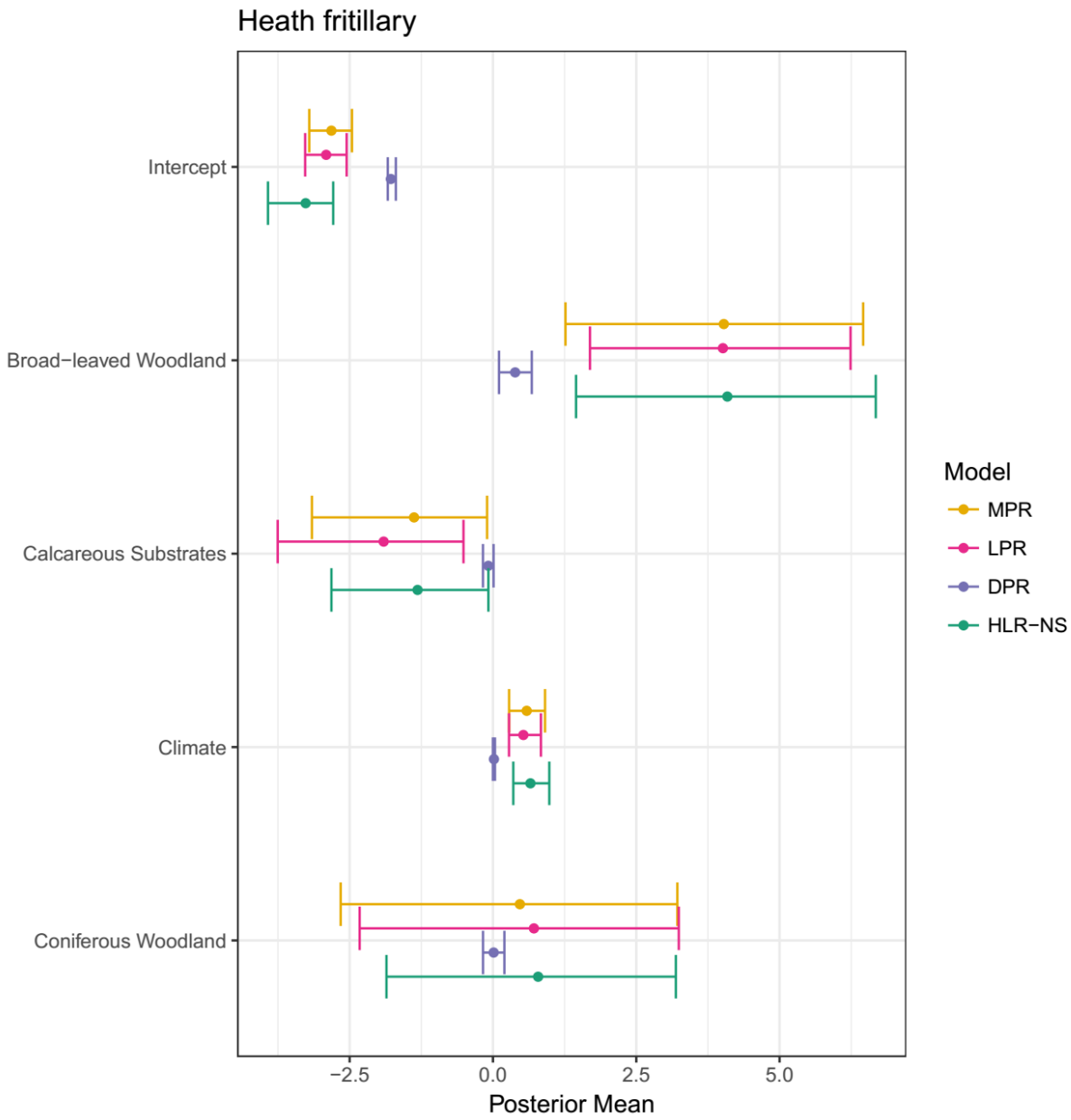


Figure 397: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the heath fritillary, *Melitaea athalia*.

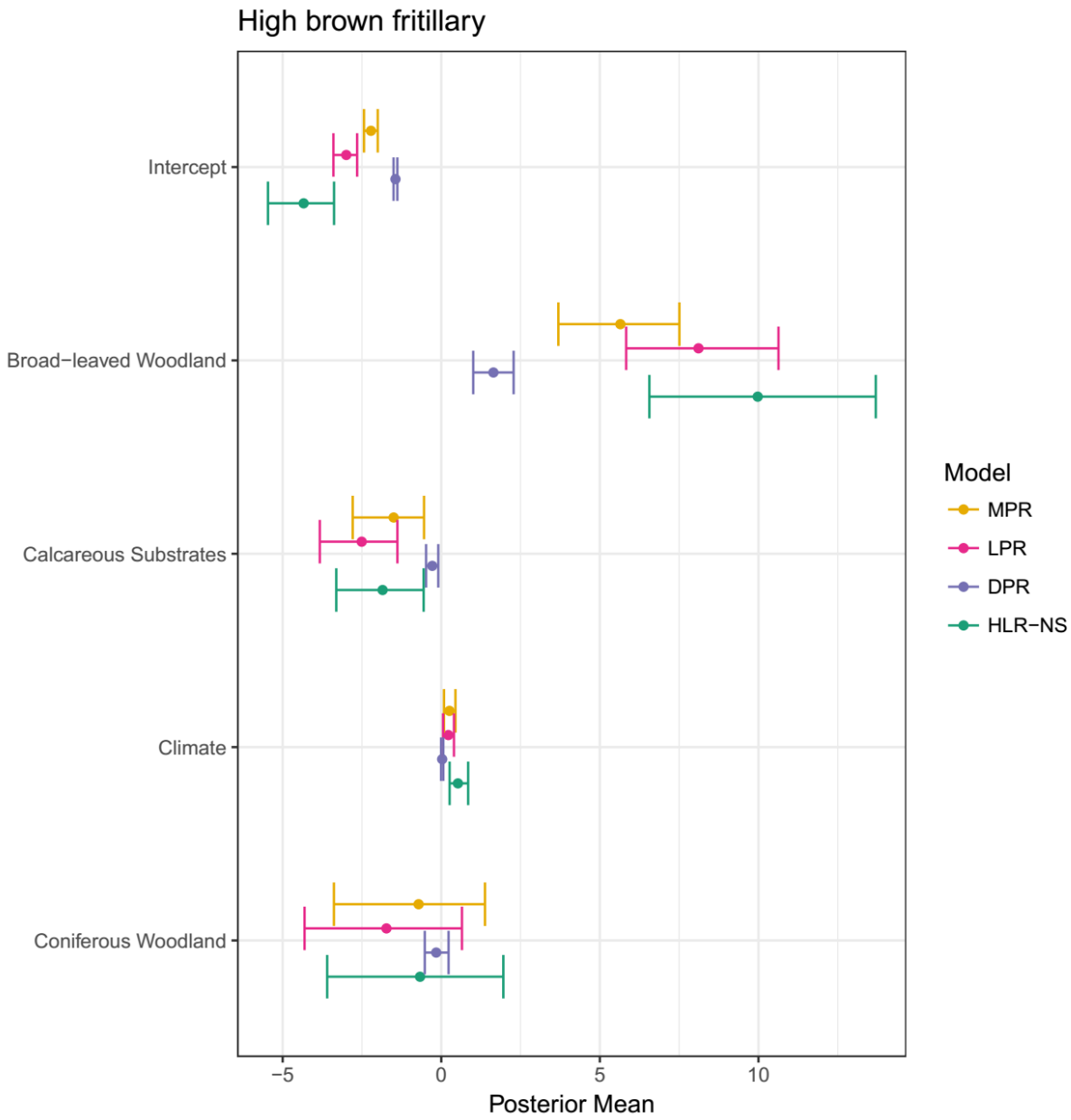


Figure 398: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the high brown fritillary, *Fabriciana adippe*.

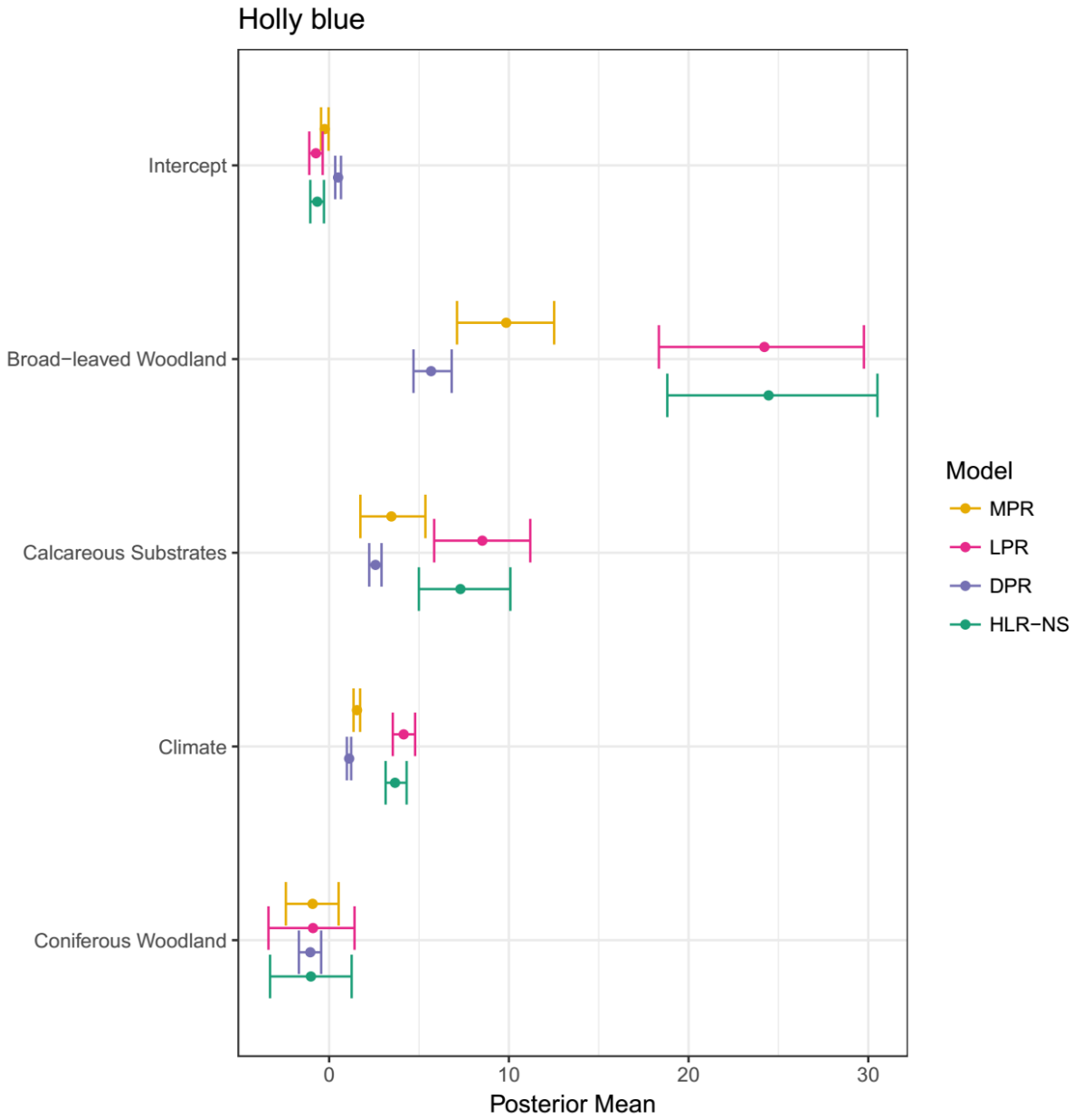


Figure 399: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the holly blue, *Celastrina argiolus*.

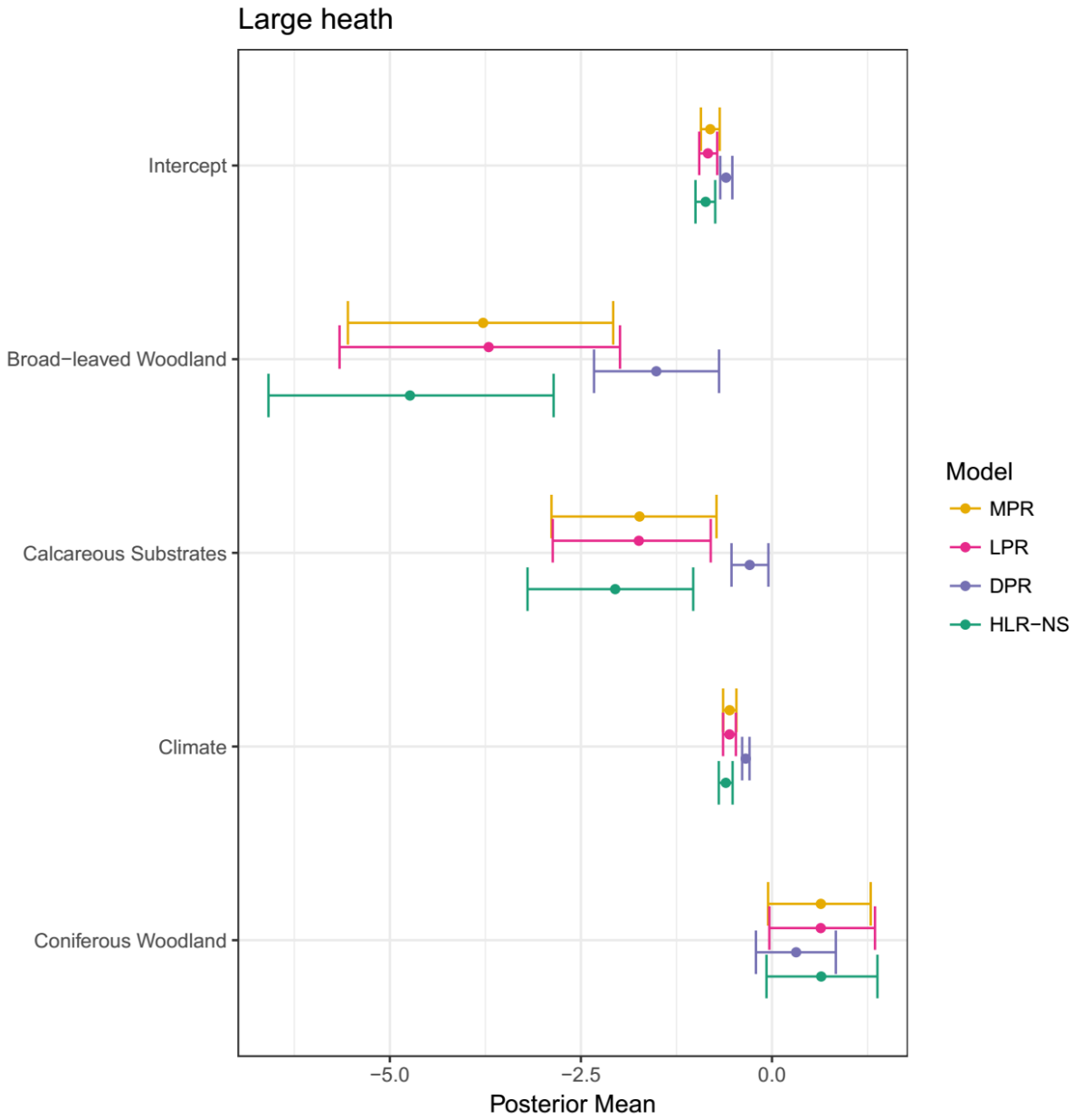


Figure 400: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the large heath, *Coenonympha tullia*.

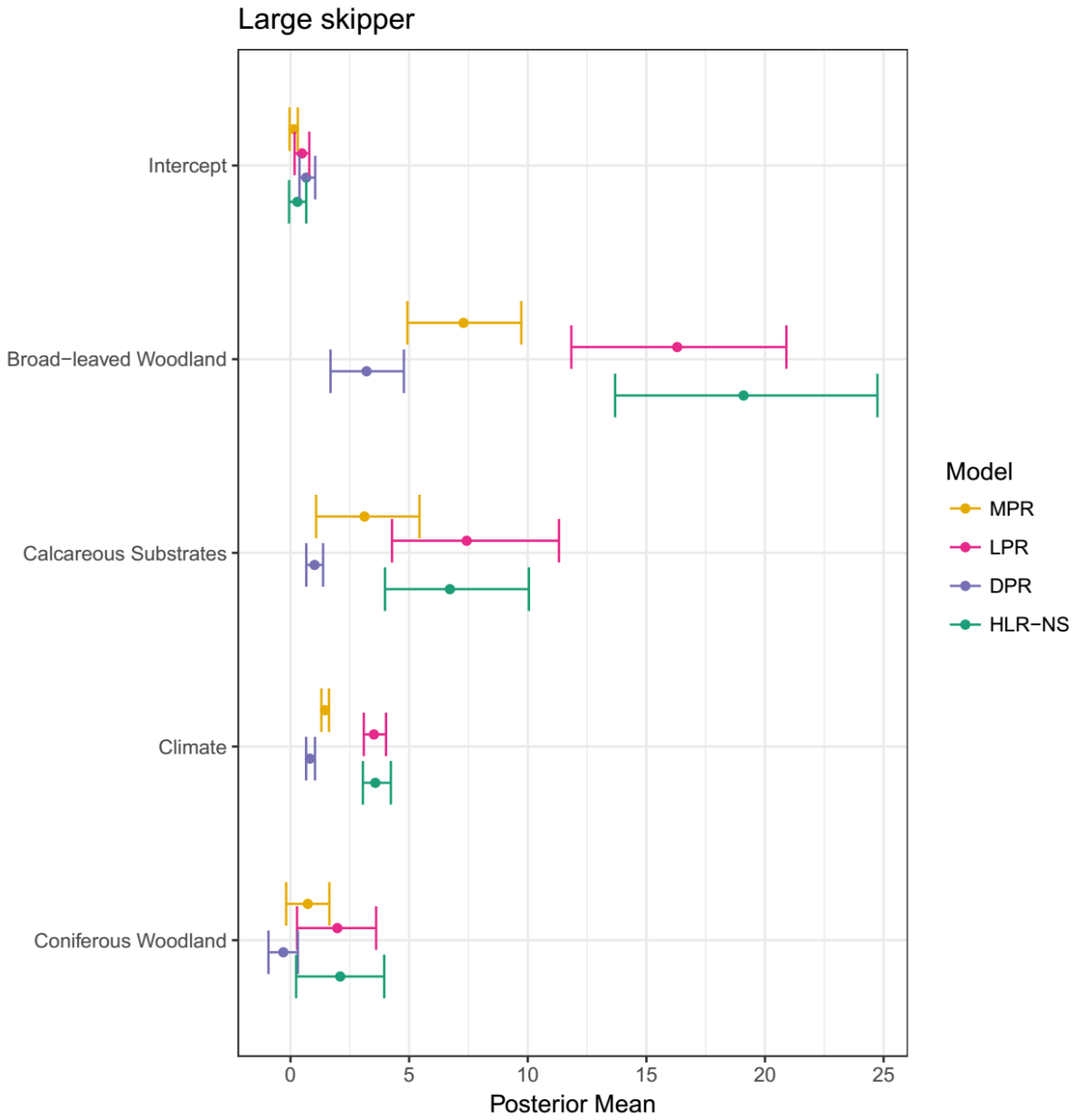


Figure 401: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the large skipper, *Ochlodes sylvanus*.

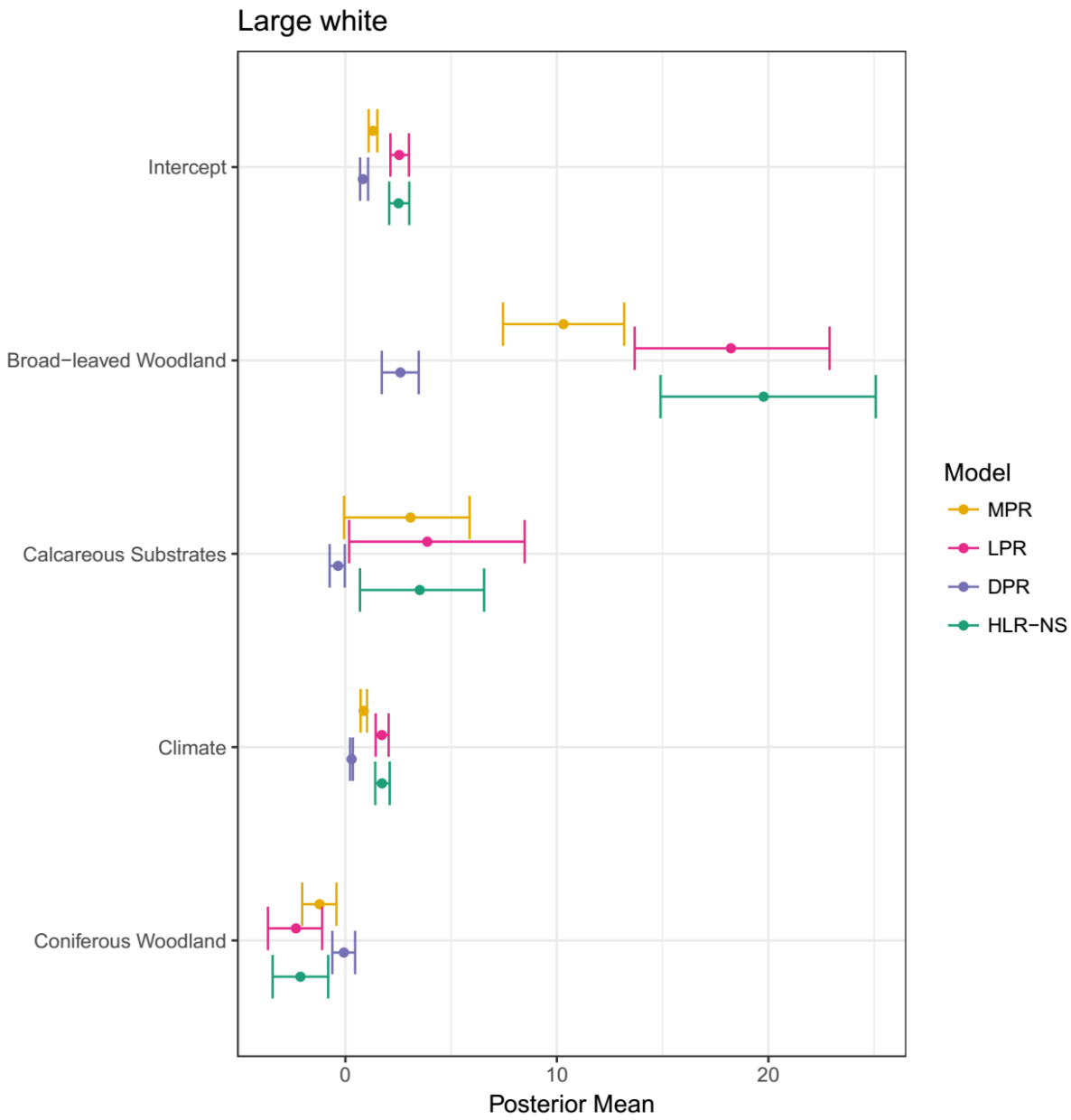


Figure 402: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the large white, *Pieris brassicae*.

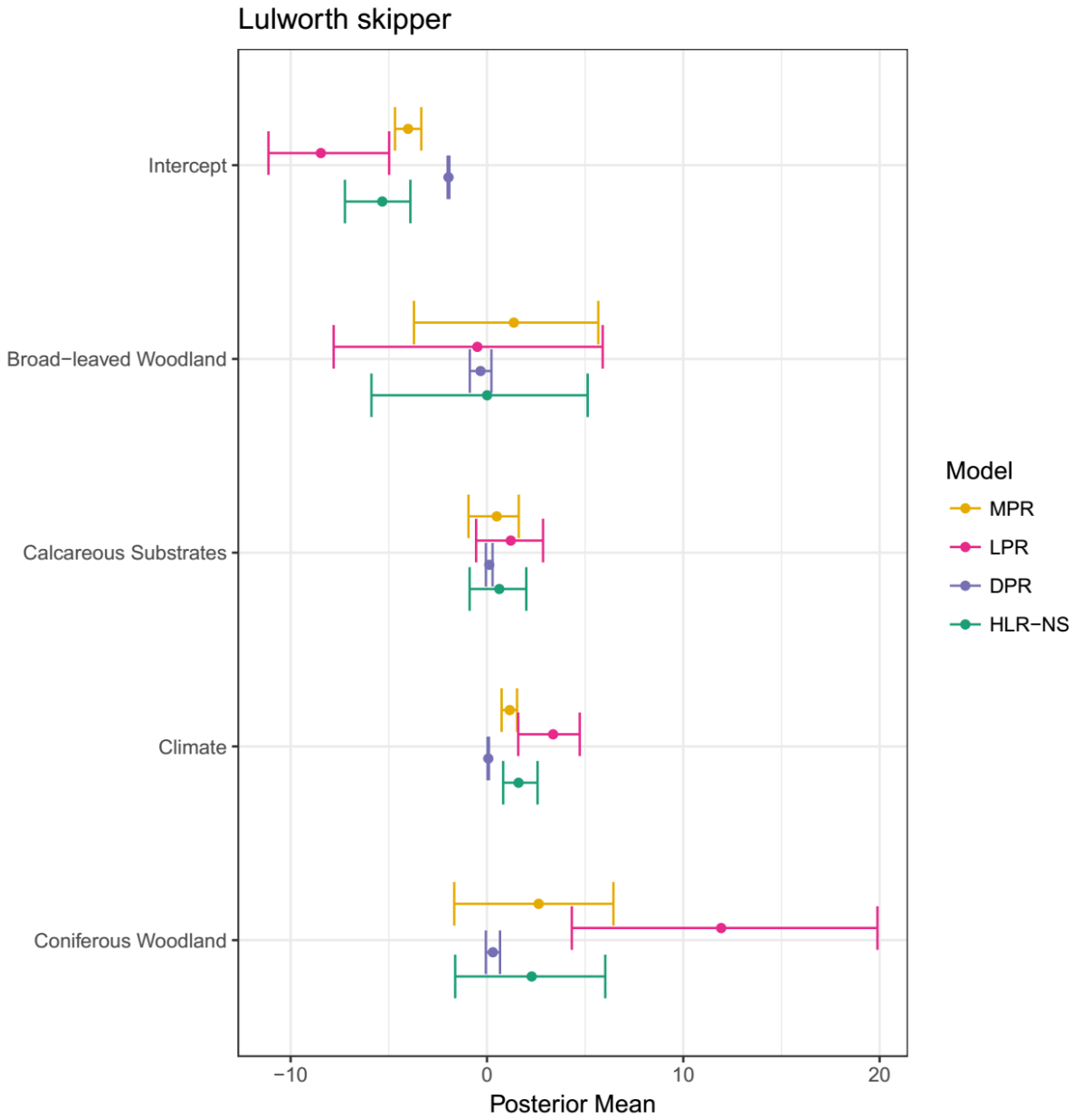


Figure 403: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Lulworth skipper, *Thymelicus acteon*.

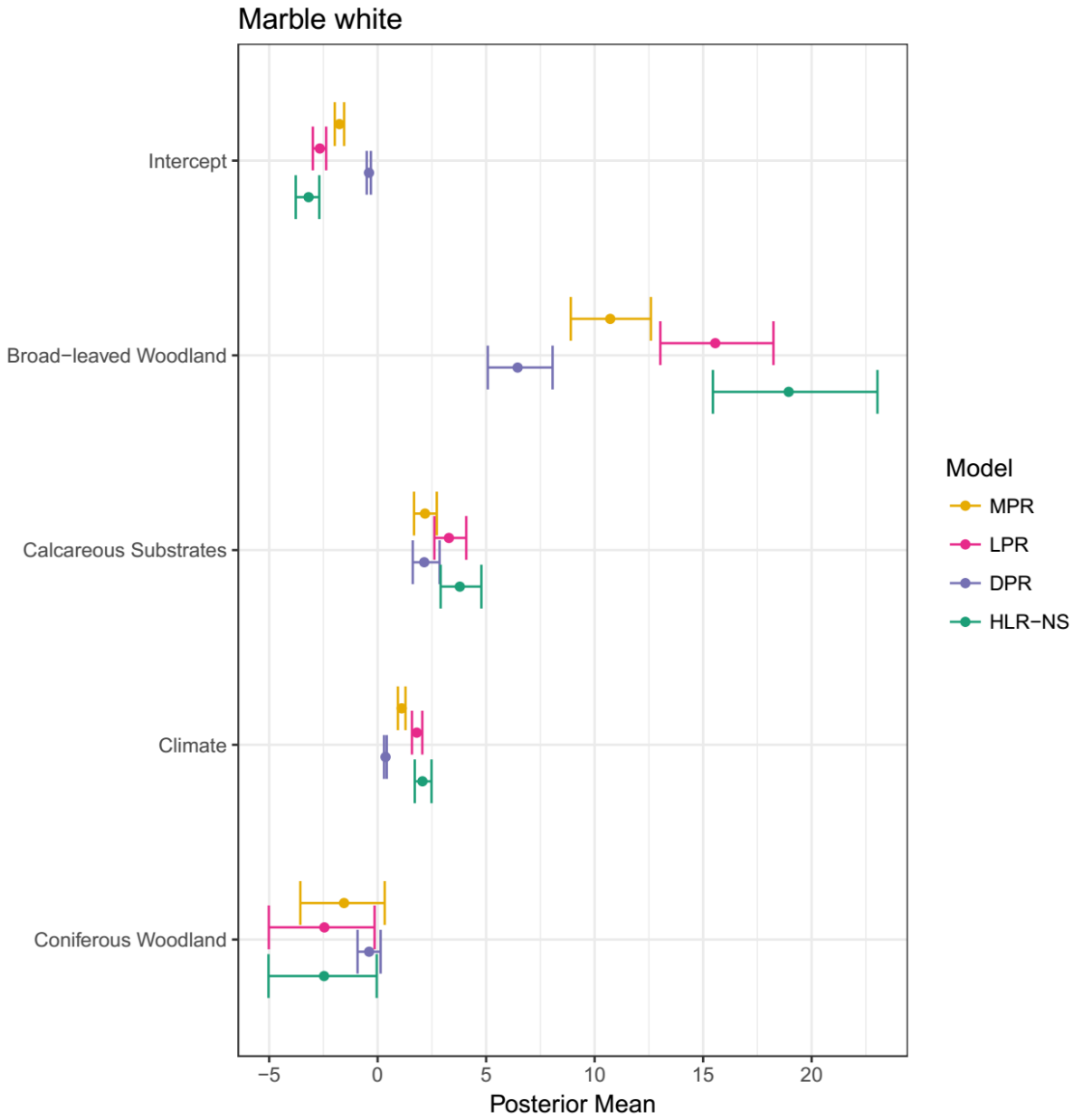


Figure 404: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the marble white, *Melanargia galathea*.

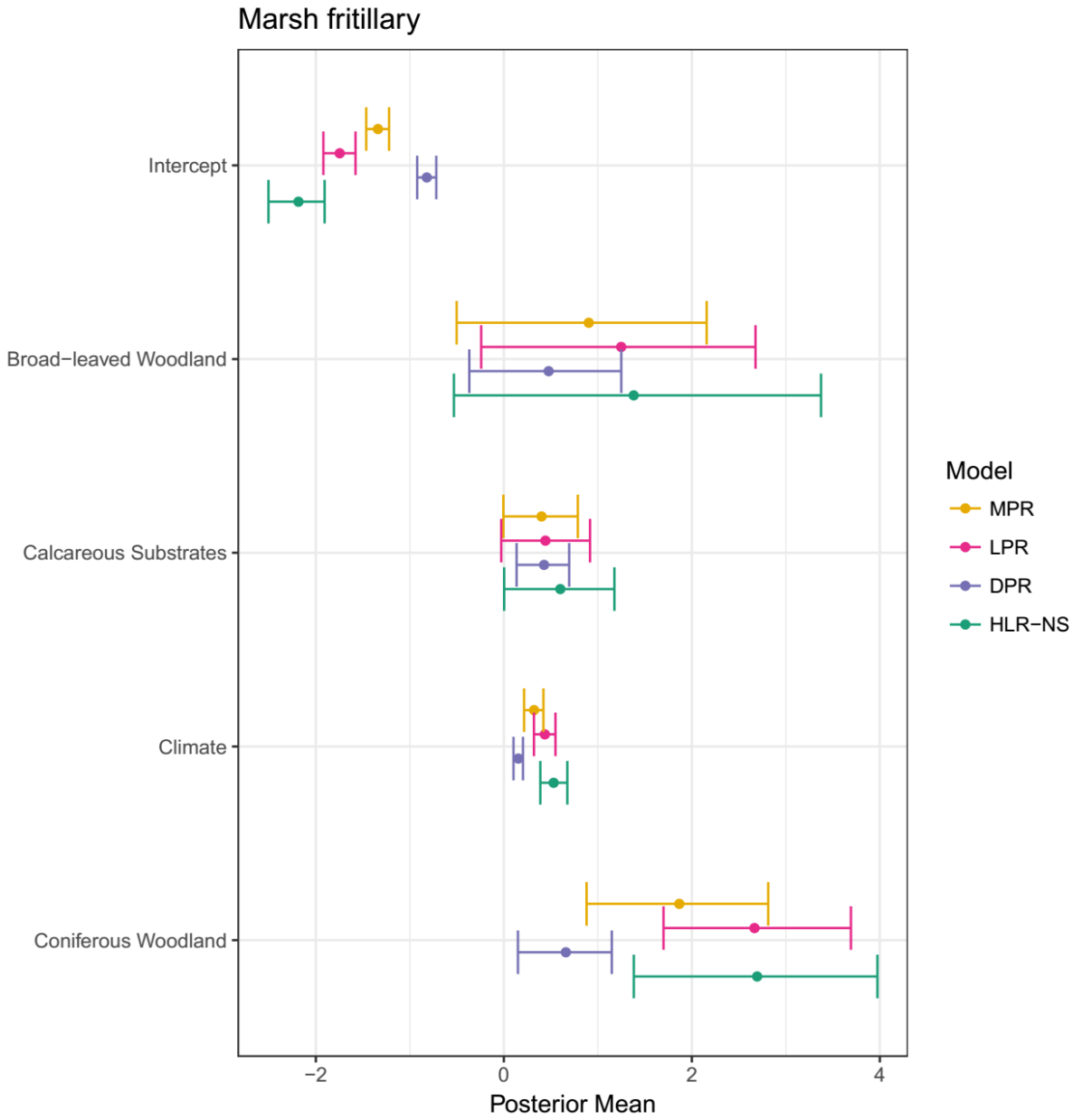


Figure 405: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the marsh fritillary, *Melanerpes formicivorus*.

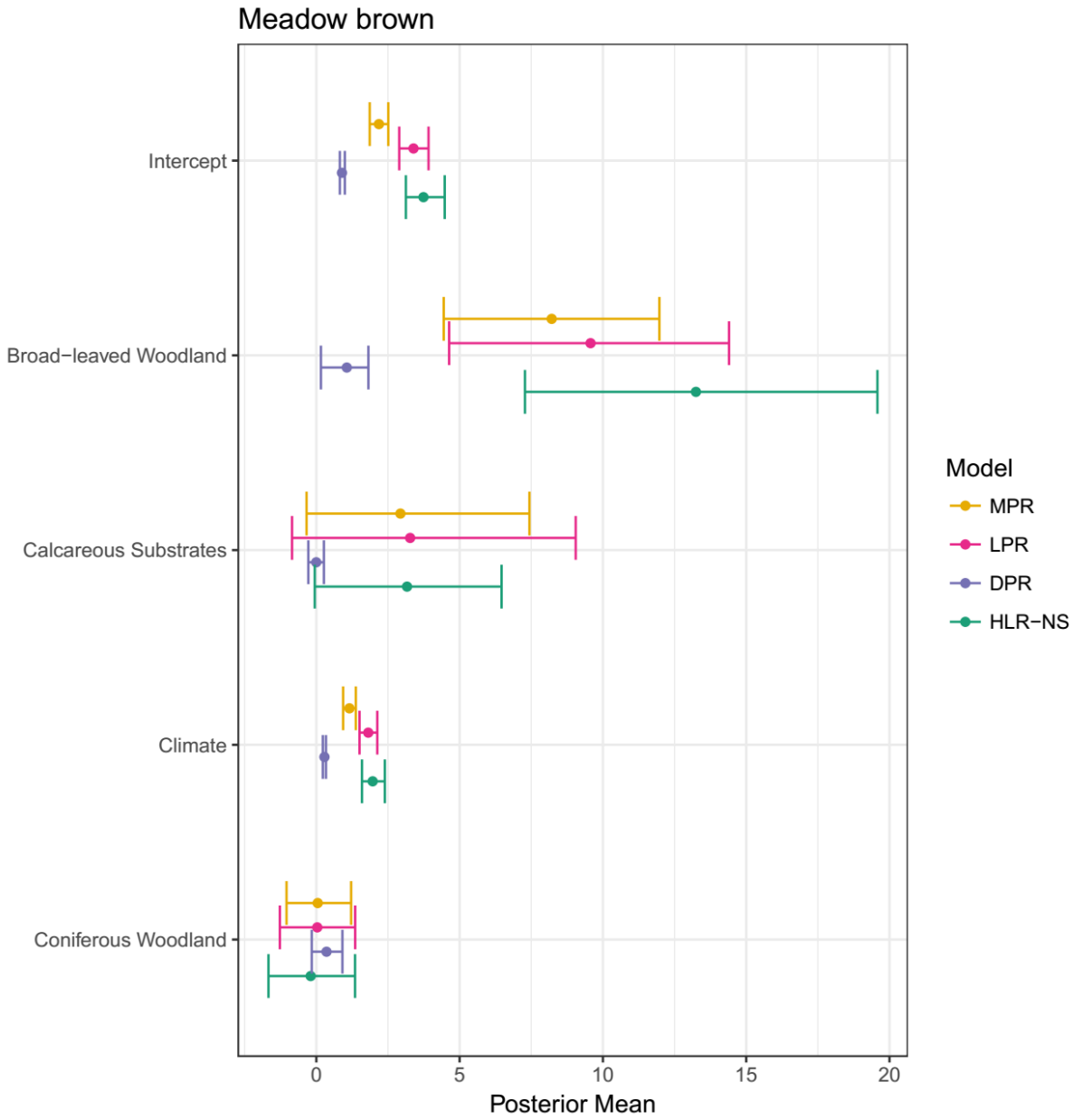


Figure 406: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the meadow brown, *Maniola jurtina*.

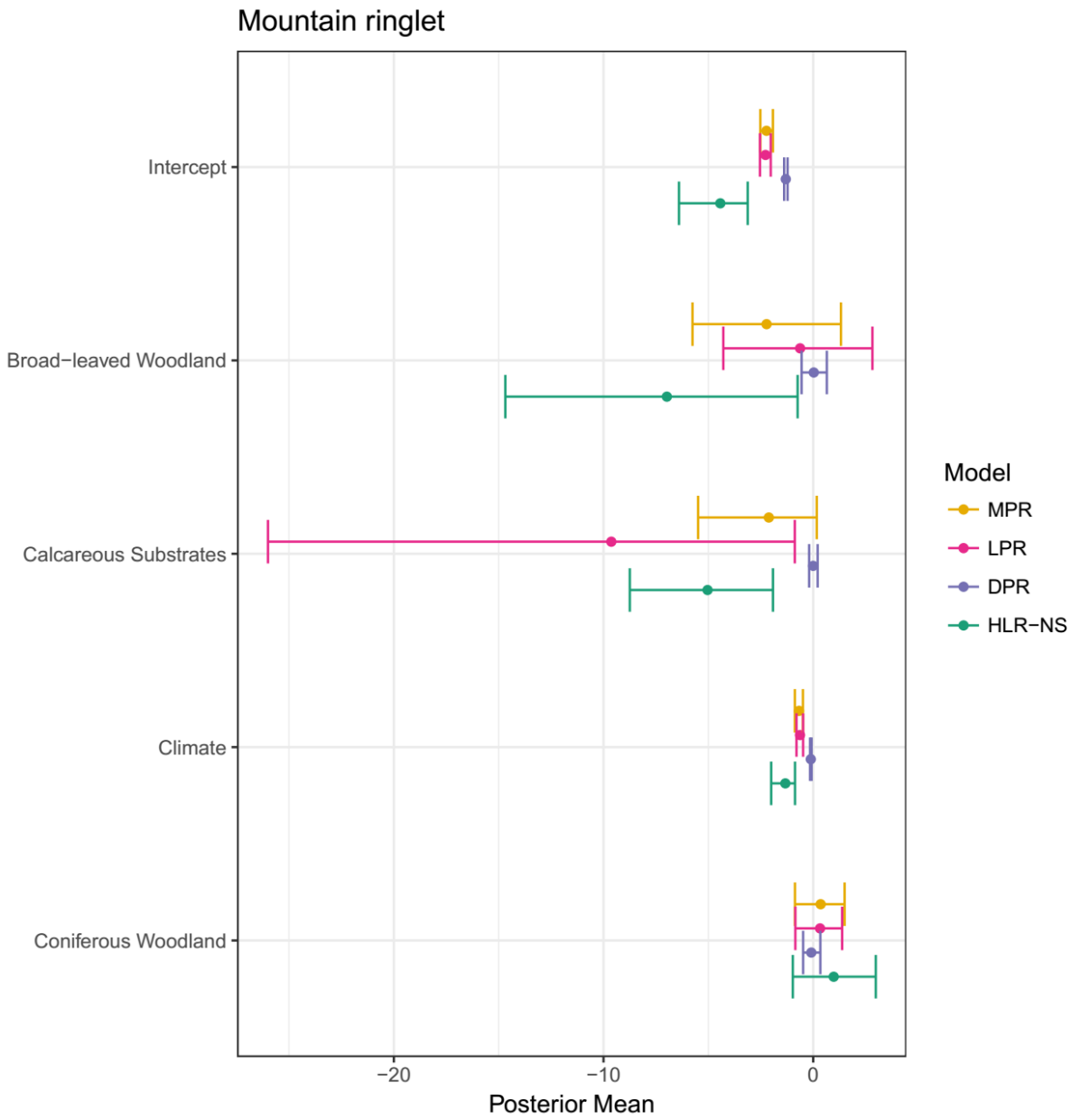


Figure 407: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the mountain ringlet, *Erebia epiphron*.

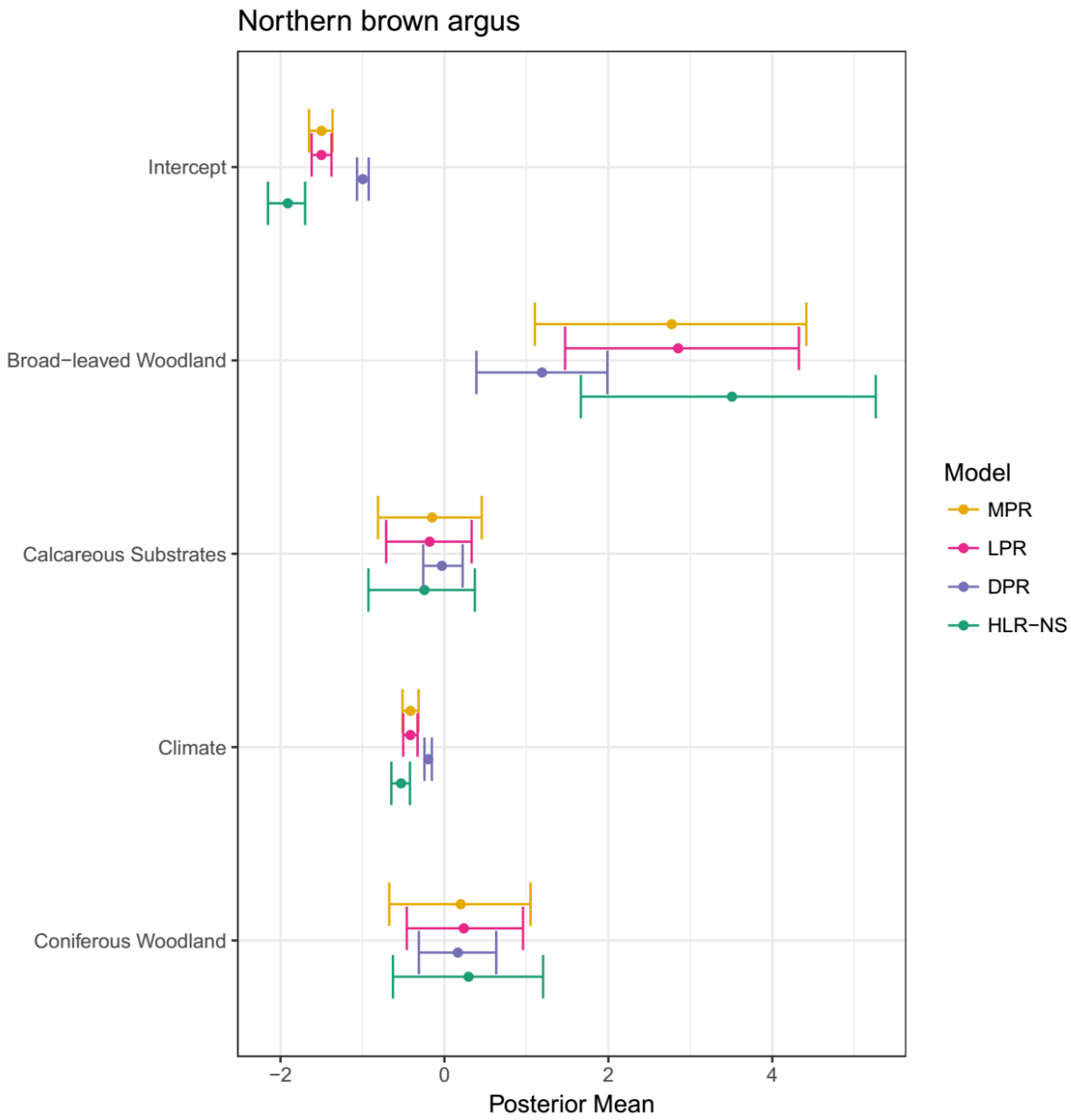


Figure 408: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the northern brown argus, *Aricia artaxerxes*.

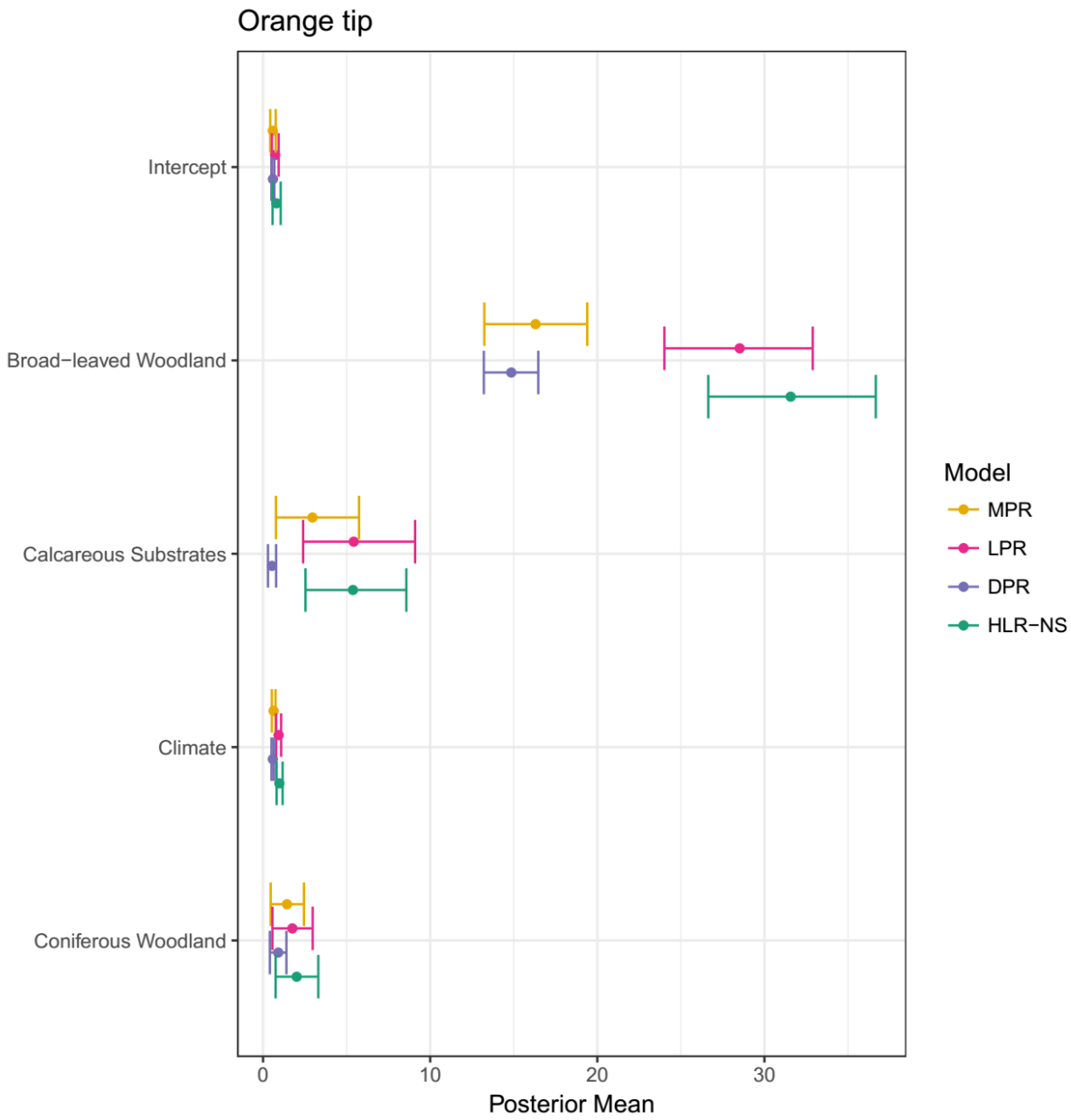


Figure 409: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the orange tip, *Anthocharis cardamines*.

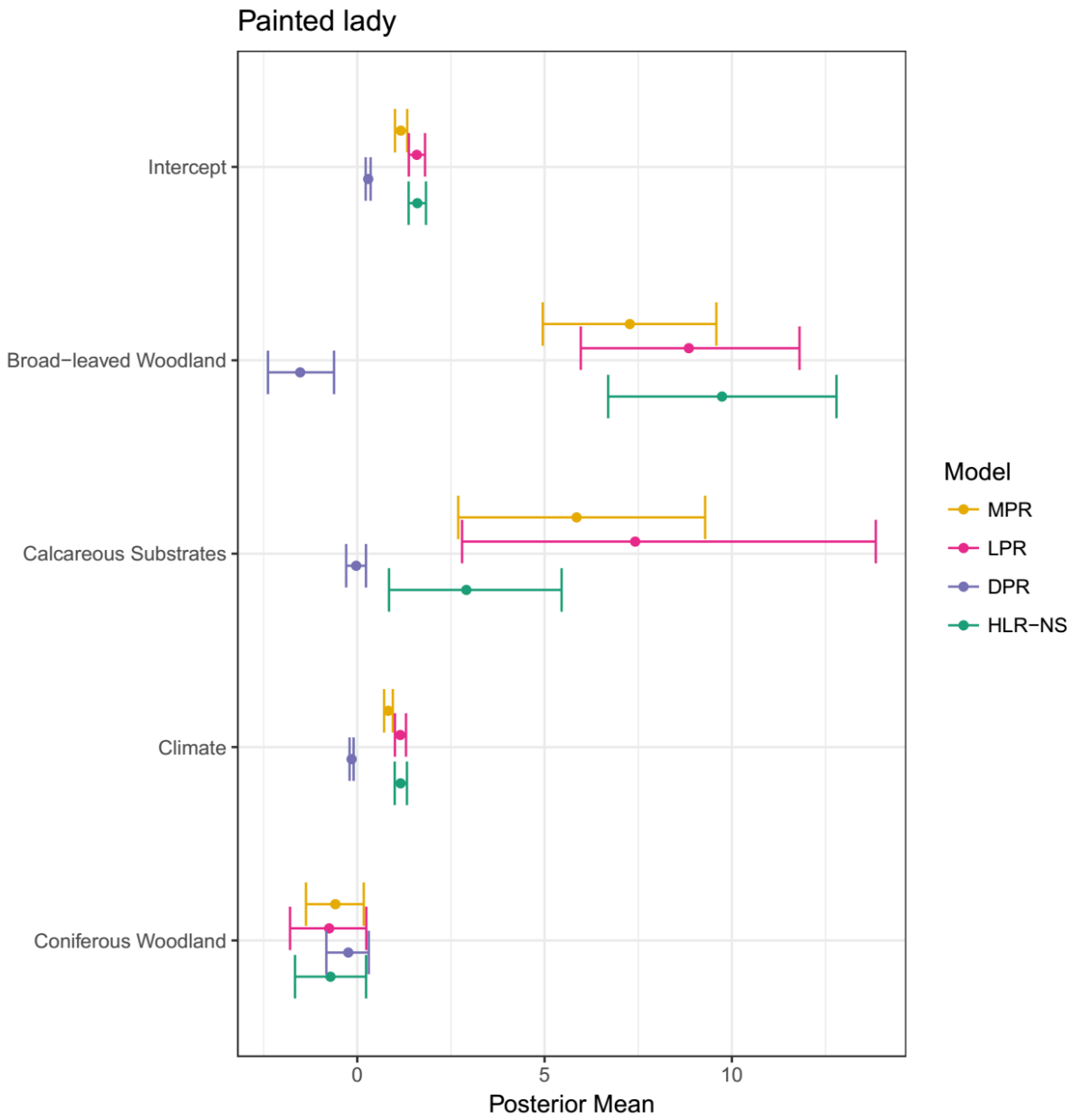


Figure 410: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the painted lady, *Vanessa cardui*.

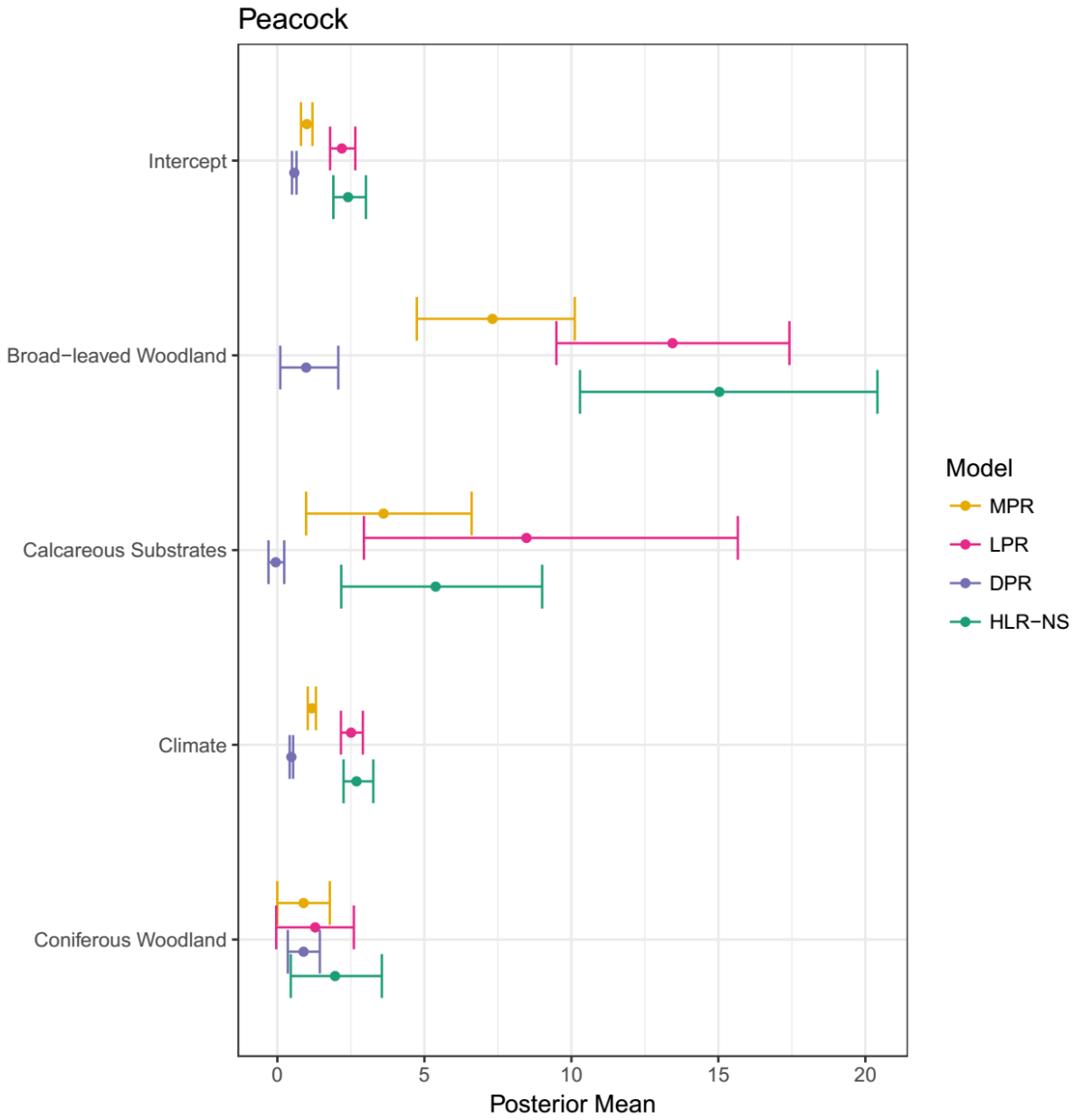


Figure 411: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the peacock, *Aglais io*.

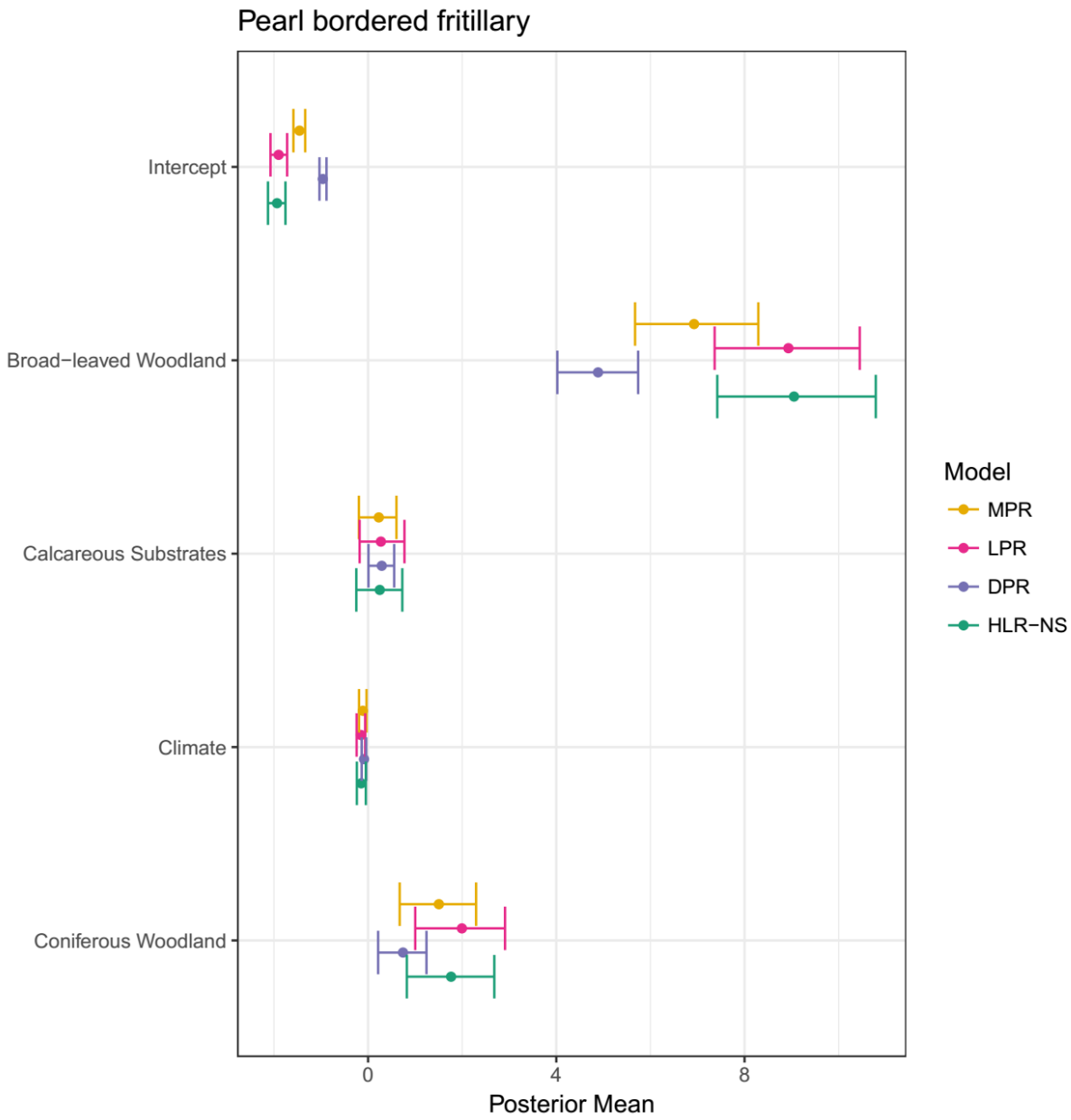


Figure 412: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pearl-bordered fritillary, *Boloria euphrosyne*.

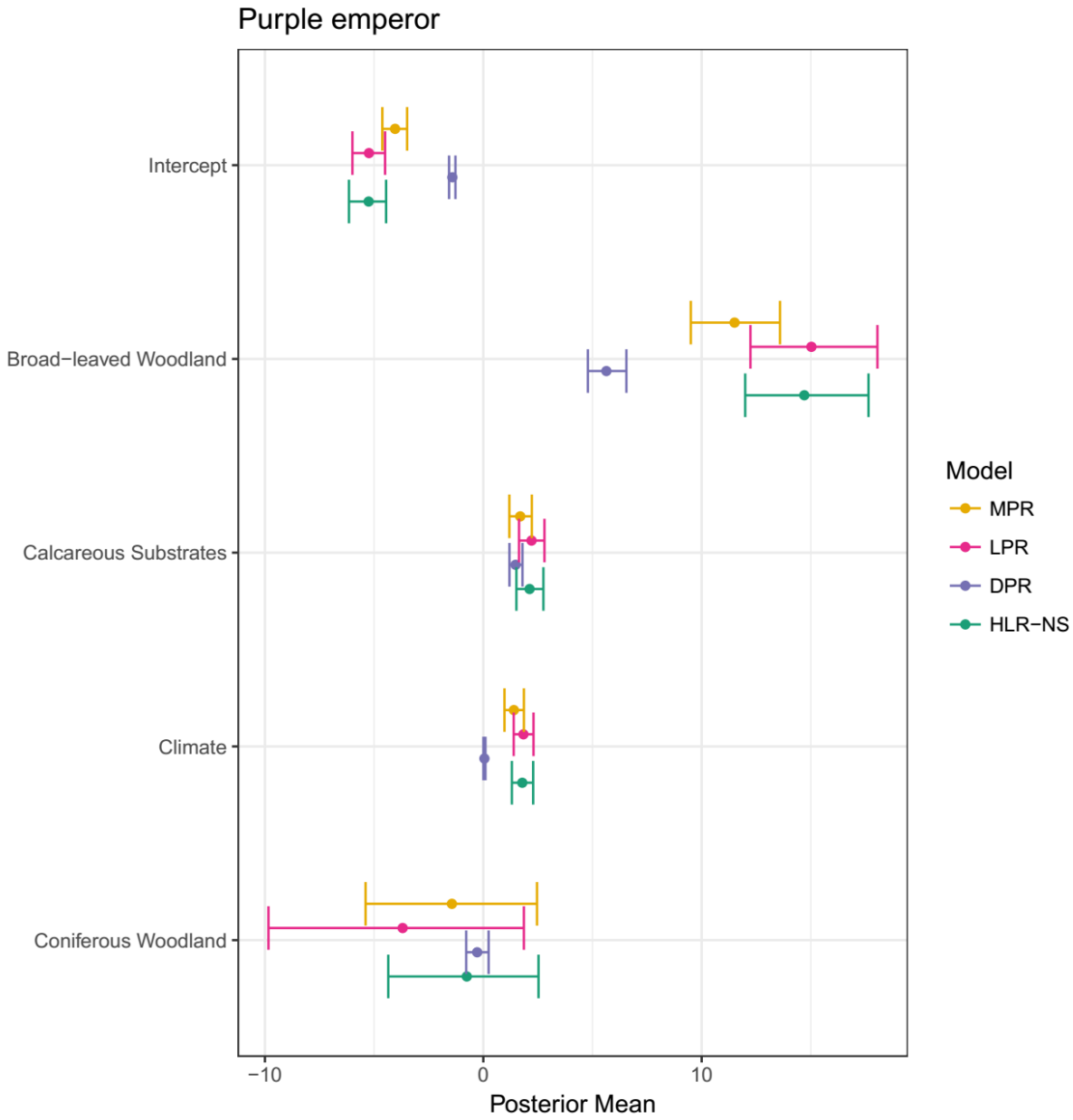


Figure 413: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the purple emperor, *Apatura iris*.

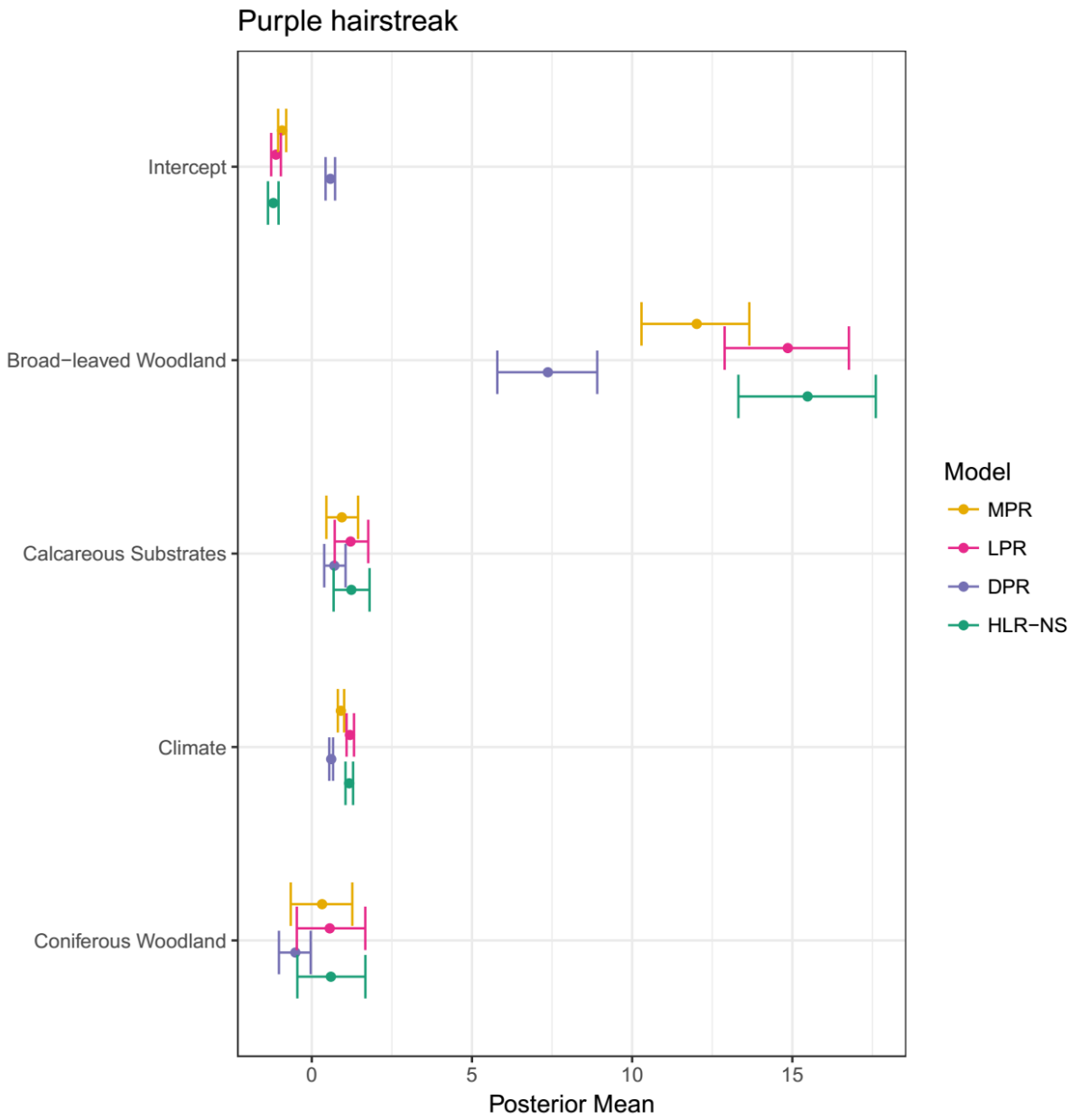


Figure 414: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the purple hairstreak, *Neozephyrus quercus*.

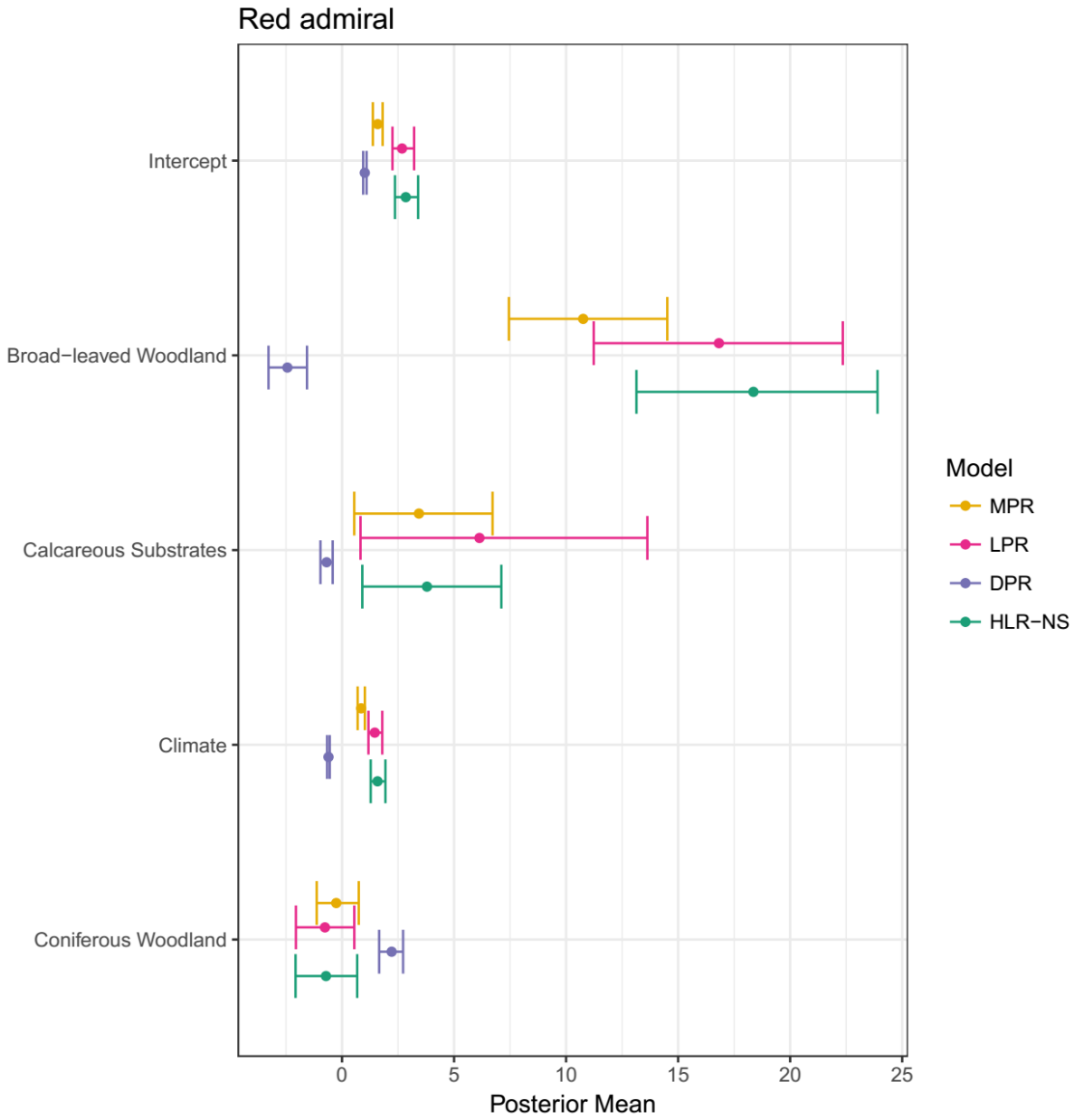


Figure 415: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the red admiral, *Vanessa atalanta*.

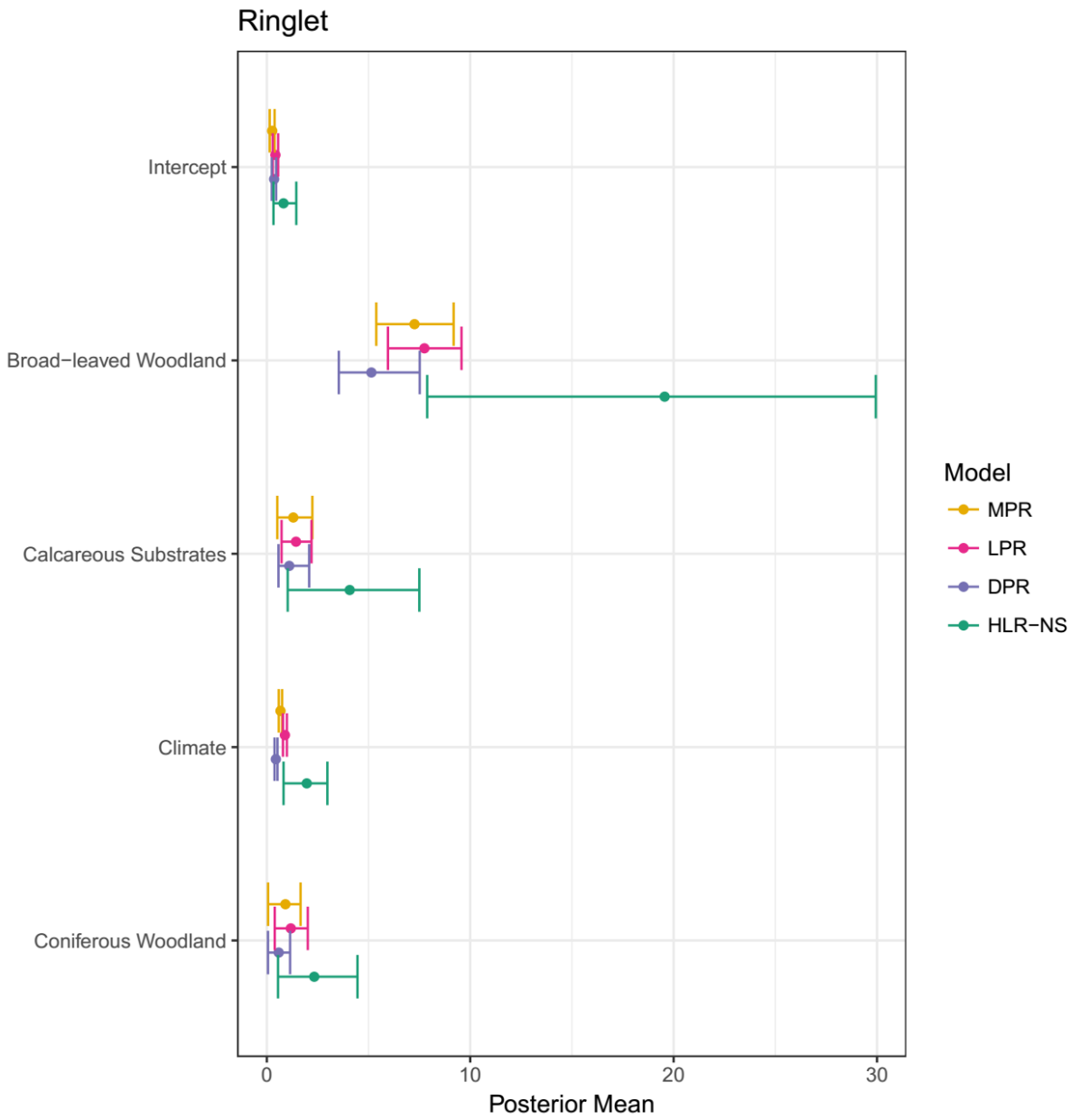


Figure 416: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ringlet, *Aphantopus hyperantus*.

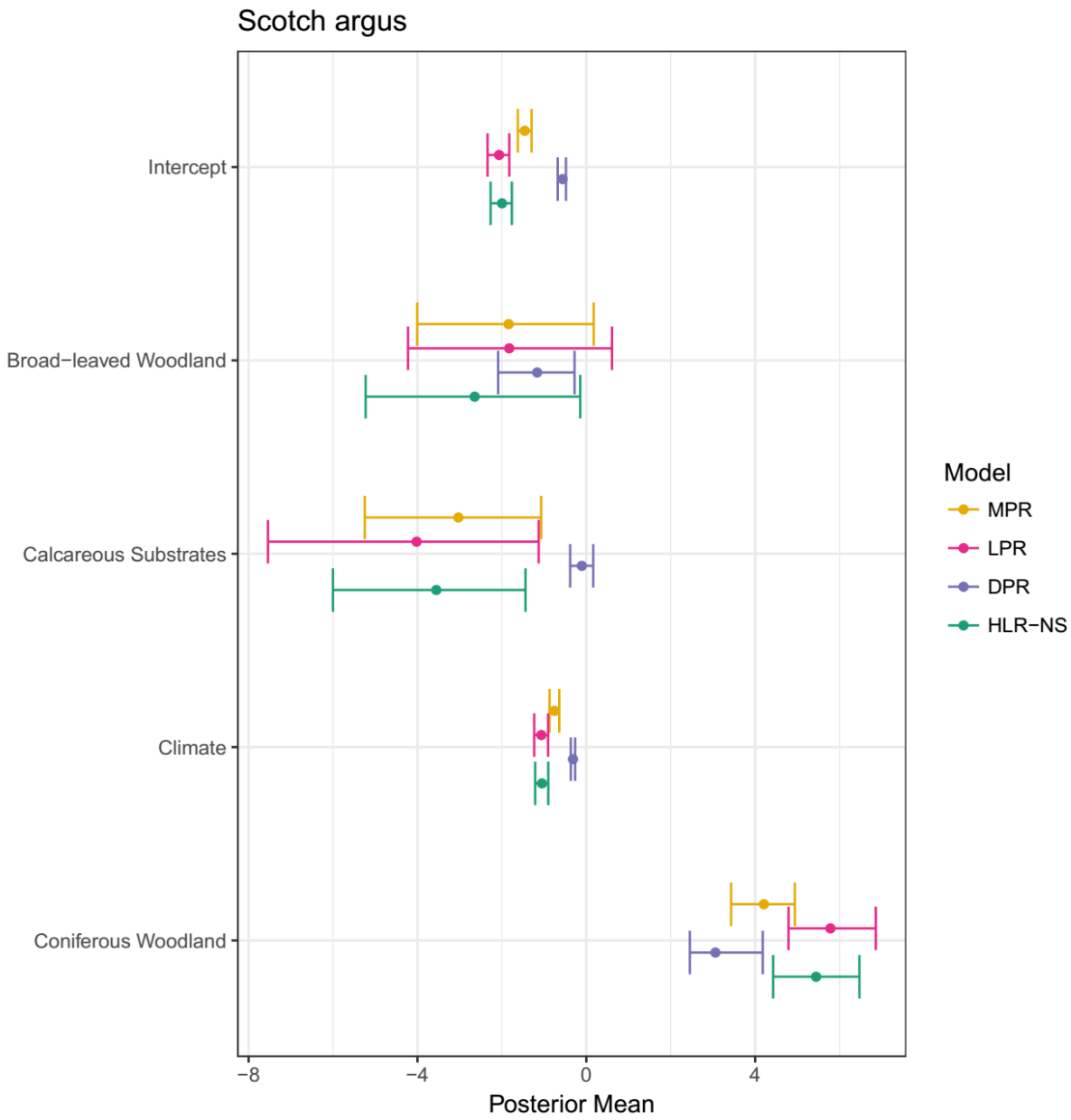


Figure 417: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Scotch argus, *Erebia aethiops*.

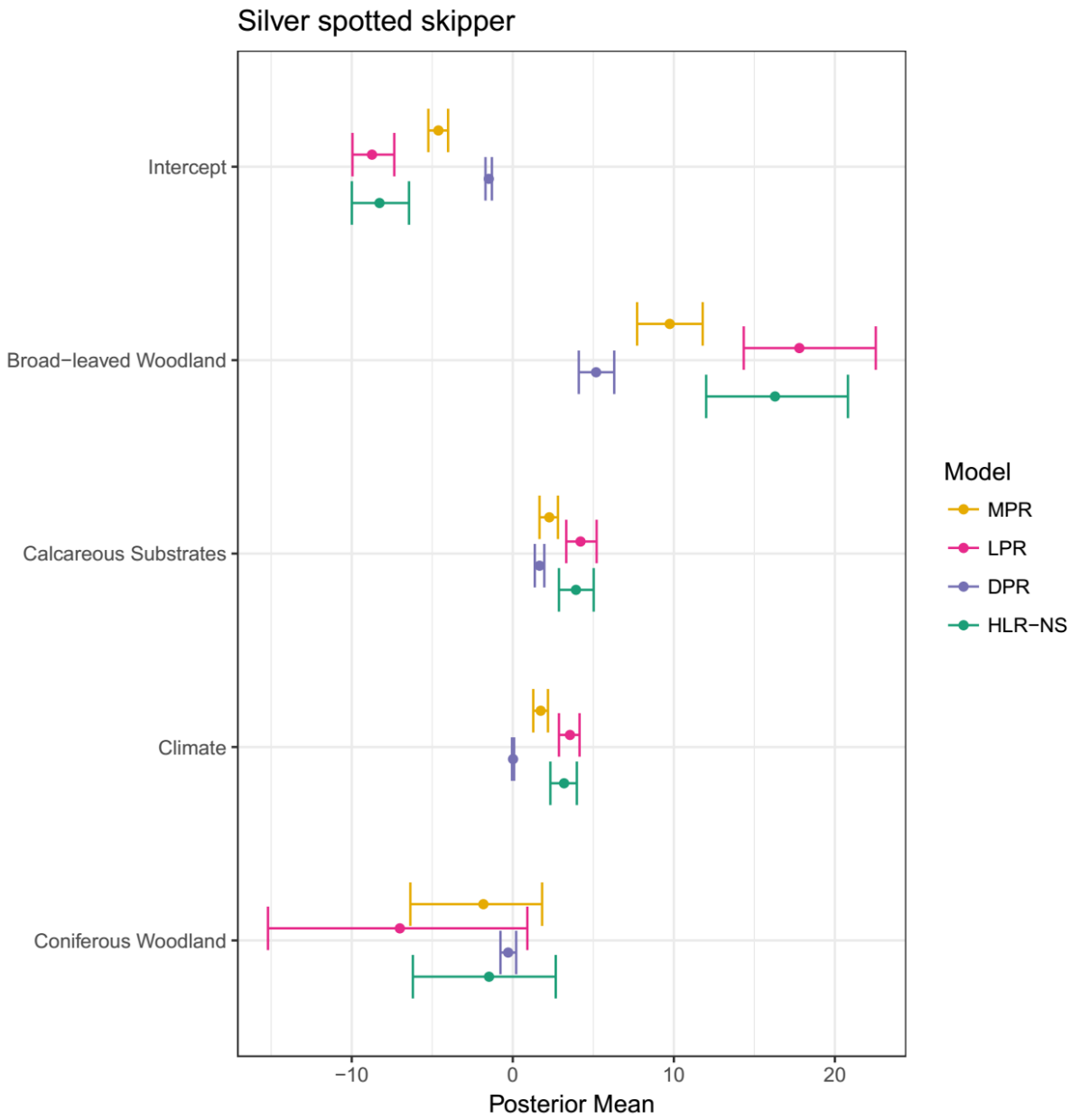


Figure 418: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the silver-spotted skipper, *Epargyreus clarus*.

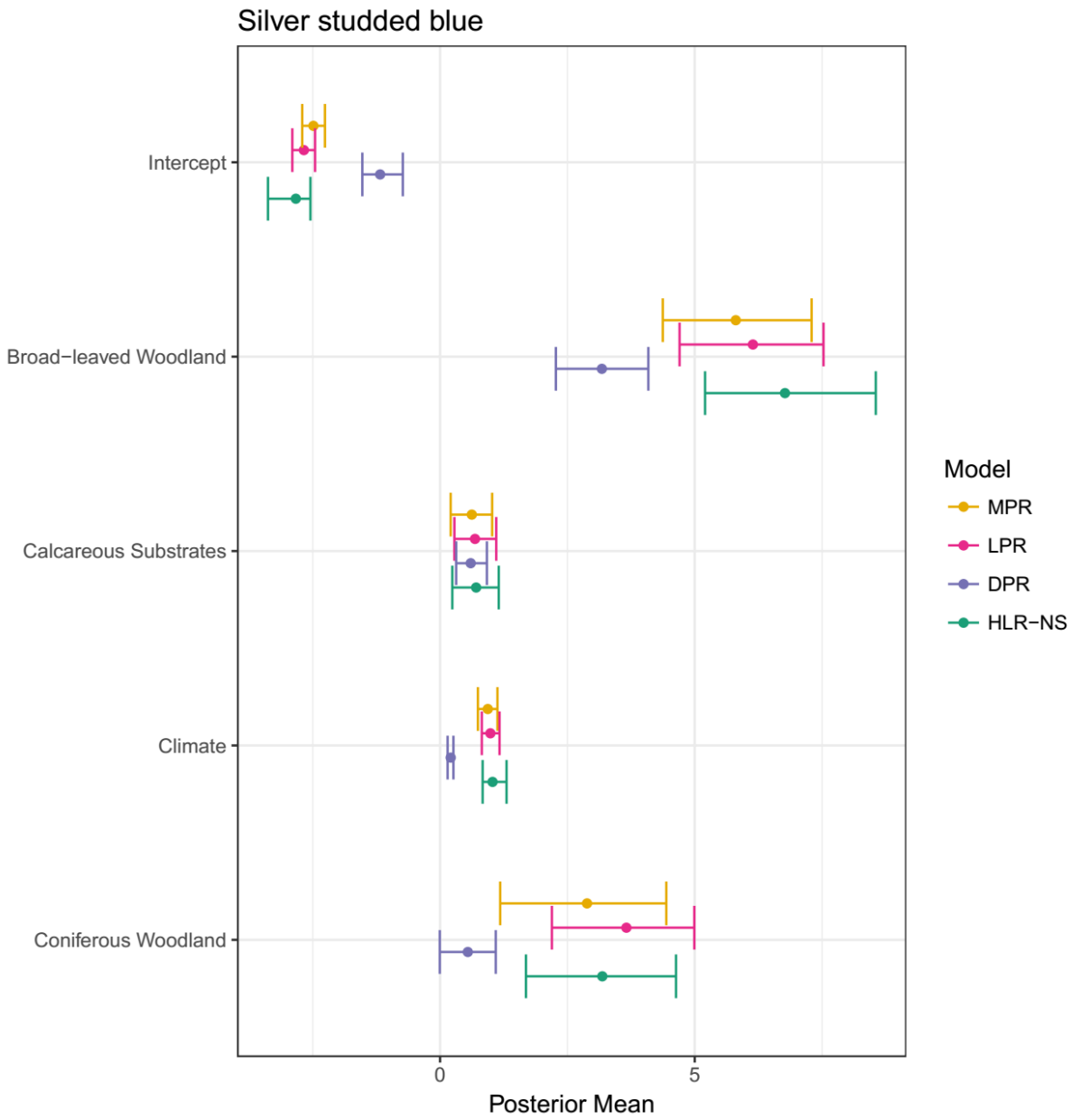


Figure 419: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the silver-studded blue, *Plebejus argus*.

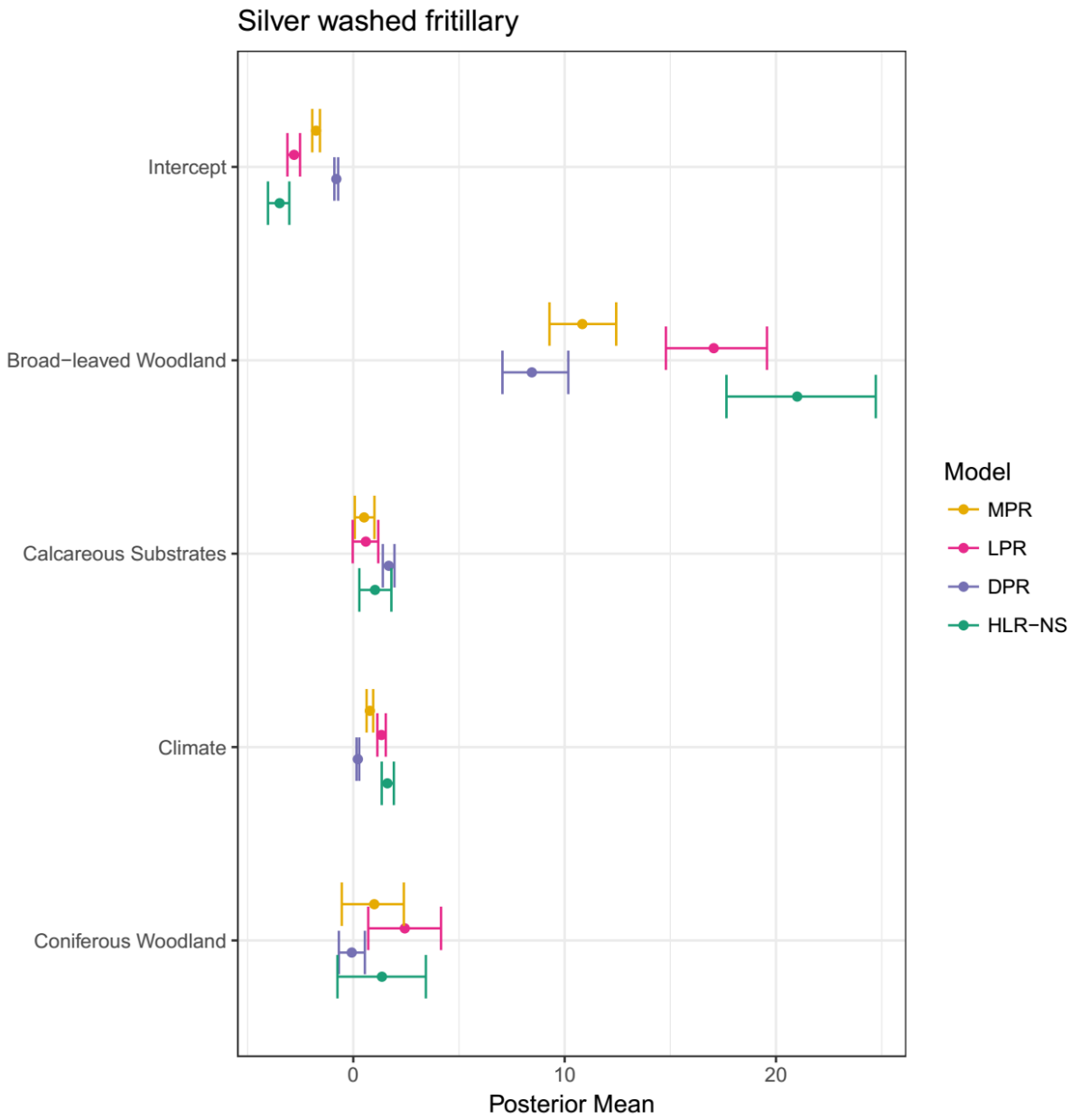


Figure 420: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the silver-washed fritillary, *Argynnis paphia*.

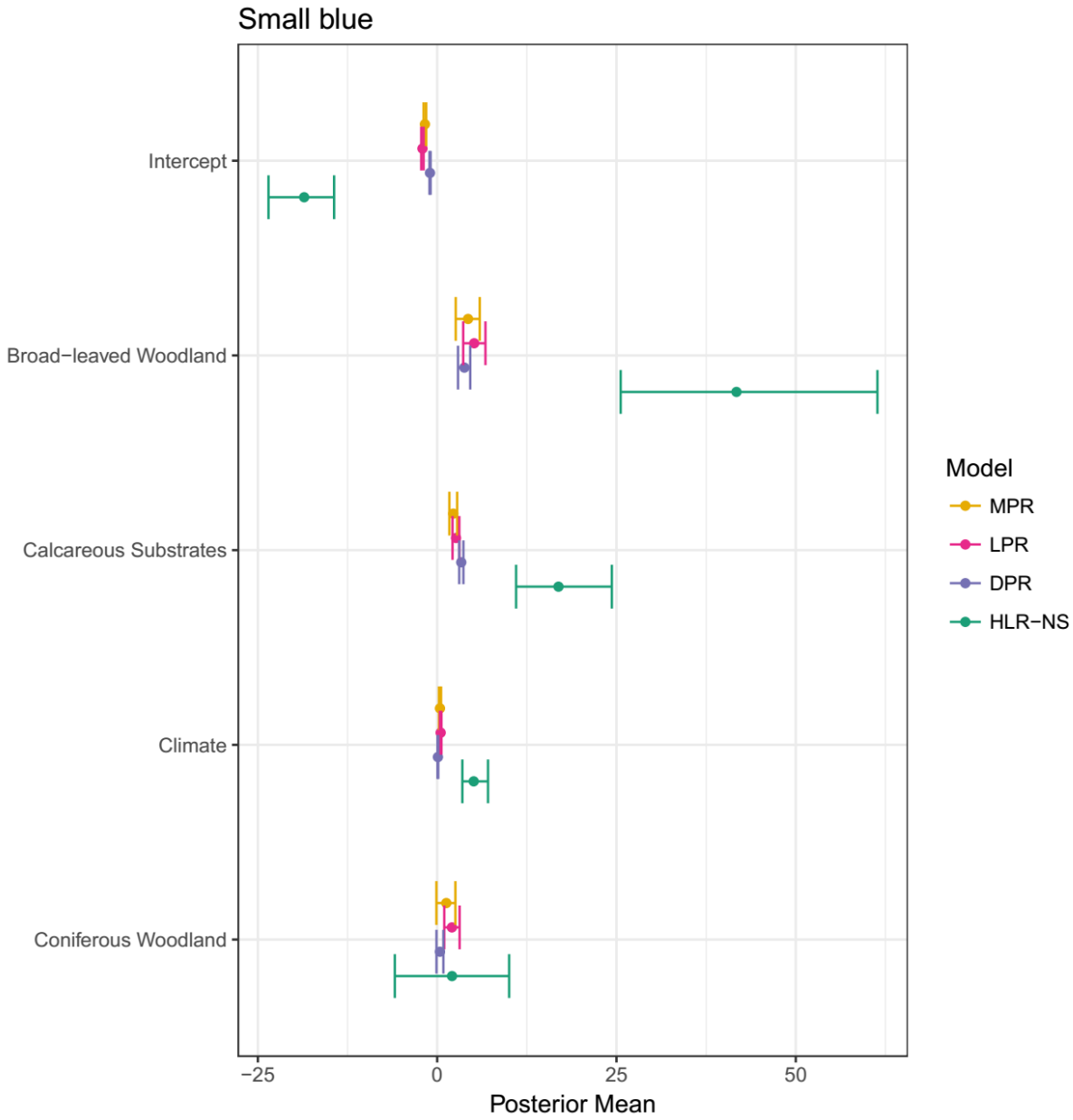


Figure 421: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small blue, *Cupido minimus*.

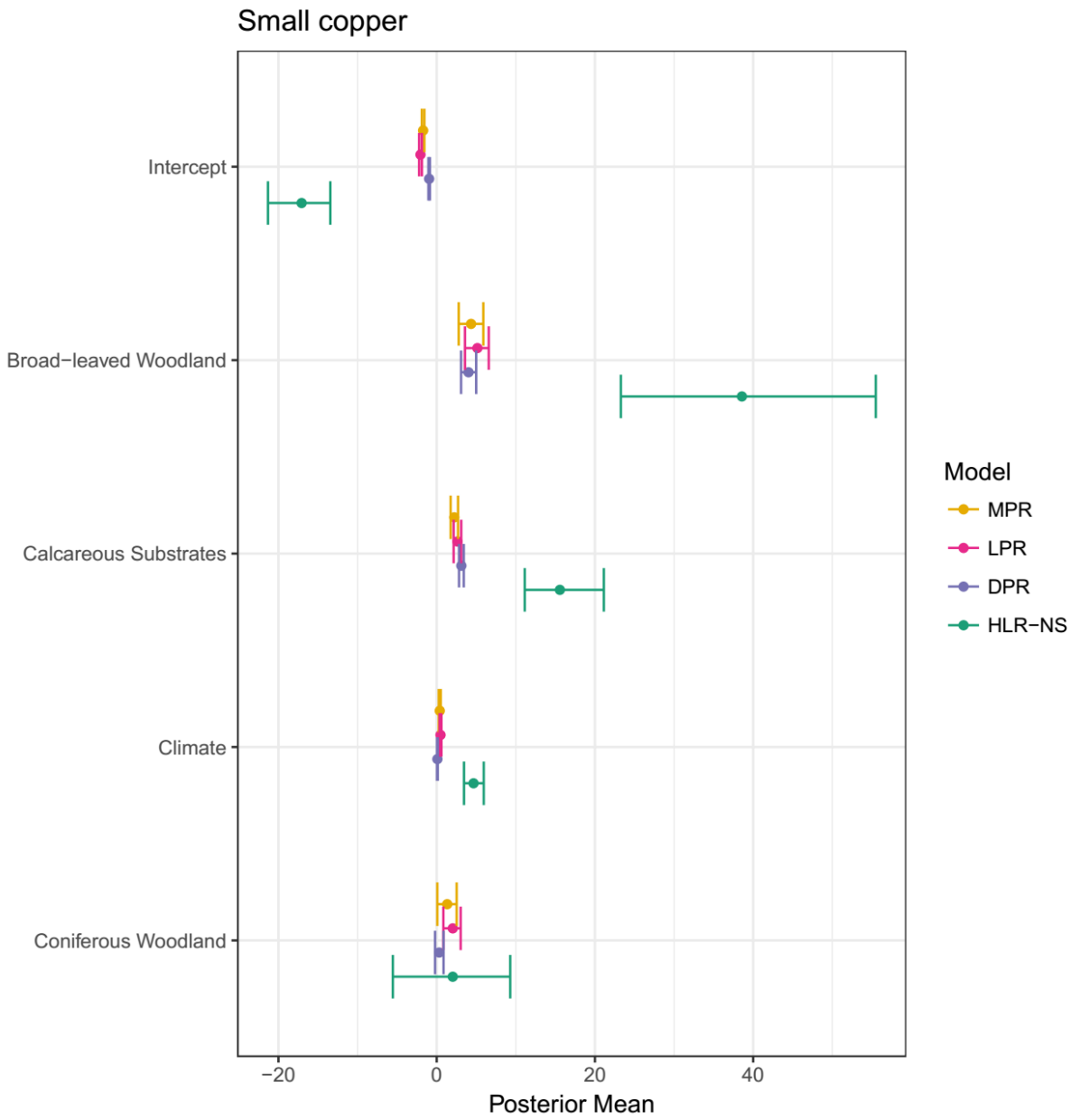


Figure 422: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small copper, *Lycaena phlaeas*.

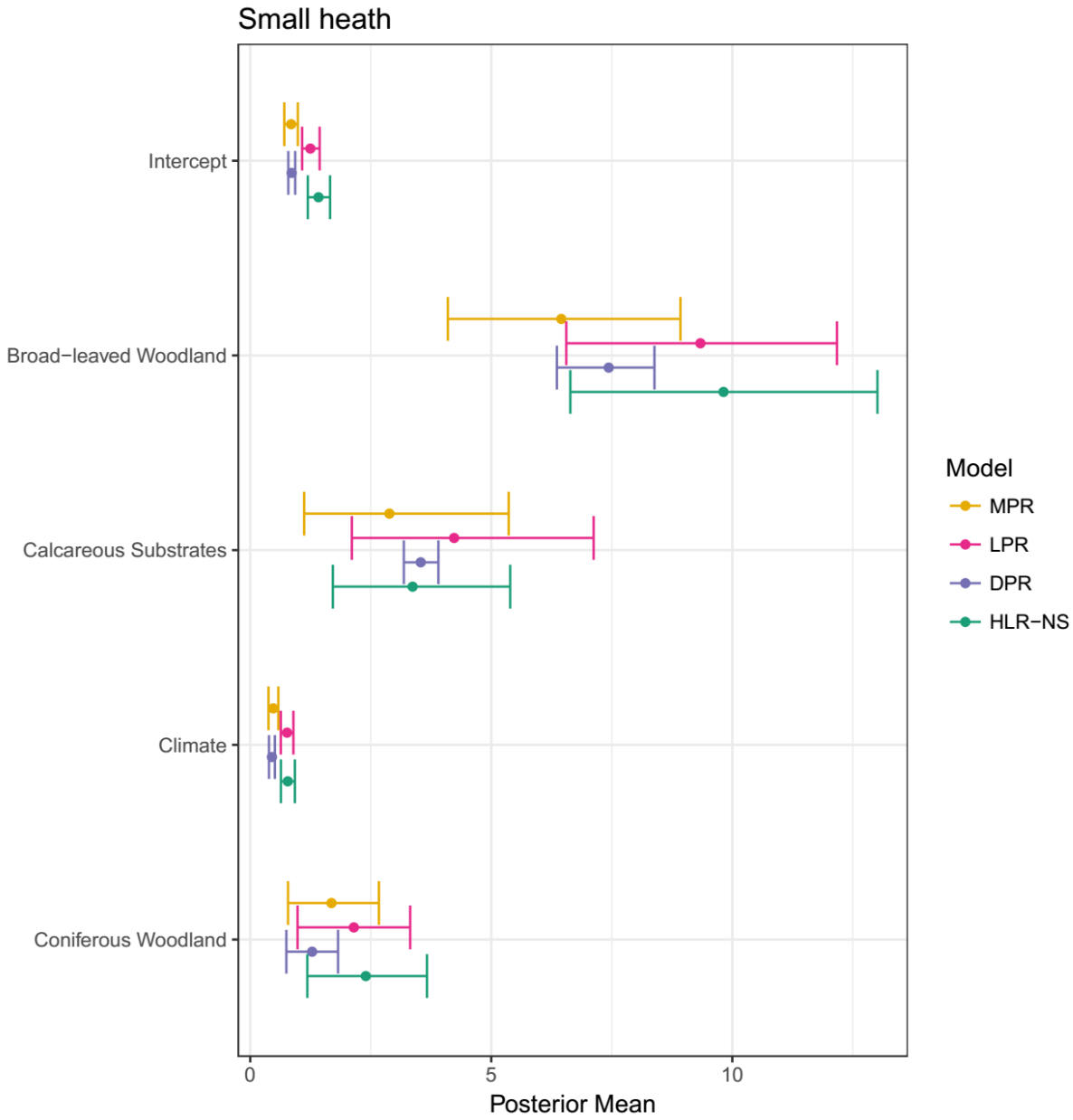


Figure 423: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small heath, *Coenonympha pamphilus*.

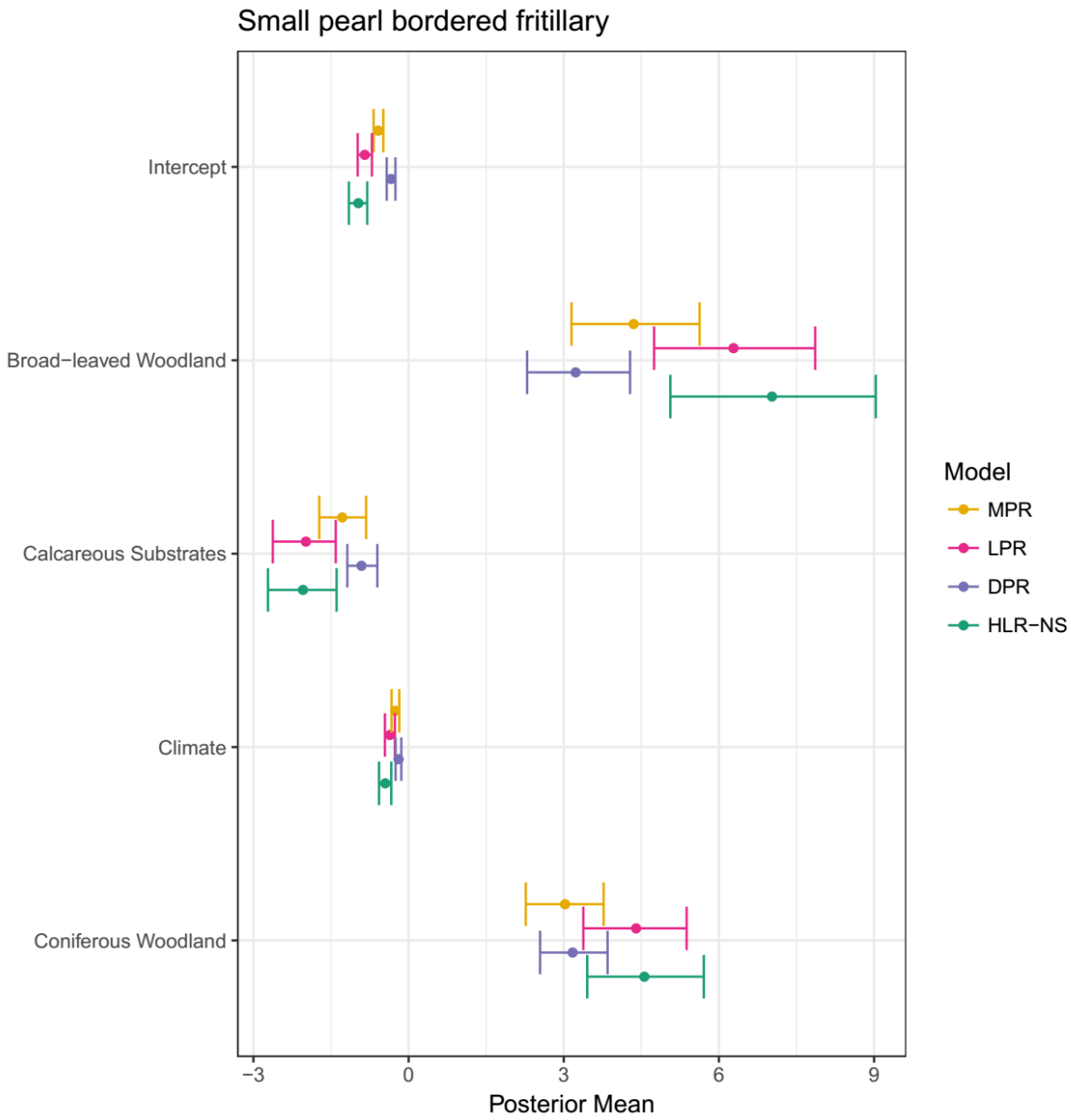


Figure 424: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small pearl-bordered fritillary, *Boloria selene*.

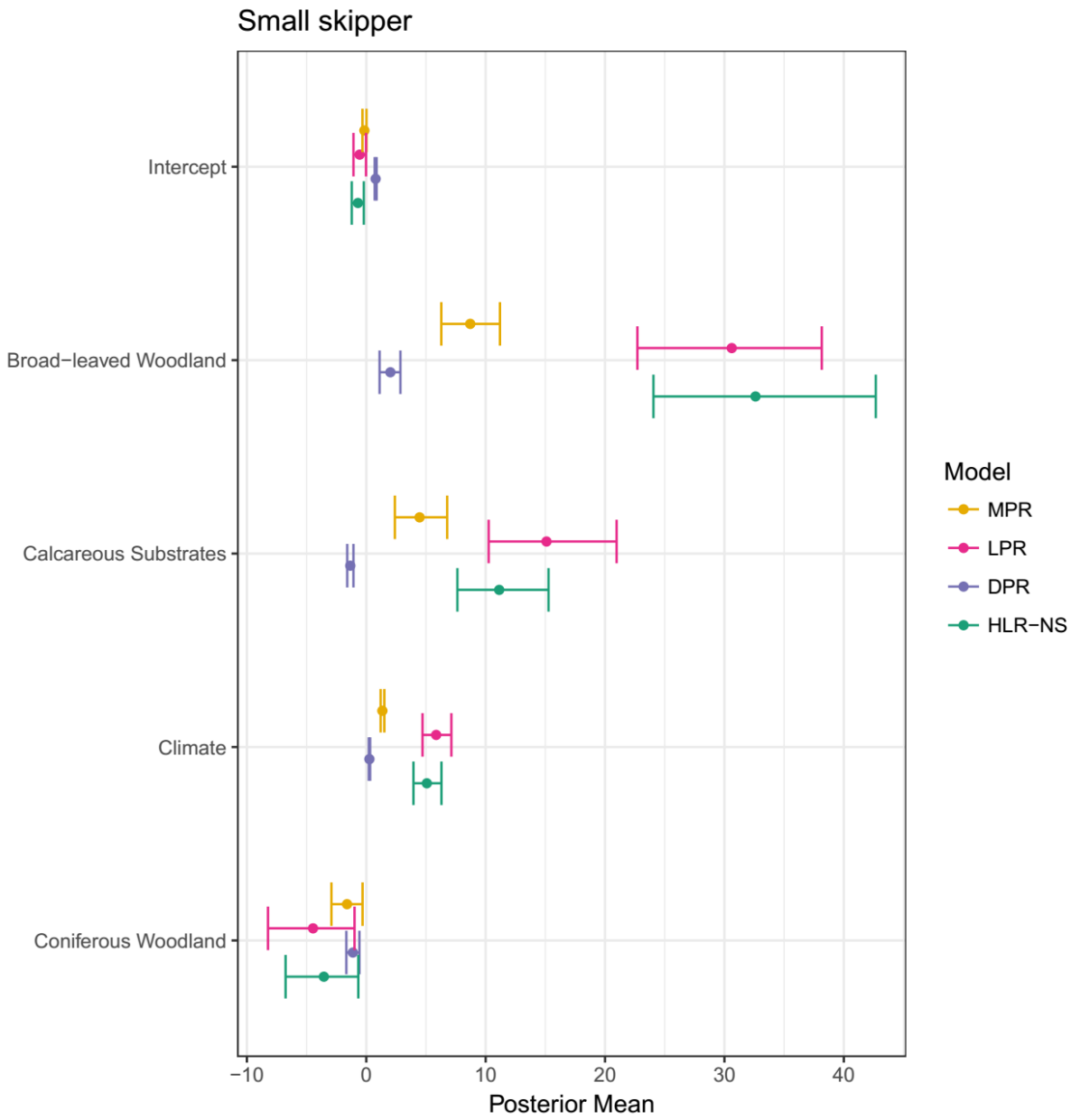


Figure 425: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small skipper, *Thymelicus sylvestris*.

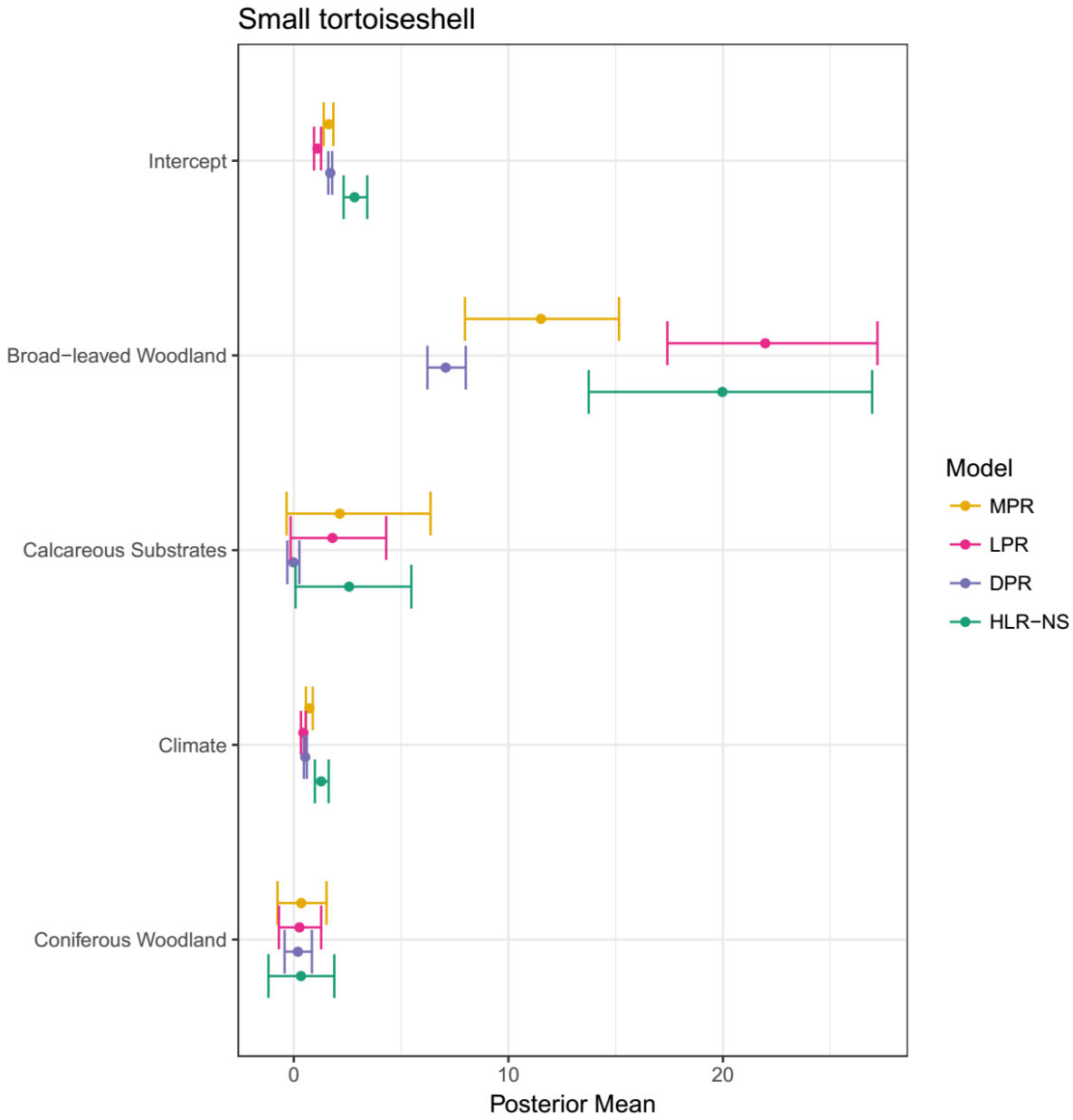


Figure 426: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small tortoiseshell, *Aglais urticae*.

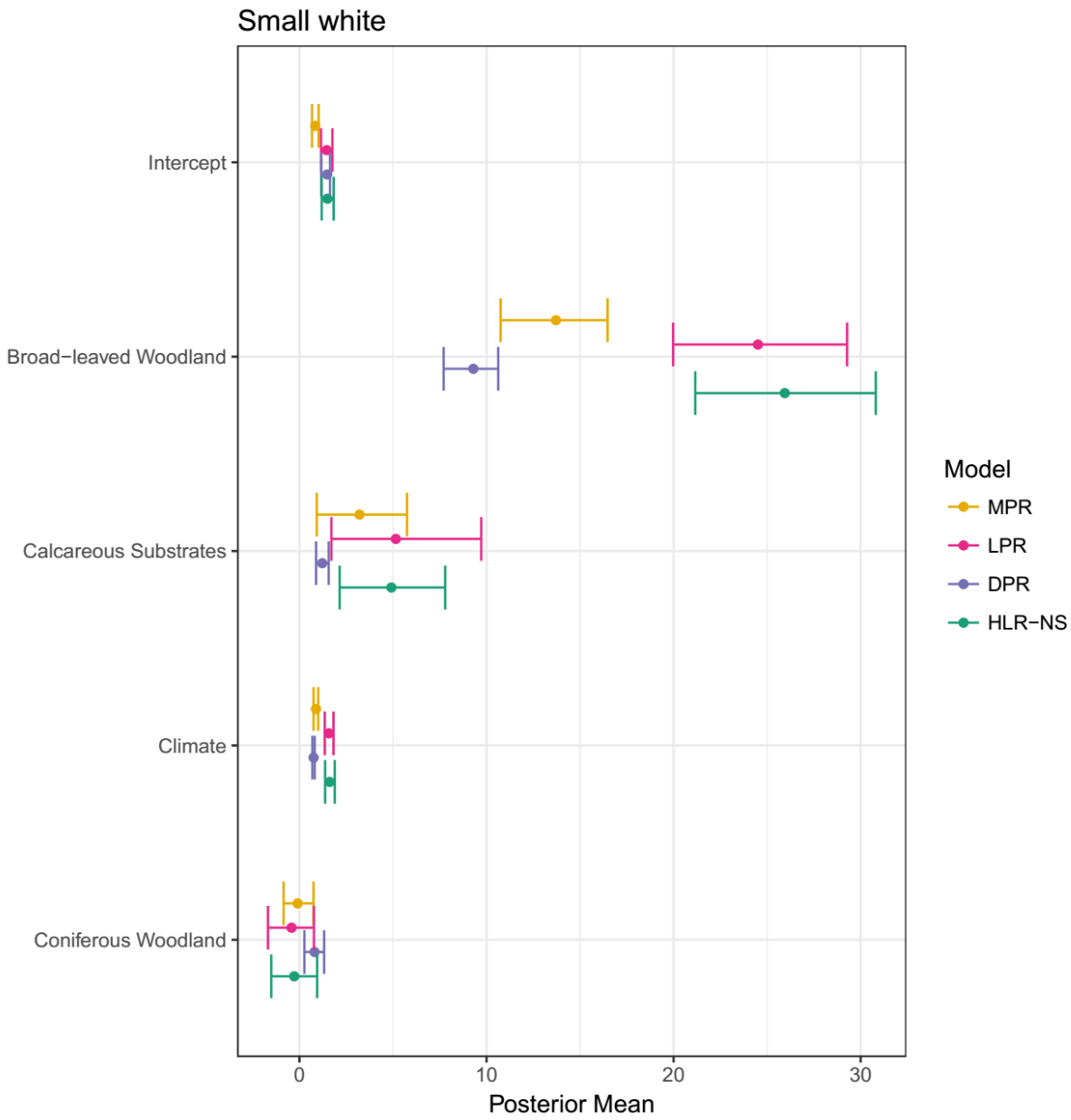


Figure 427: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the small white, *Pieris rapae*.

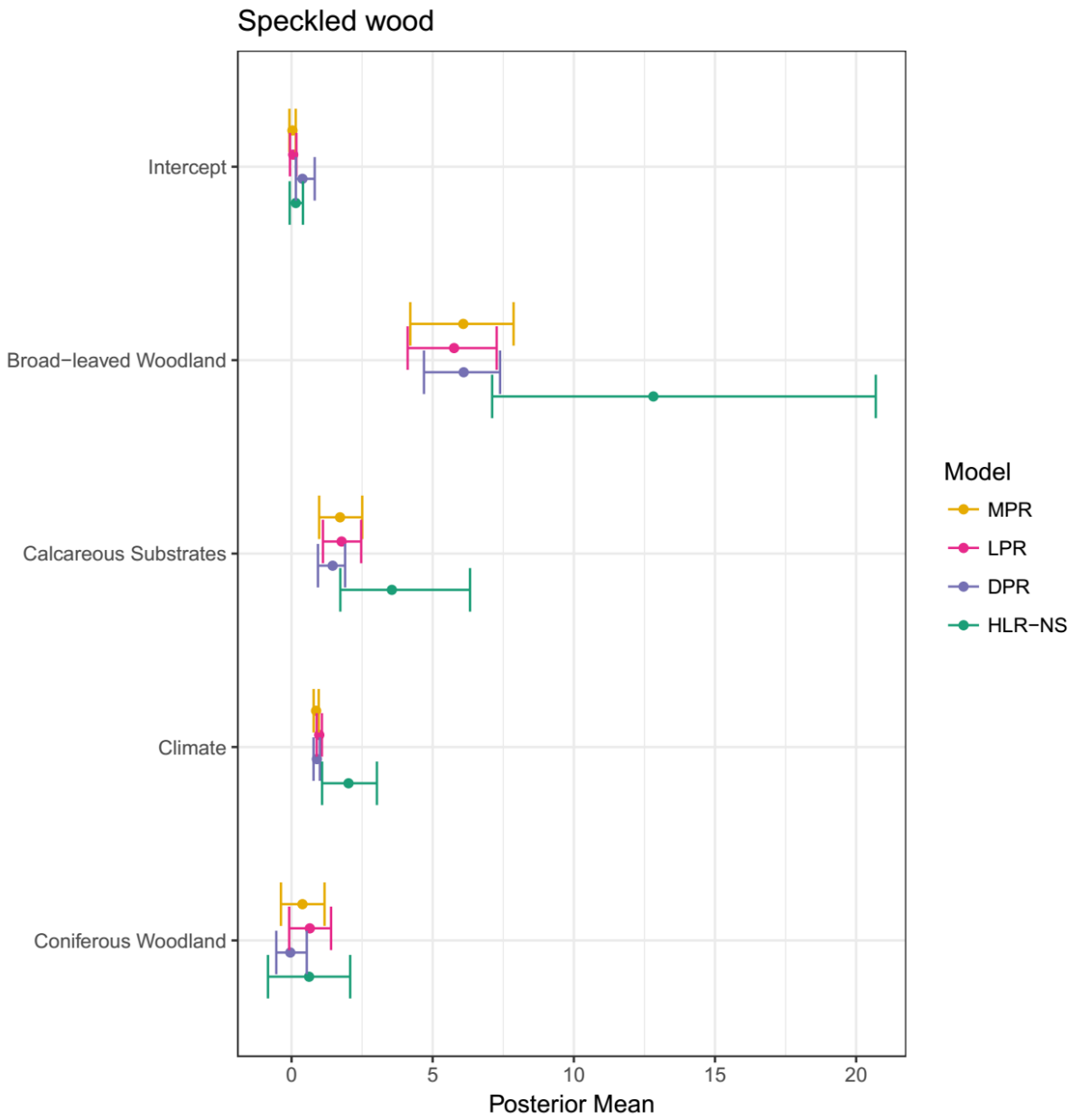


Figure 428: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the speckled wood, *Pararge aegeria*.

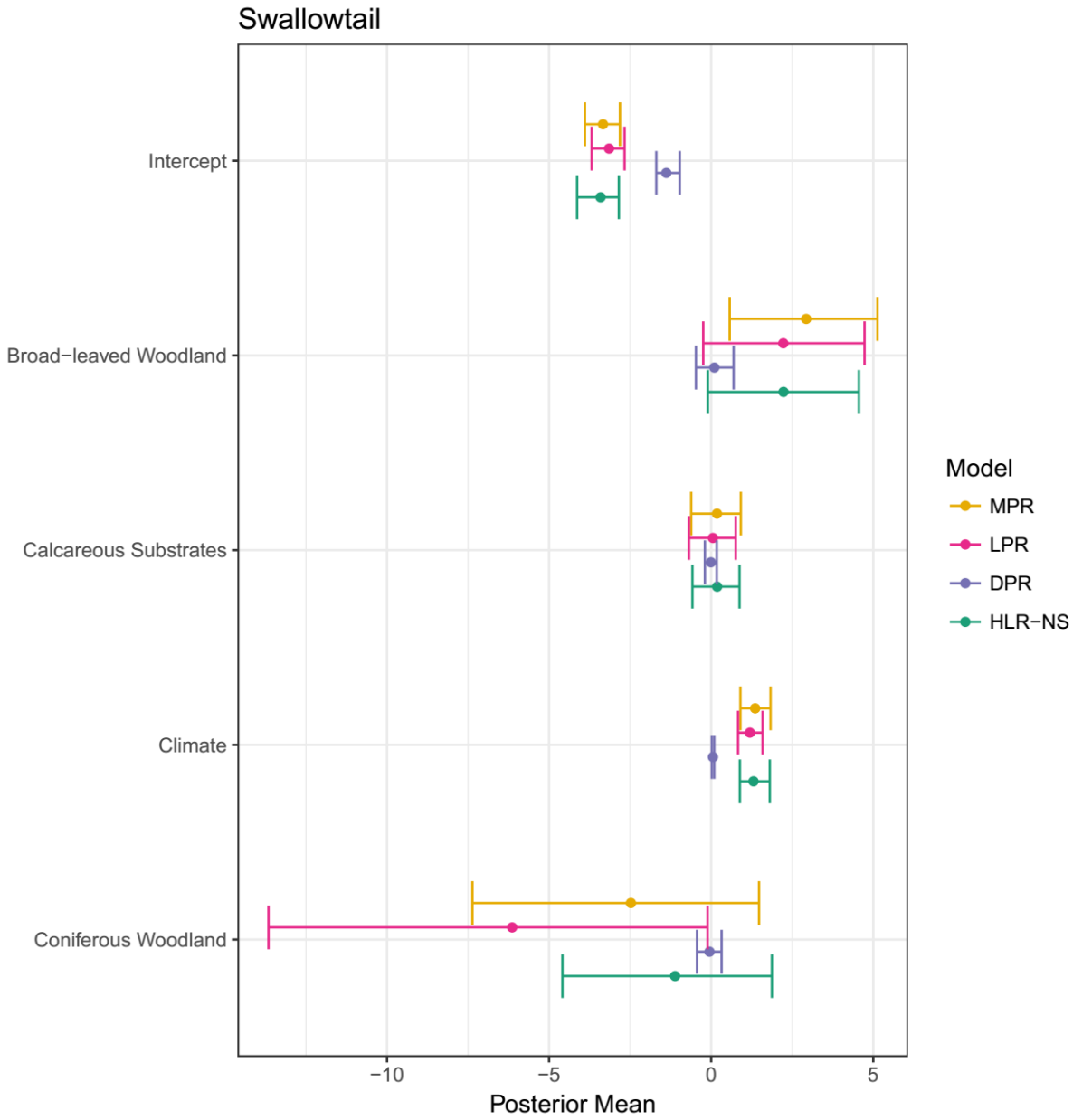


Figure 429: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for swallowtail butterflies, *Papilionidae*.

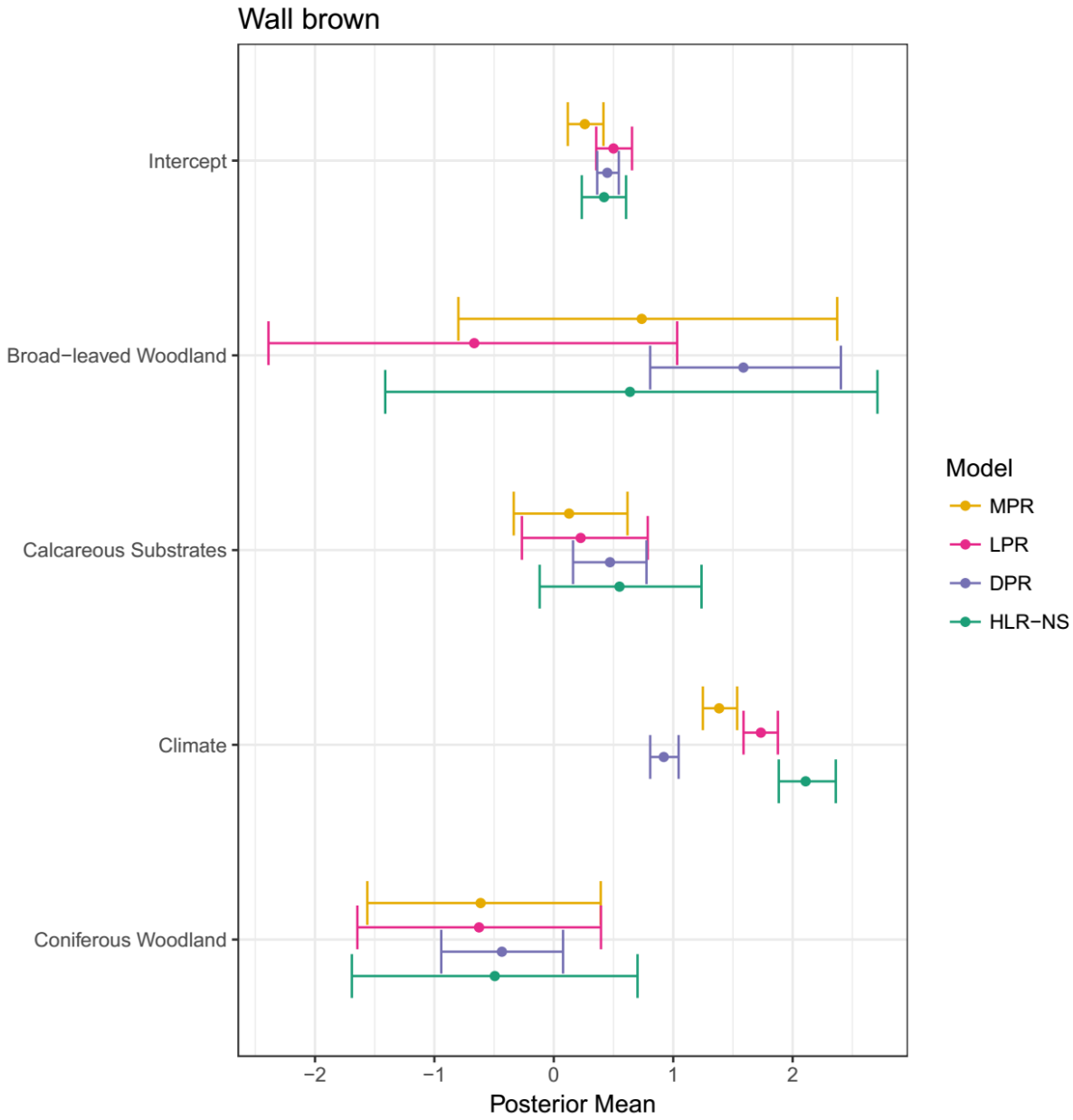


Figure 430: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the wall brown, *Lasiommata megera*.

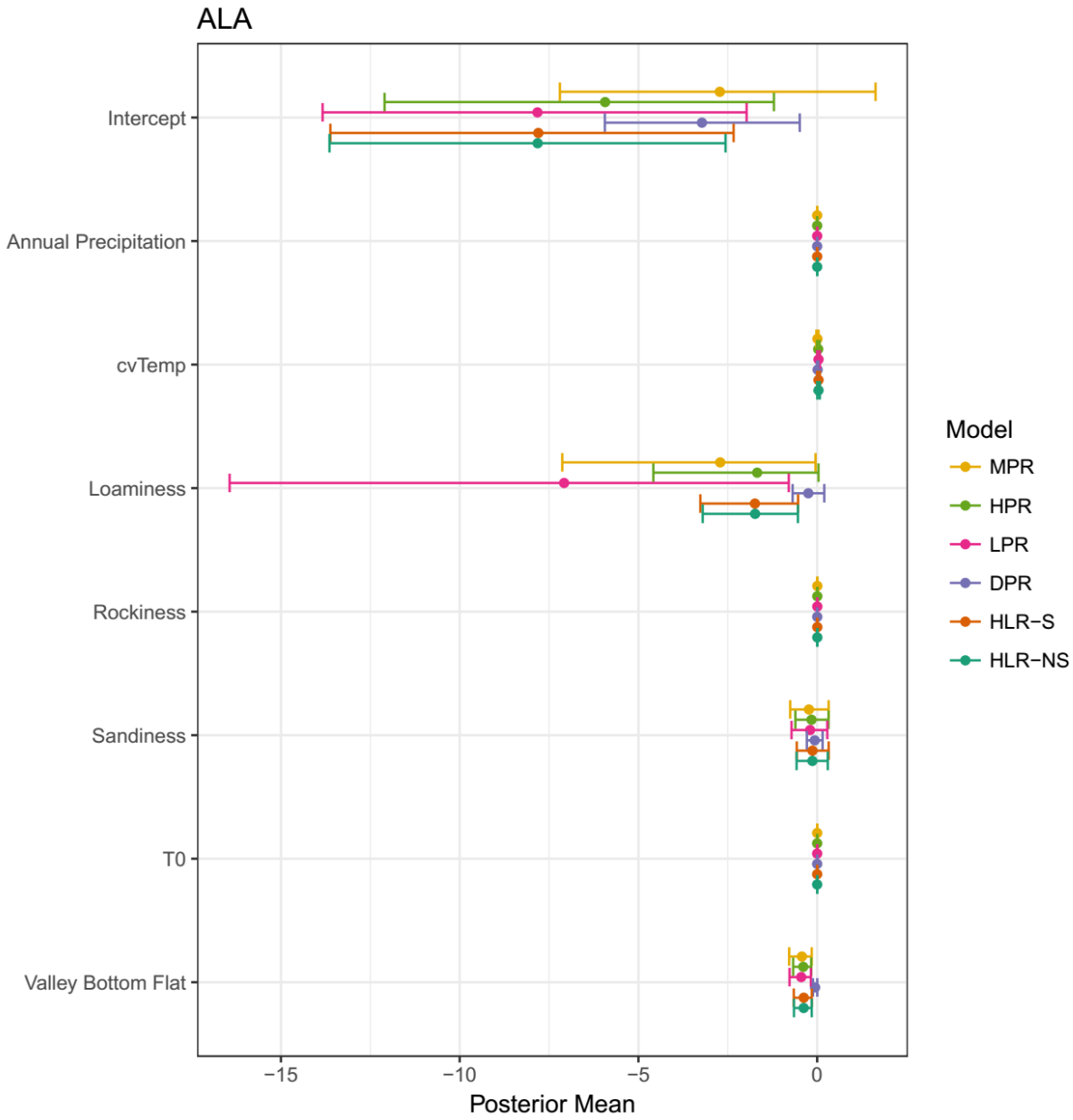


Figure 431: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Grampians gum, *Eucalyptus alaticaulis*.

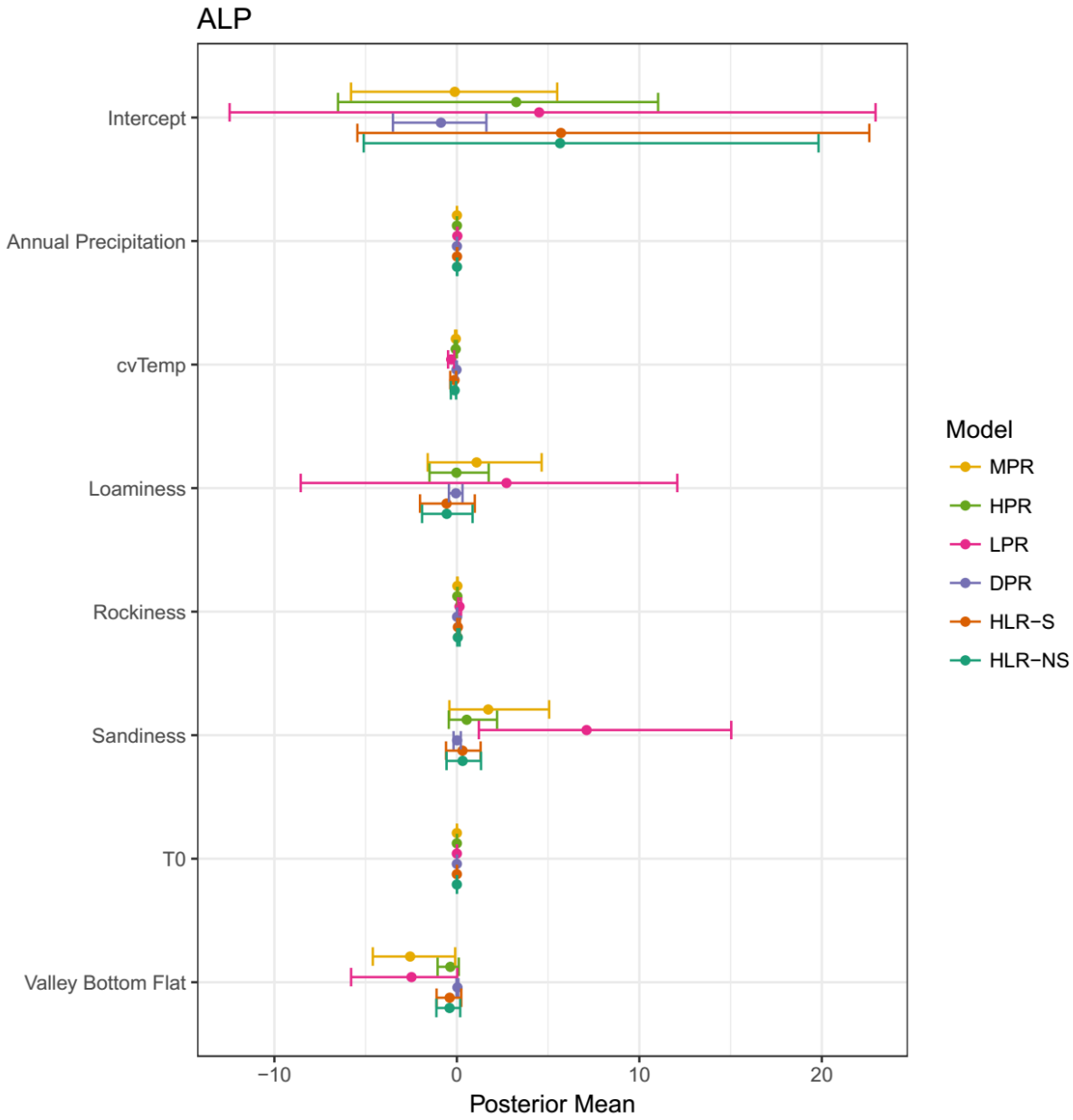


Figure 432: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Serra Range stringybark, *Eucalyptus sarraensis*, Mt. Abrupt stringybark, *Eucalyptus verrucata*, and Victoria Range stringybark, *Eucalyptus victoriana*.

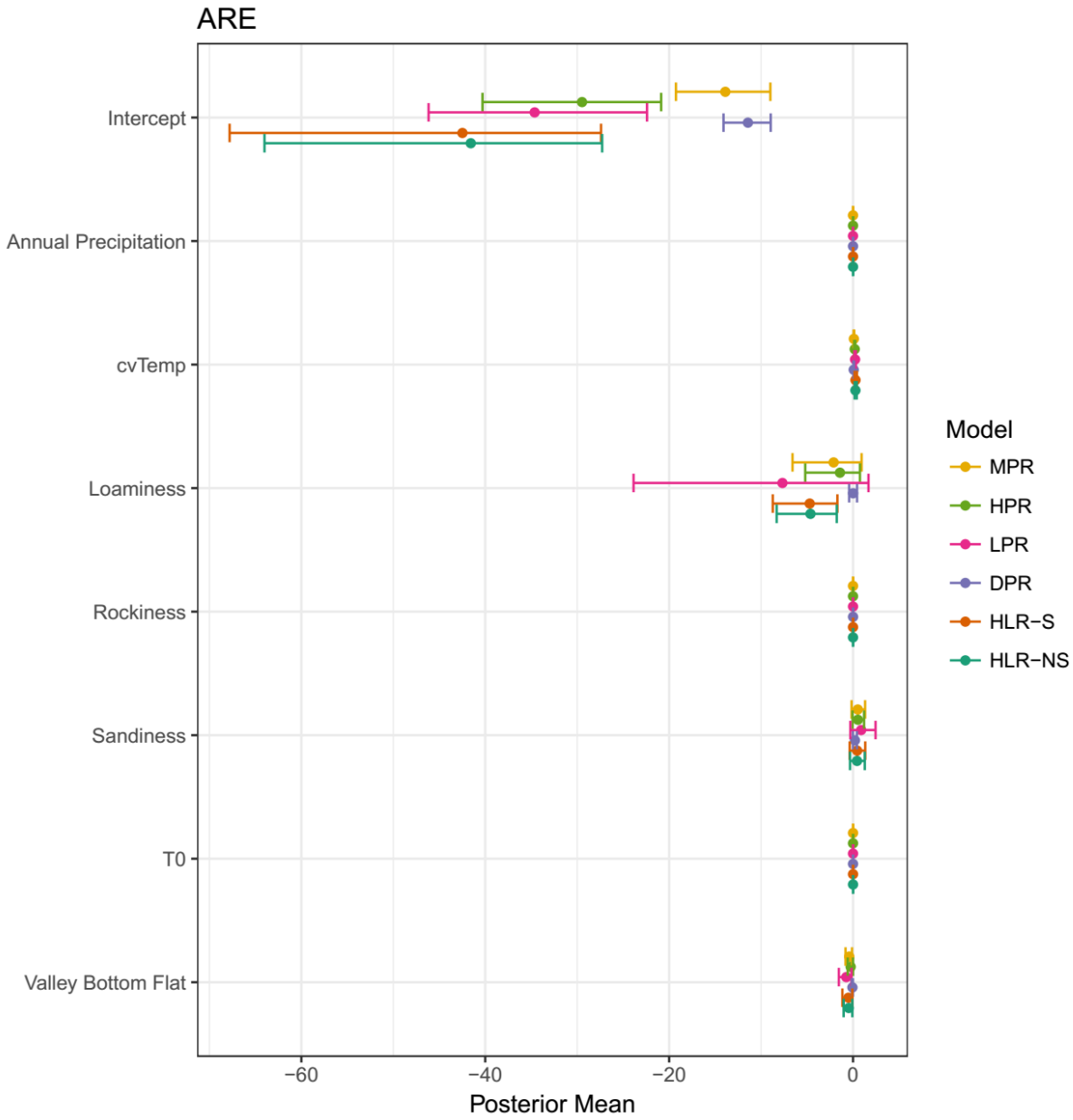


Figure 433: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the sand stringybark, *Eucalyptus arenacea*.

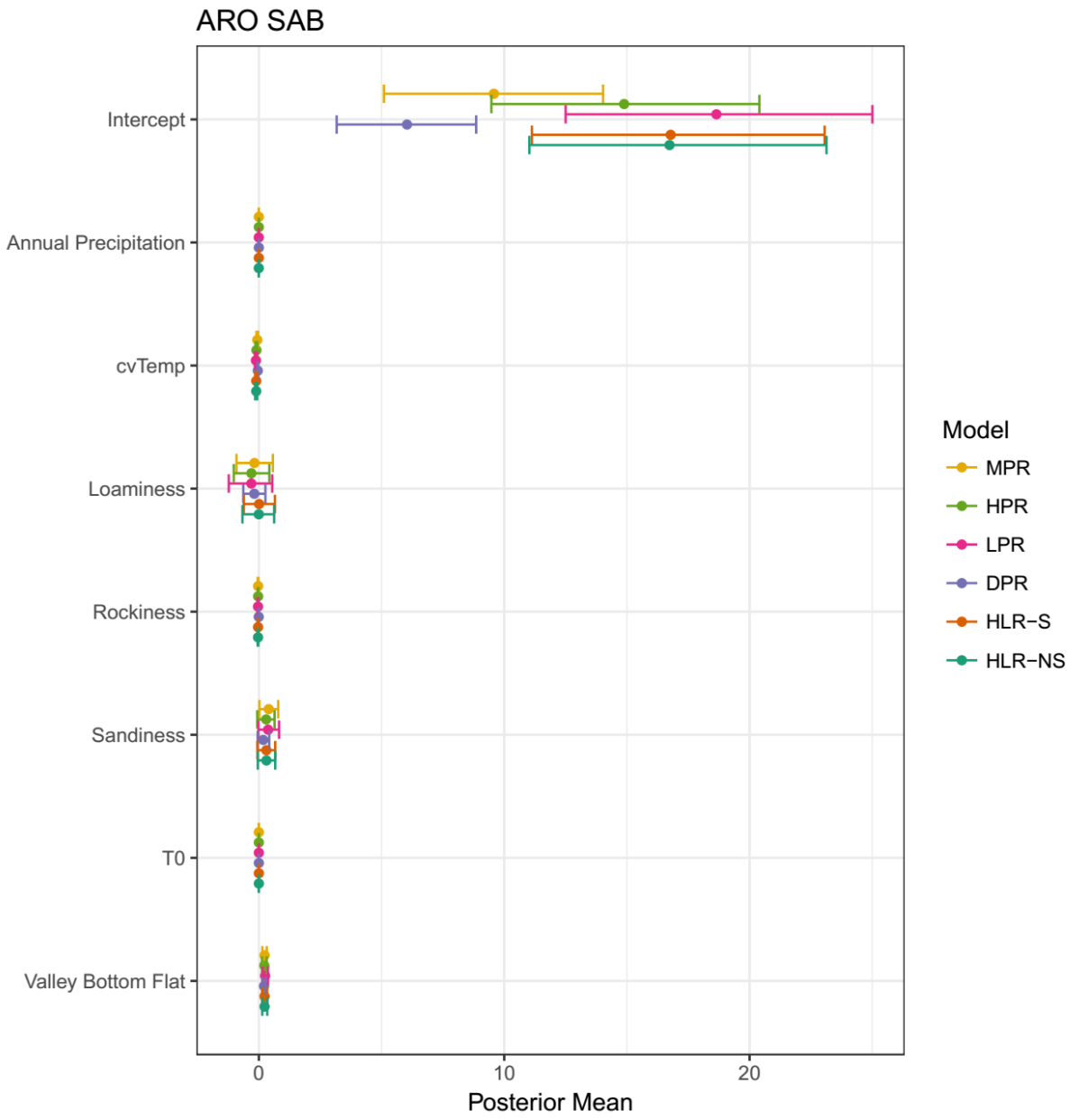


Figure 434: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Creswick apple-box, *Eucalyptus aromaploia*, and western scent-bark, *Eucalyptus sabulosa*.

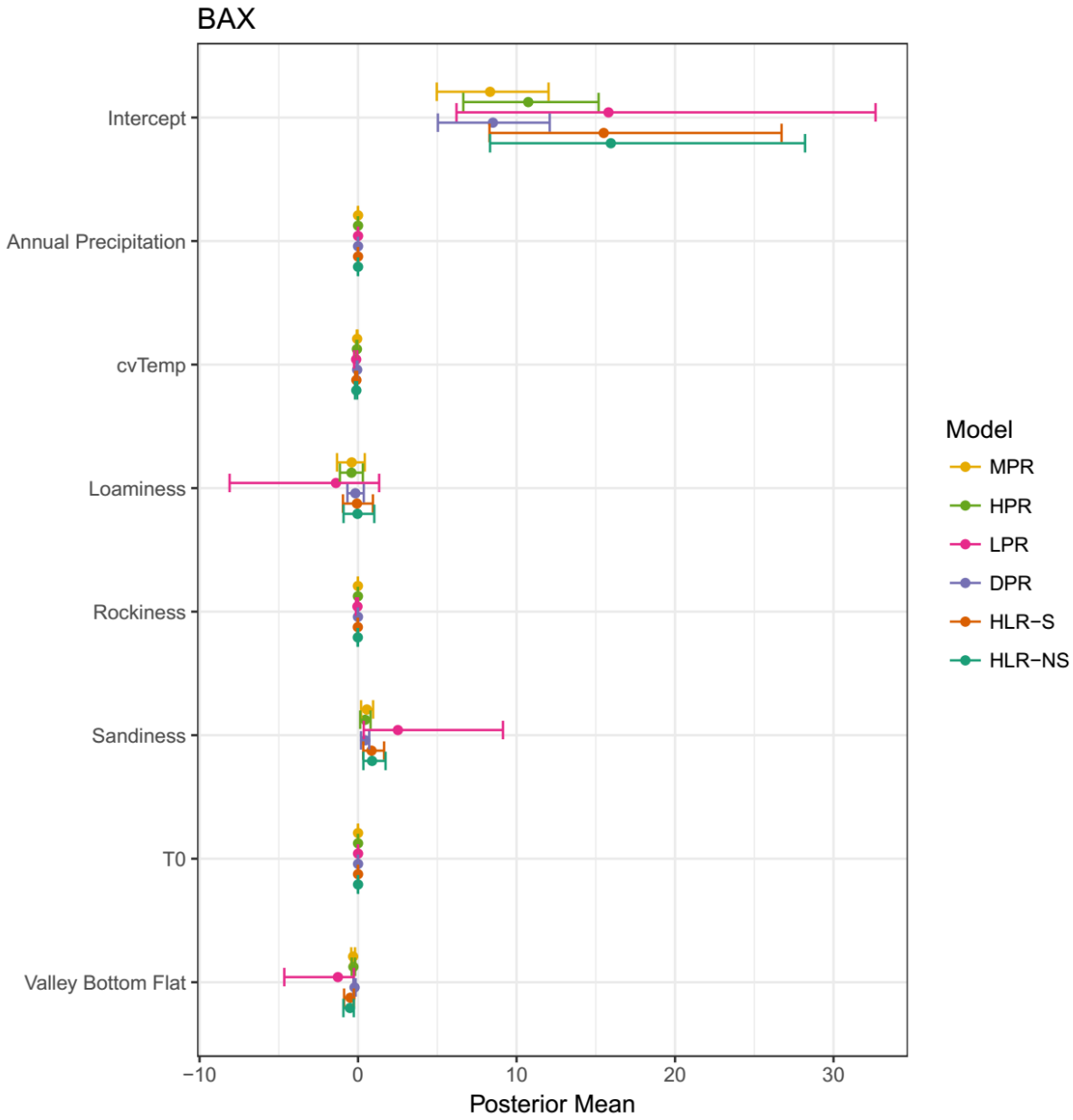


Figure 435: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the brown stringybark, *Eucalyptus baxteri*.

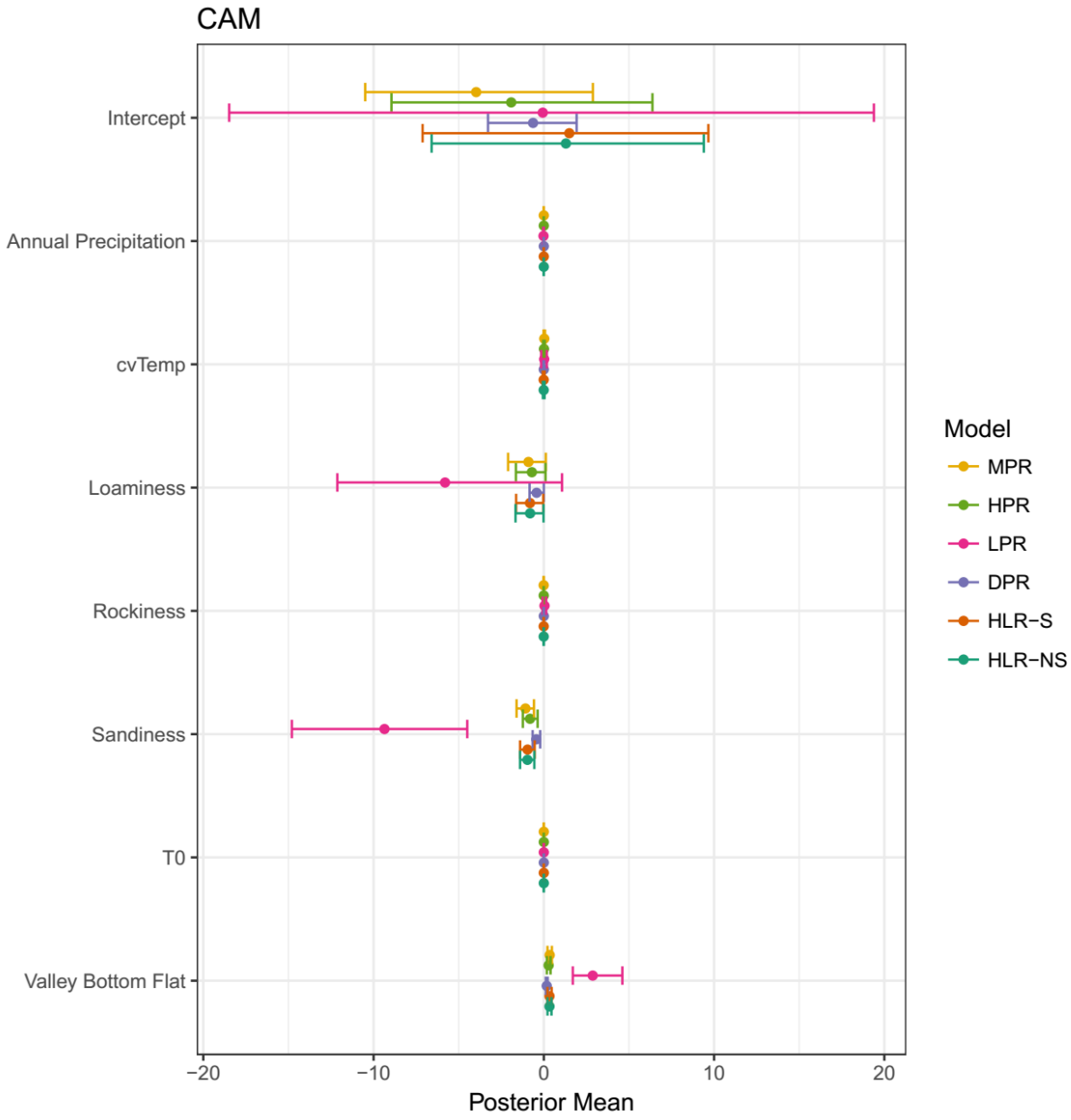


Figure 436: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the river red gum, *Eucalyptus camaldulensis*.

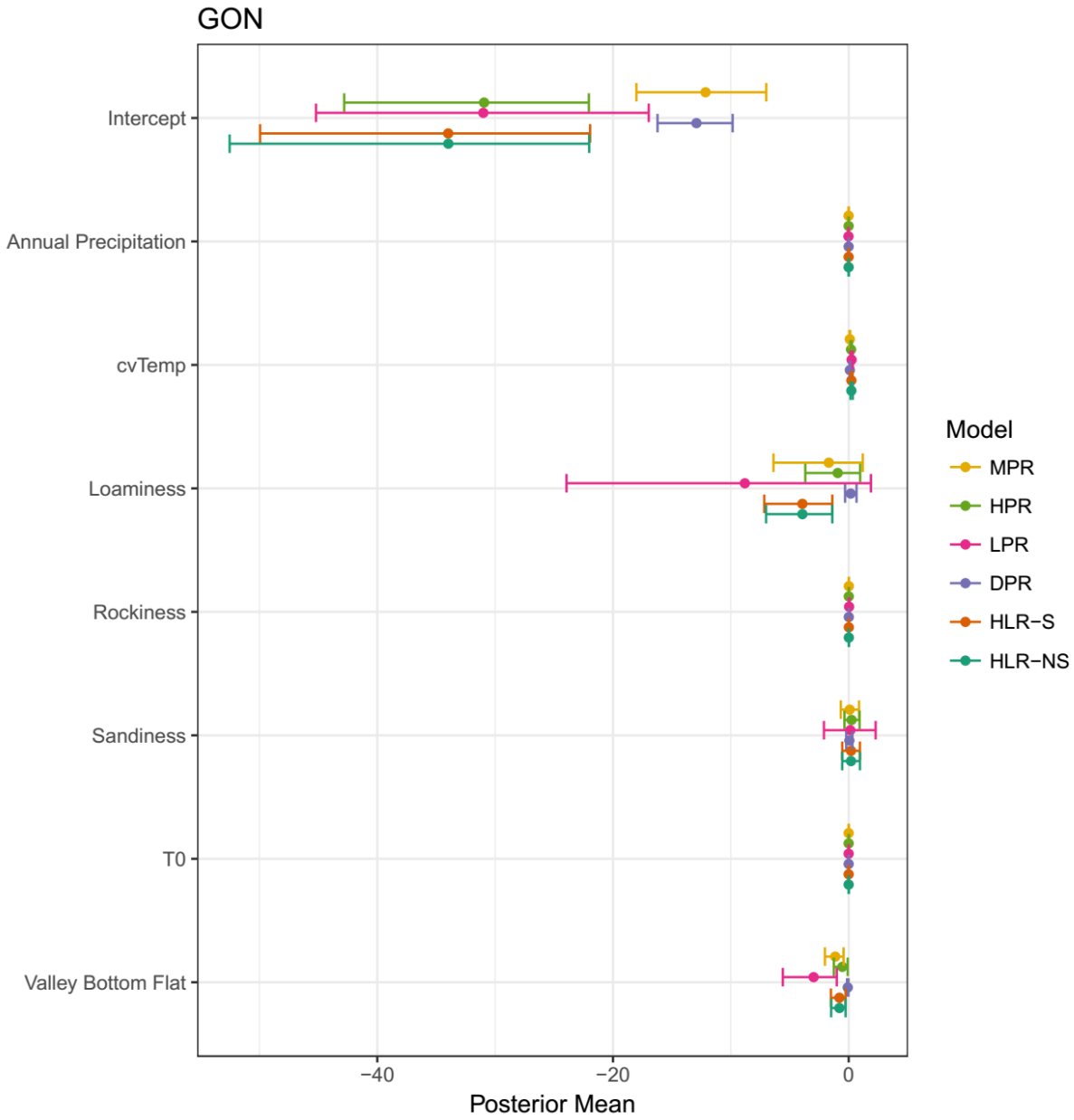


Figure 437: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the long-leaved box, *Eucalyptus goniocalyx*.

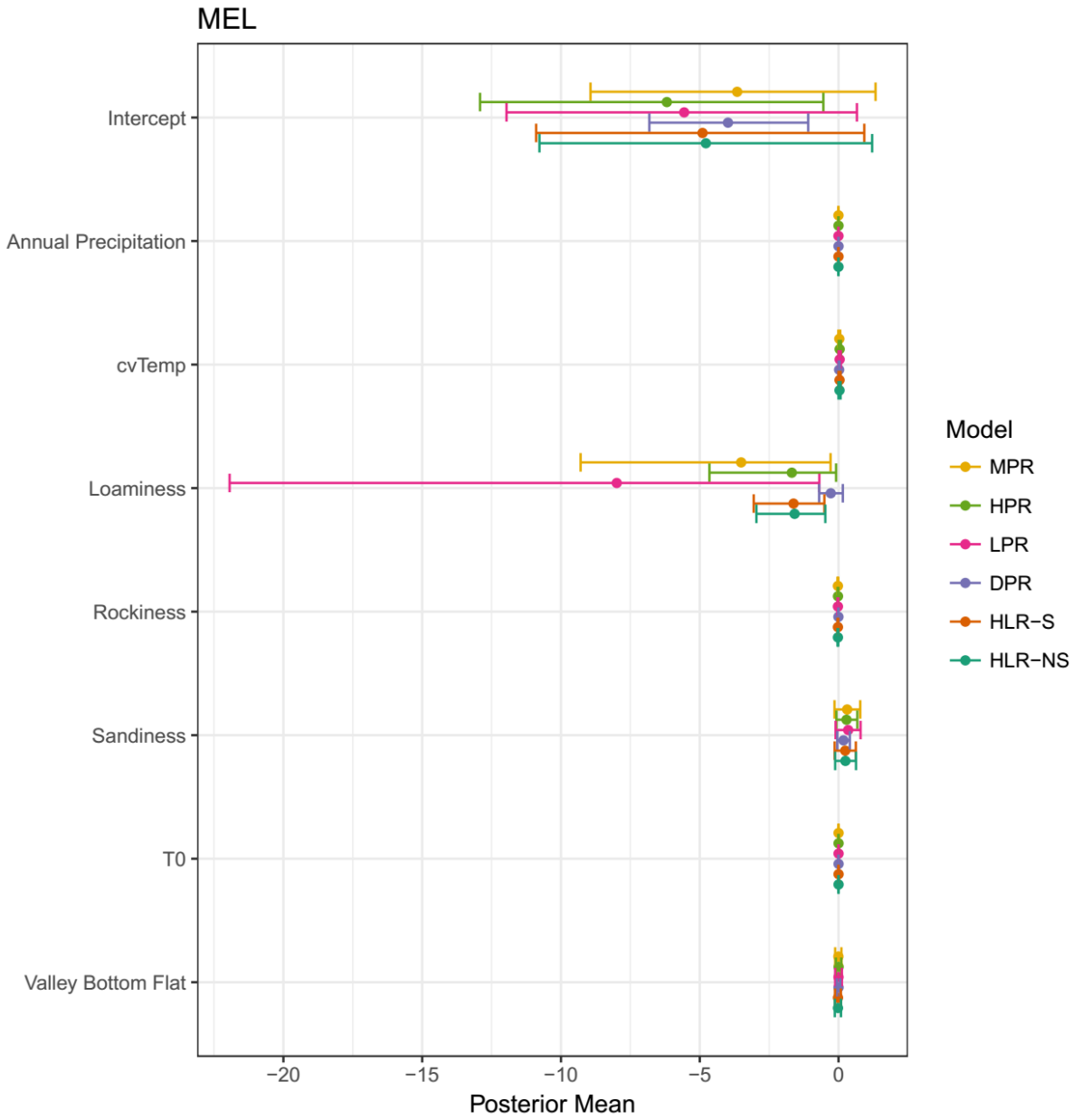


Figure 438: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the yellow box, *Eucalyptus melliodora*.

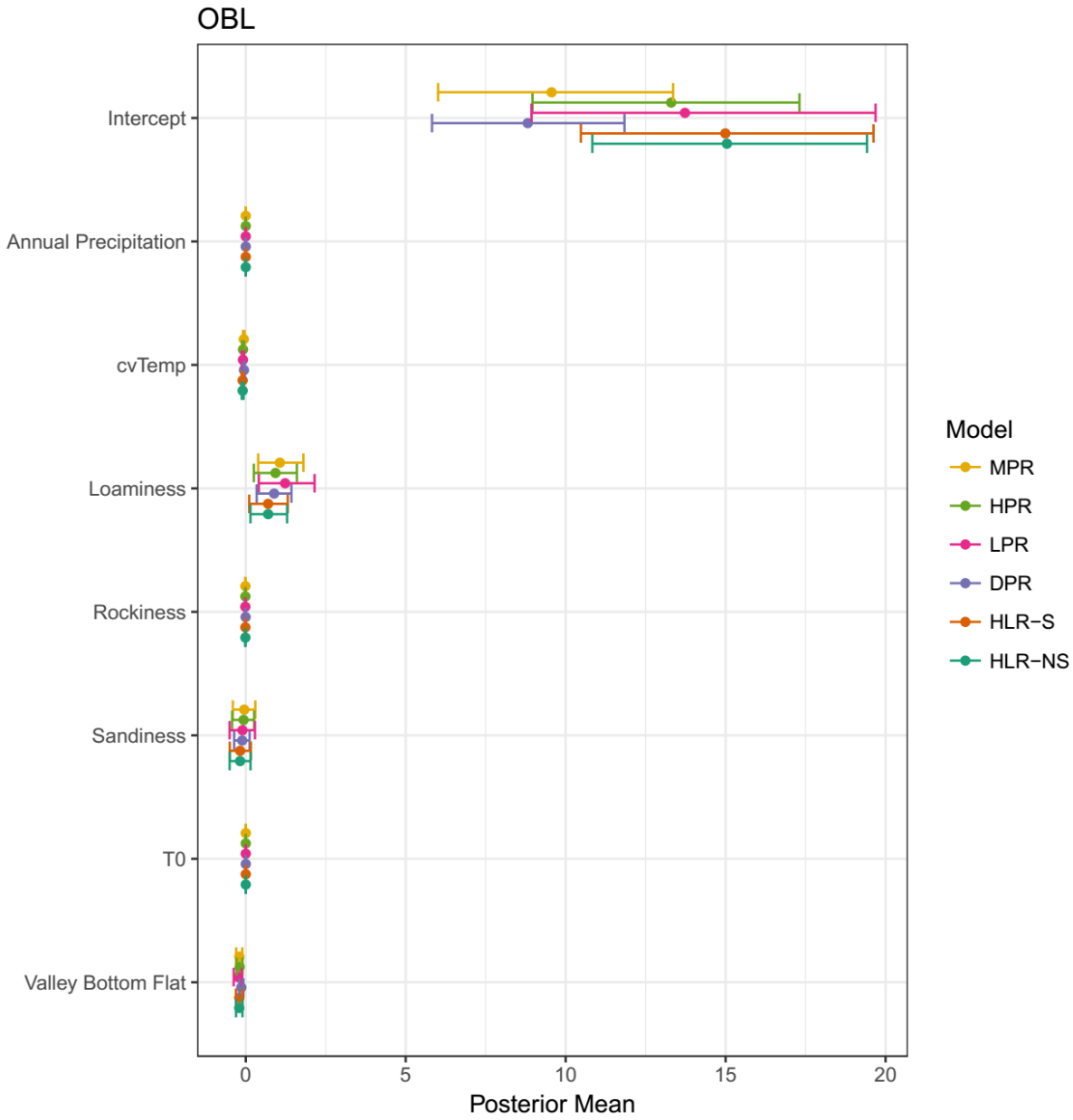


Figure 439: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the messmate stringybark, *Eucalyptus oblique*.

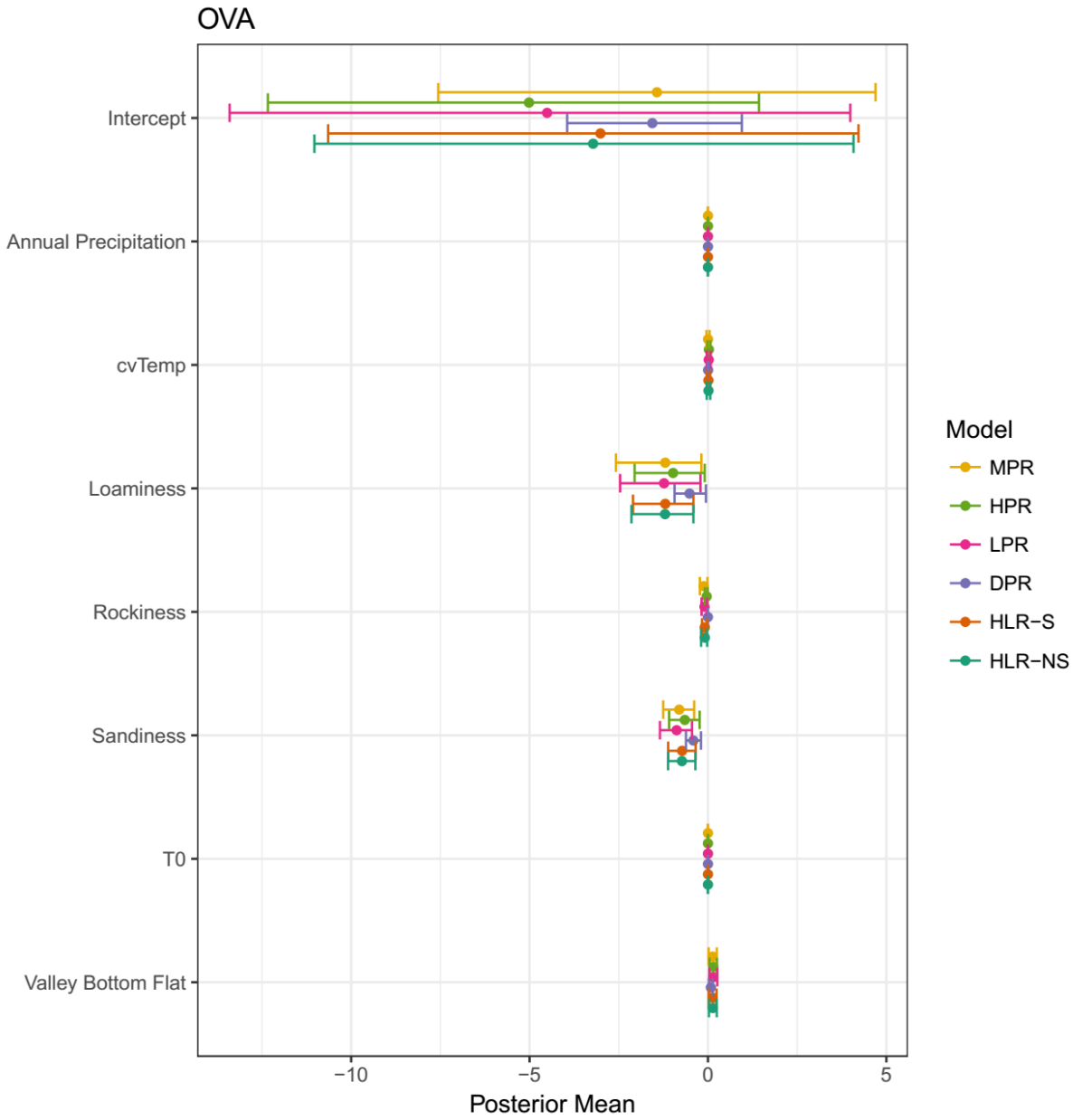


Figure 440: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the swamp gum, *Eucalyptus ovata*.

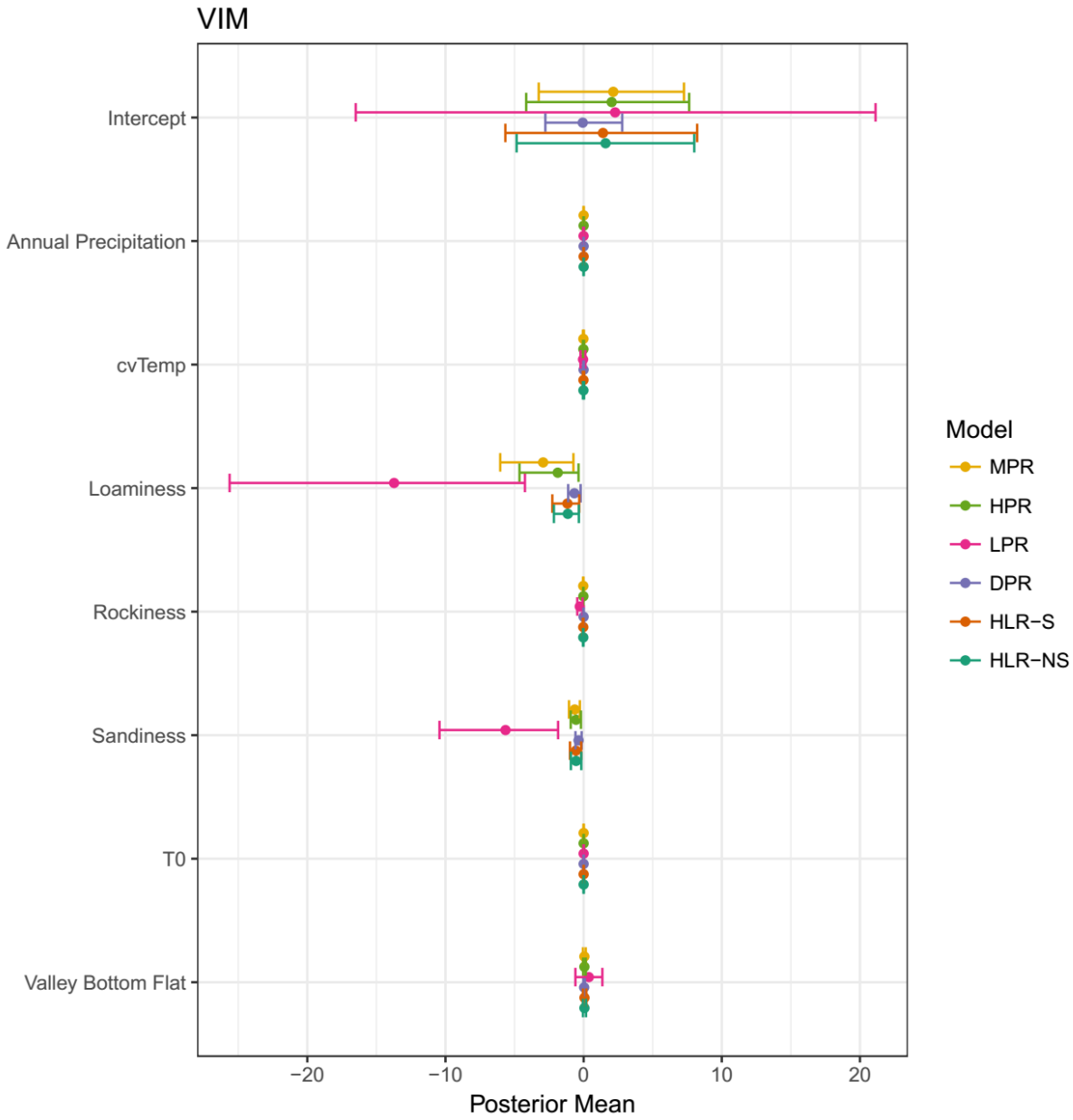


Figure 441: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the manna gum, *Eucalyptus viminalis viminalis*, and the rough-barked manna gum, *Eucalyptus viminalis cygnetensis*.

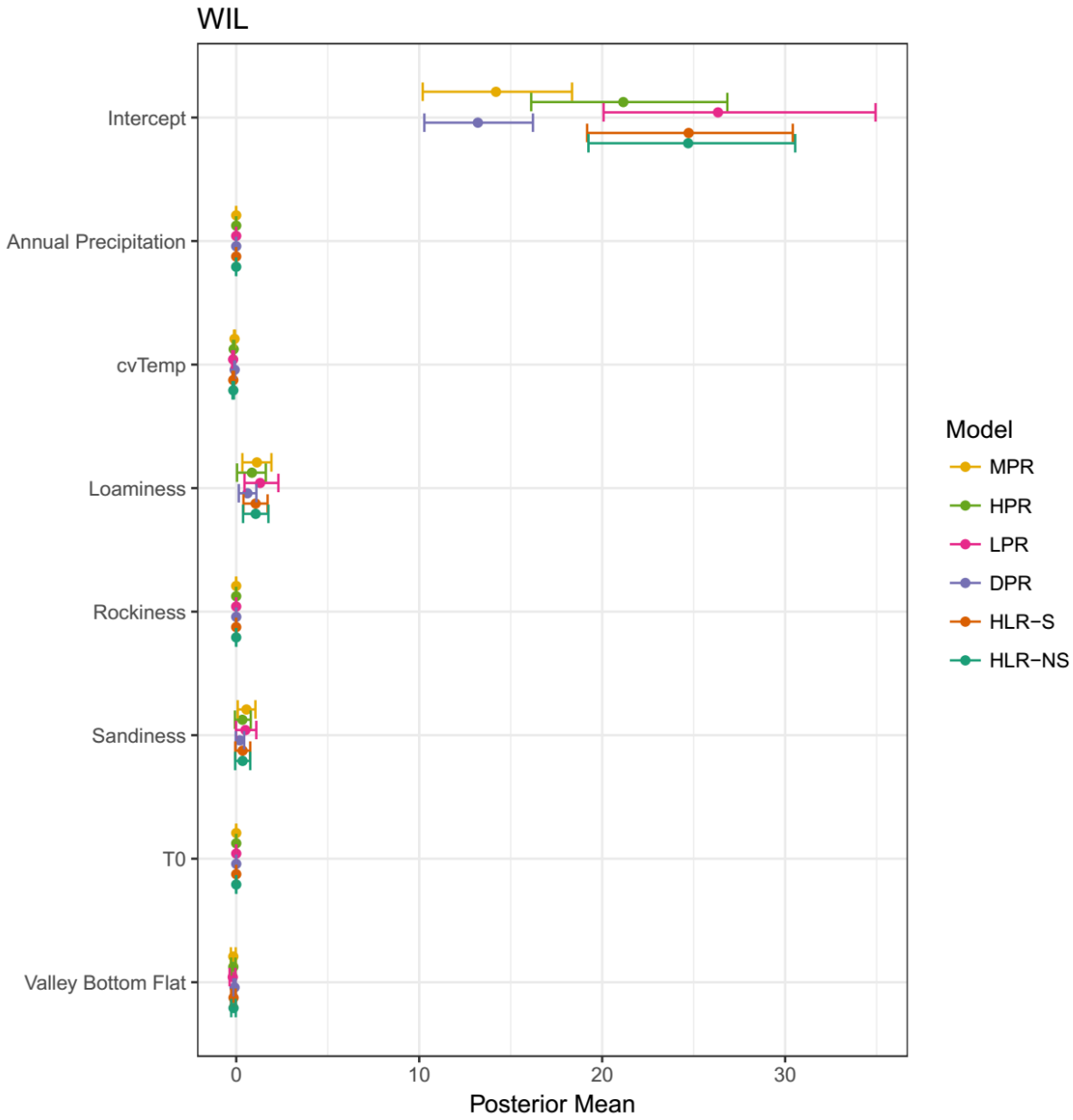


Figure 442: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Mt. William snow gum, *Eucalyptus willissi falciformis*.

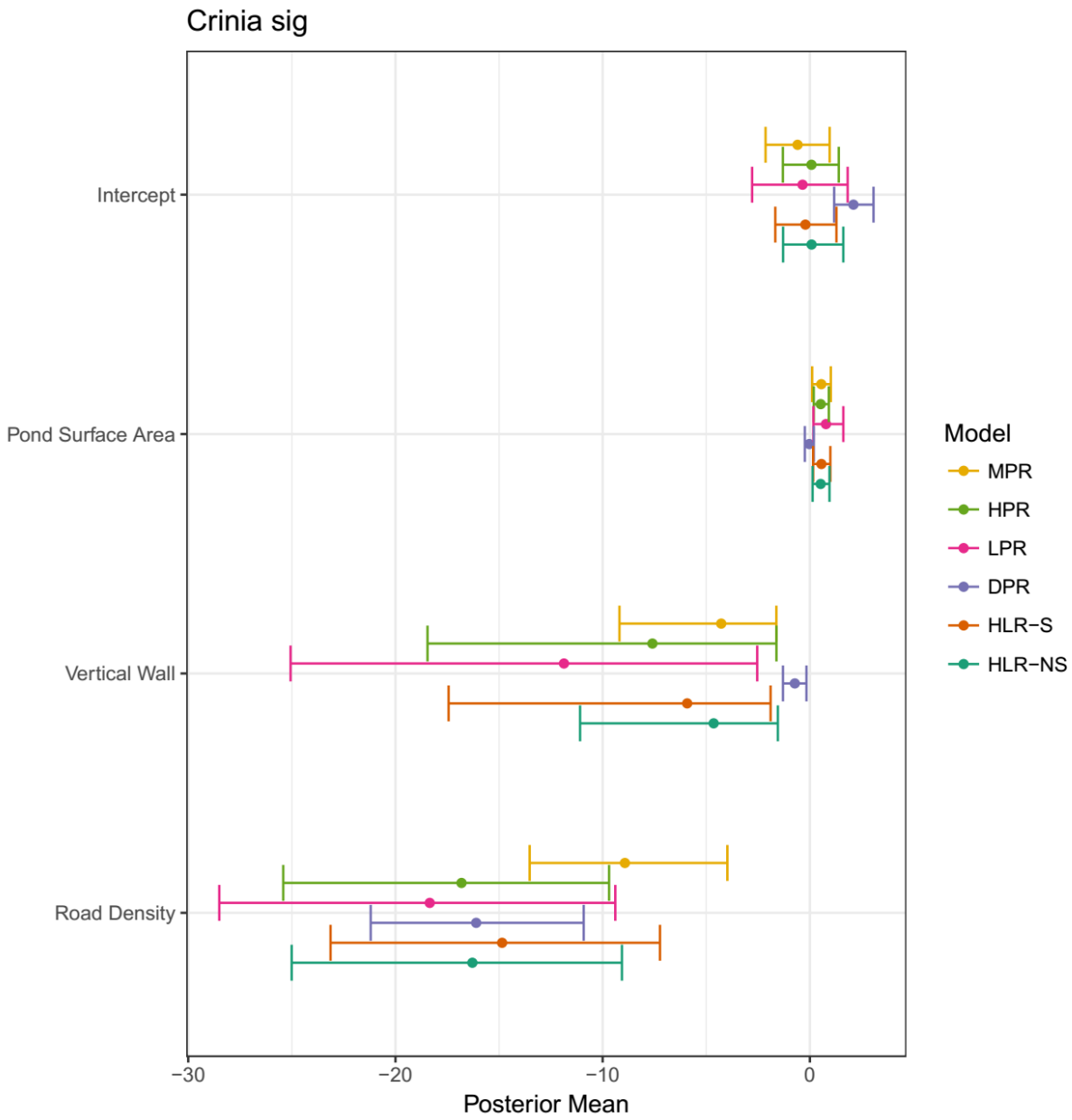


Figure 443: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the common eastern froglet, *Crinia signifera*.

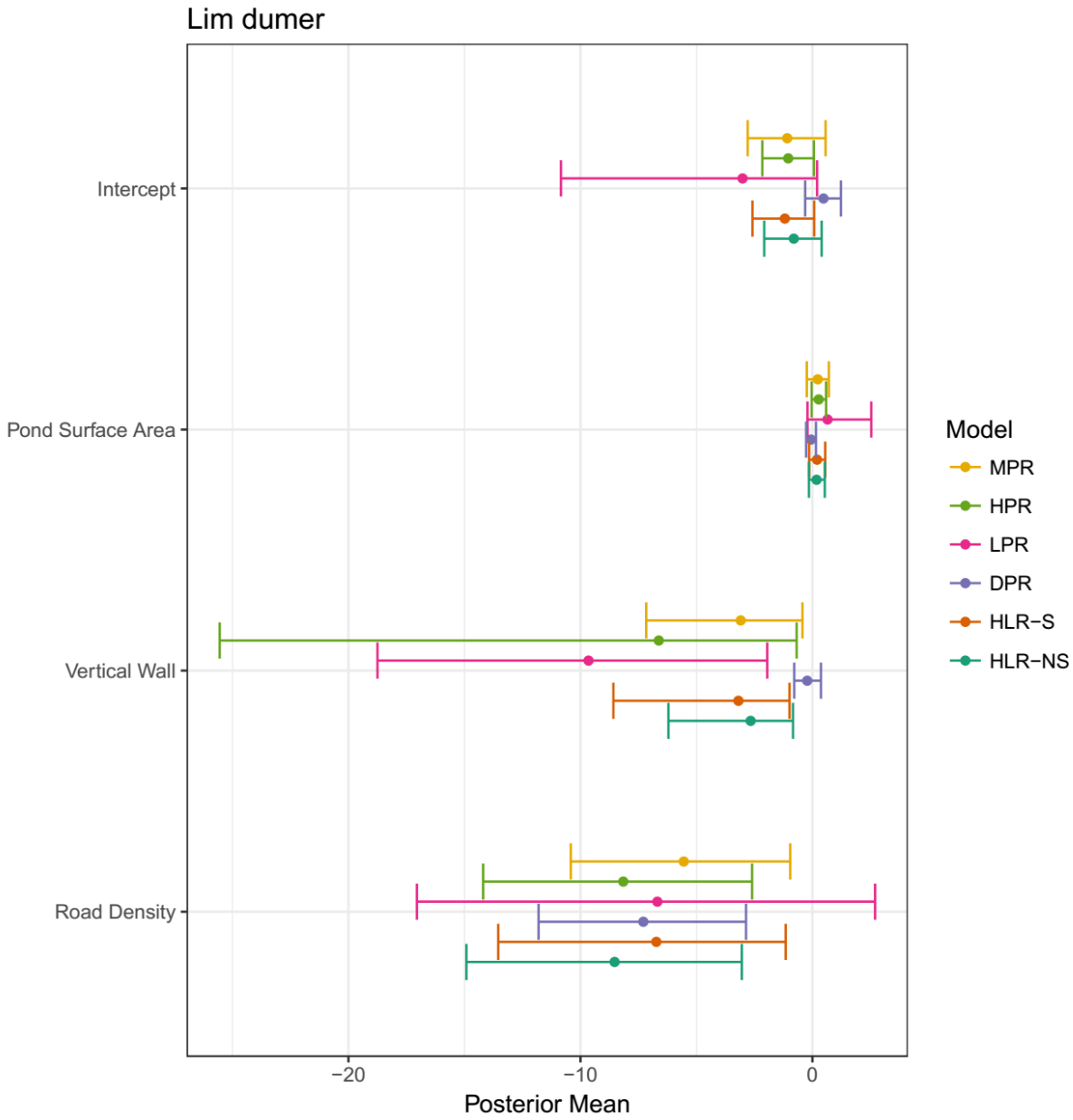


Figure 444: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the pobblebonk, *Limnodynastes dumerilii*.

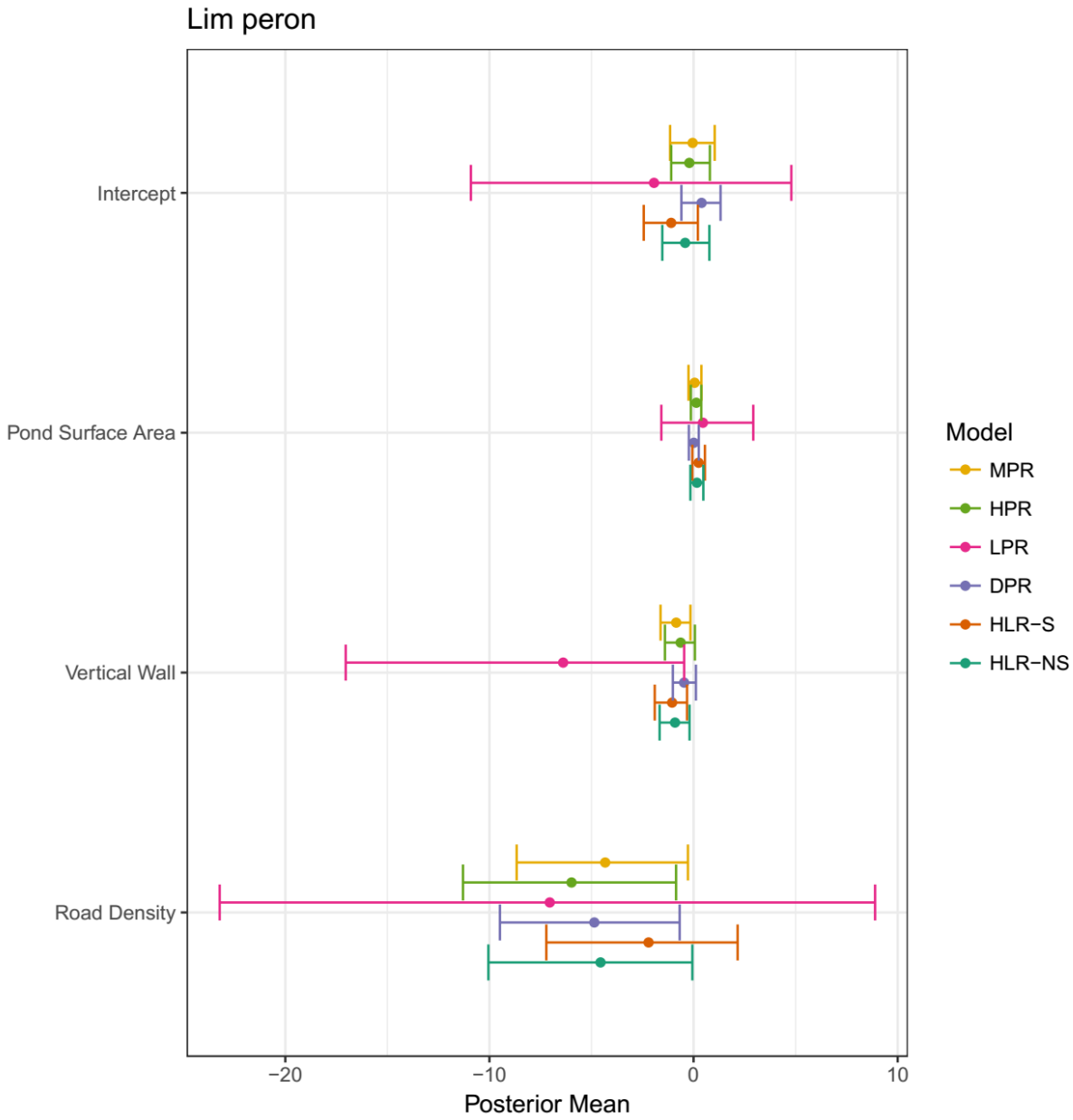


Figure 445: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the striped marsh frog, *Limnodynastes peronii*.

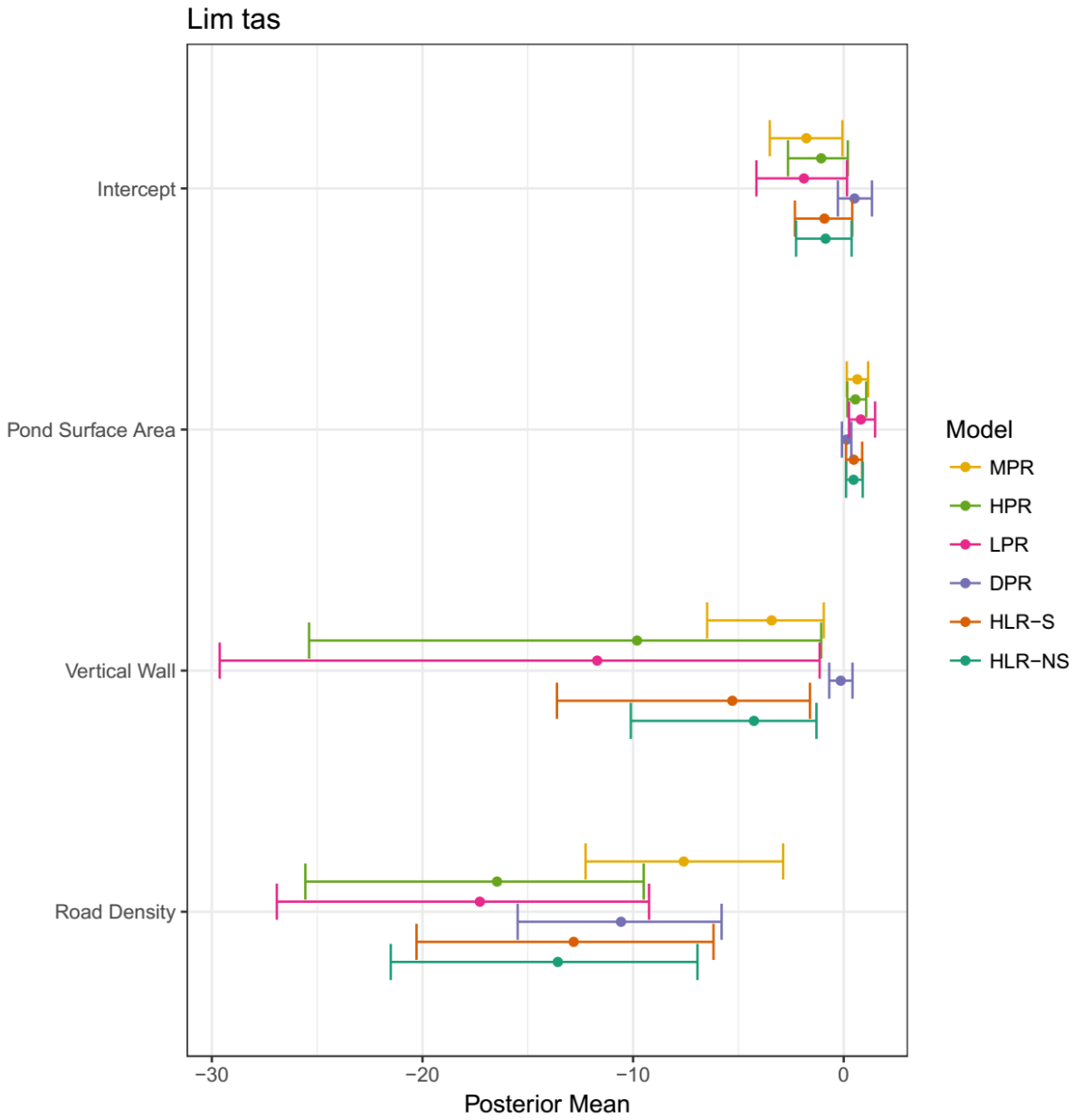


Figure 446: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the spotted grass frog, *Limnodynastes tasmaniensis*.

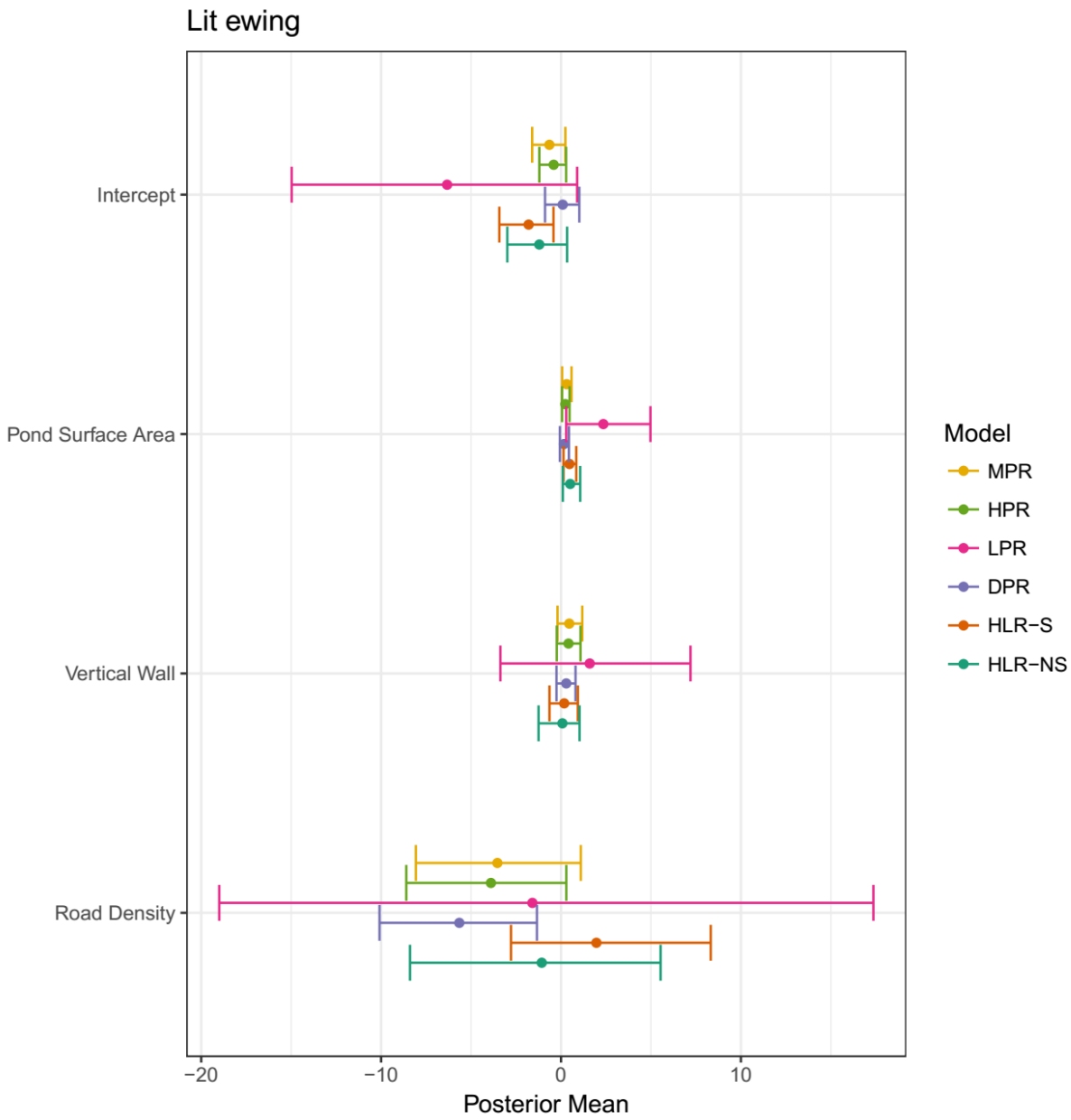


Figure 447: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the southern brown tree frog, *Litoria ewingii*.

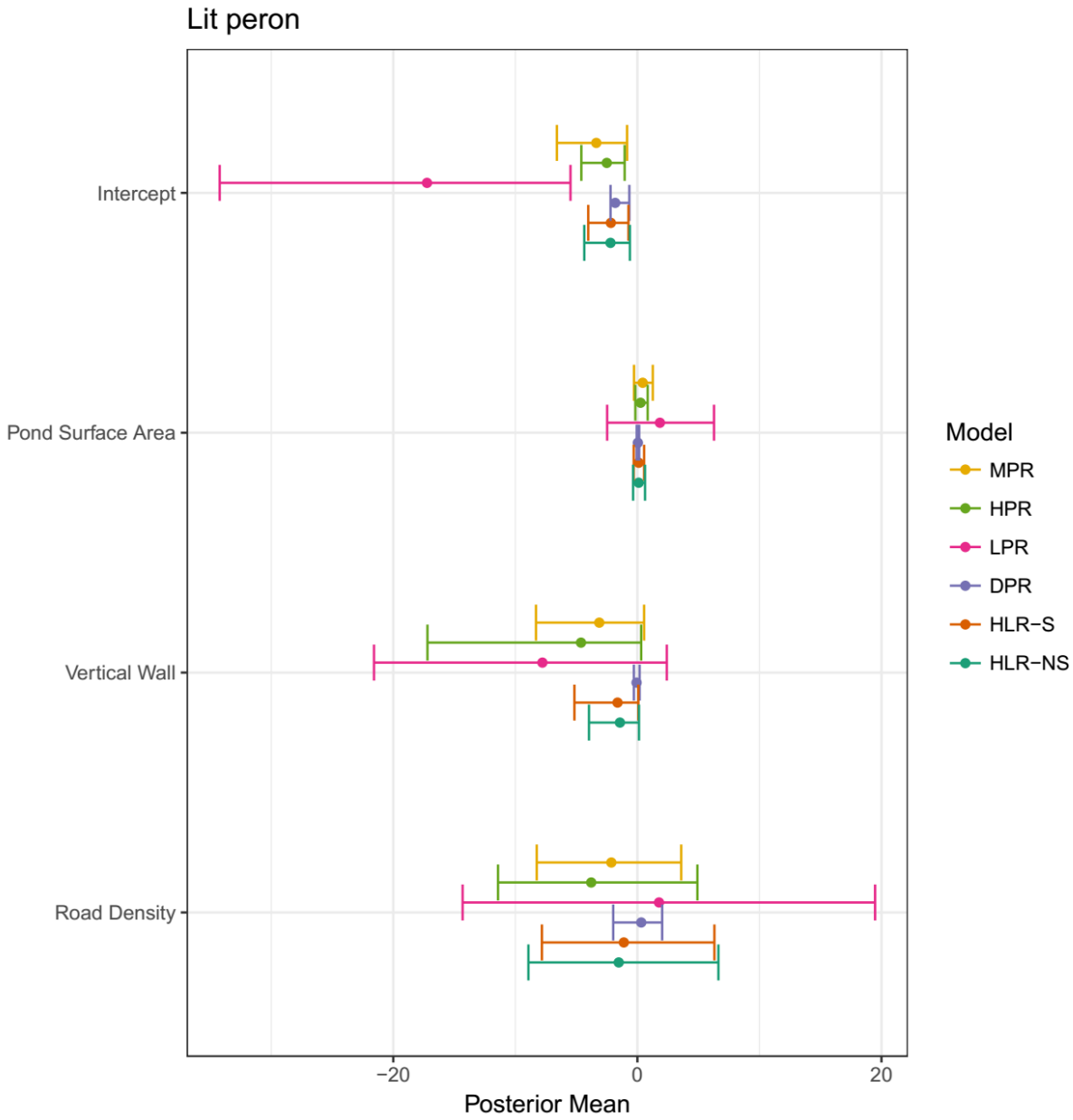


Figure 448: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Peron's tree frog, *Litoria peronii*.

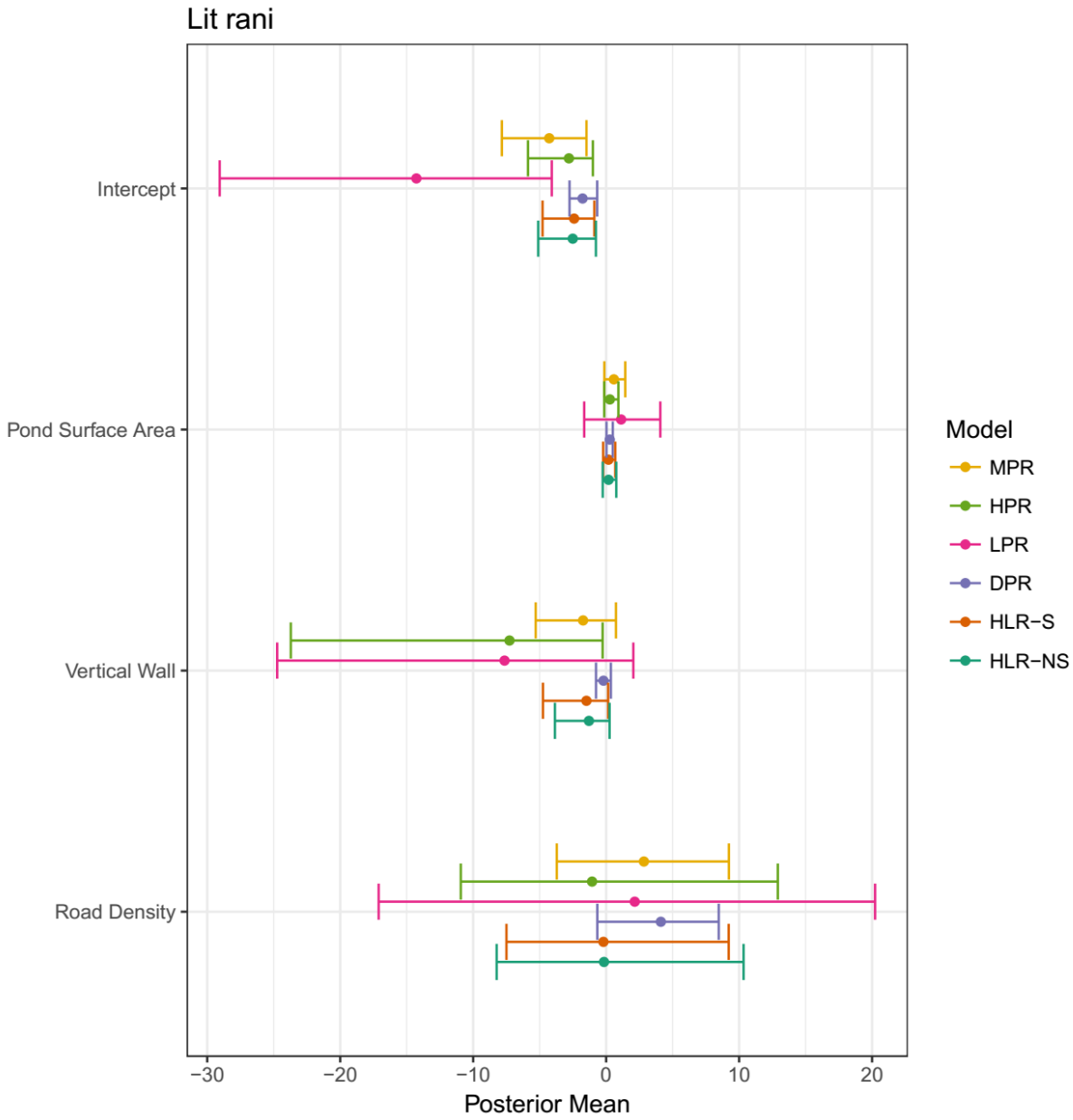


Figure 449: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the growling grass frog, *Litoria raniformis*.

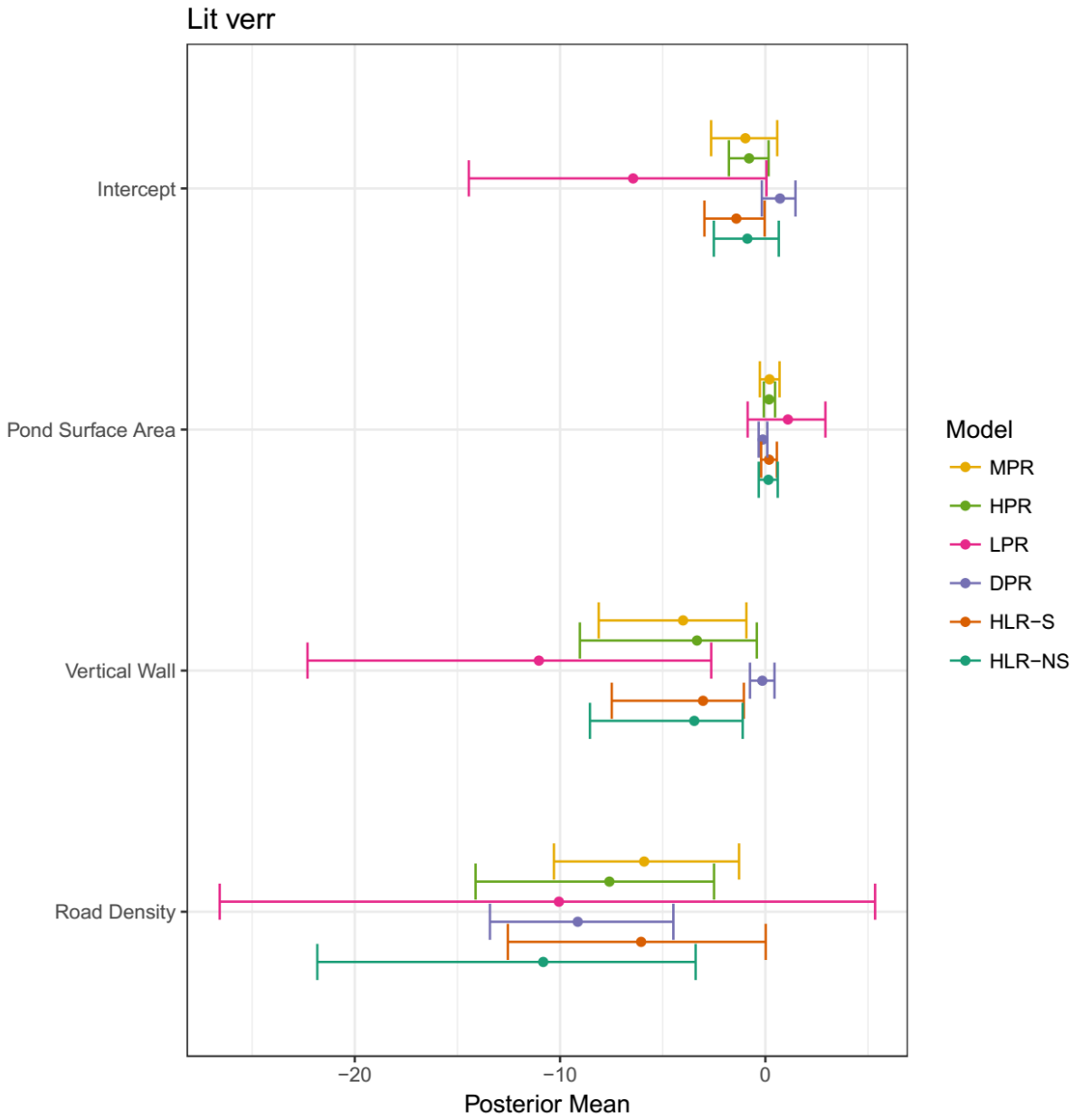


Figure 450: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the whistling tree frog, *Litoria verreauxii*.

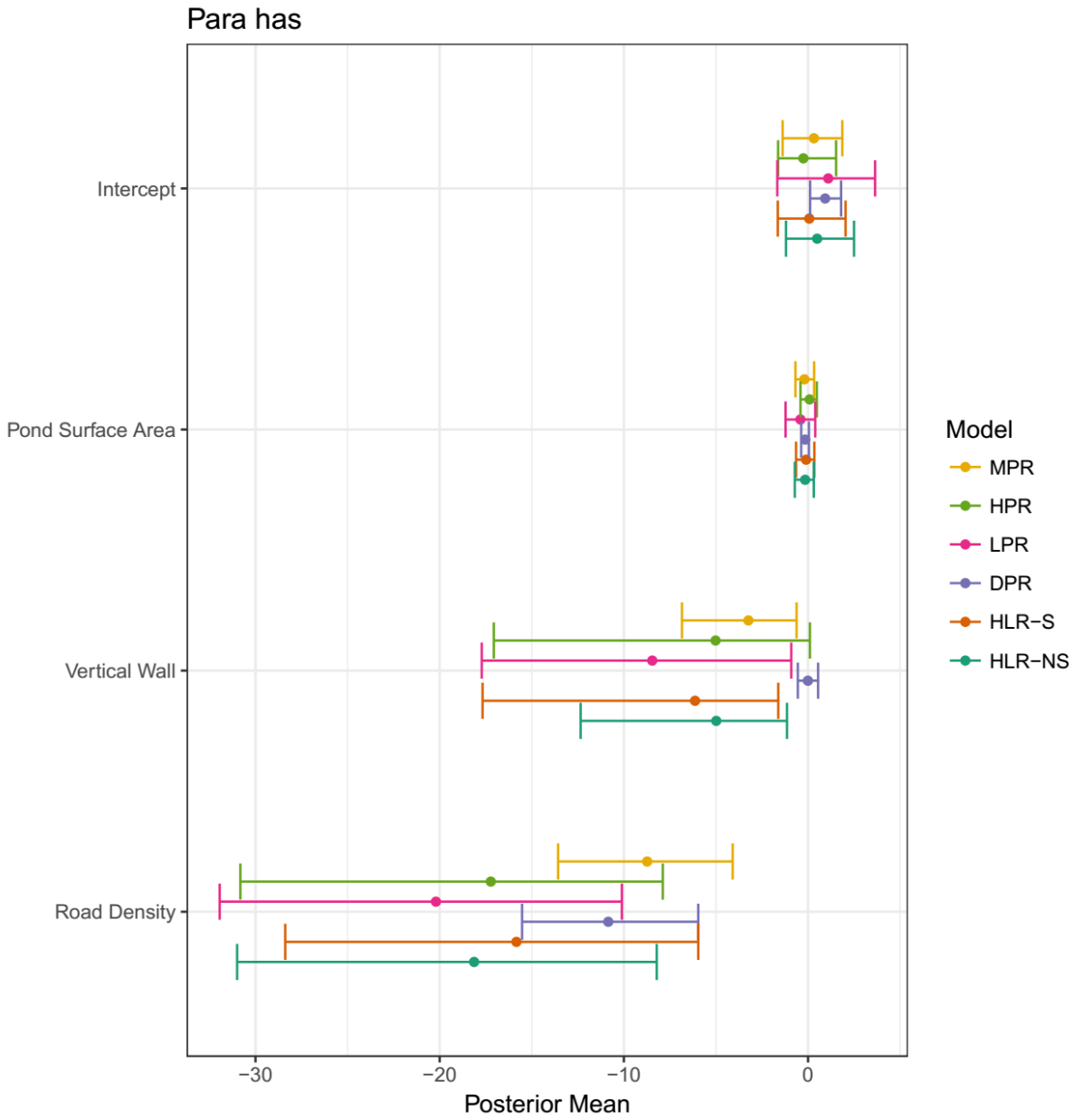


Figure 451: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the Haswell's froglet, *Paracrinia haswelli*.

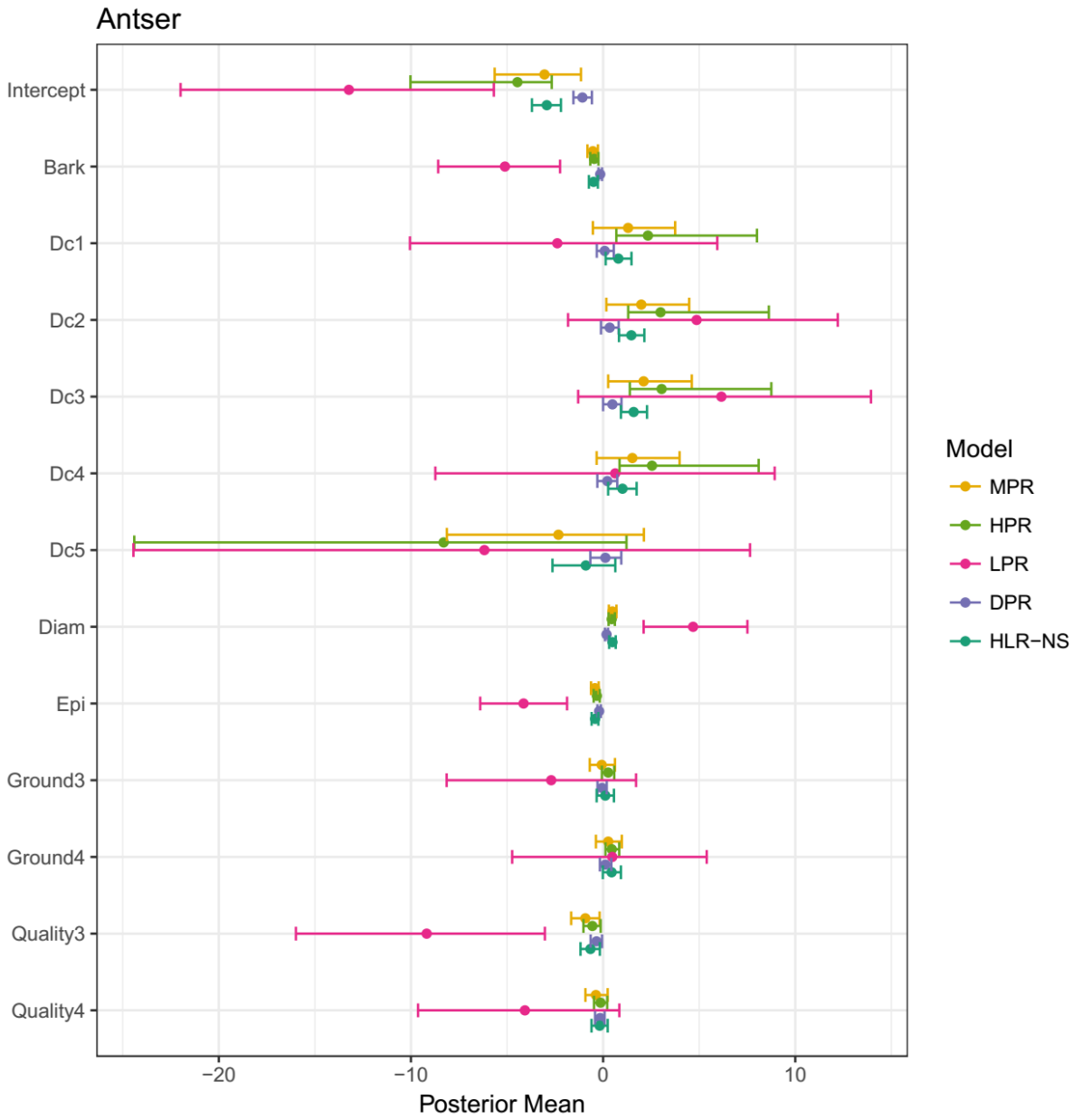


Figure 452: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Antrodia serialis*.

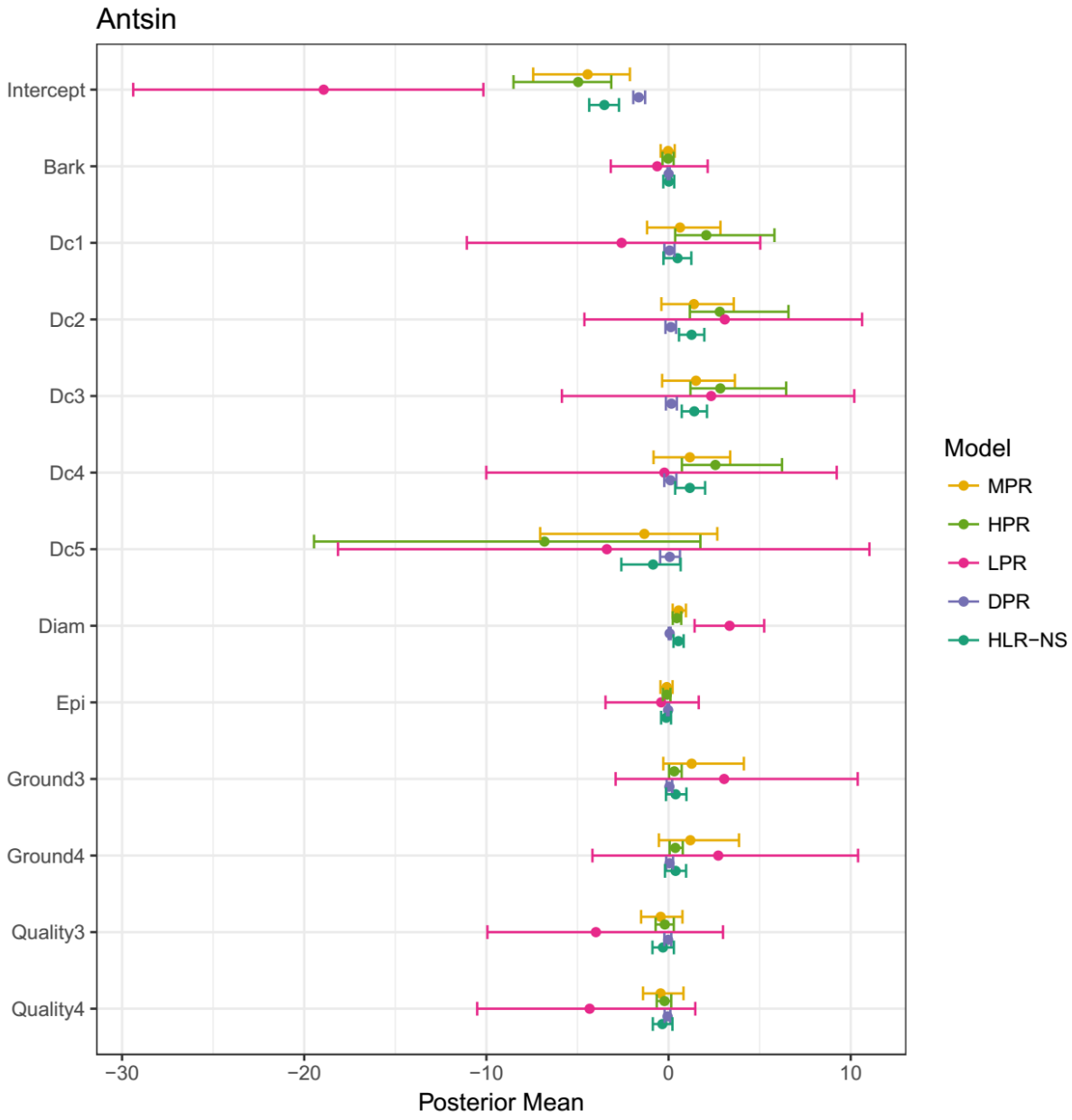


Figure 453: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Antrodia sinuosa*.

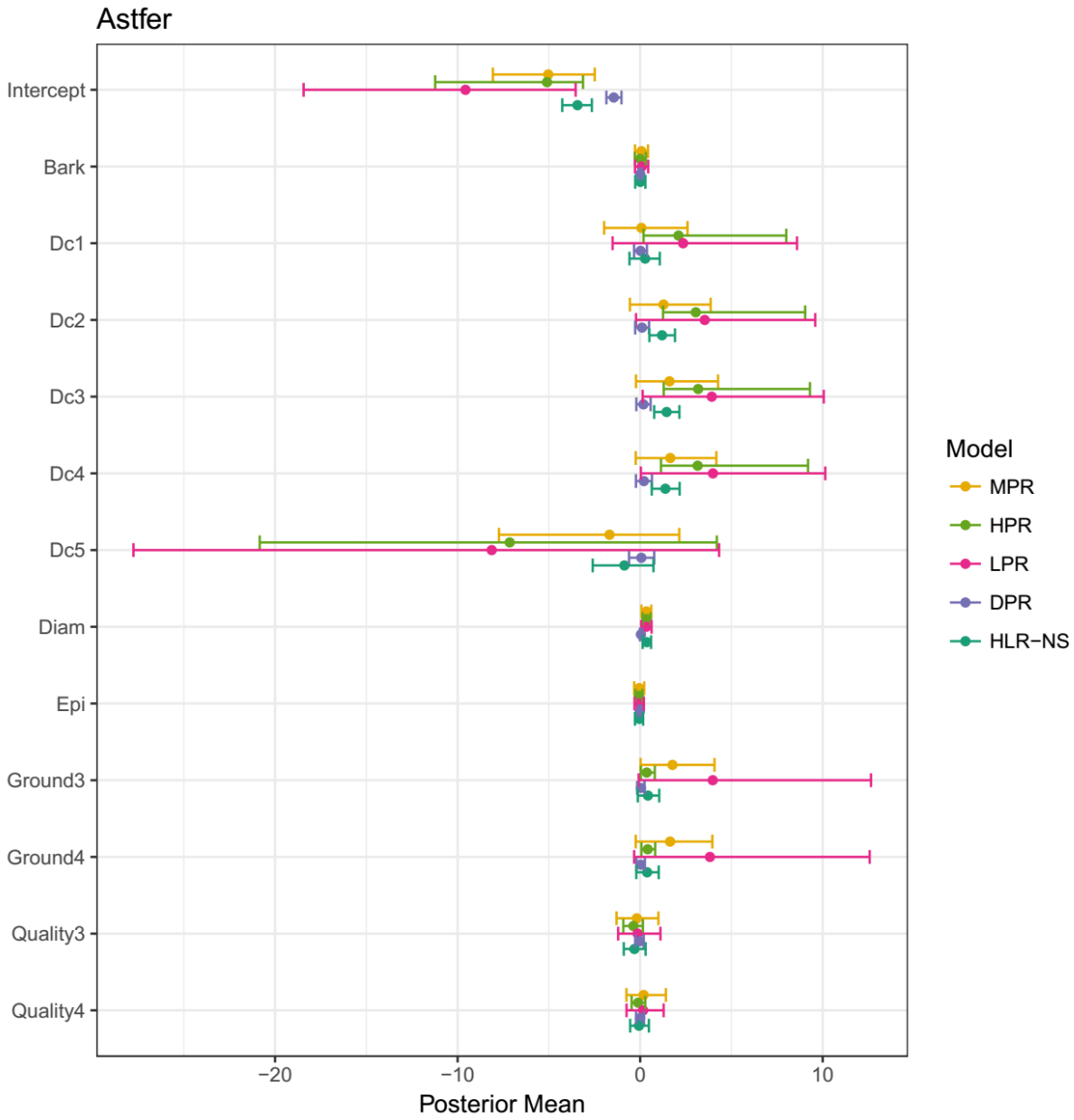


Figure 454: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Asterodon ferruginosus*.

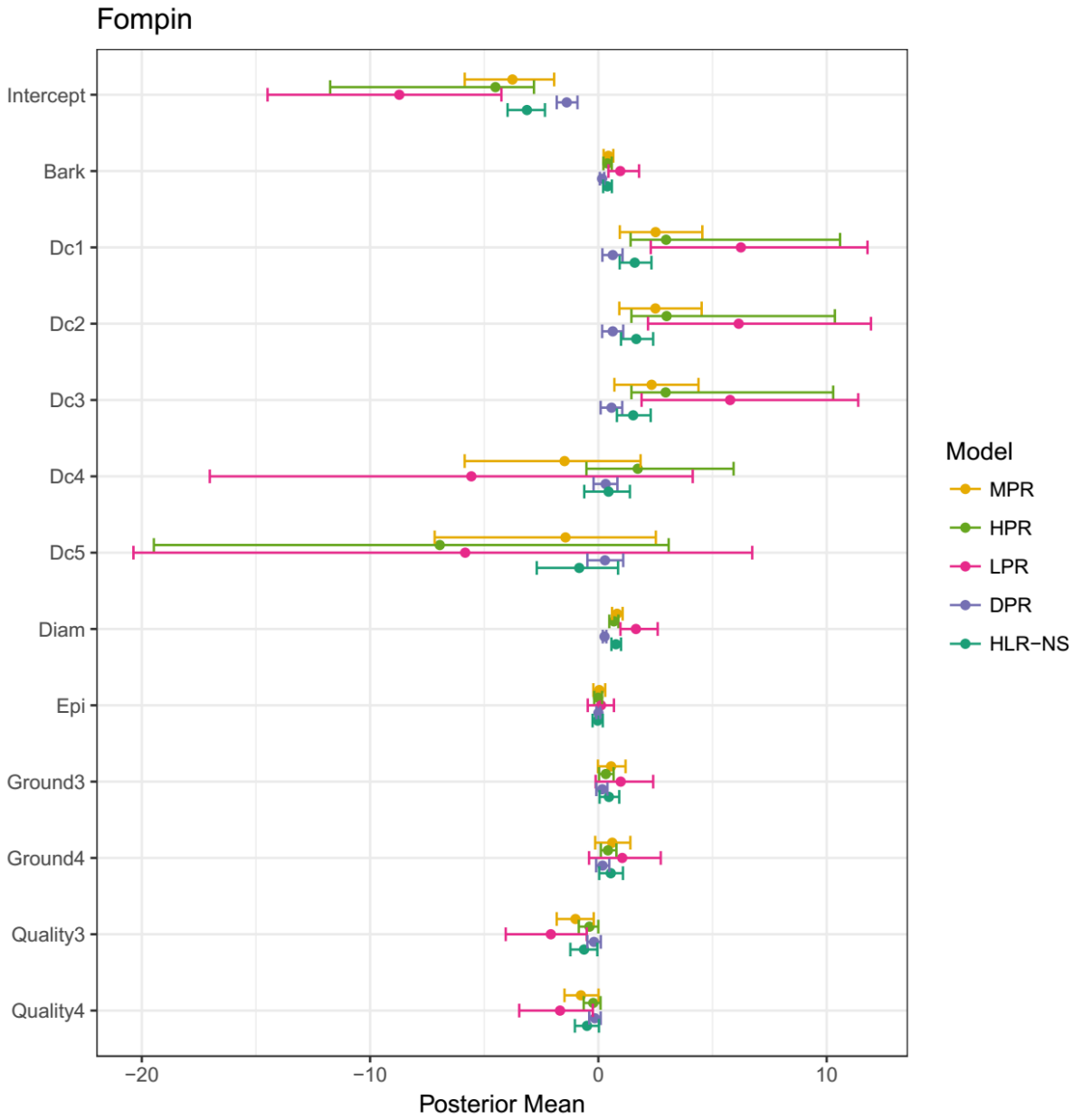


Figure 455: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Fomitopsis pinicola*.

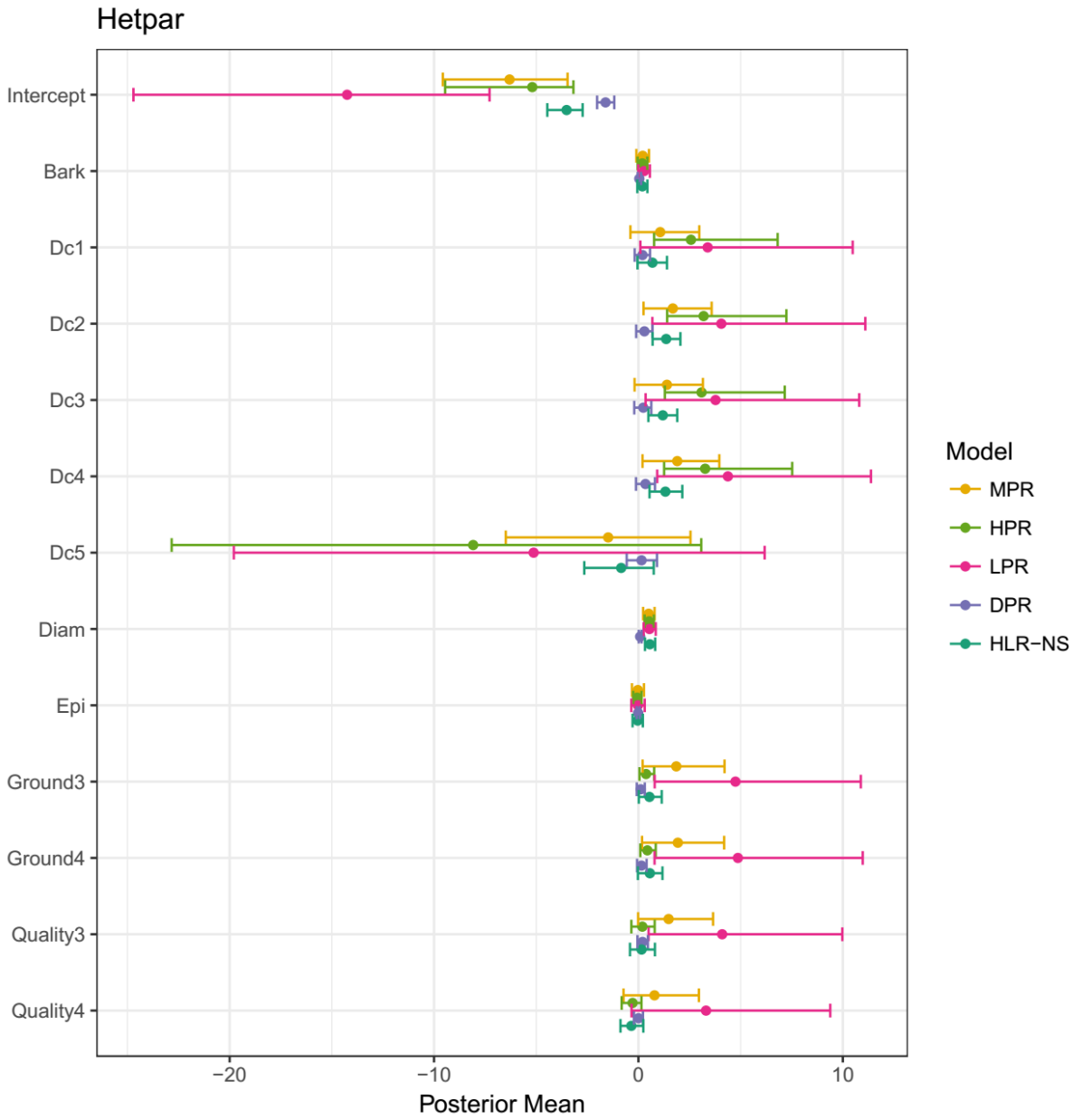


Figure 456: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Heterobasidion parviporum*.

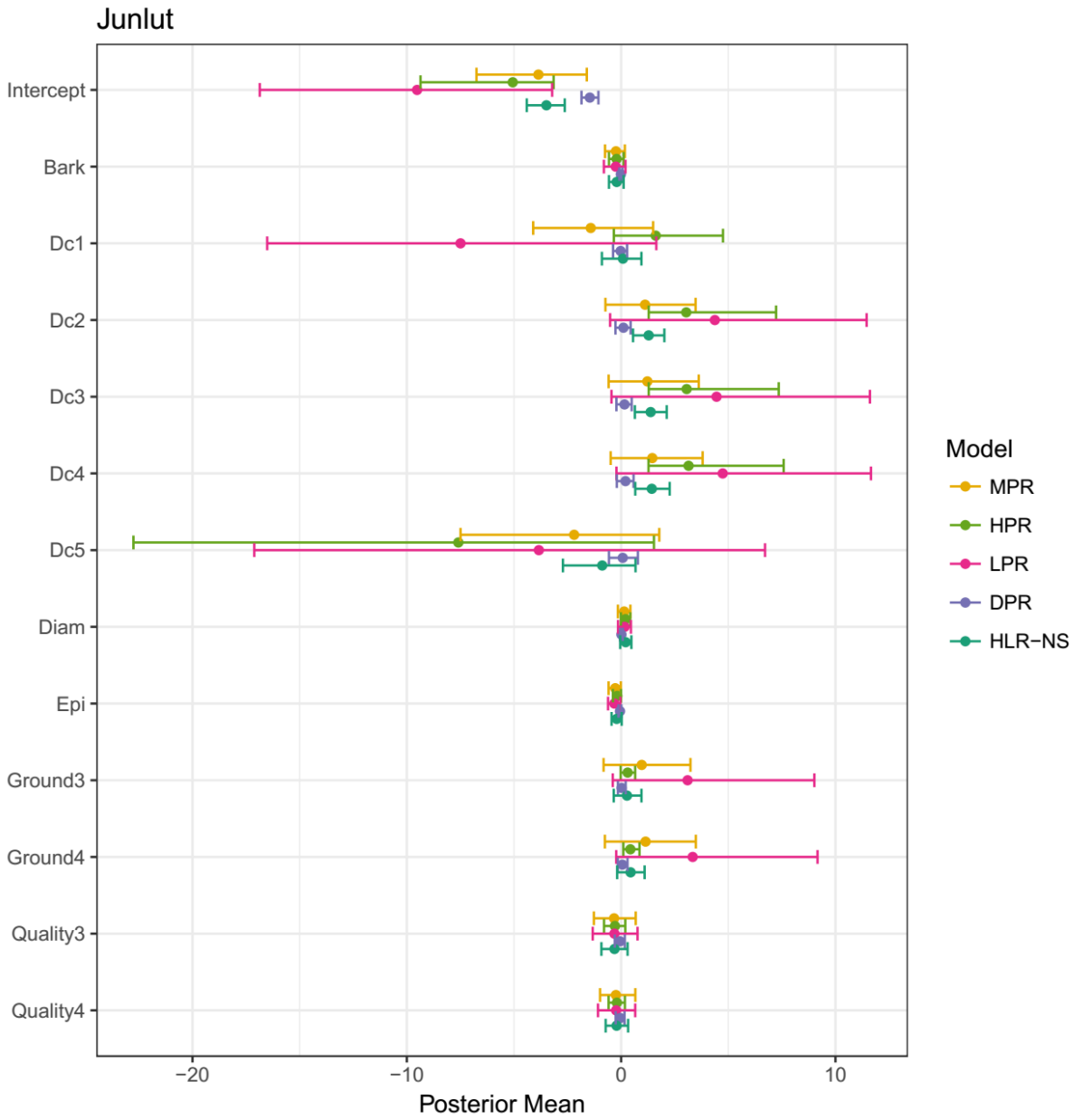


Figure 457: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Junghuhnia luteoalba*.

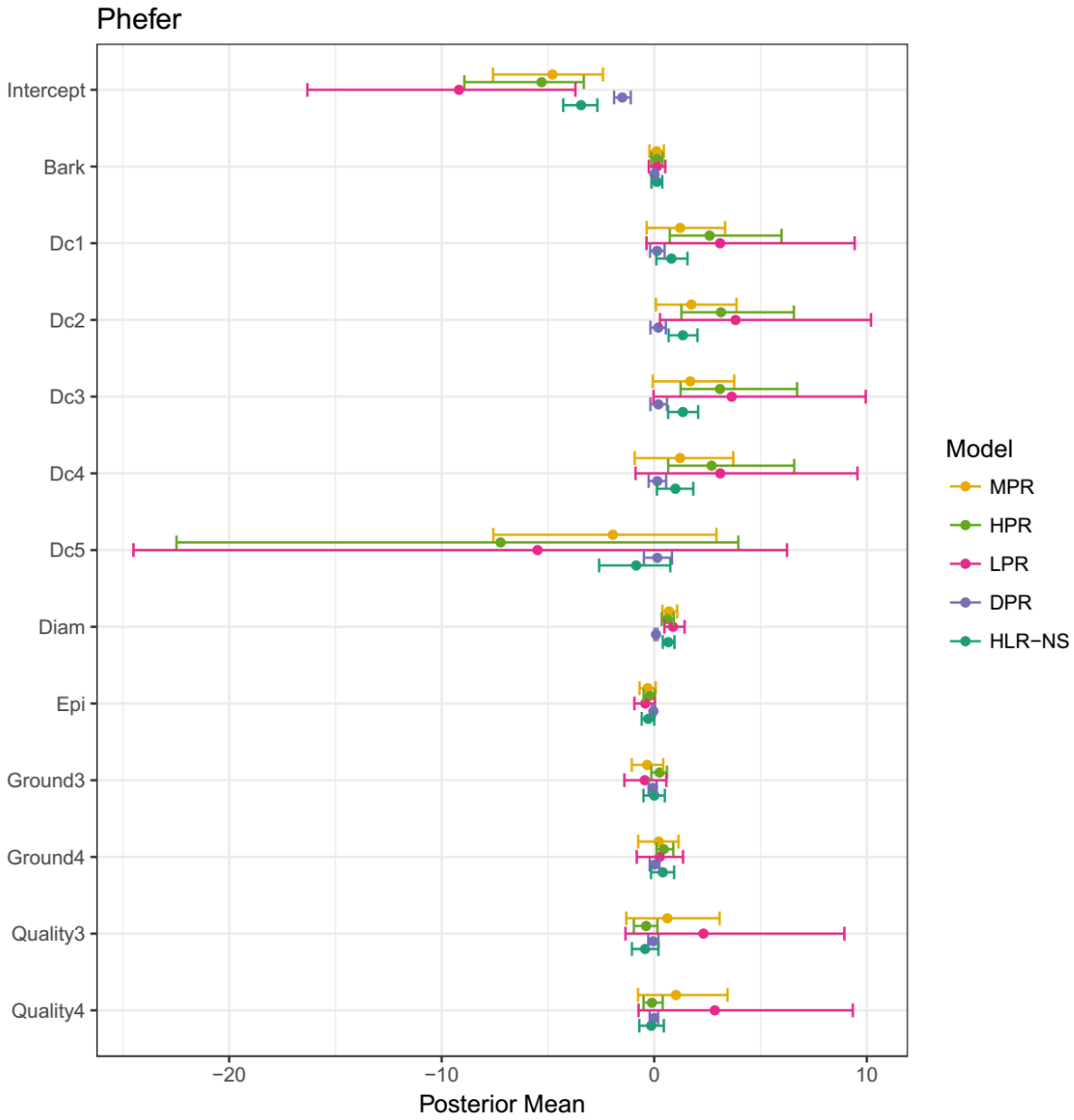


Figure 458: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Phellinus ferrugineofuscus*.

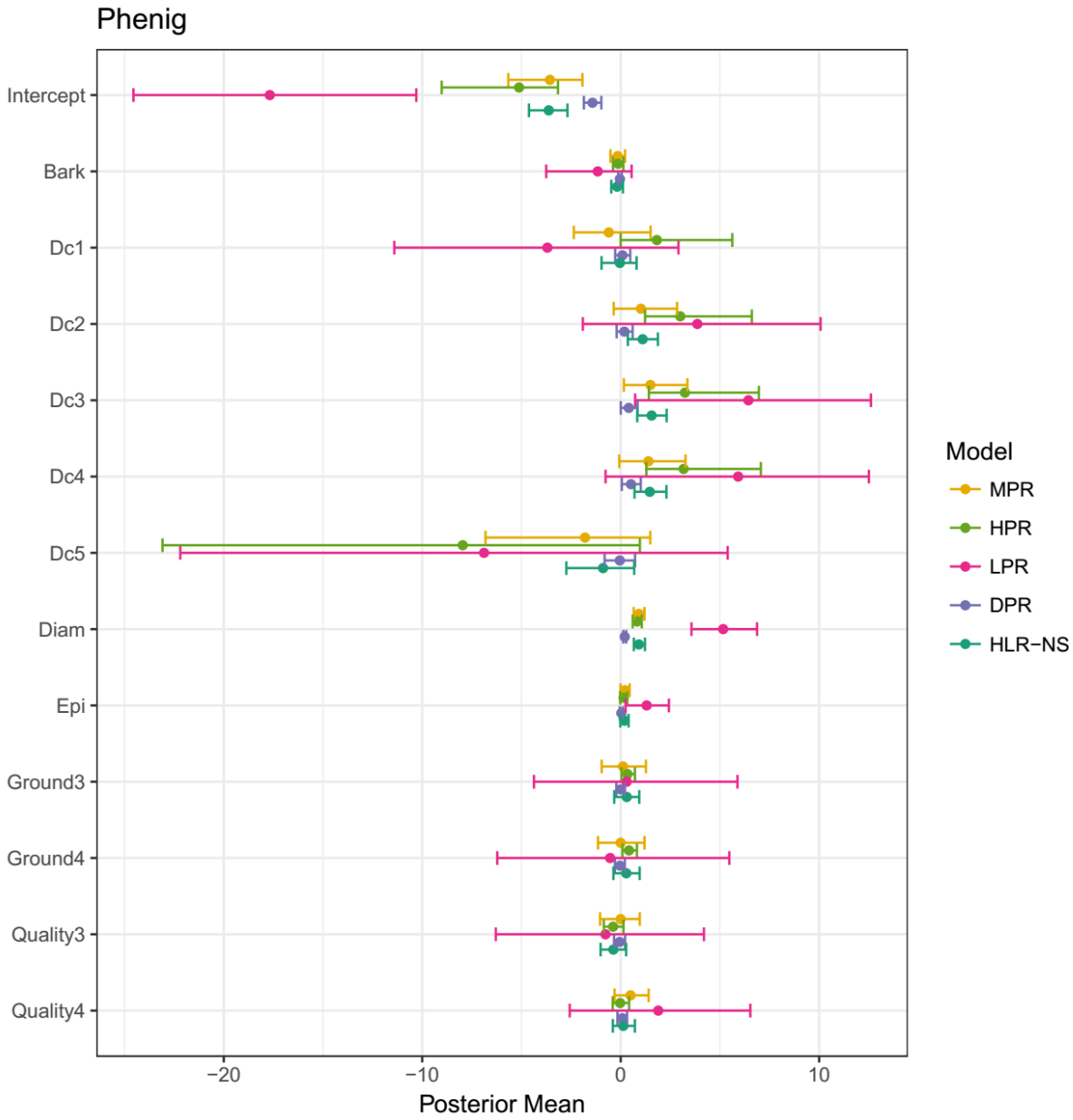


Figure 459: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Phellinus nigrolimitatus*.

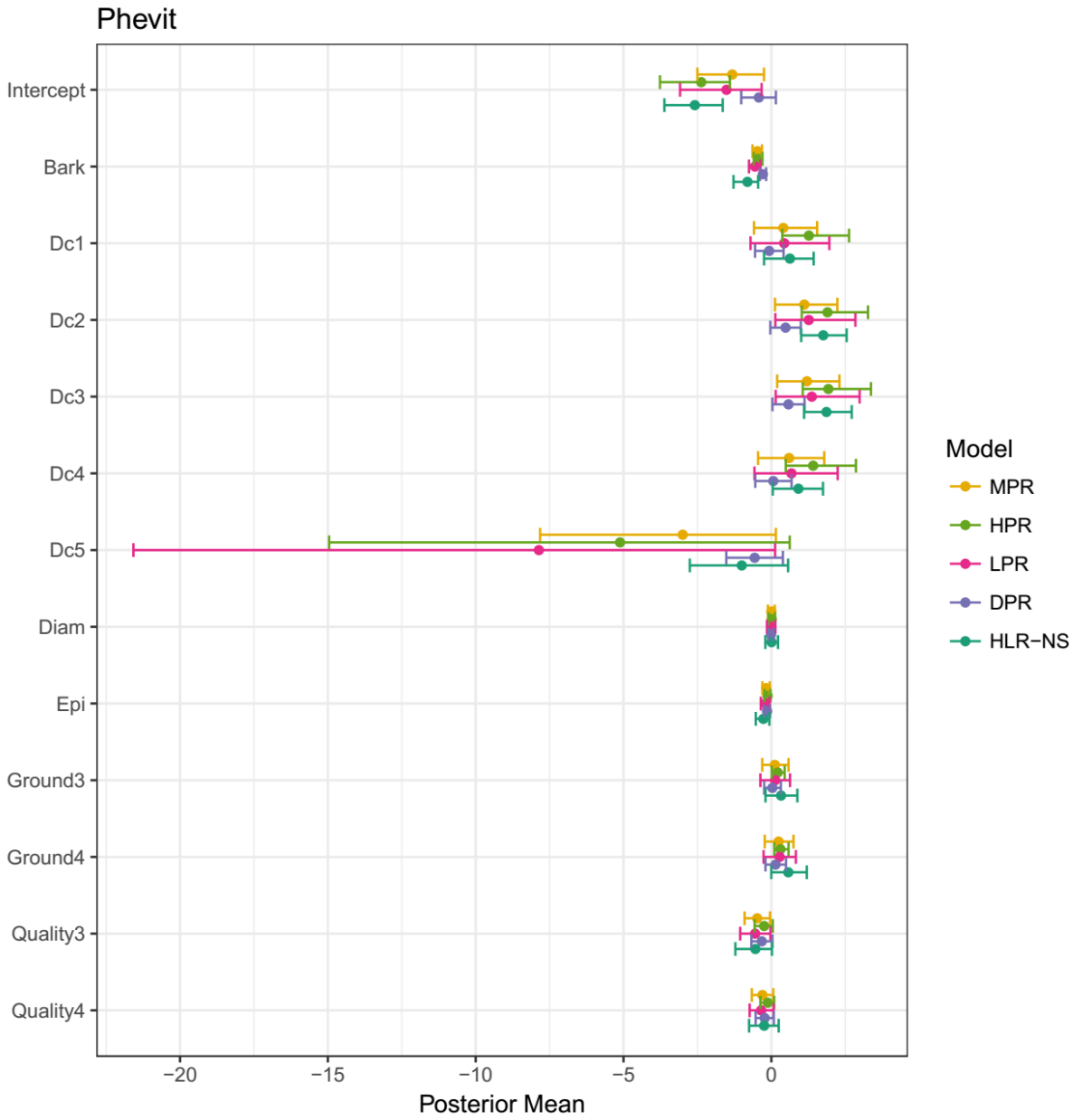


Figure 460: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Phellinus viticola*.

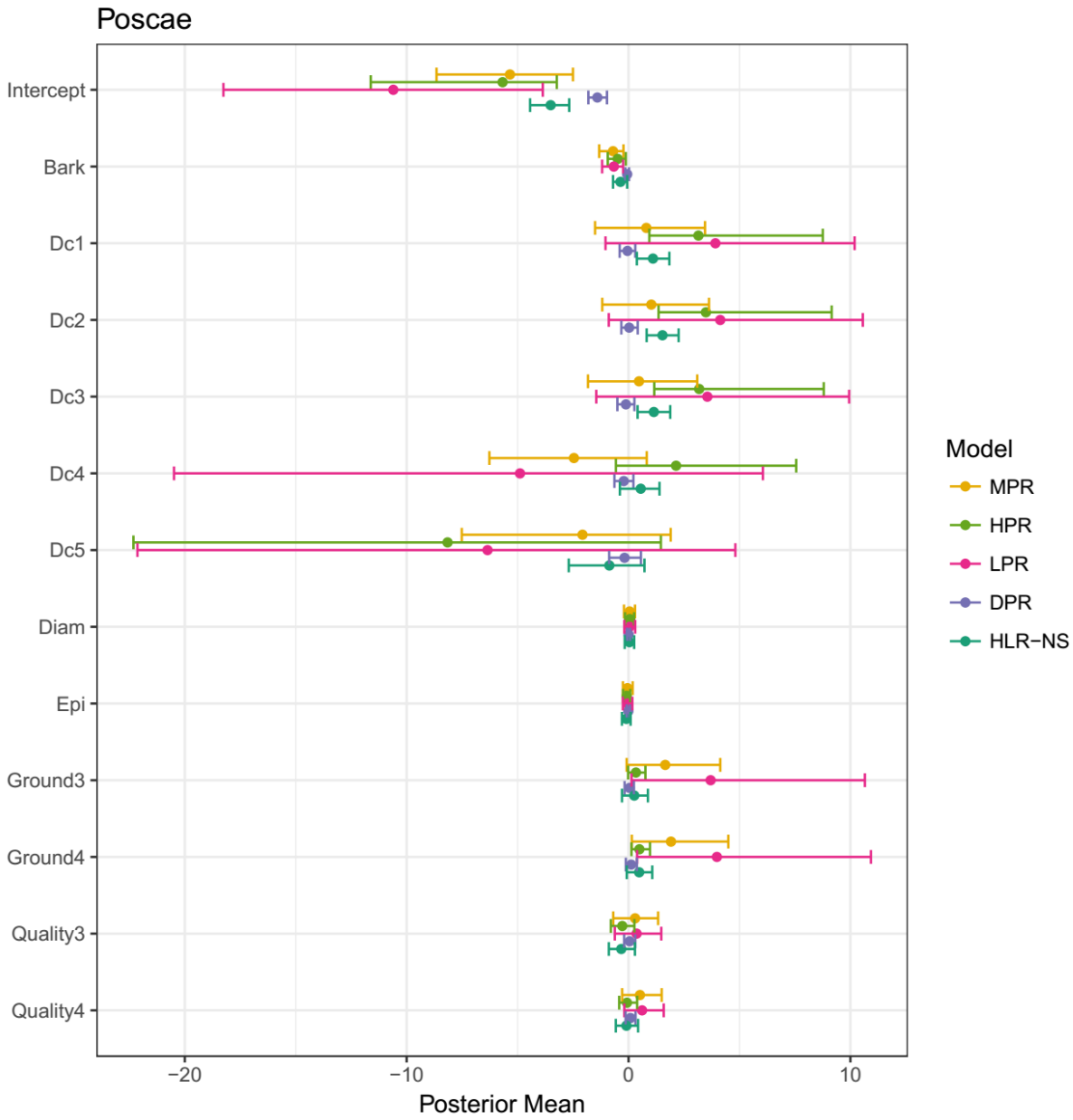


Figure 461: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Postia caesia*.

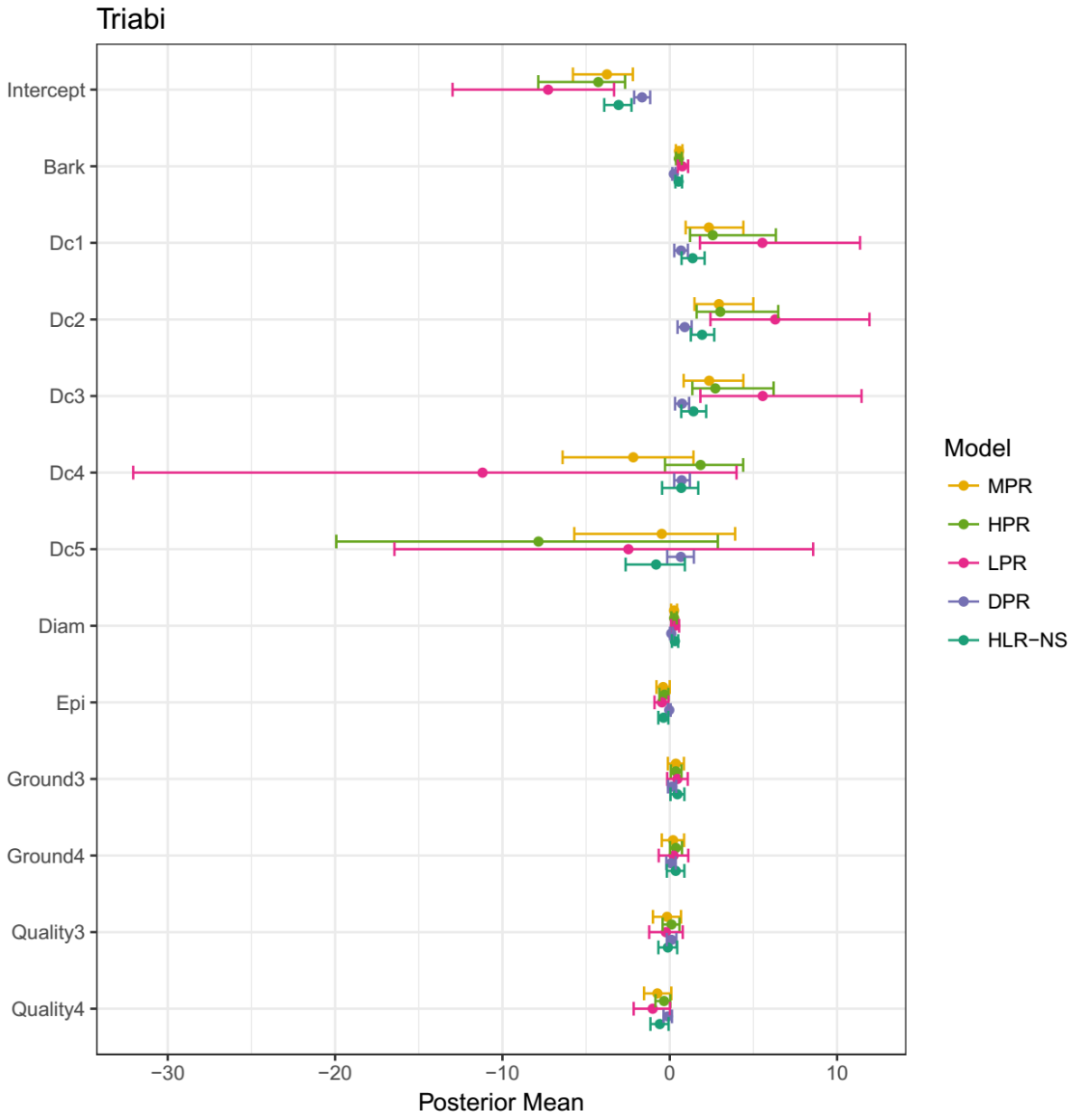


Figure 462: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Trichaptum abietinum*.

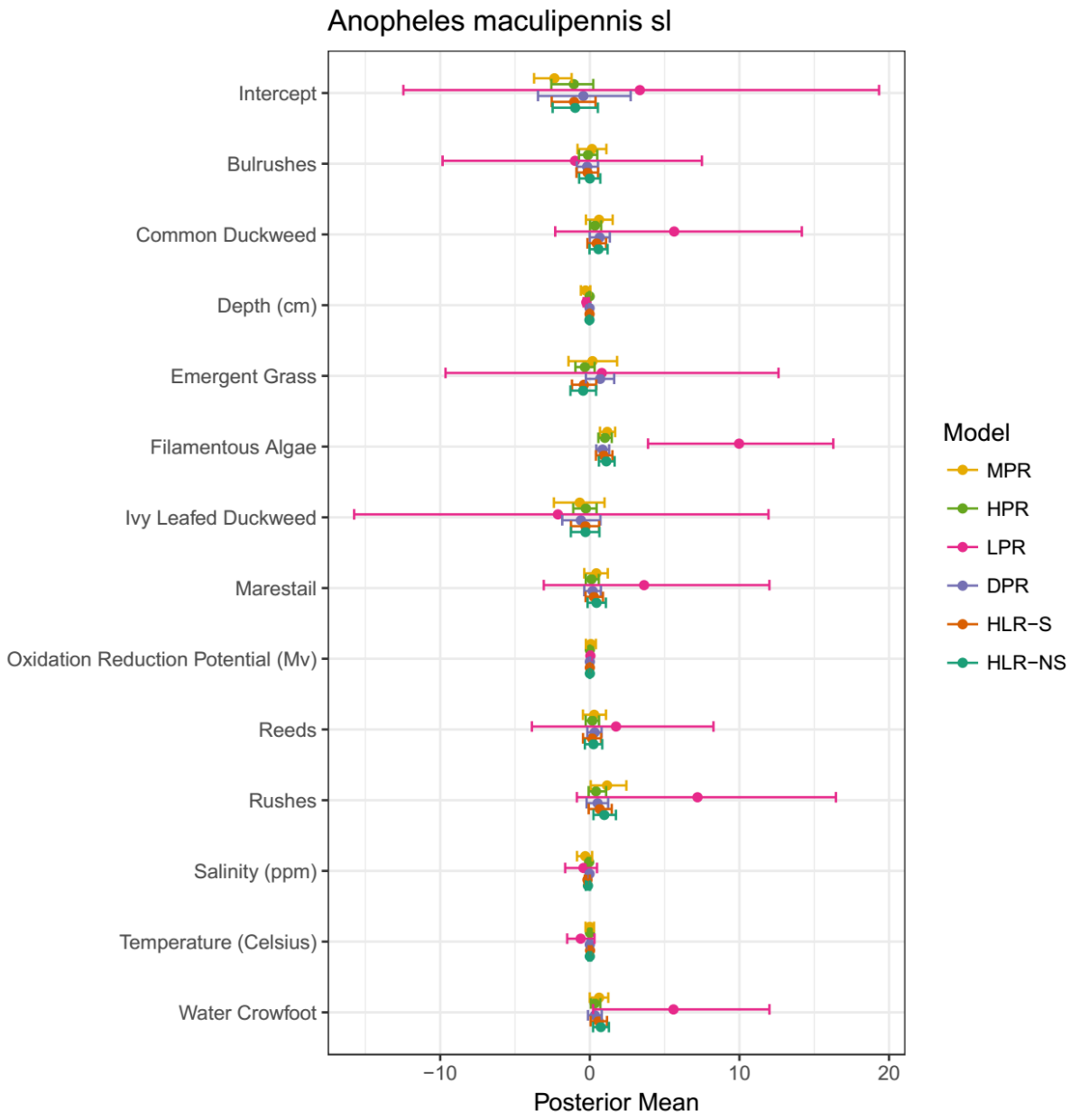


Figure 463: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Anopheles maculipennis s.l.*

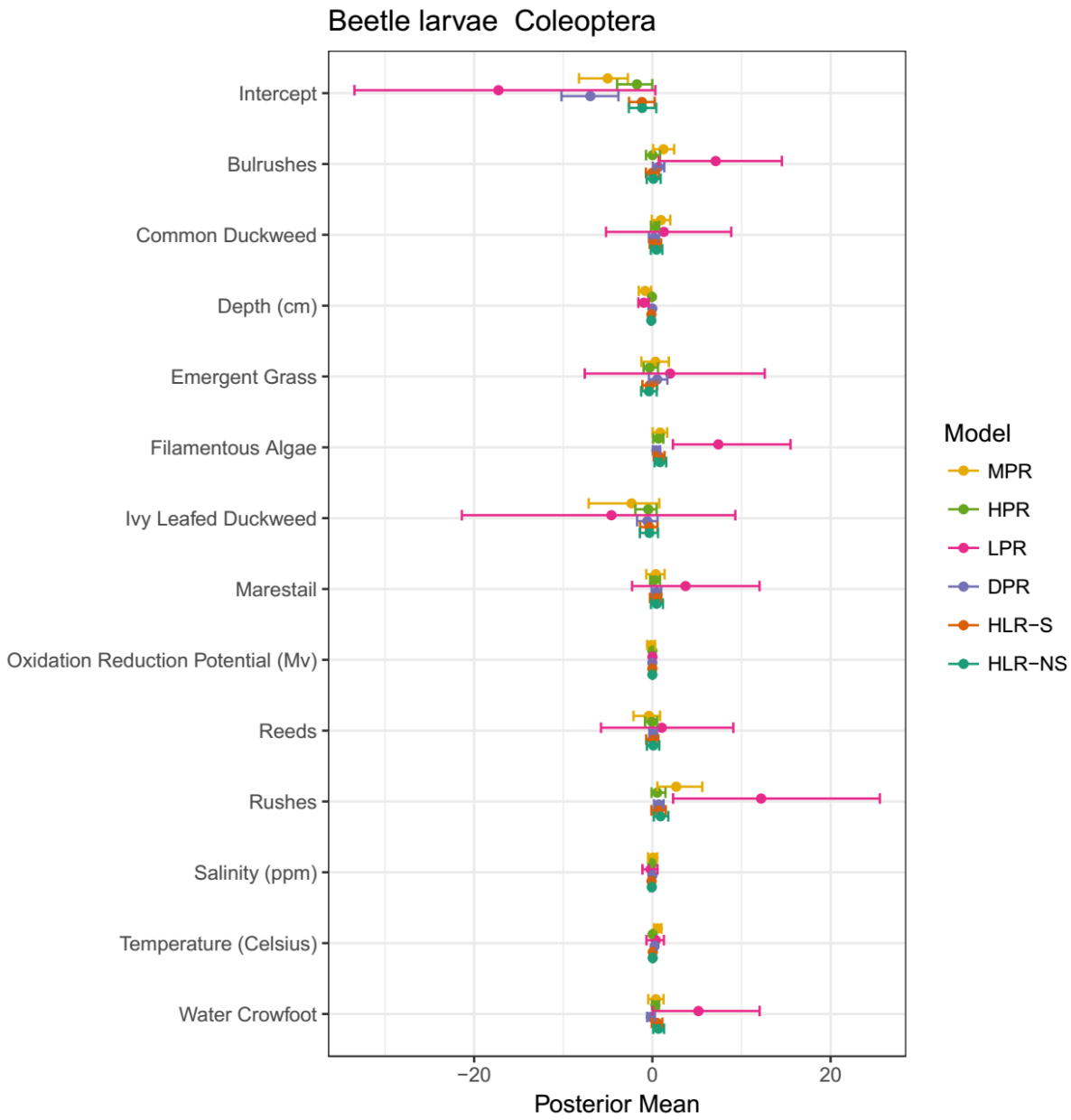


Figure 464: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for Coleoptera beetle larvae.

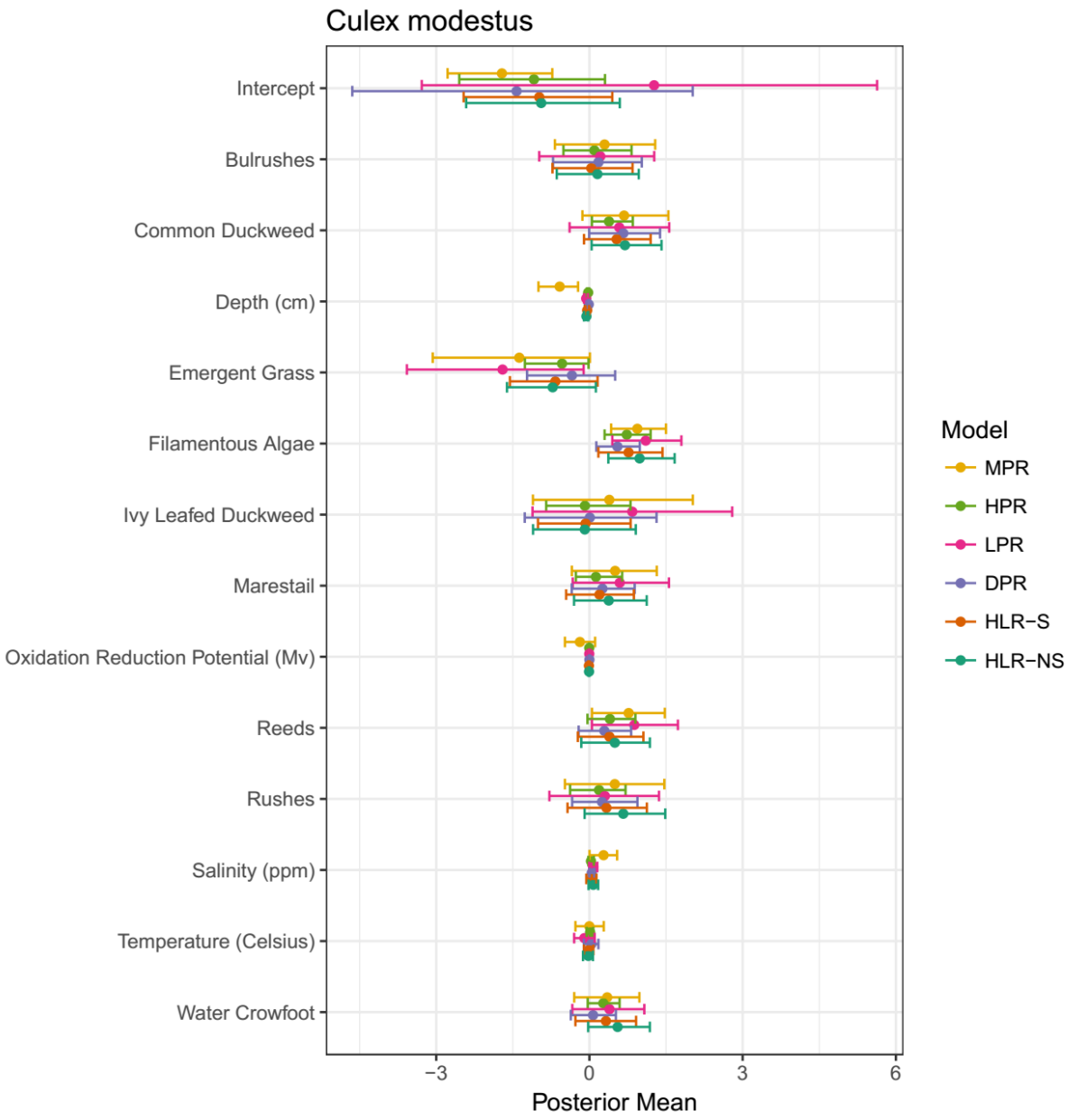


Figure 465: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Culex modestus*.

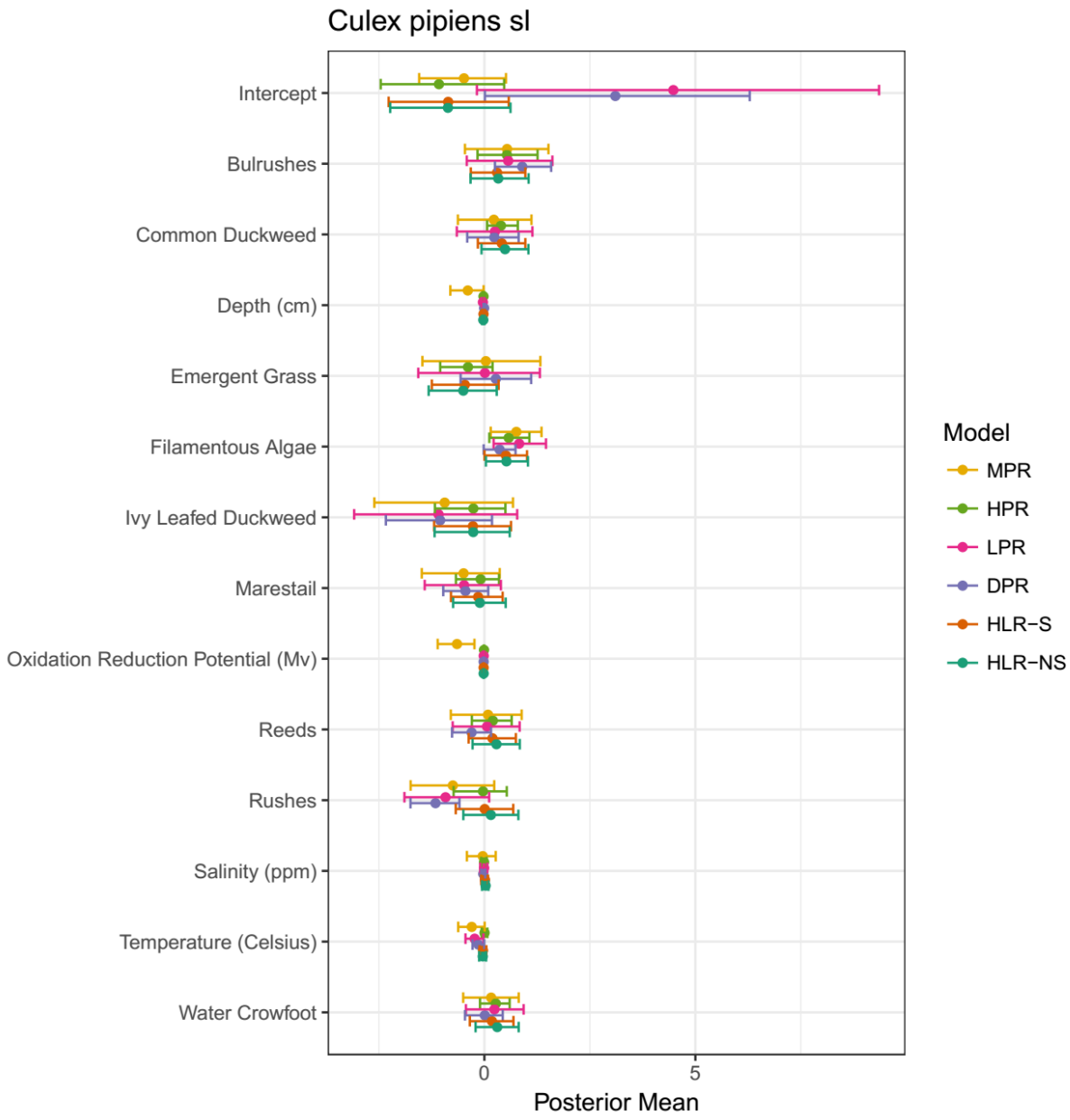


Figure 466: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Culex pipiens s.l.*

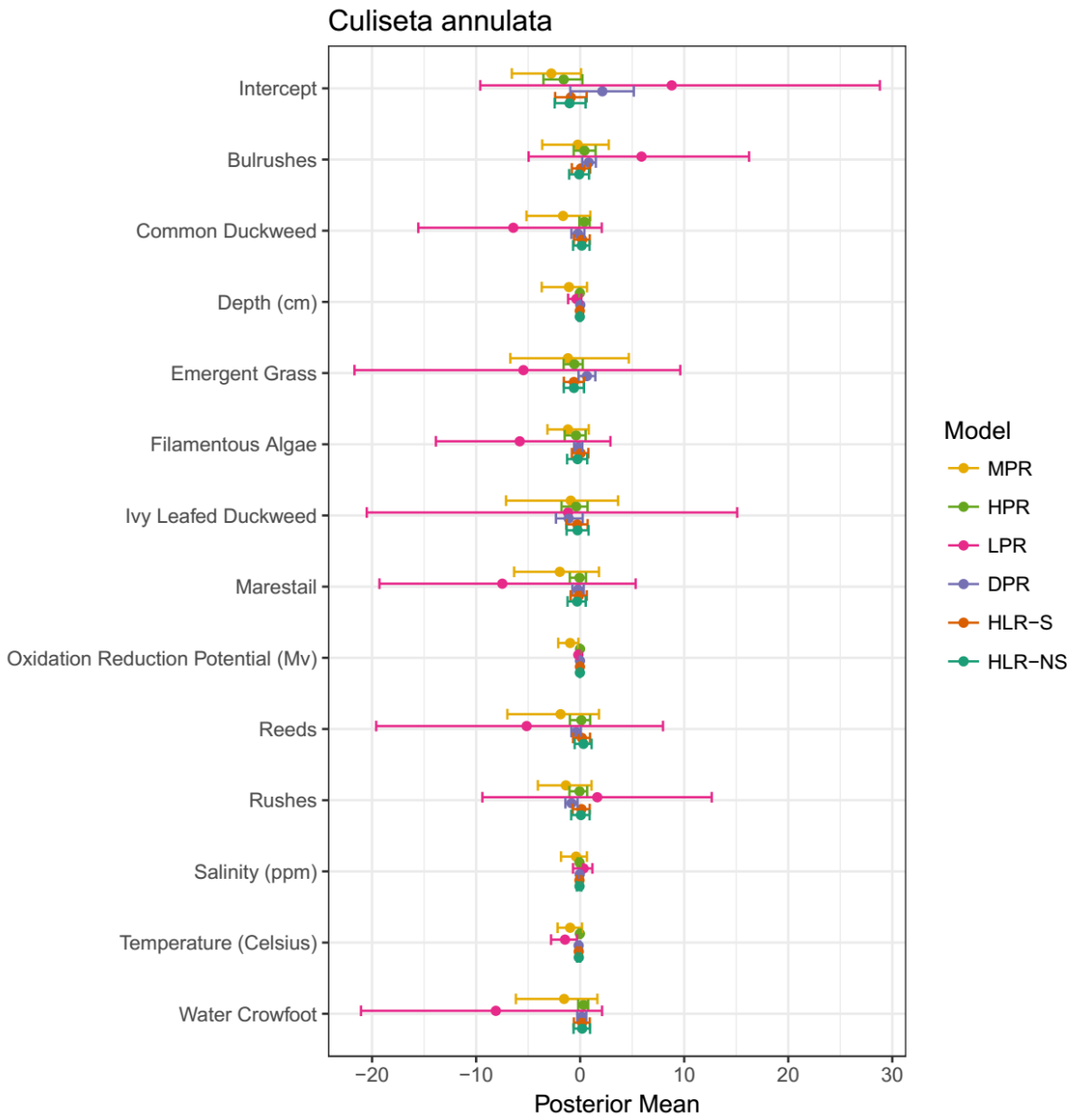


Figure 467: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Culiseta annulata*.

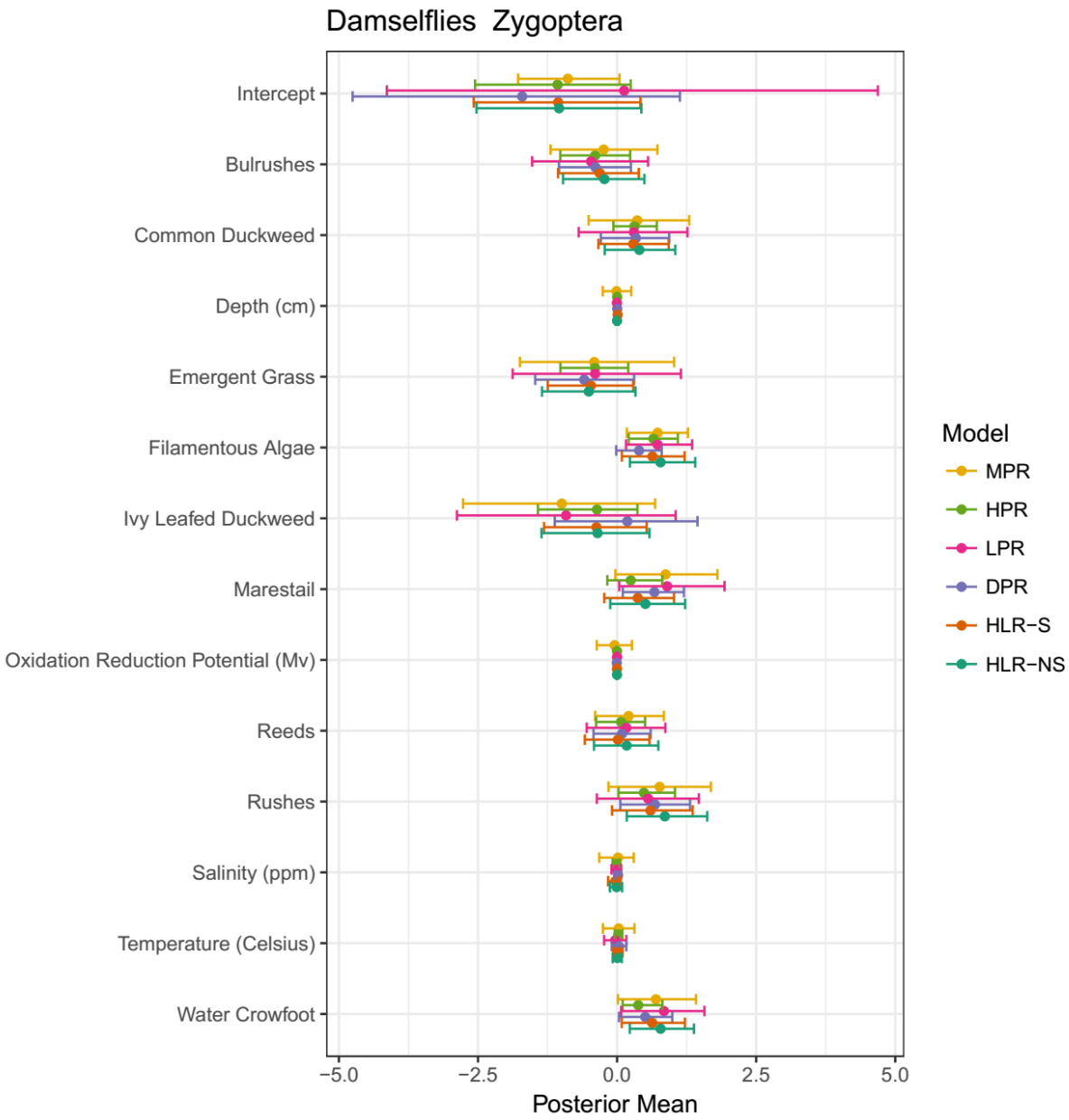


Figure 468: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for Damselflies of the Zygoptera sub-Order.

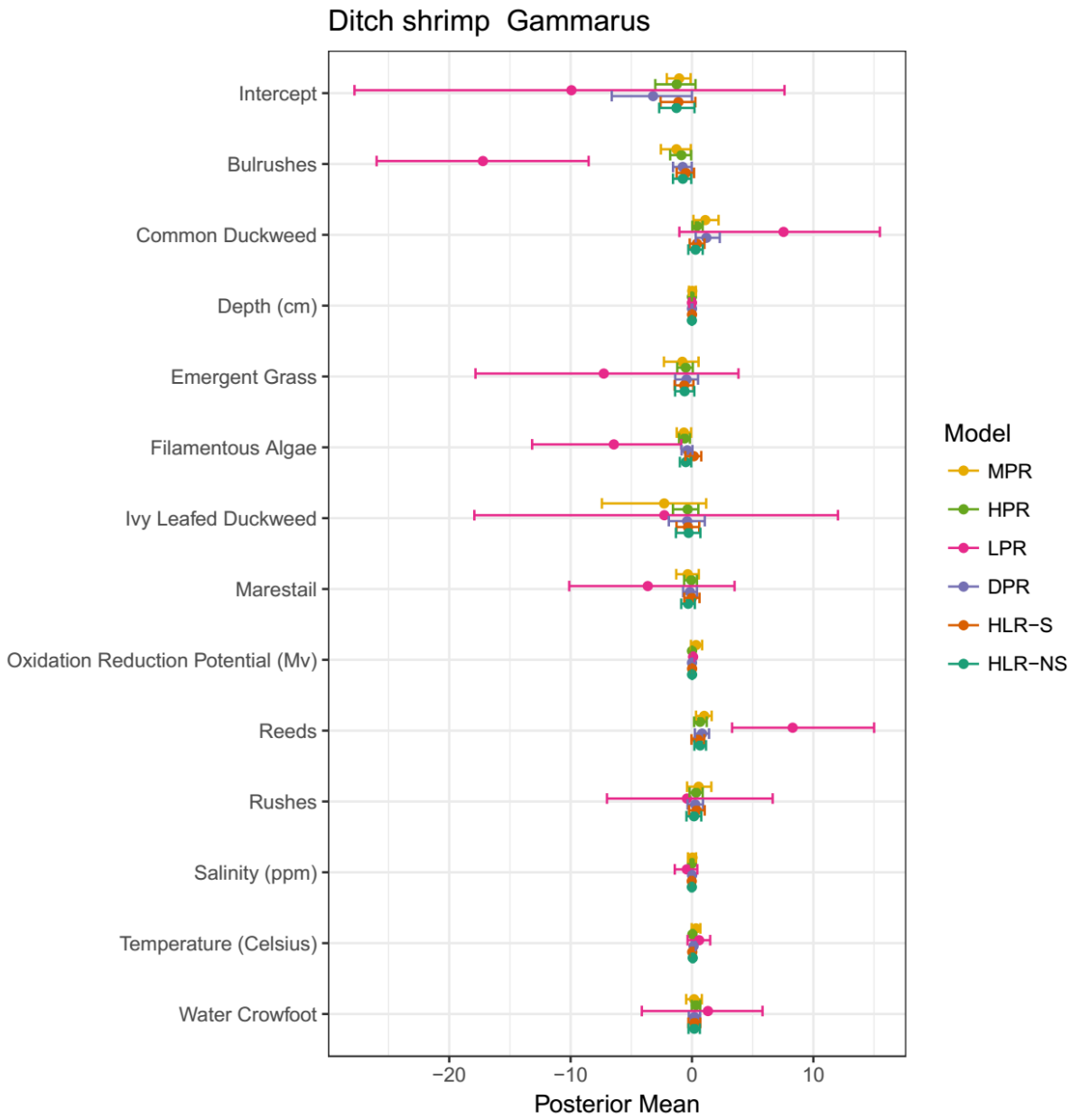


Figure 469: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ditch shrimp, *Gammarus pulex*.

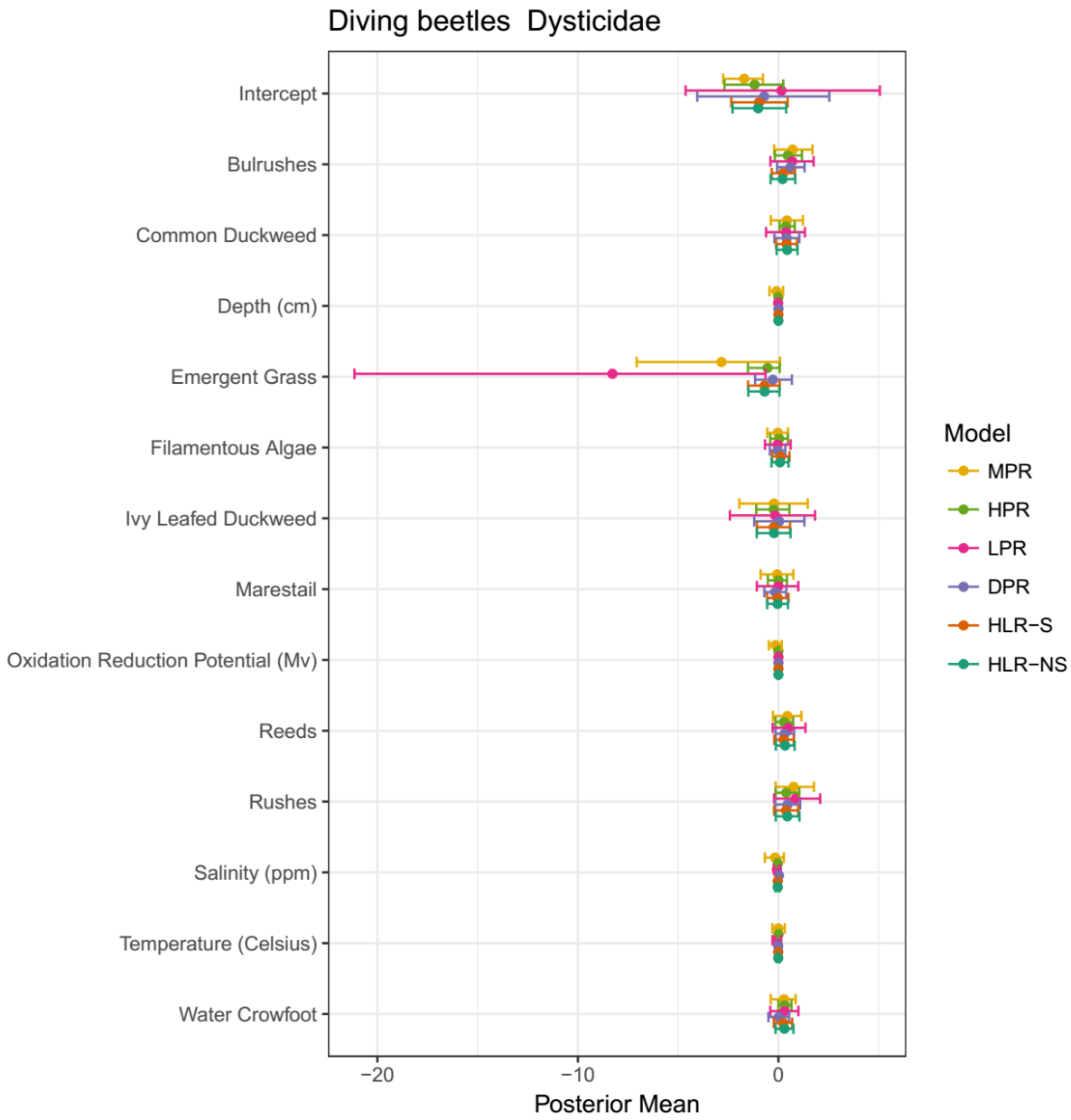


Figure 470: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for diving beetles of the Dytiscidae Family.

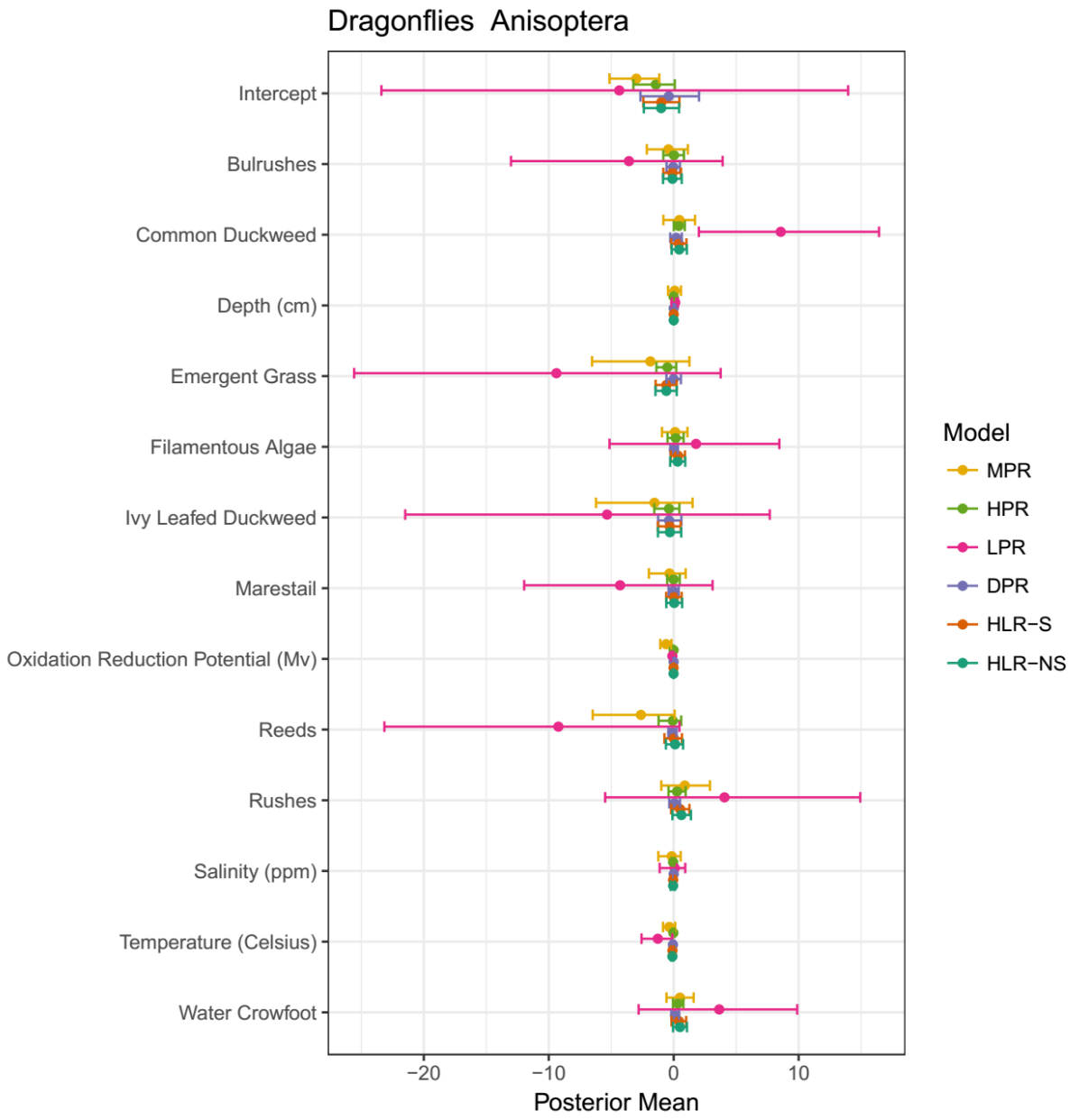


Figure 471: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for dragonflies of the Anisoptera infraorder.

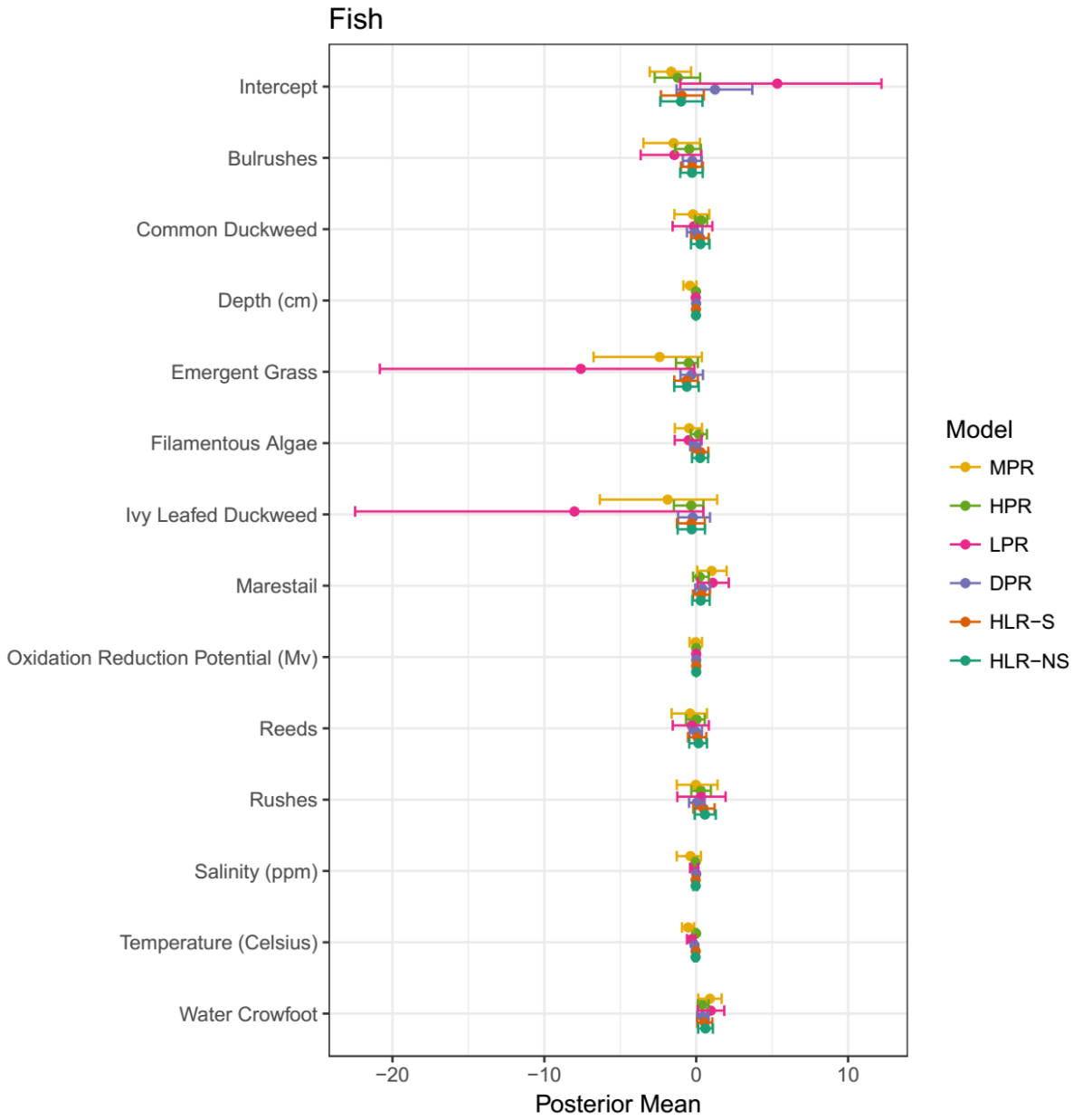


Figure 472: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for freshwater fish of Class Pisces.

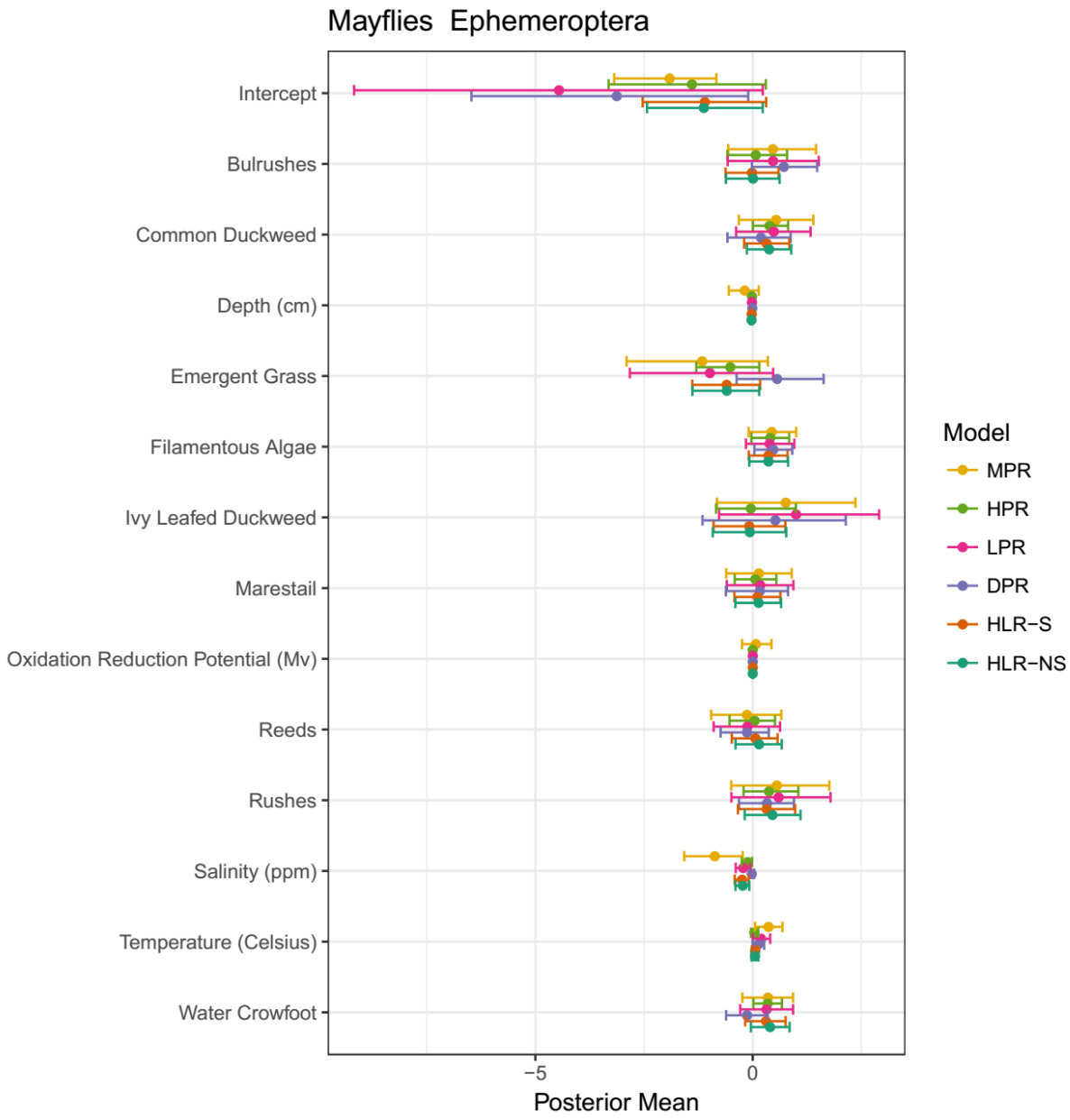


Figure 473: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for mayflies of the order Ephemeroptera.

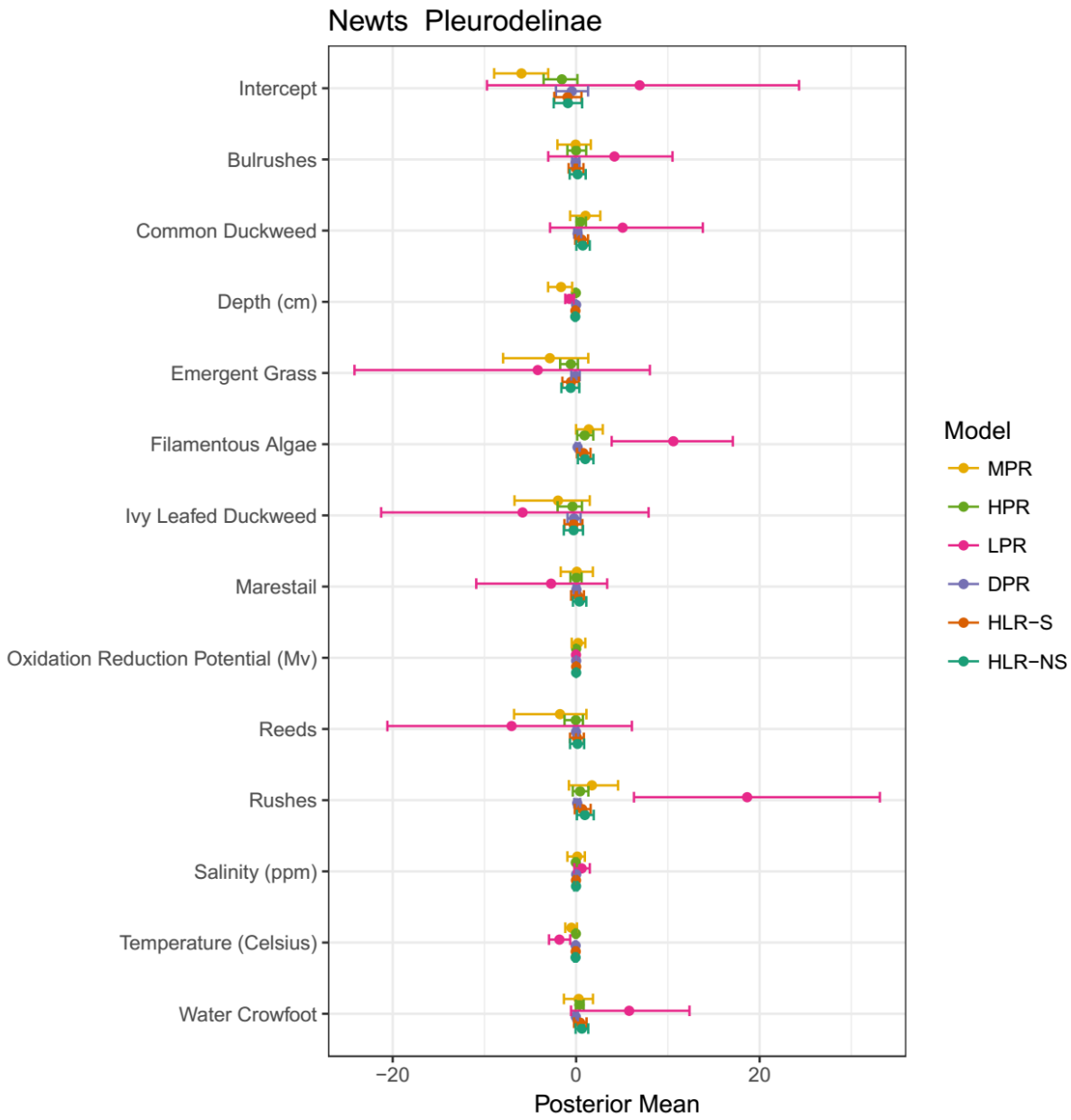


Figure 474: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for newts of the subfamily Pleurodelinae.

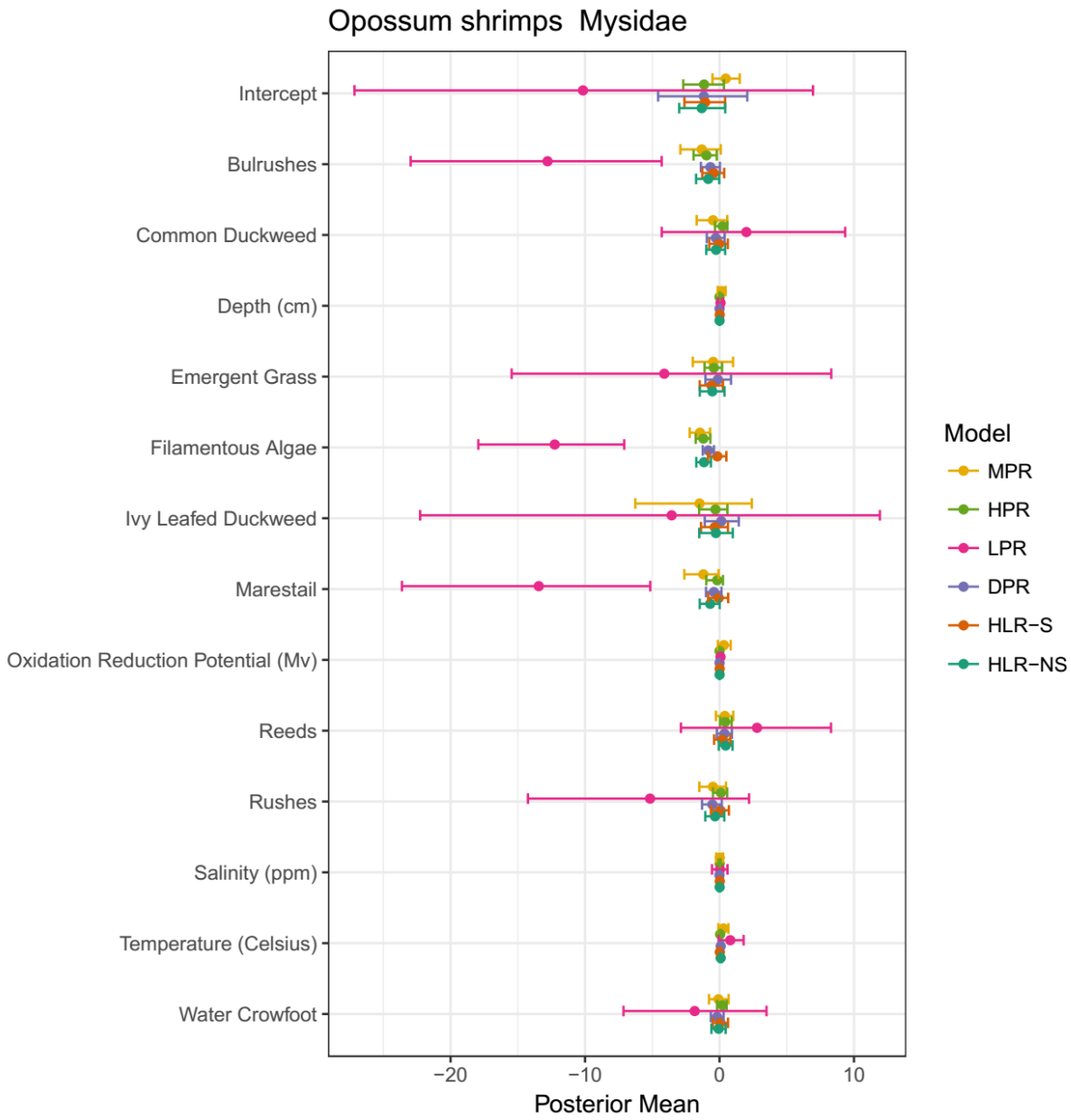


Figure 475: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for opossum shrimp of the order Mysidae.

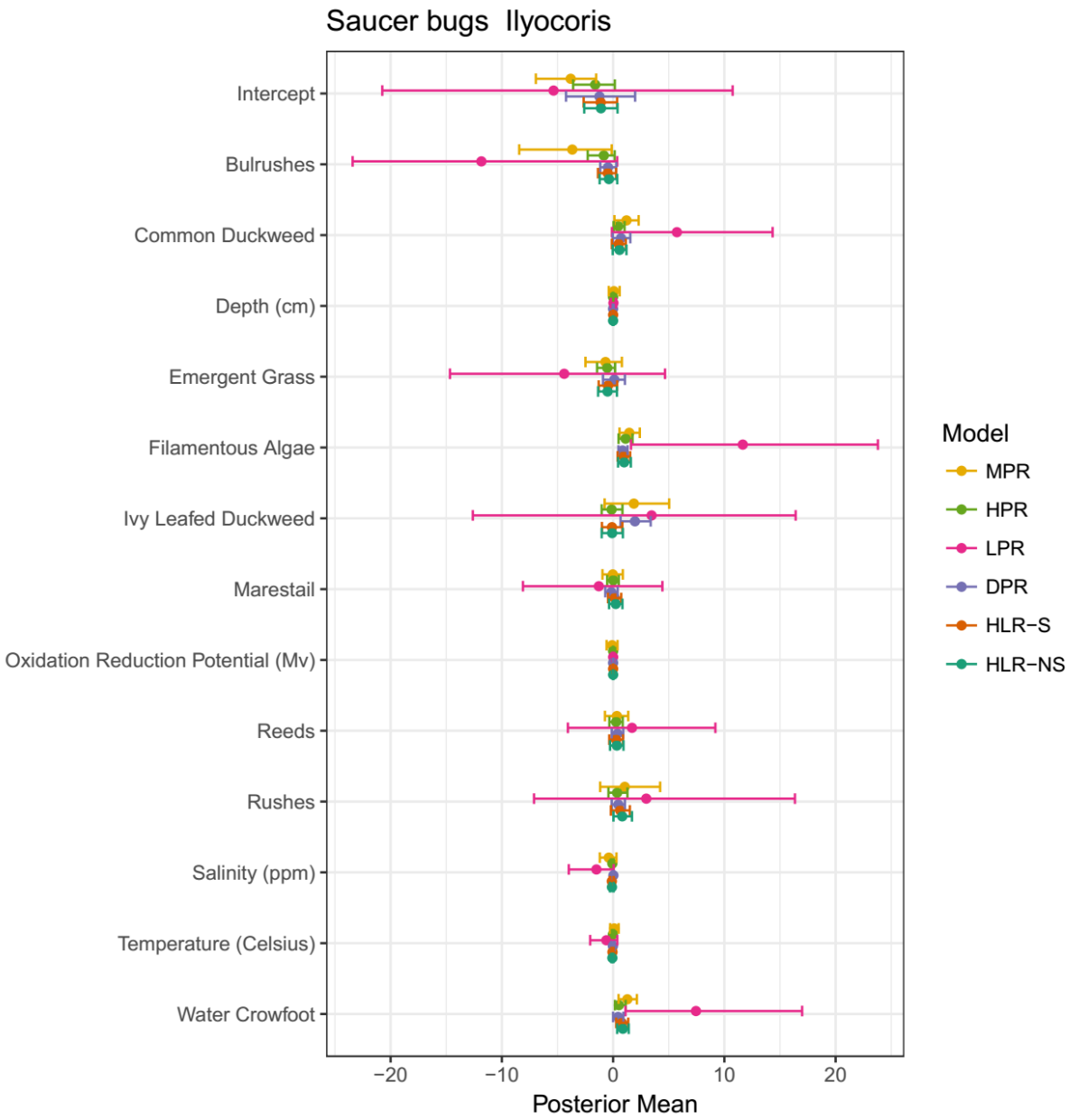


Figure 476: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for saucer bugs of the genus Ilyocoris.

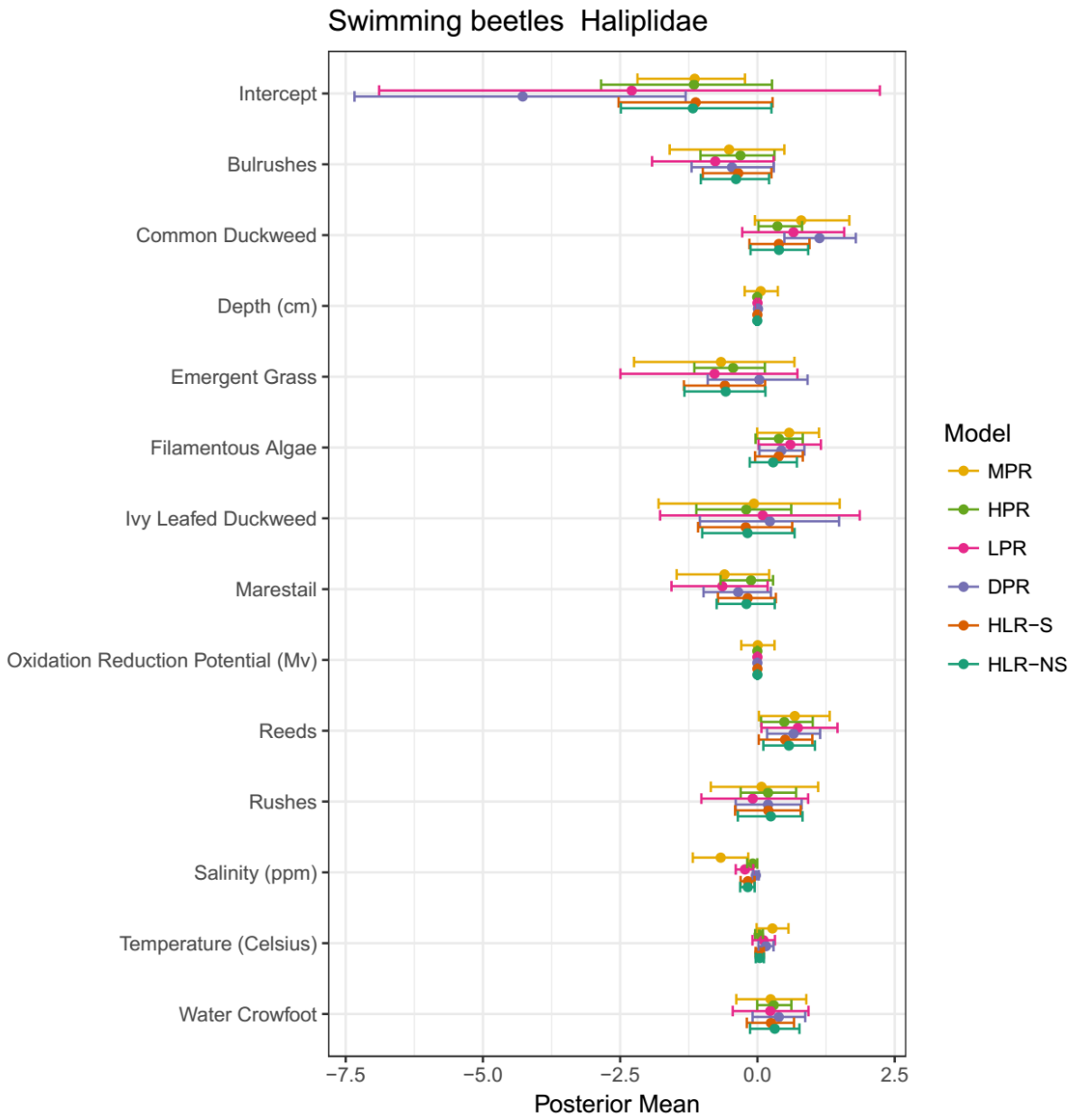


Figure 477: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for swimming beetles of the family Haliplidae.

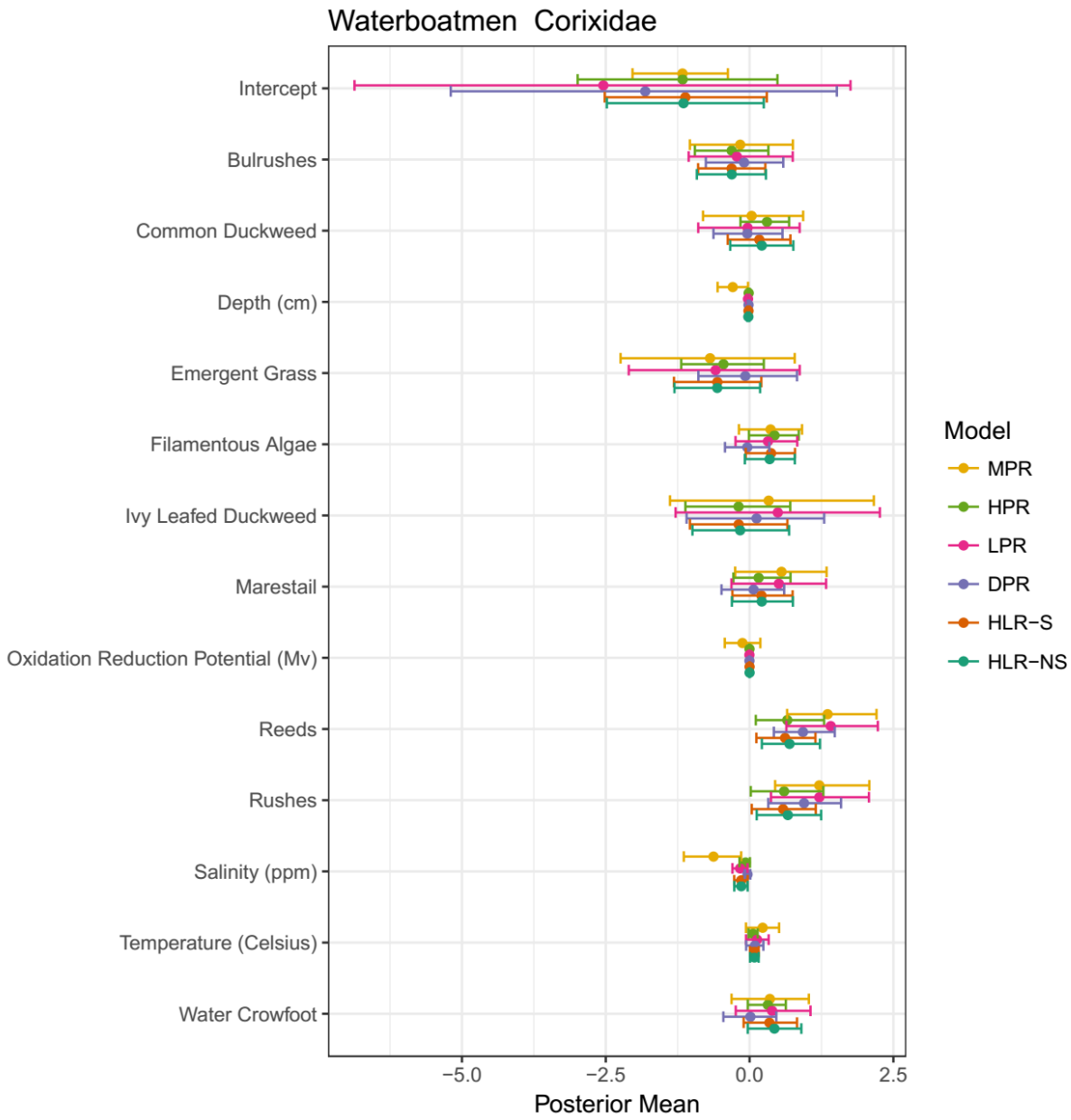


Figure 478: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for waterboatmen of the family Corixidae.

APPENDIX C

LPR FITTING ISSUE

To combat the issue of non-convergence due to issues of complete separation in binary variables for the LPR model we have fit it again to the mosquito dataset using a more informative prior on the regression coefficients. The original prior was $\beta \sim N(0, 10)$ and it has been replaced with a more informative $\beta \sim N(0, 1)$ on the non-intercept regression coefficients. No other models were re-fit for this comparison, so these figures can be directly compared with Appendix S1Figures 458 – 473 to examine the difference in LPR estimates under different priors.

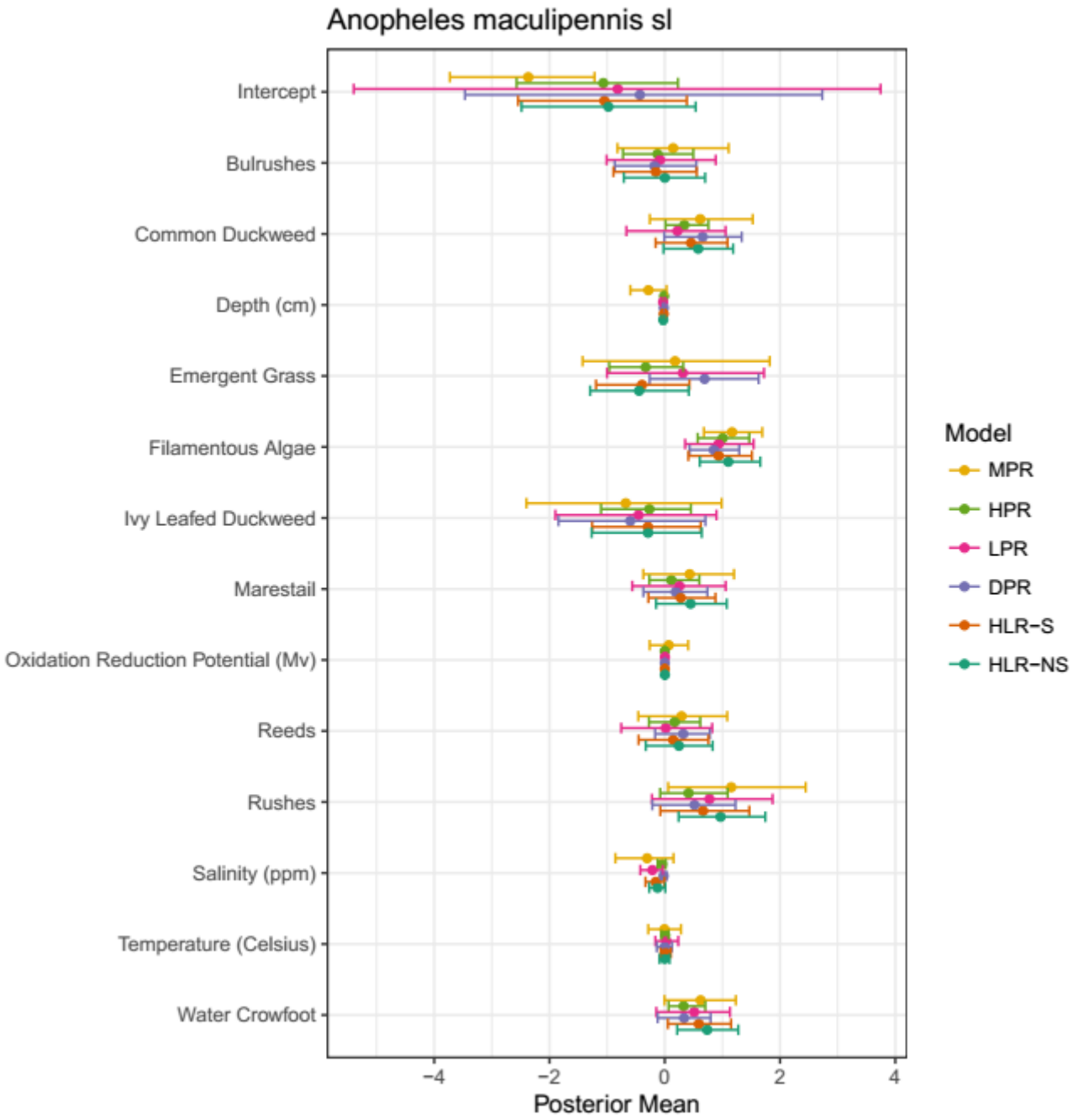


Figure 1479: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Anopheles maculipennis s.l.*

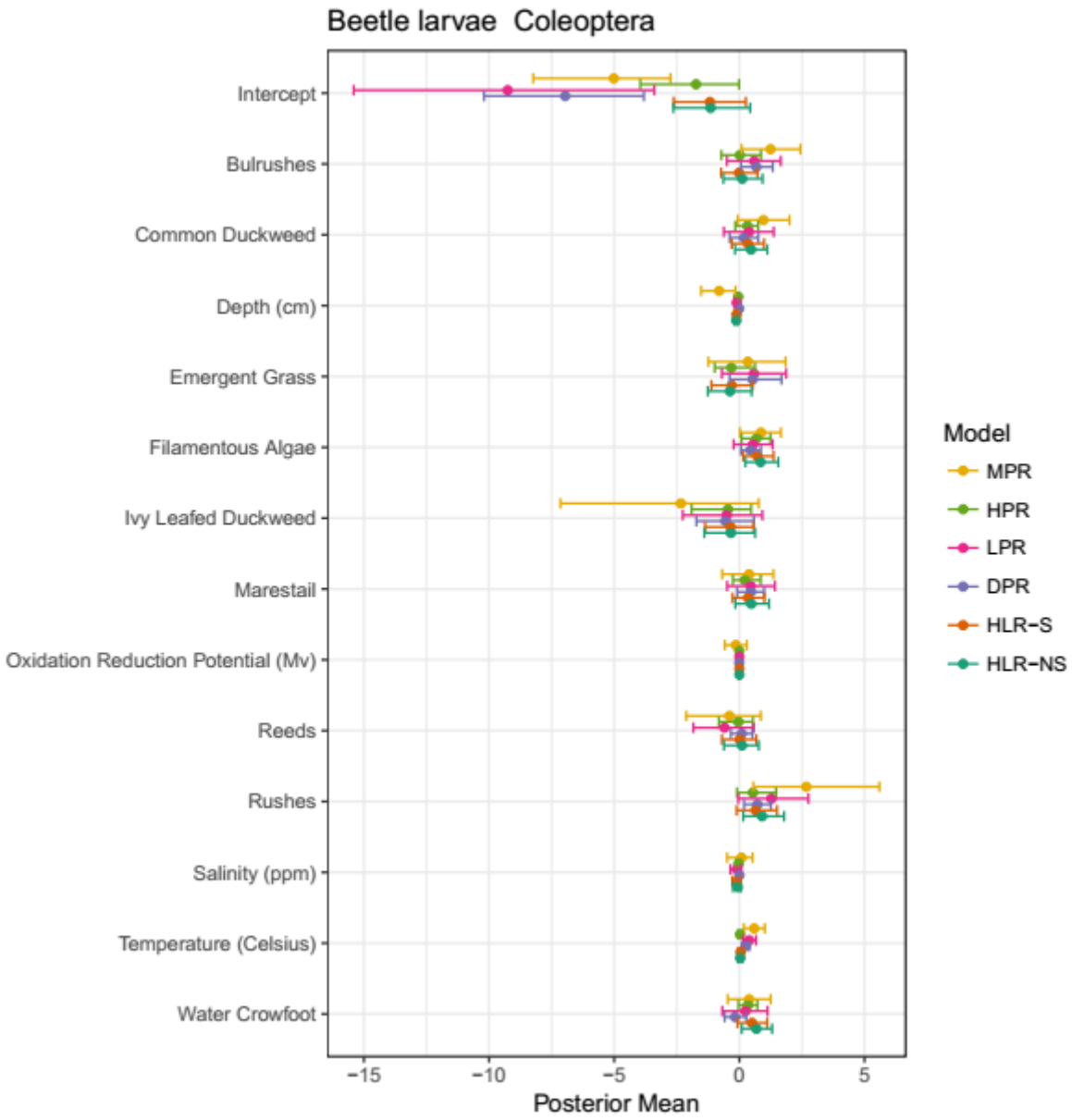


Figure 2: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for Coleoptera beetle larvae.

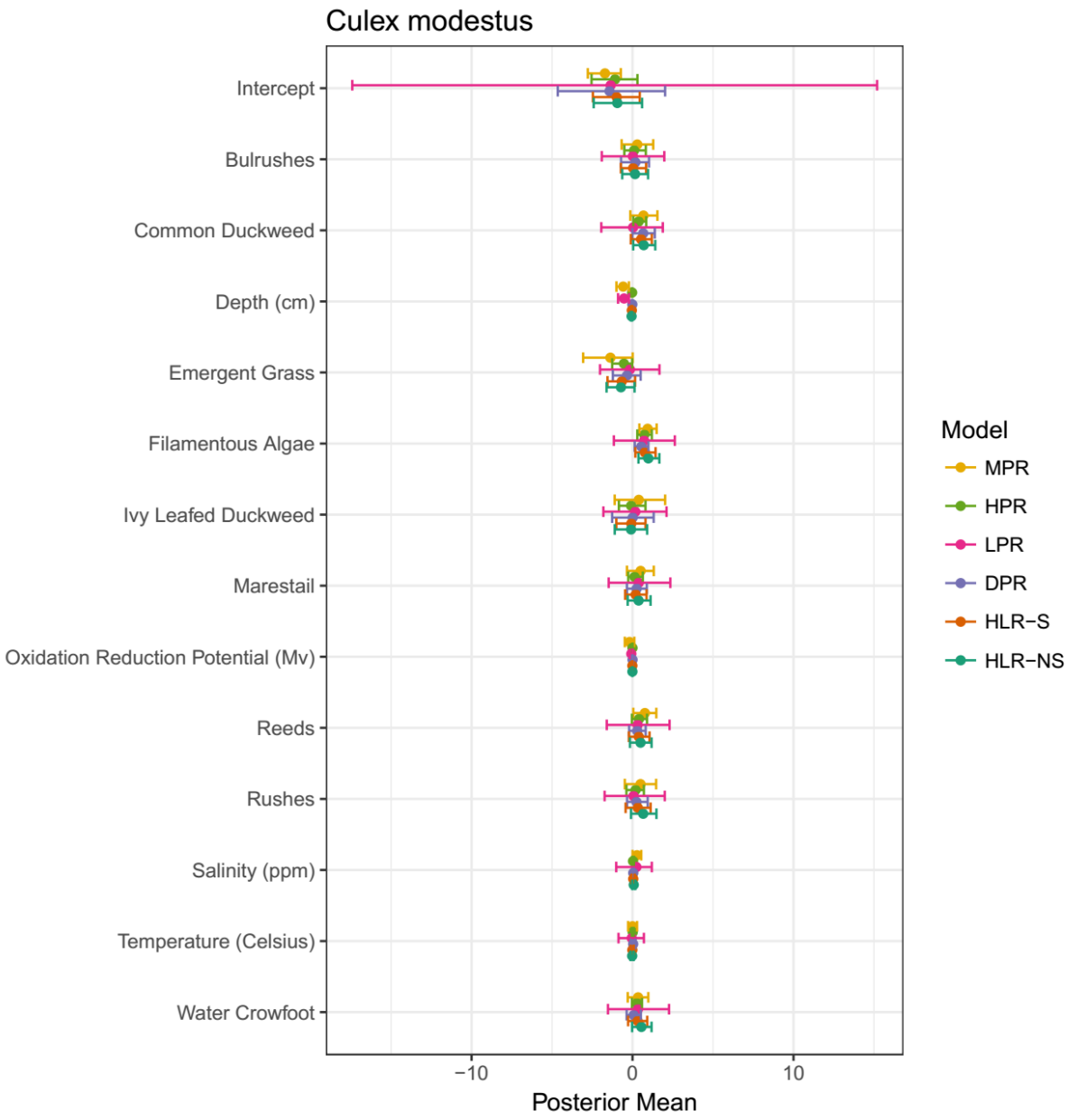


Figure 3: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Culex modestus*.

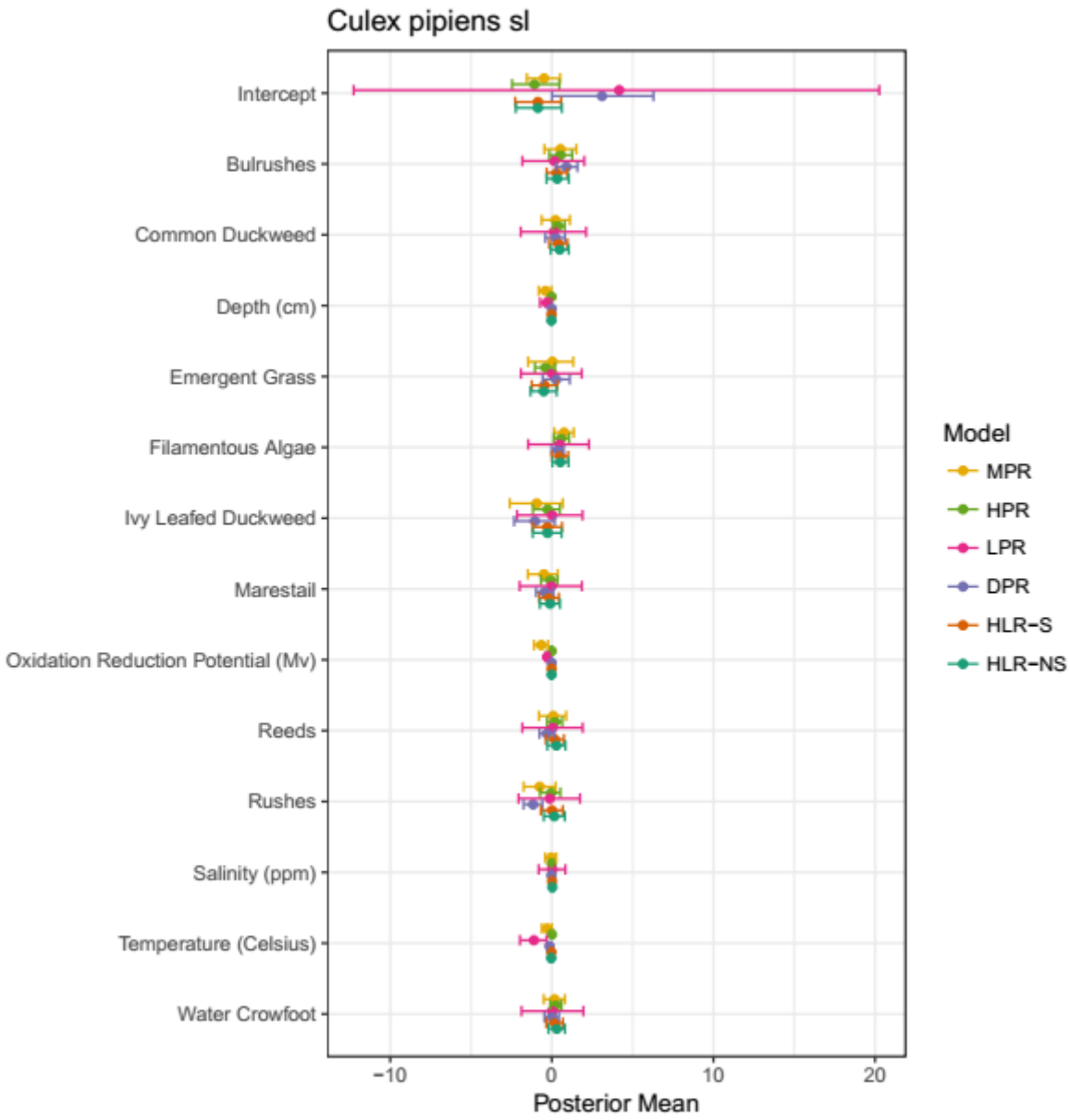


Figure 4: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Culex pipiens s.l.*

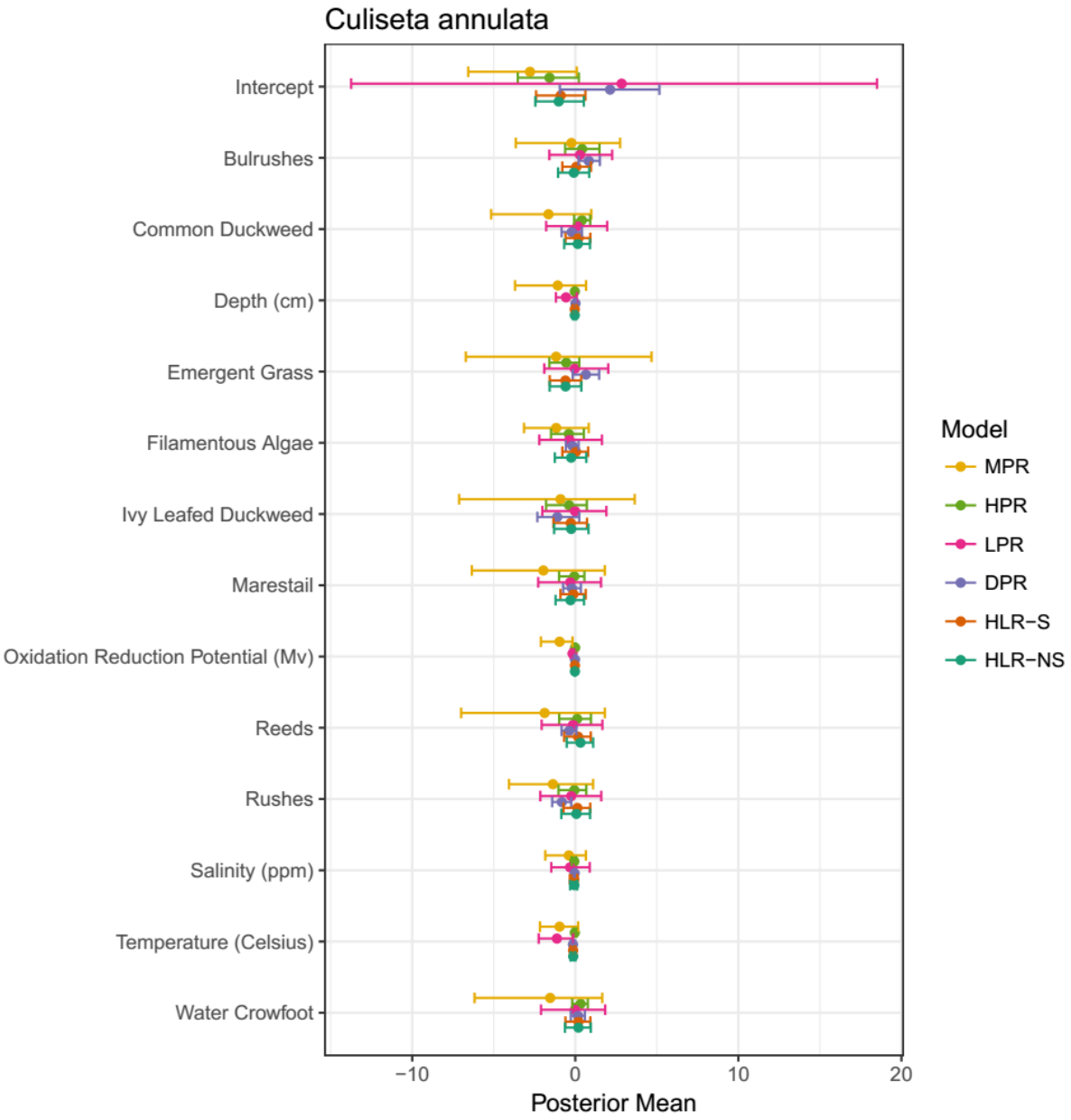


Figure 5: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for *Culiseta annulata*.

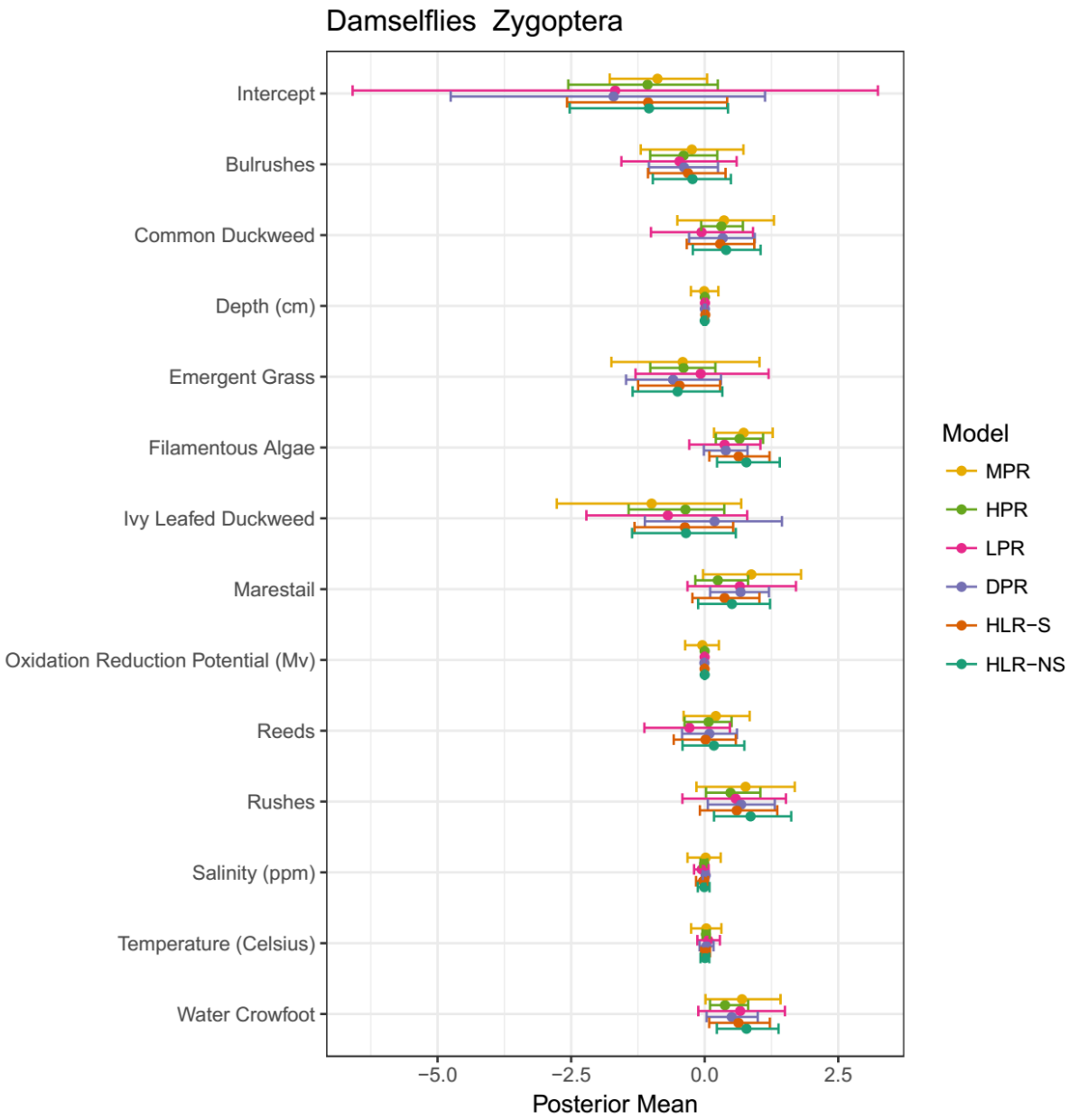


Figure 6: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for Damselflies of the Zygoptera sub-Order.

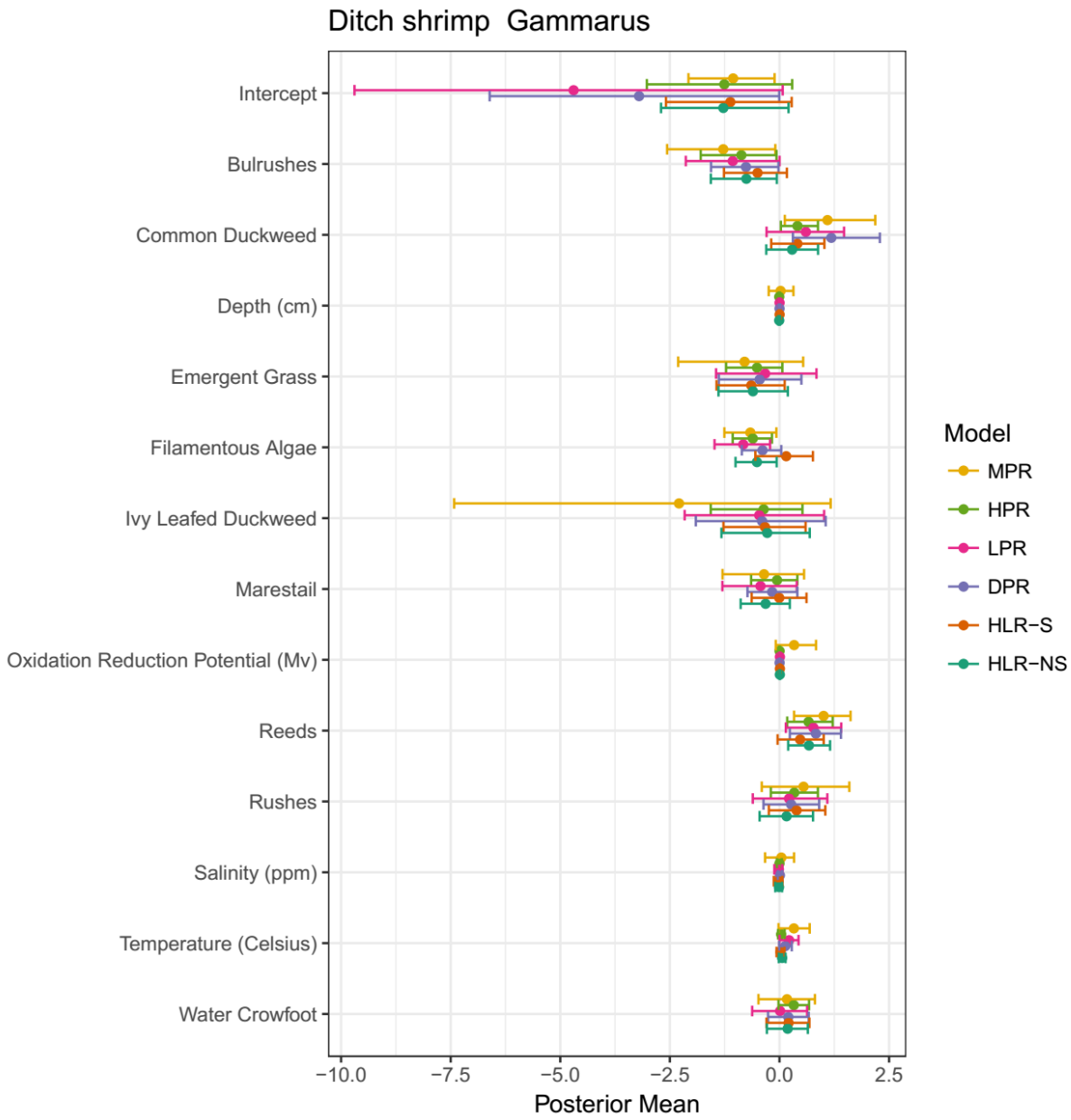


Figure 7: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for the ditch shrimp, *Gammarus pulex*.

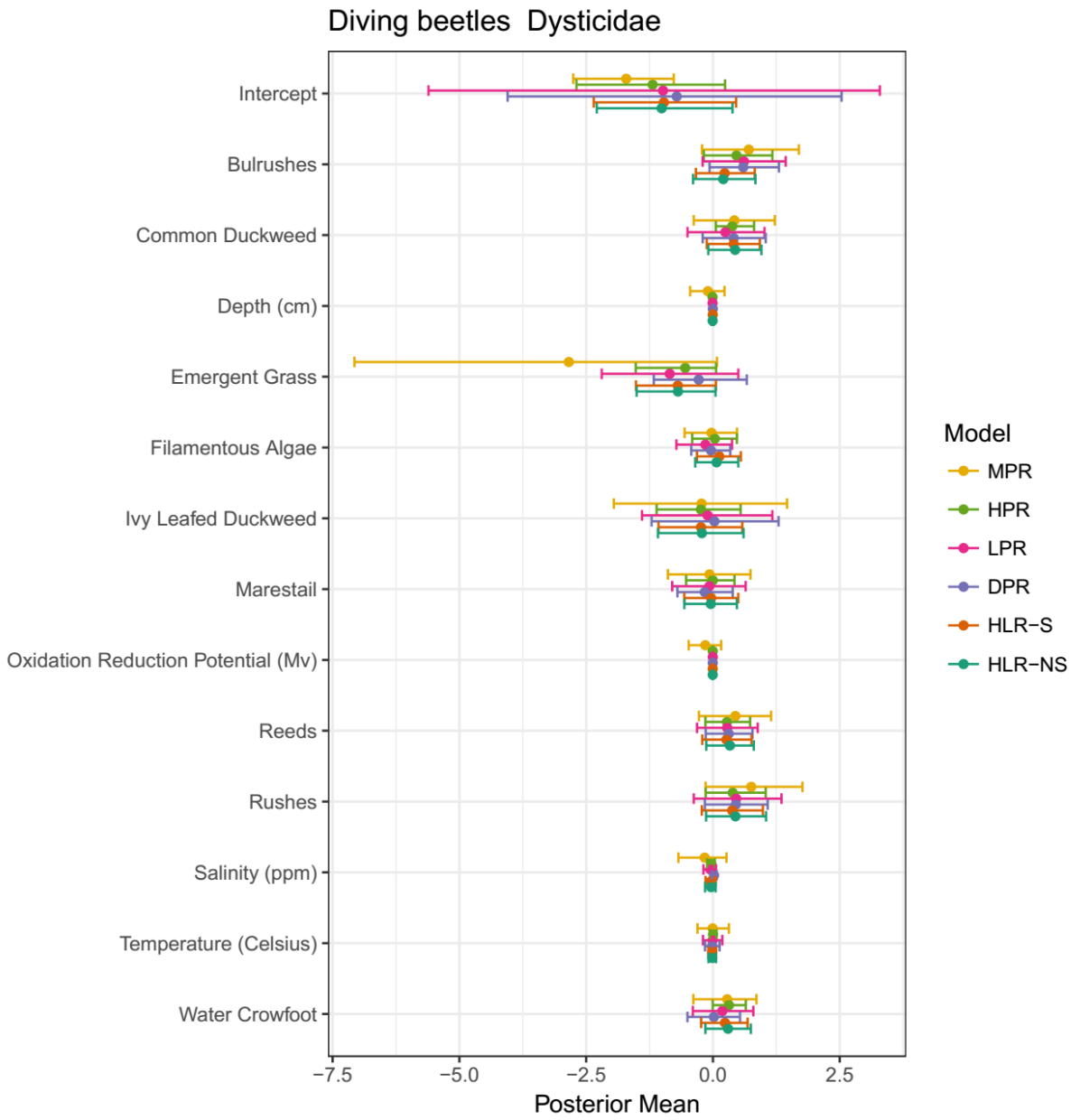


Figure 8: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for diving beetles of the Dytiscidae Family.

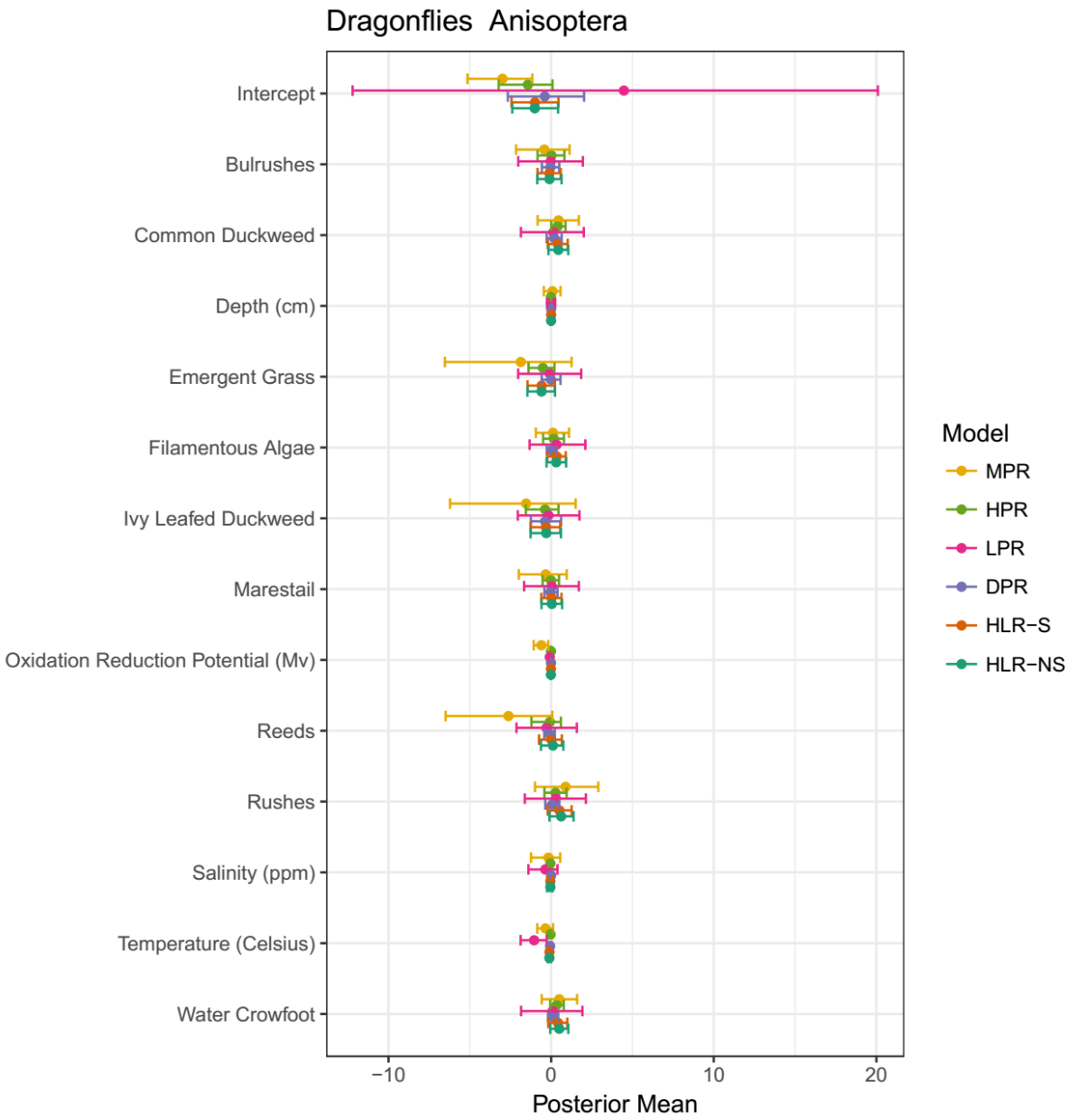


Figure 9: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for dragonflies of the Anisoptera infraorder.

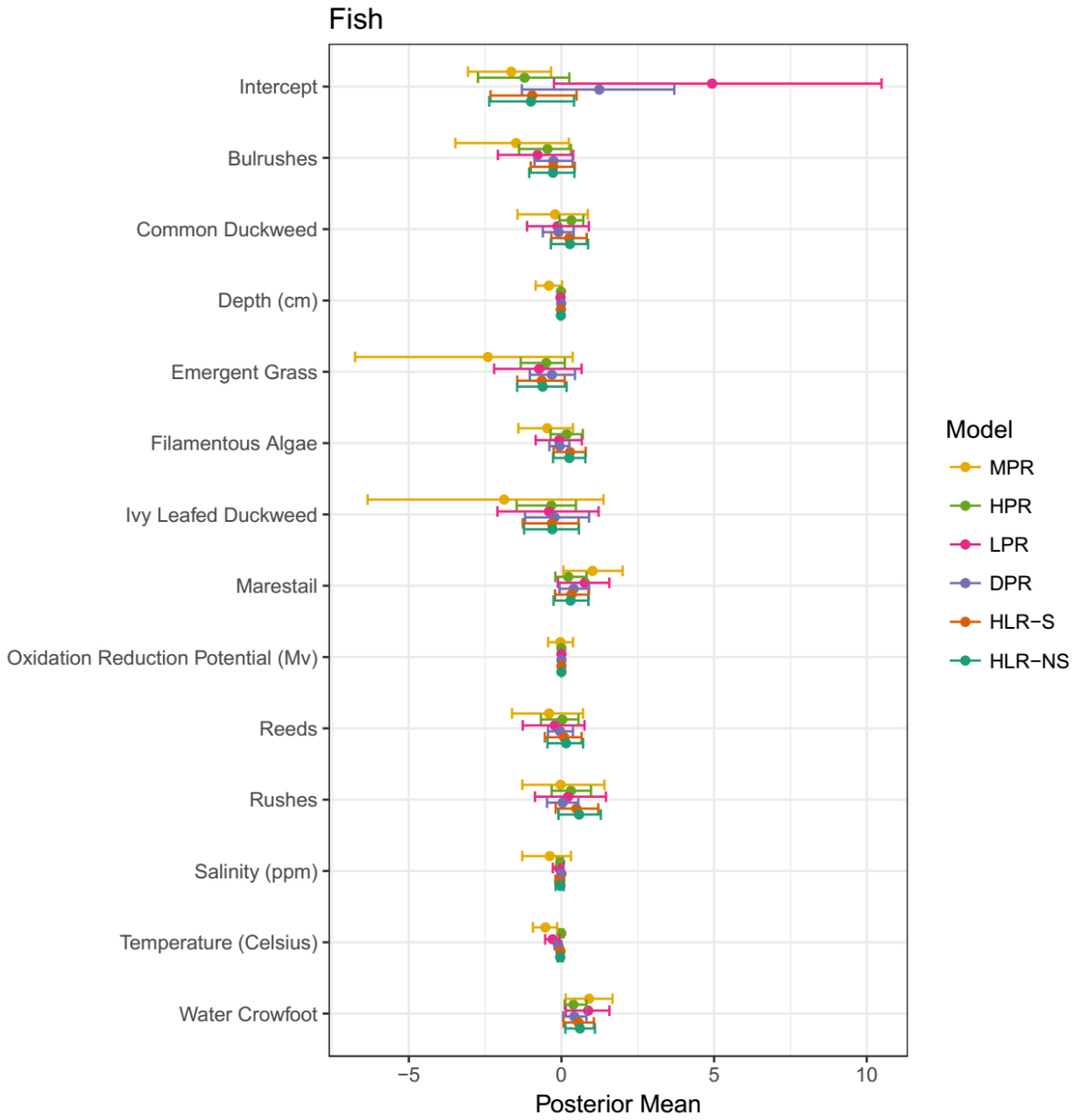


Figure 10: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for freshwater fish of Class Pisces.

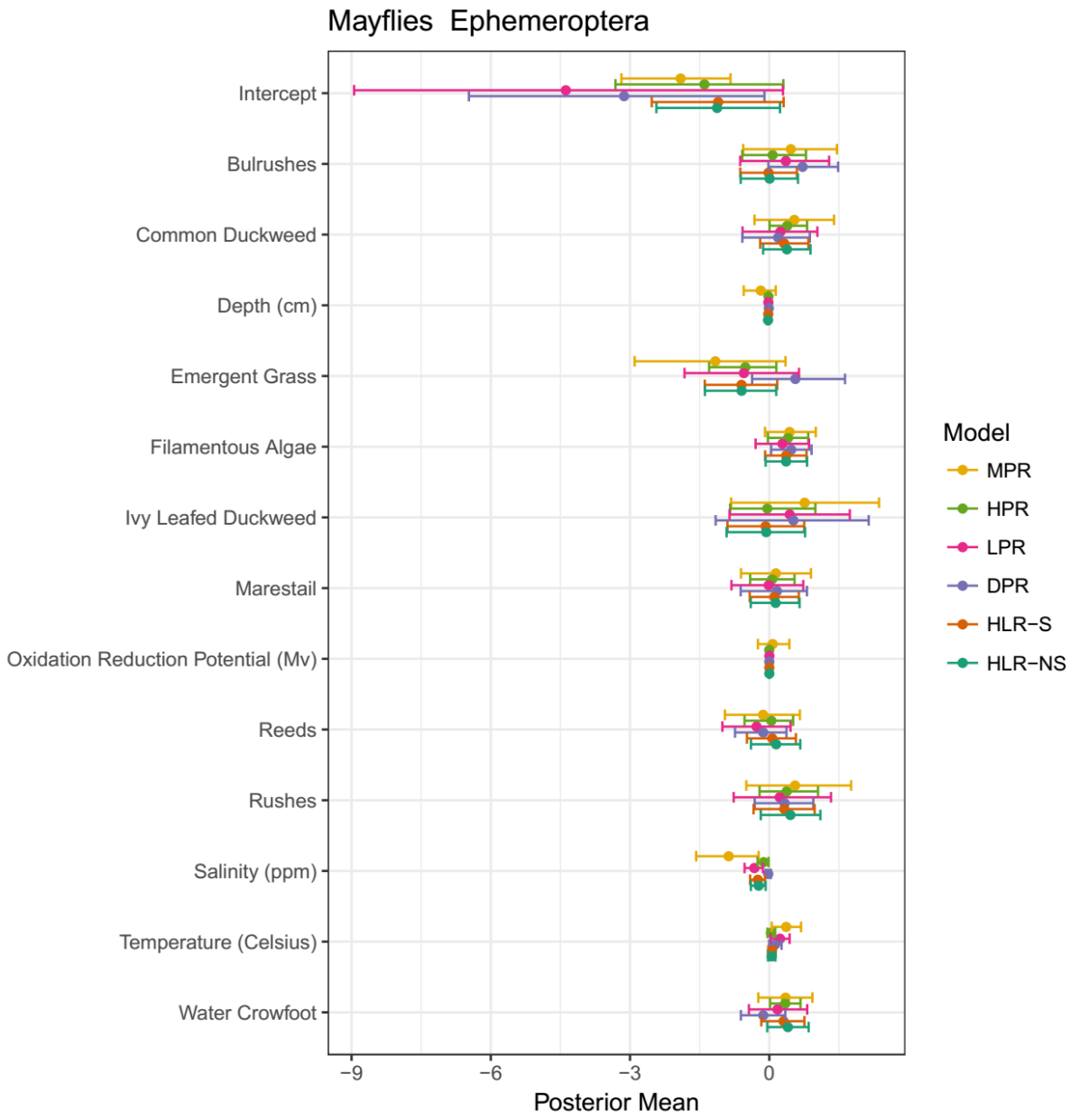


Figure 11: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for mayflies of the order Ephemeroptera.

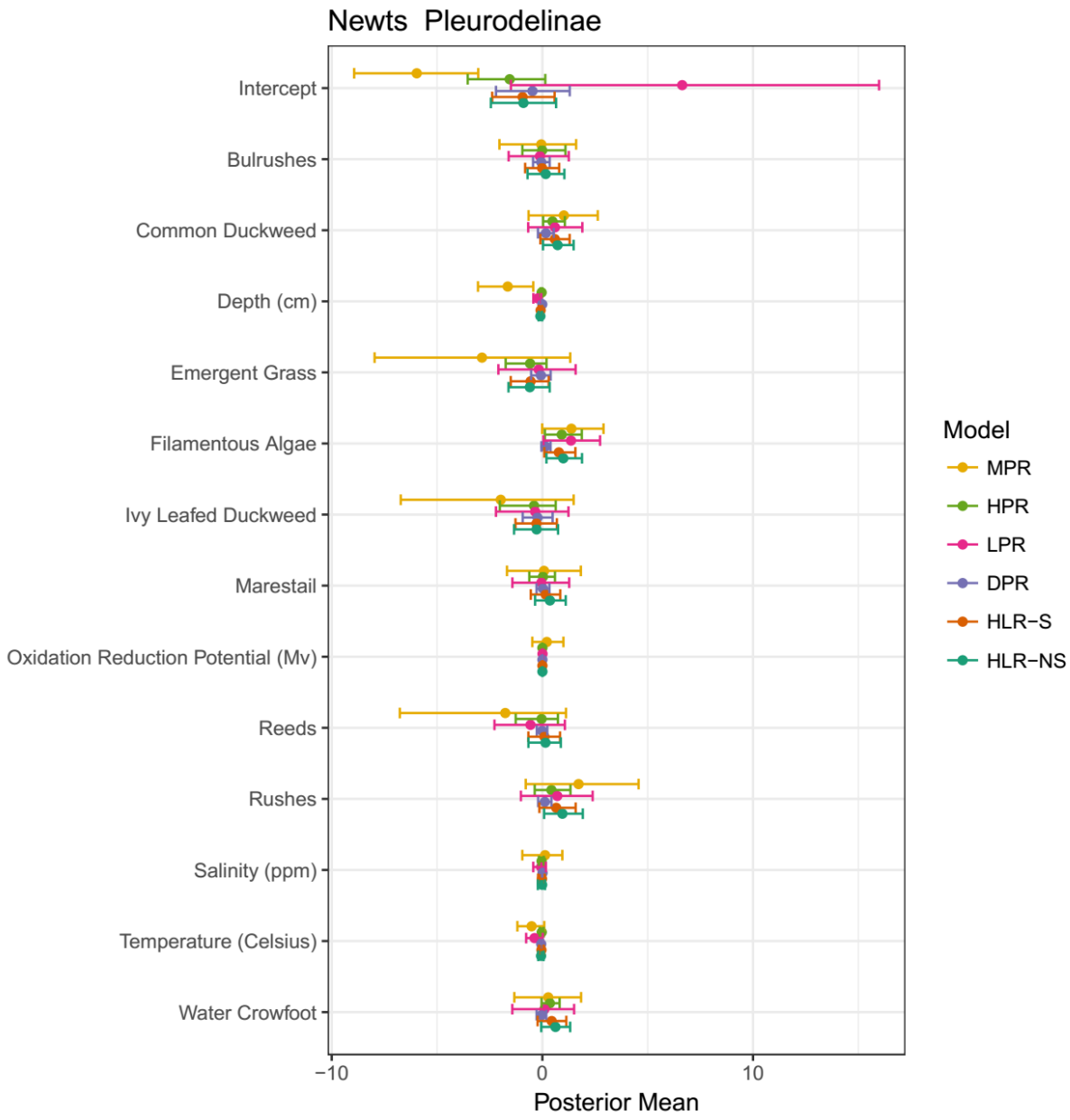


Figure 12: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for newts of the subfamily Pleurodelinae.

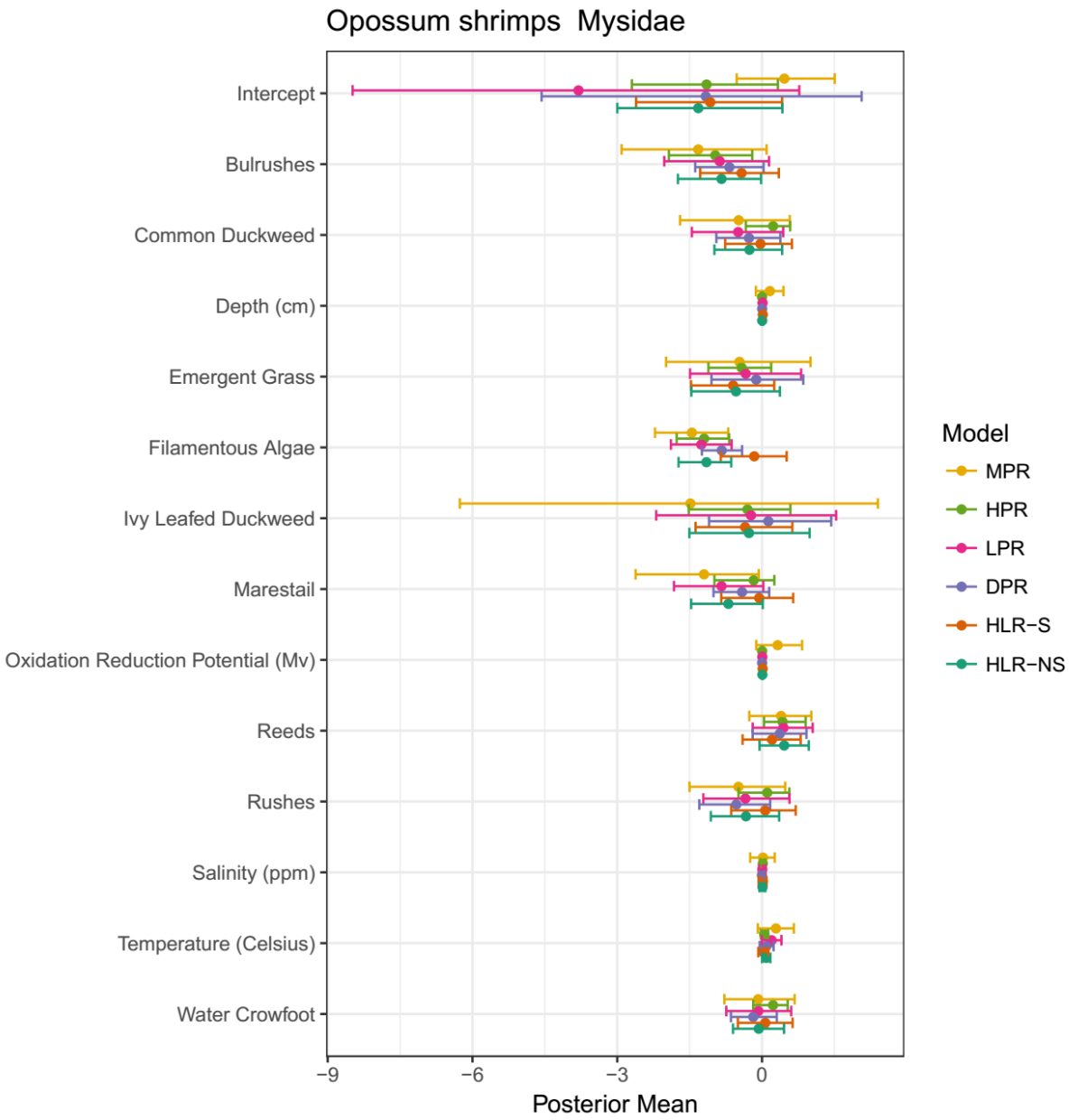


Figure 13: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for opossum shrimp of the order Mysidae.

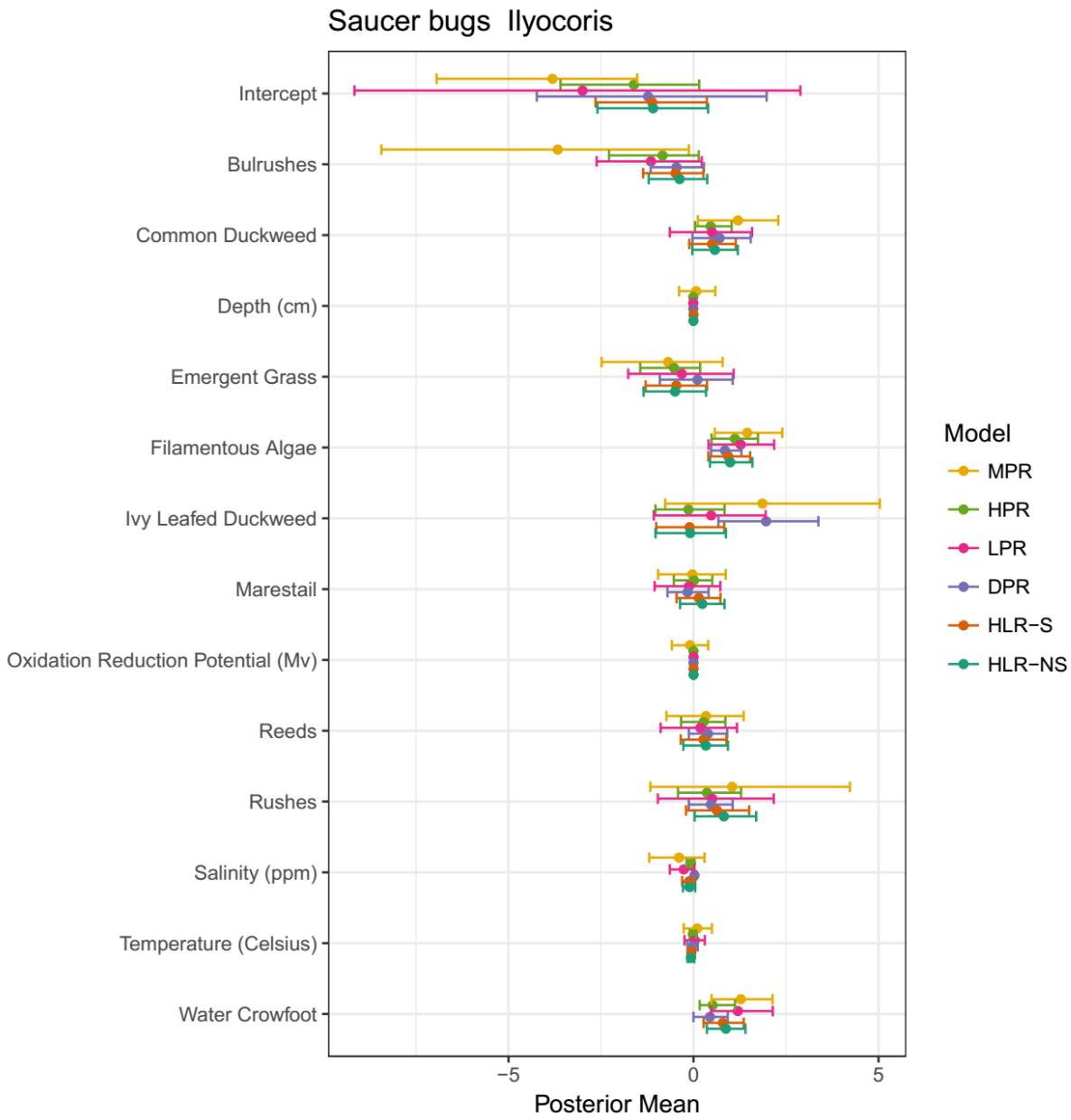


Figure 14: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for saucer bugs of the genus Ilyocoris.

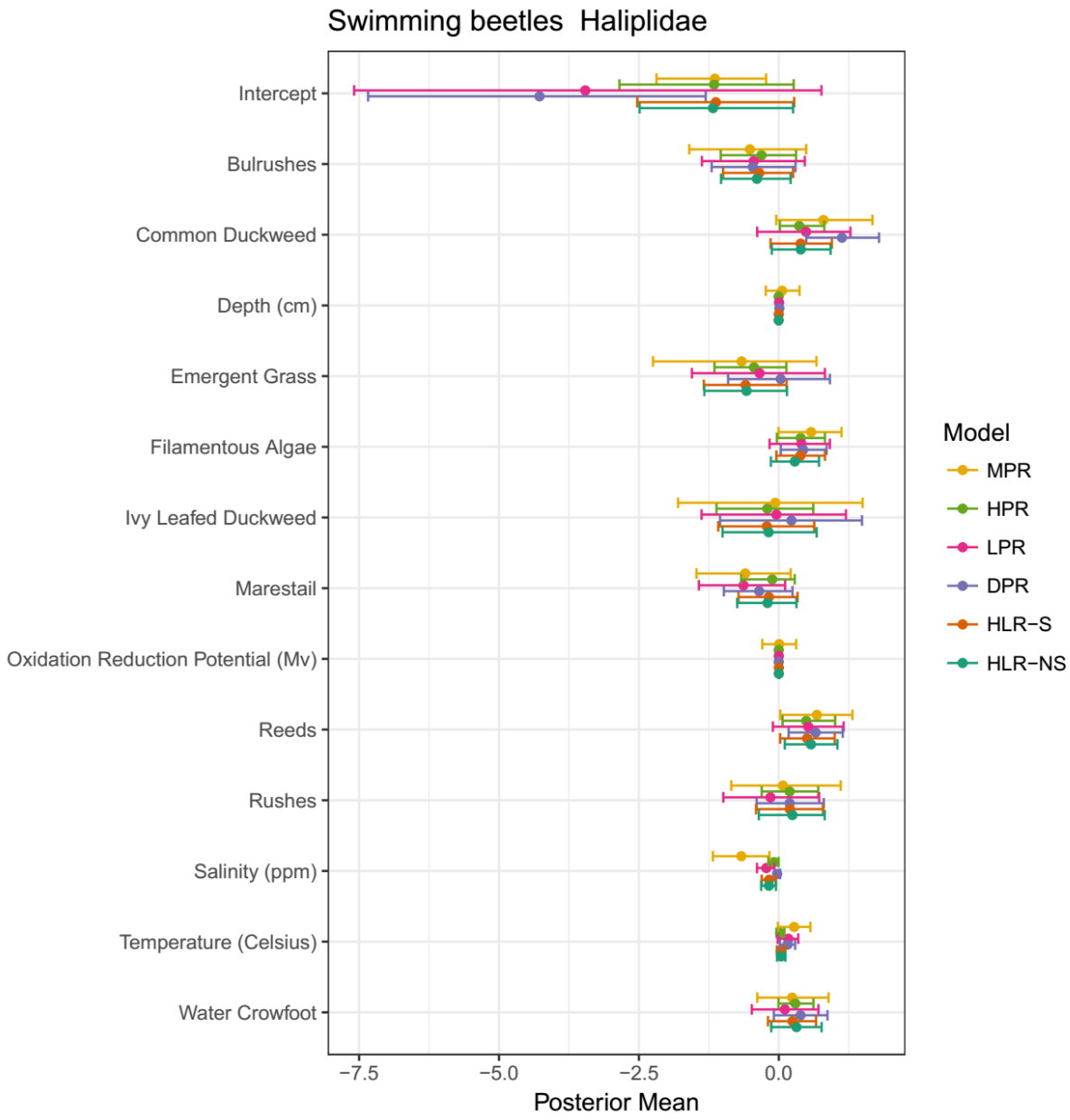


Figure 15: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for swimming beetles of the family Haliplidae.

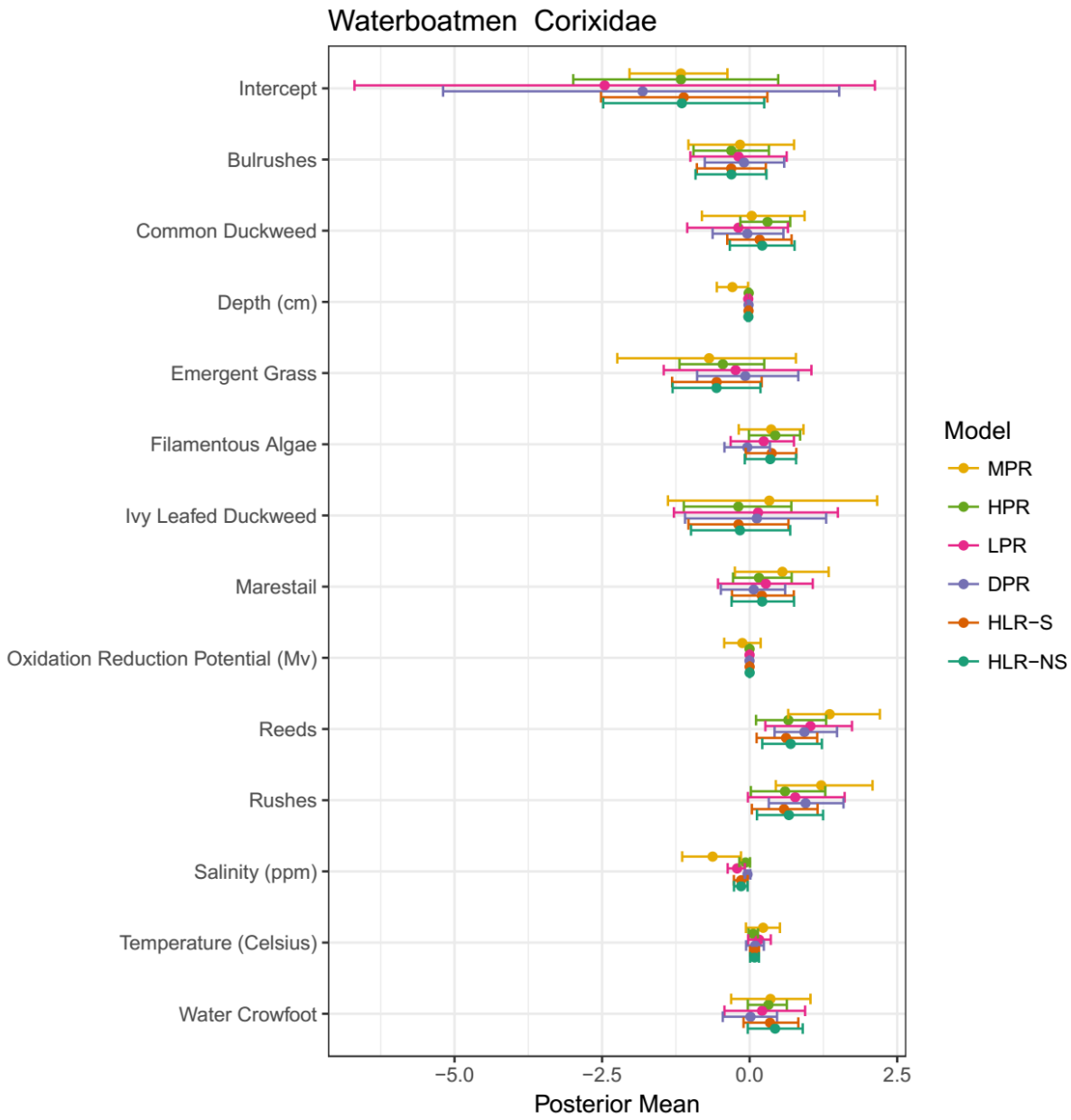


Figure 16: Posterior means and 95% credible intervals of regression coefficient estimates for all compatible datasets for waterboatmen of the family Corixidae.

APPENDIX D

REGRESSION COEFFICIENT UNCERTAINTY METRICS

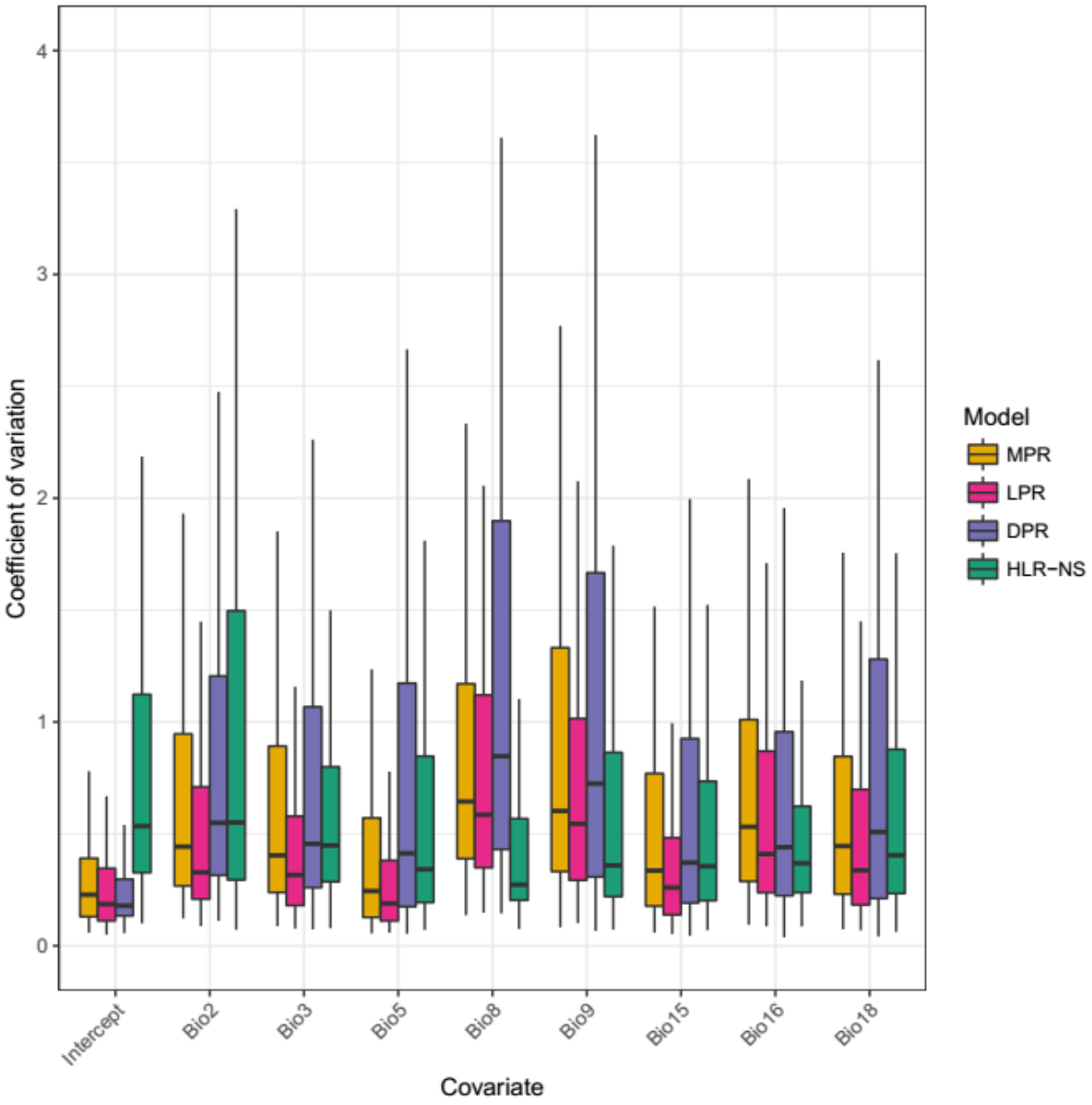


Figure 480: Coefficient of variation of the regression coefficient estimates of models on the bird dataset.

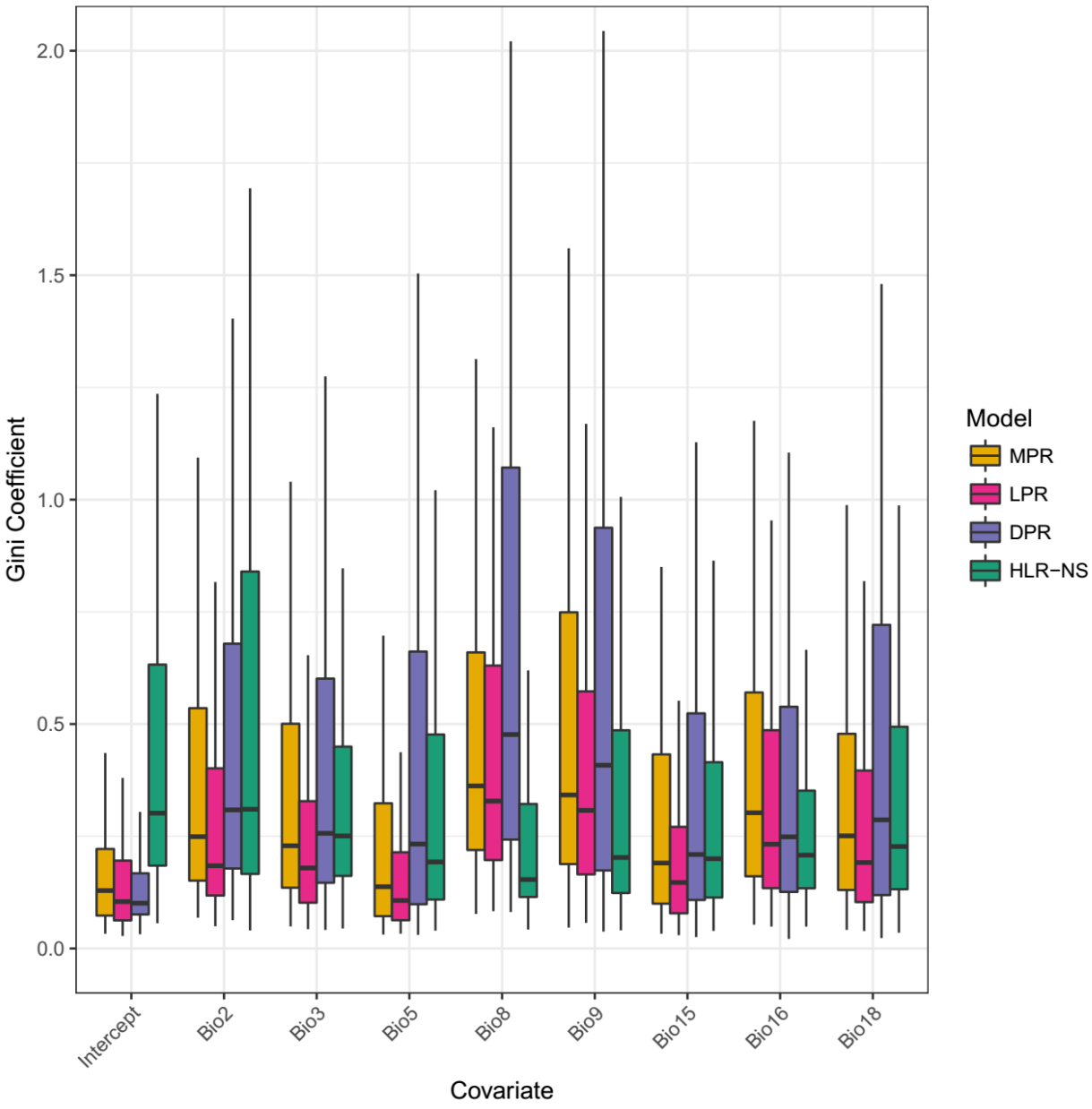


Figure 481: Gini coefficient of the regression coefficient estimates of models on the bird dataset.

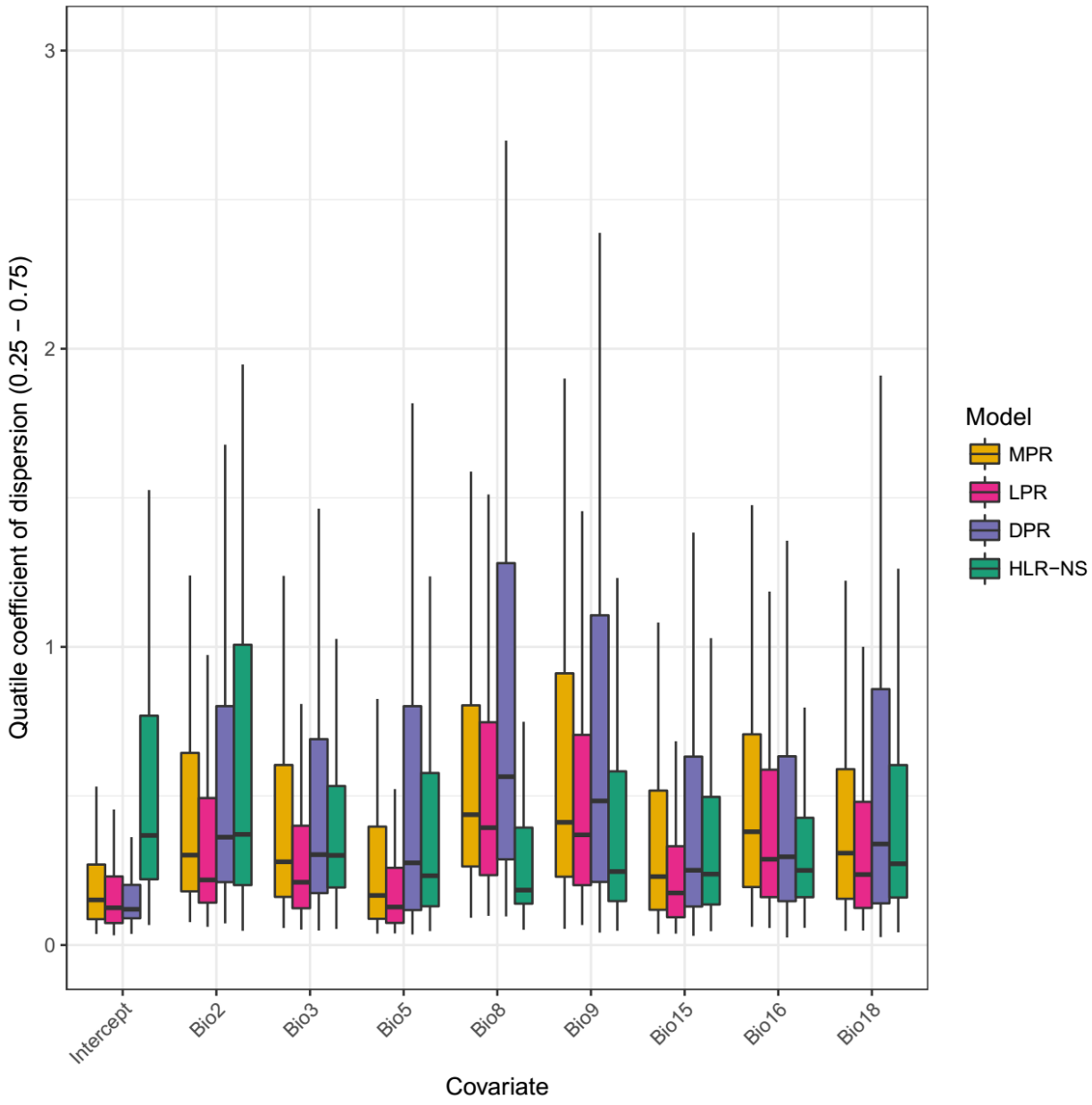


Figure 482: Quantile coefficient of dispersion of the regression coefficient estimates of models on the bird dataset.

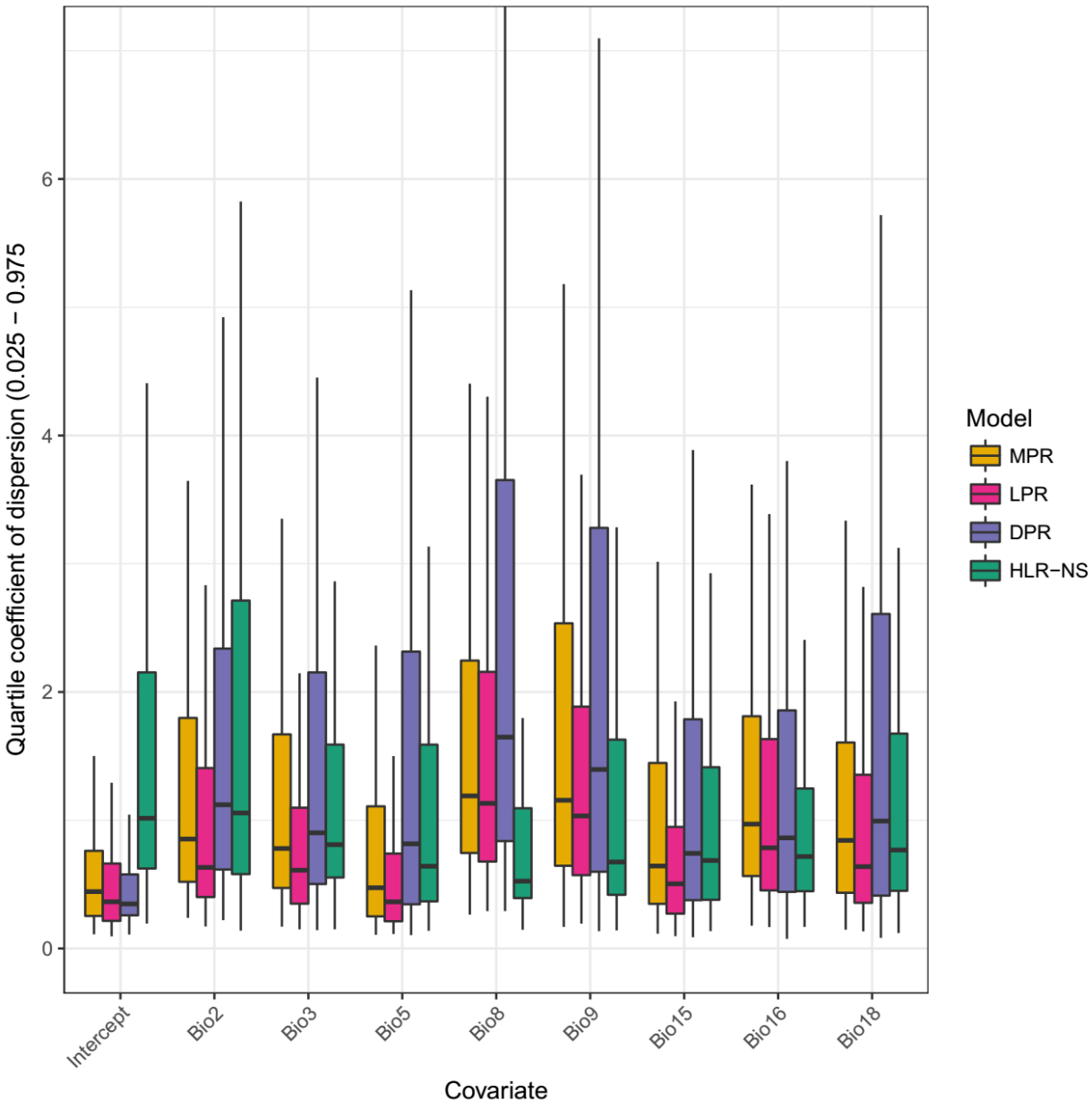


Figure 483: 90% quantile coefficient of dispersion of the regression coefficient estimates of models on the bird dataset.

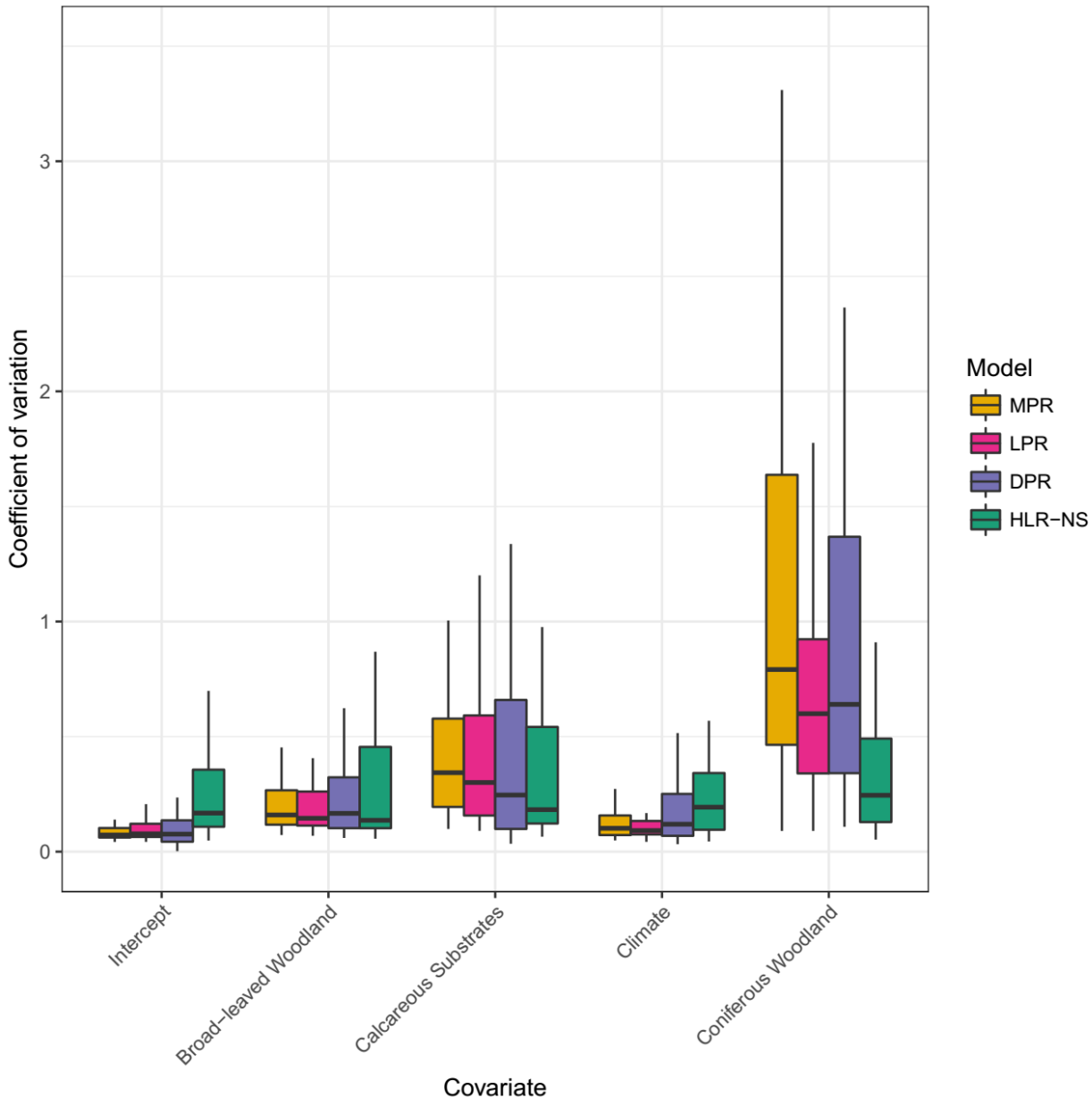


Figure 484: Coefficient of variation of the regression coefficient estimates of models on the butterfly dataset.

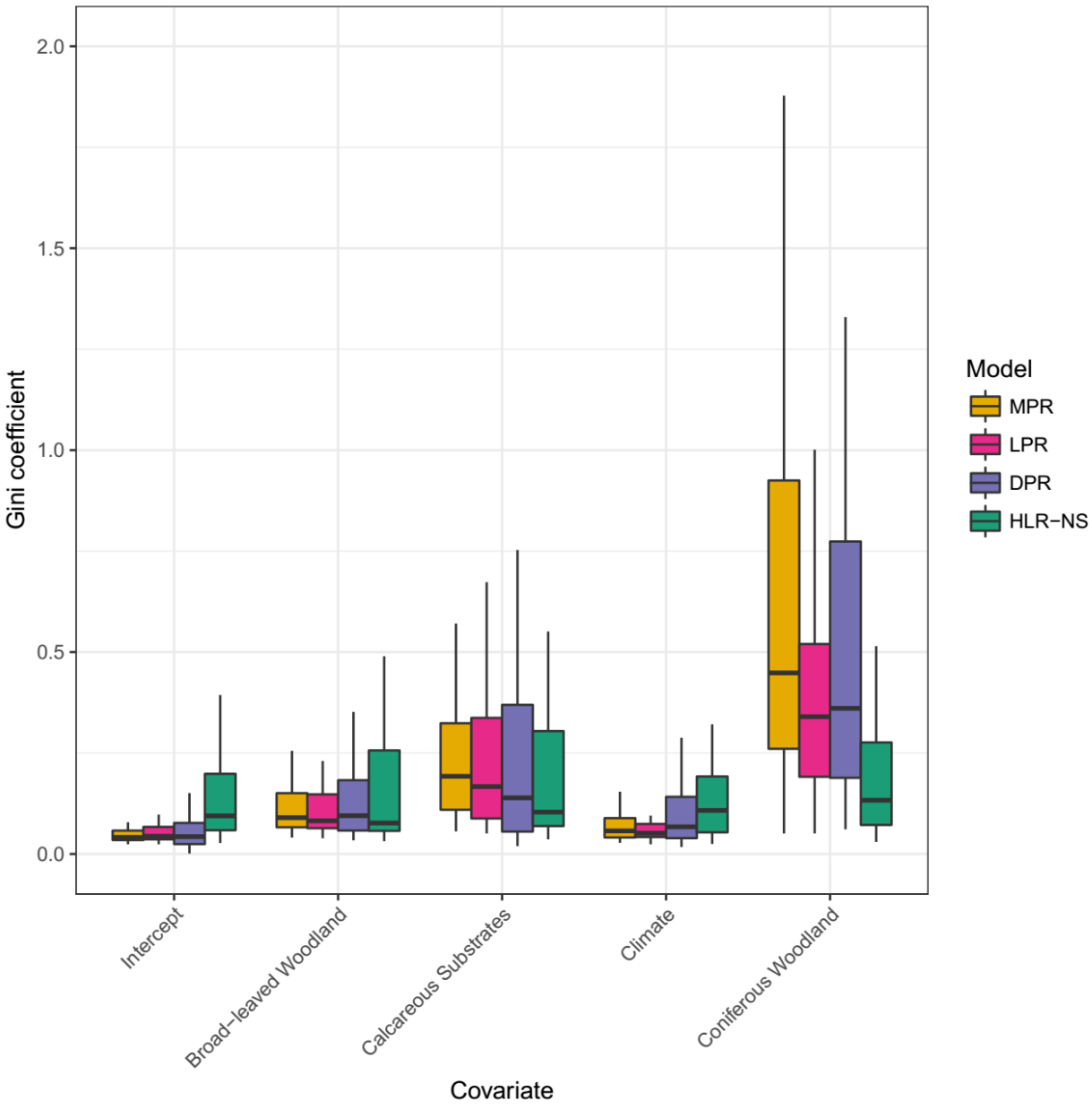


Figure 485: Gini coefficient of the regression coefficient estimates of models on the butterfly dataset.

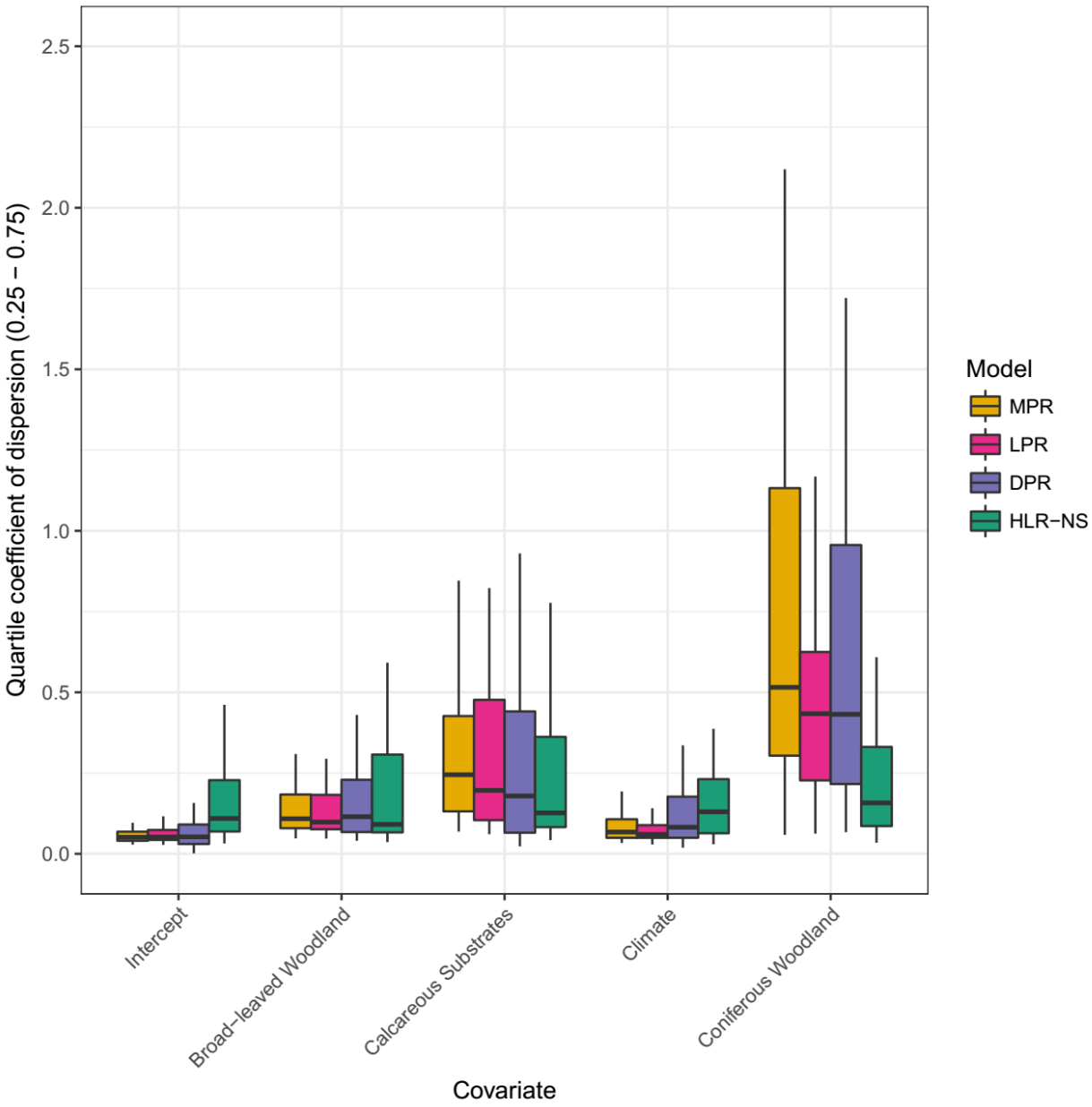


Figure 486: Quartile coefficient of dispersion of the regression coefficient estimates of models on the butterfly dataset.

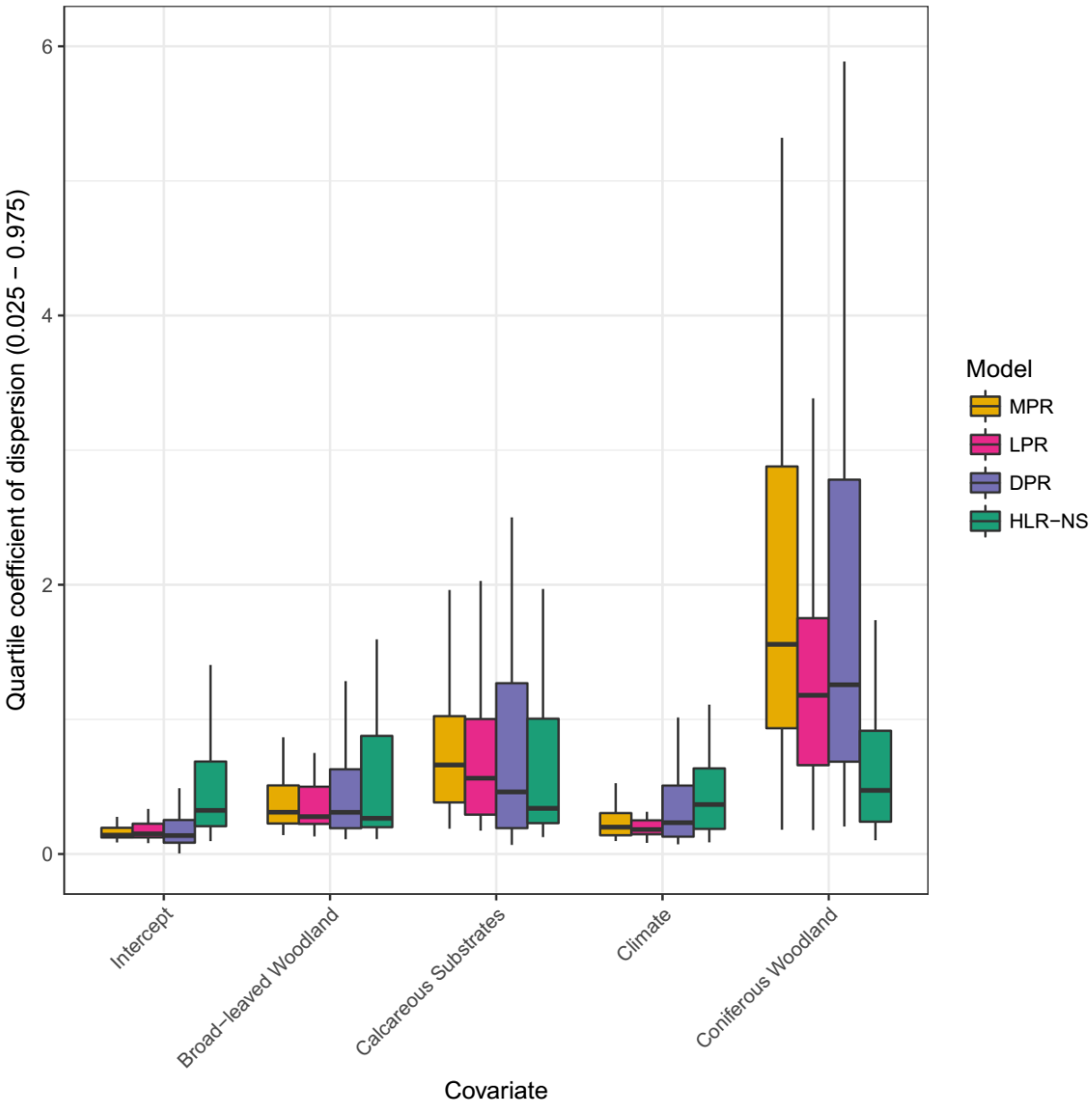


Figure 487: 90% quantile coefficient of dispersion of the regression coefficient estimates of models on the butterfly dataset.

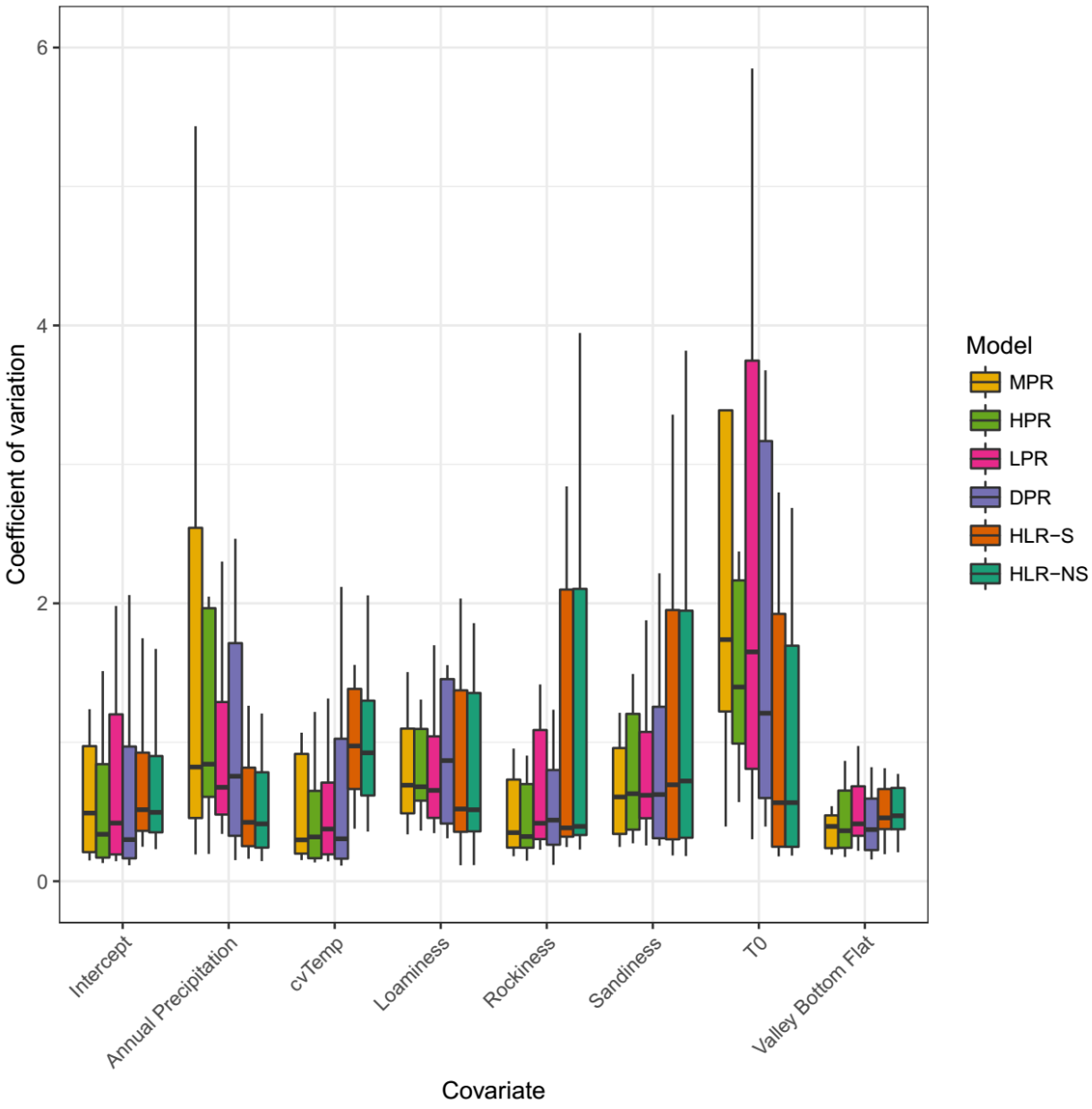


Figure 488: Coefficient of dispersion of the regression coefficient estimates of models on the eucalypt dataset.

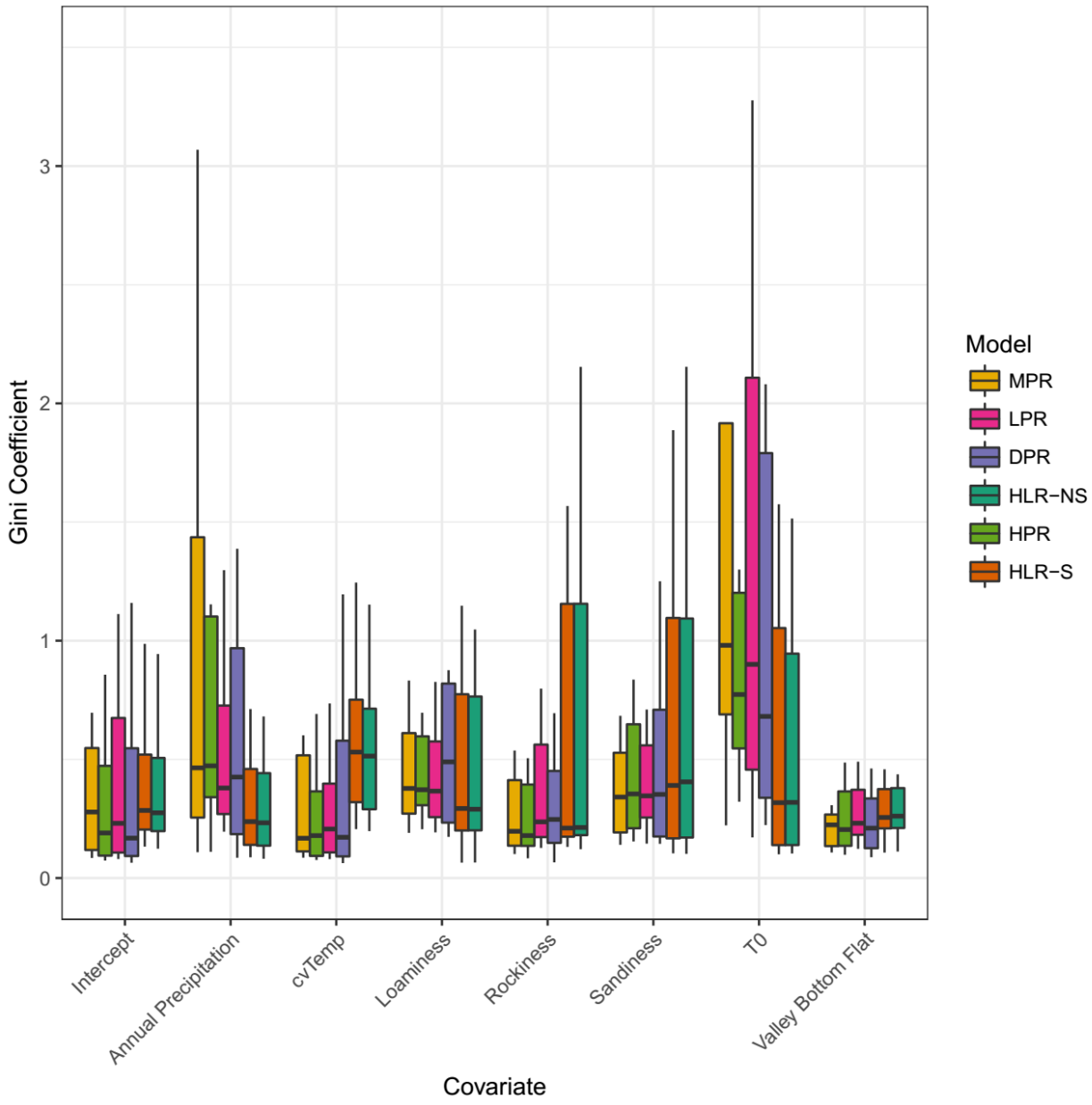


Figure 489: Gini coefficient of the regression coefficient estimates of models on the eucalypt dataset.

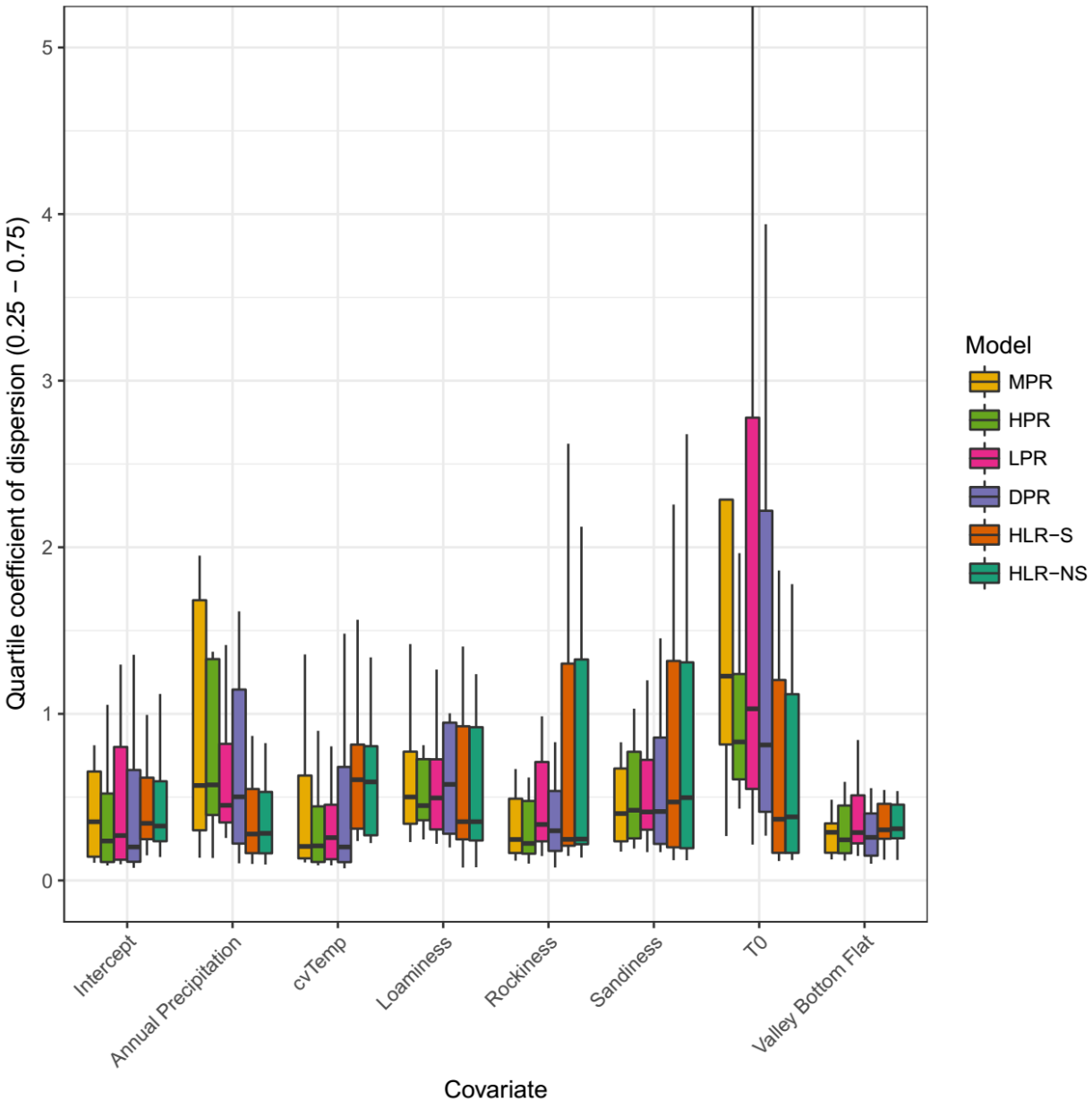


Figure 490: Quartile coefficient of dispersion of the regression coefficient estimates of models on the eucalypt dataset.

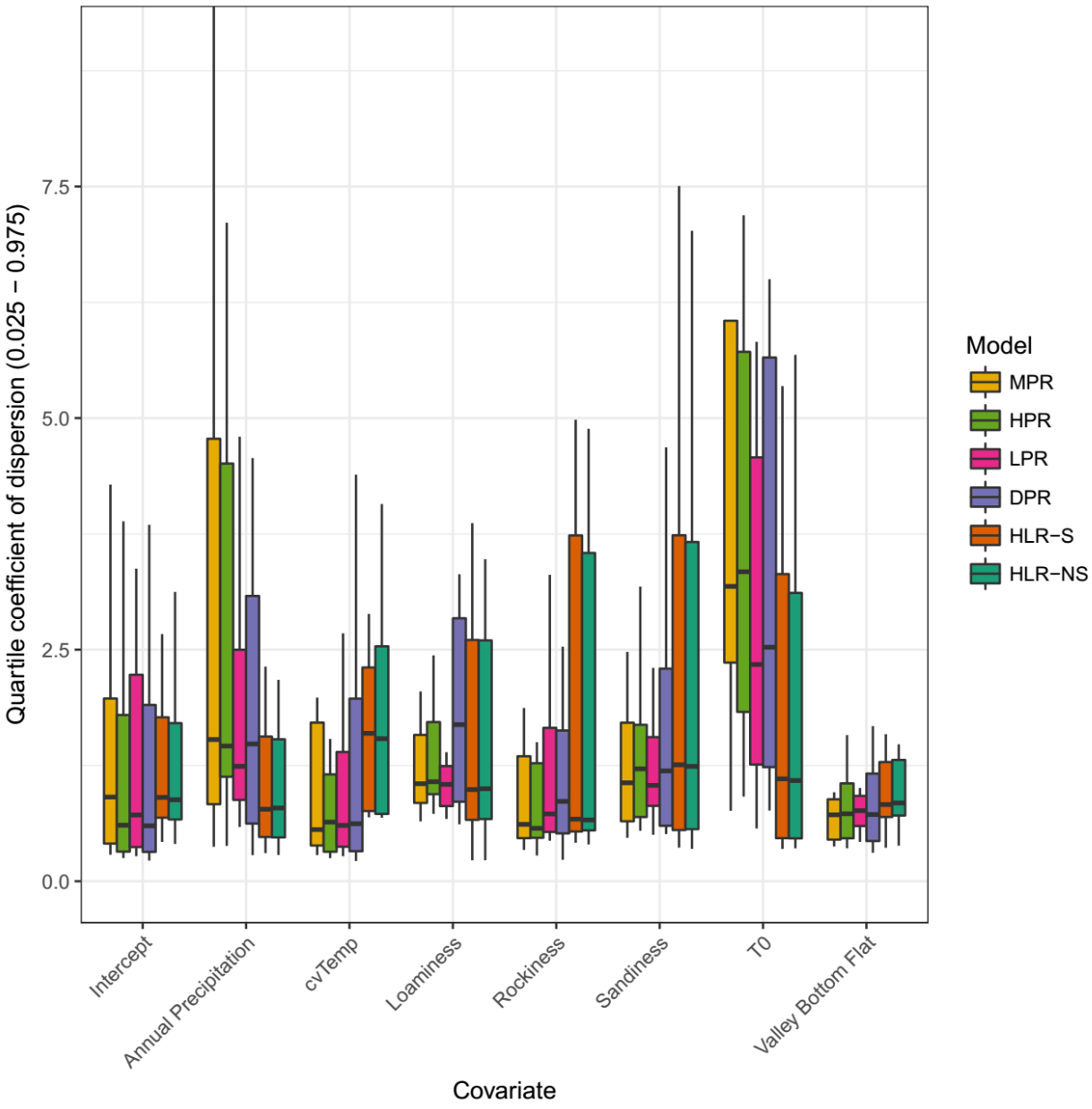


Figure 491: 90% quantile coefficient of dispersion of the regression coefficient estimates of models on the eucalypt dataset.

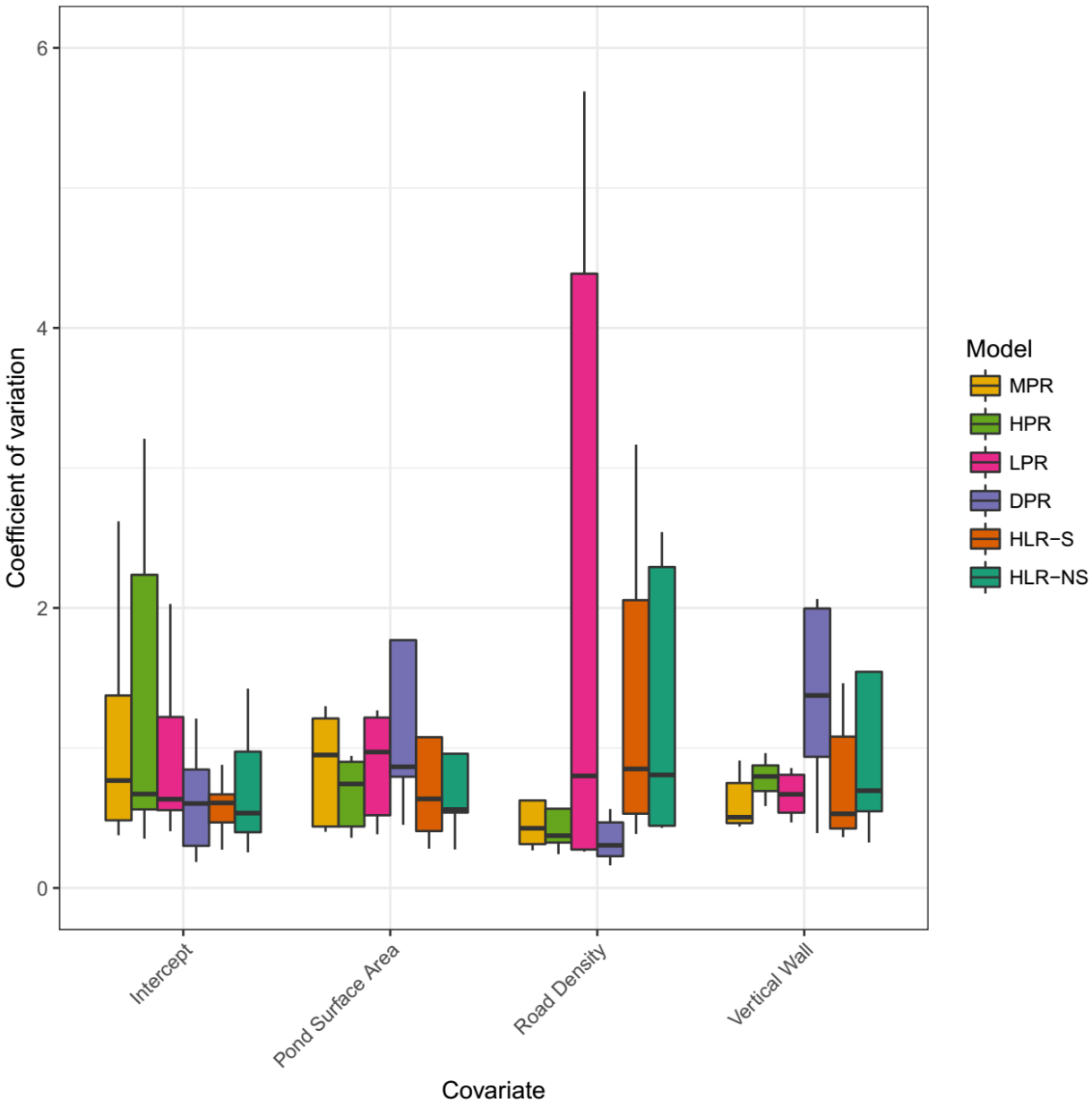


Figure 492: Coefficient of variation of the regression coefficient estimates of models on the frog dataset.

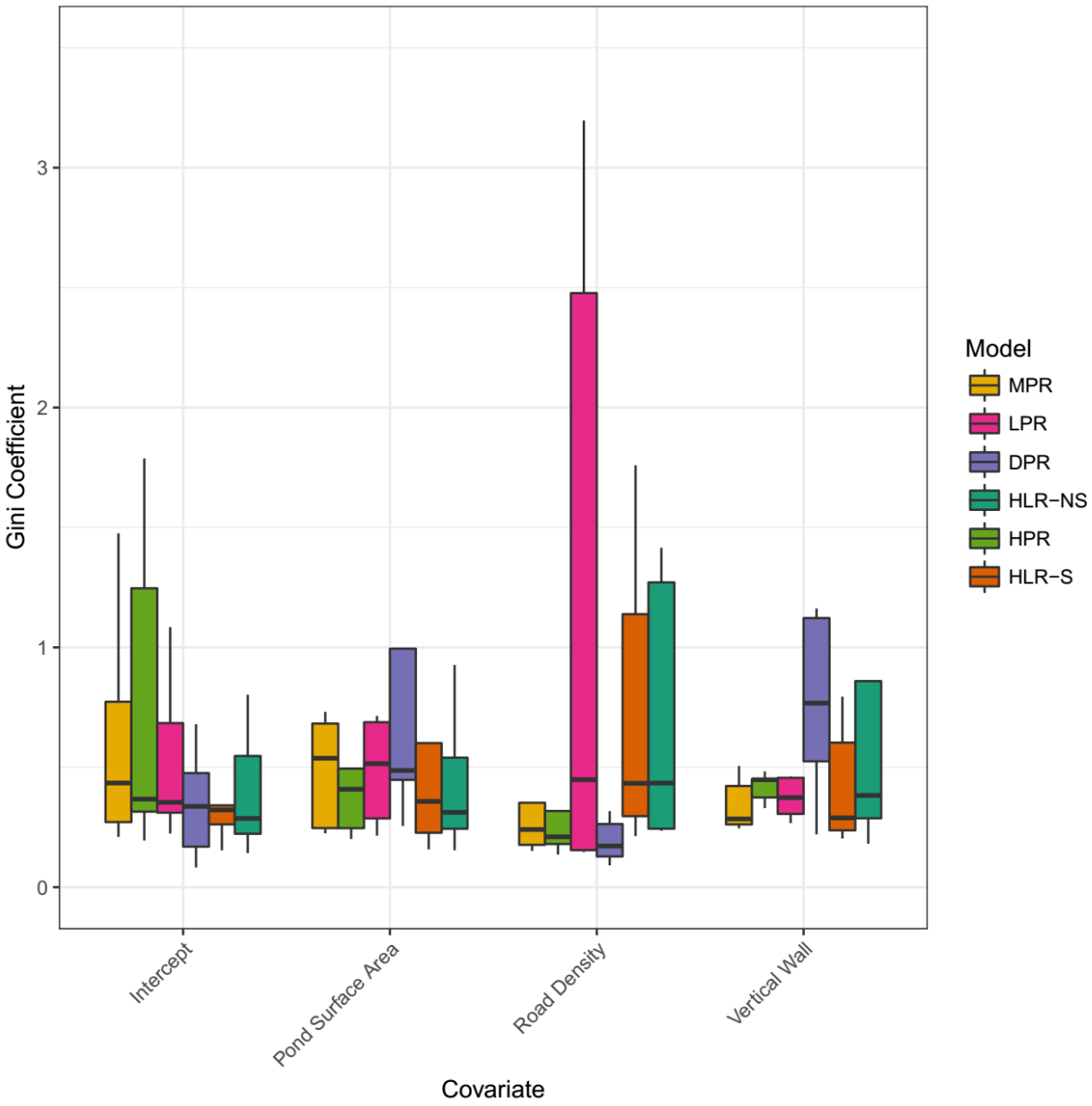


Figure 493: Gini coefficient of the regression coefficient estimates of models on the frog dataset.

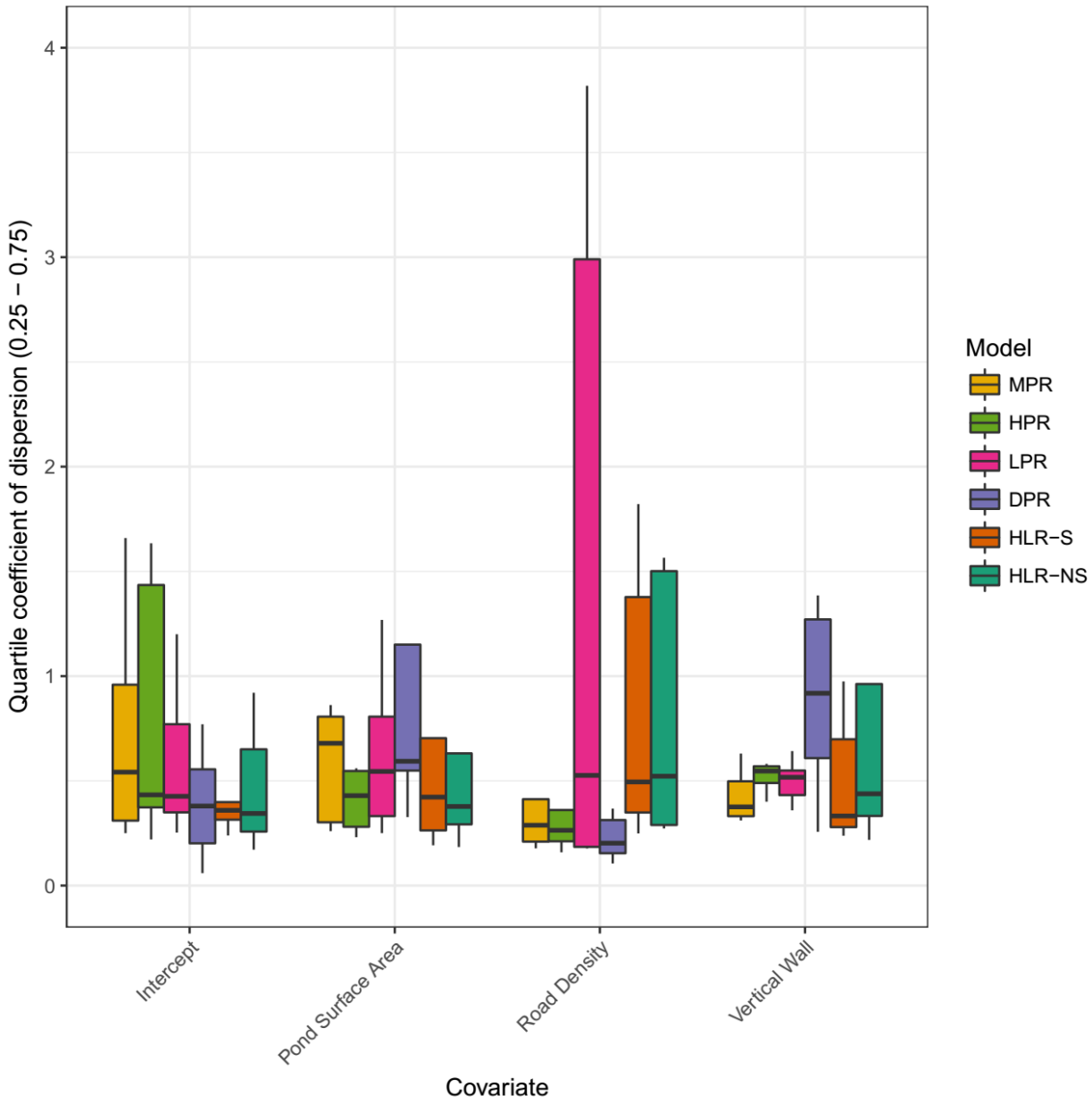


Figure 494: Quartile coefficient of dispersion of the regression coefficient estimates of models on the frog dataset.

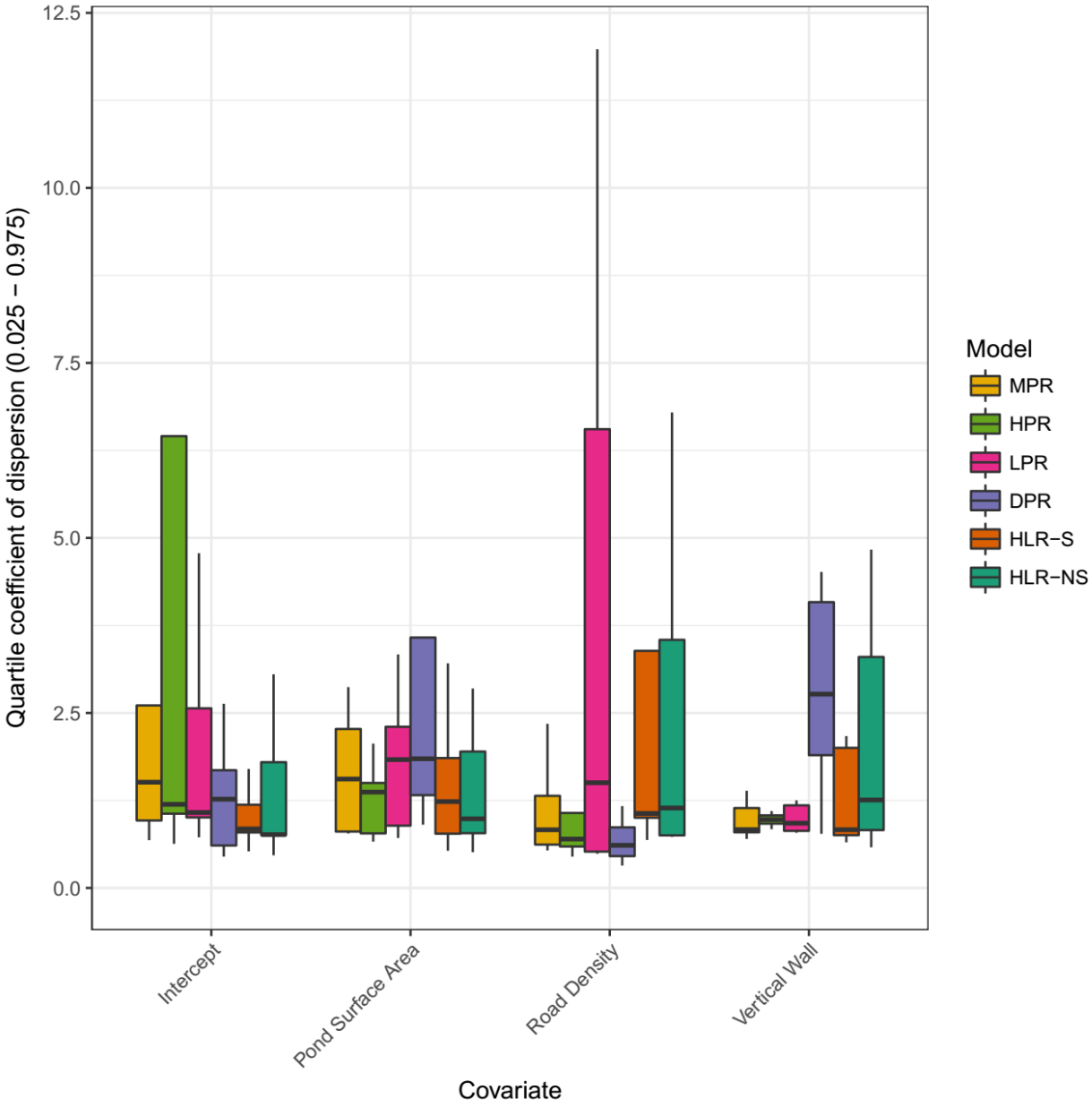


Figure 495: 90% quantile coefficient of dispersion of the regression coefficient estimates of models on the frog dataset.

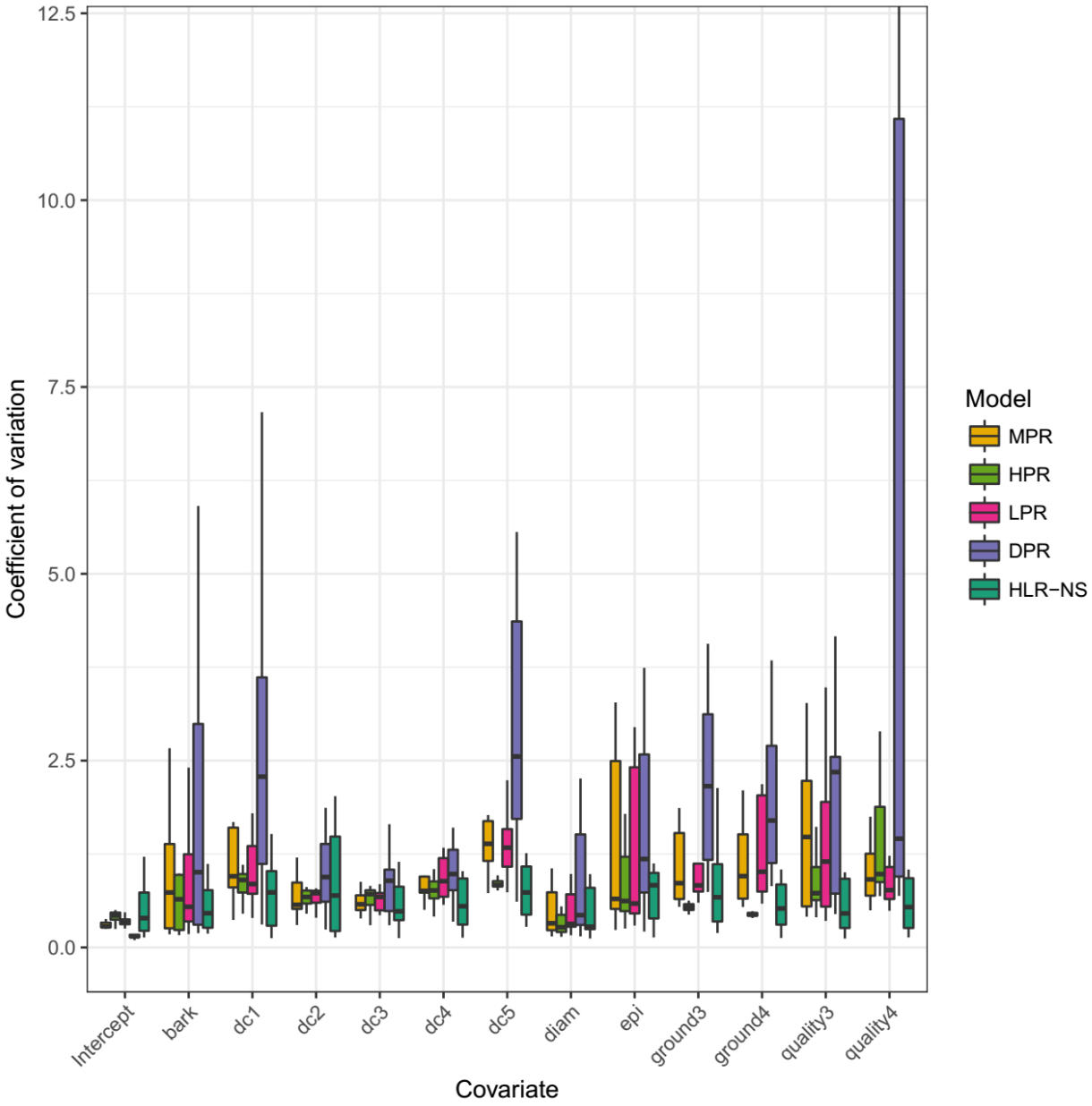


Figure 496: Coefficient of dispersion of the regression coefficient estimates of models on the fungi dataset.

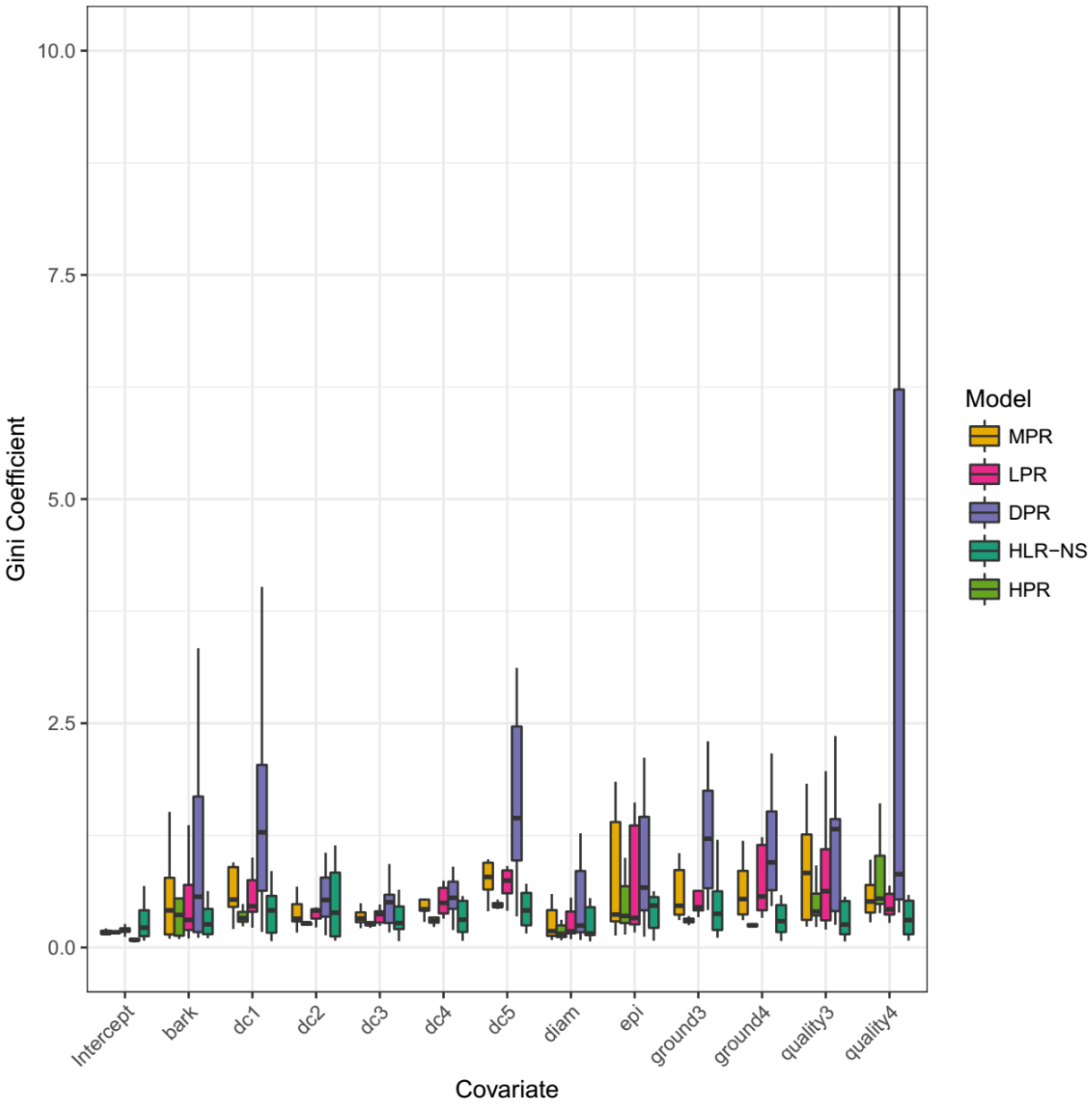


Figure 497: Gini coefficient of the regression coefficient estimates of models on the fungi dataset.

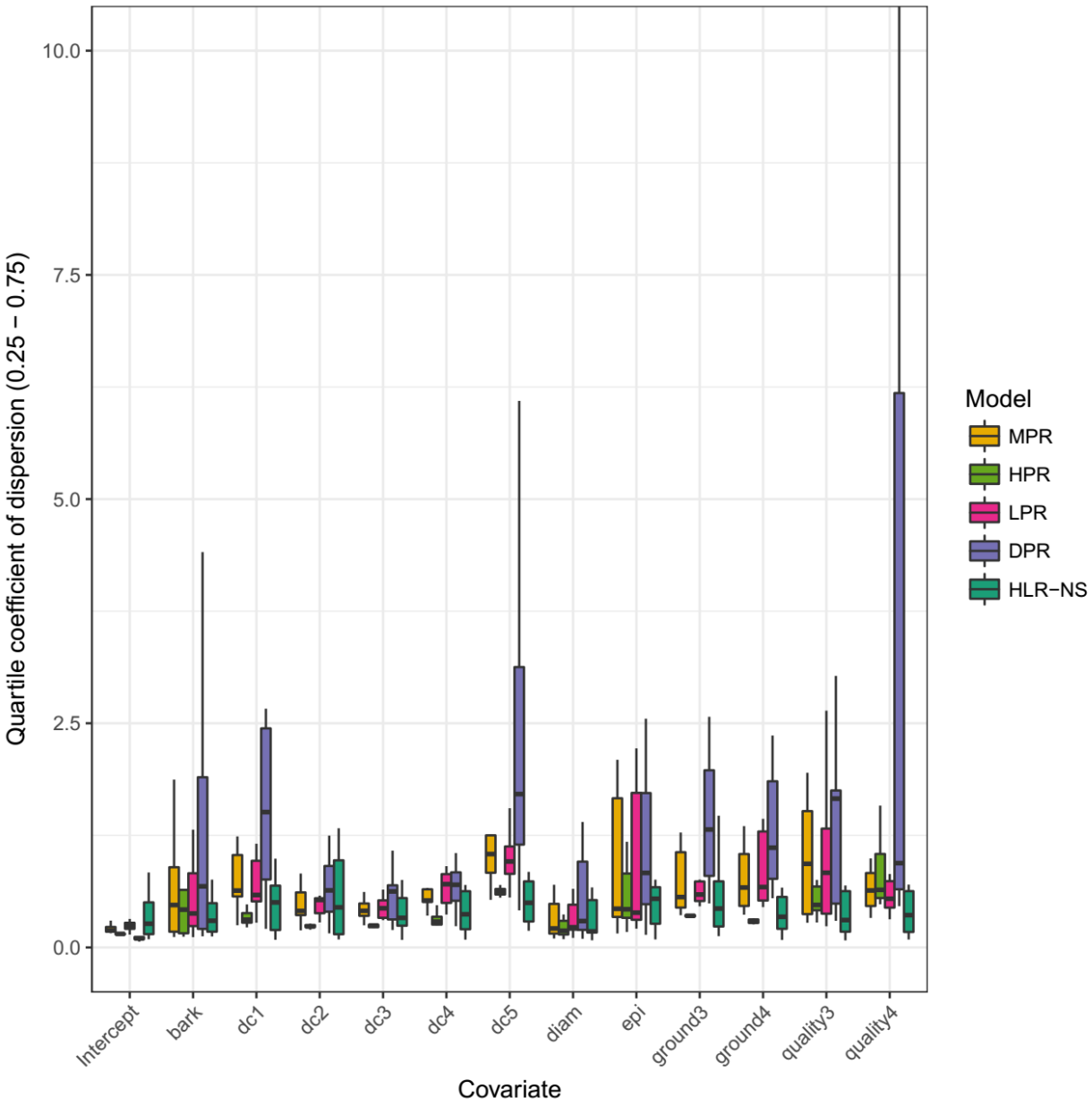


Figure 498: Quartile coefficient of dispersion of the regression coefficient estimates of models on the fungi dataset.

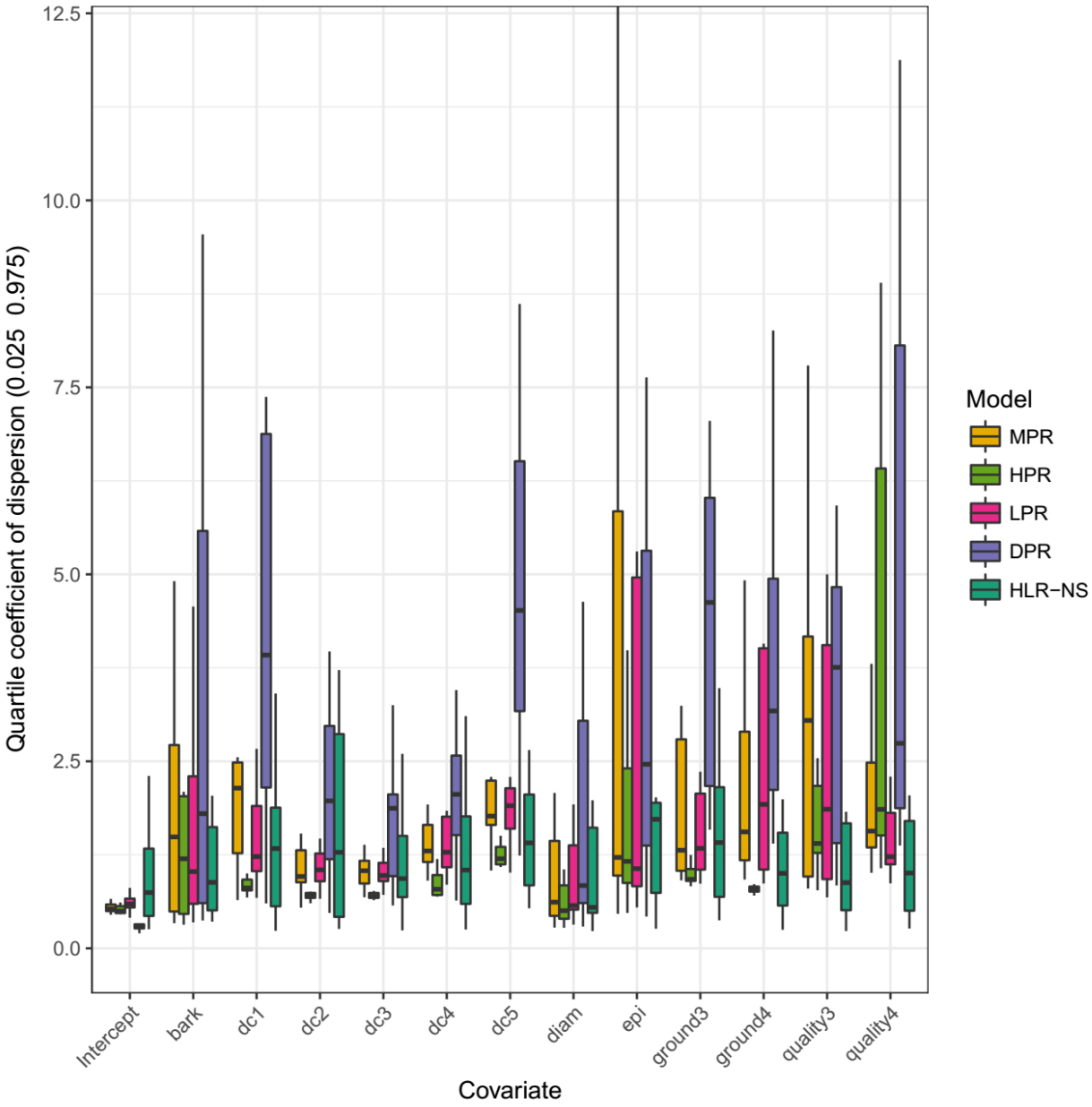


Figure 499: 90% quantile coefficient of dispersion of the regression coefficient estimates of models on the fungi dataset.

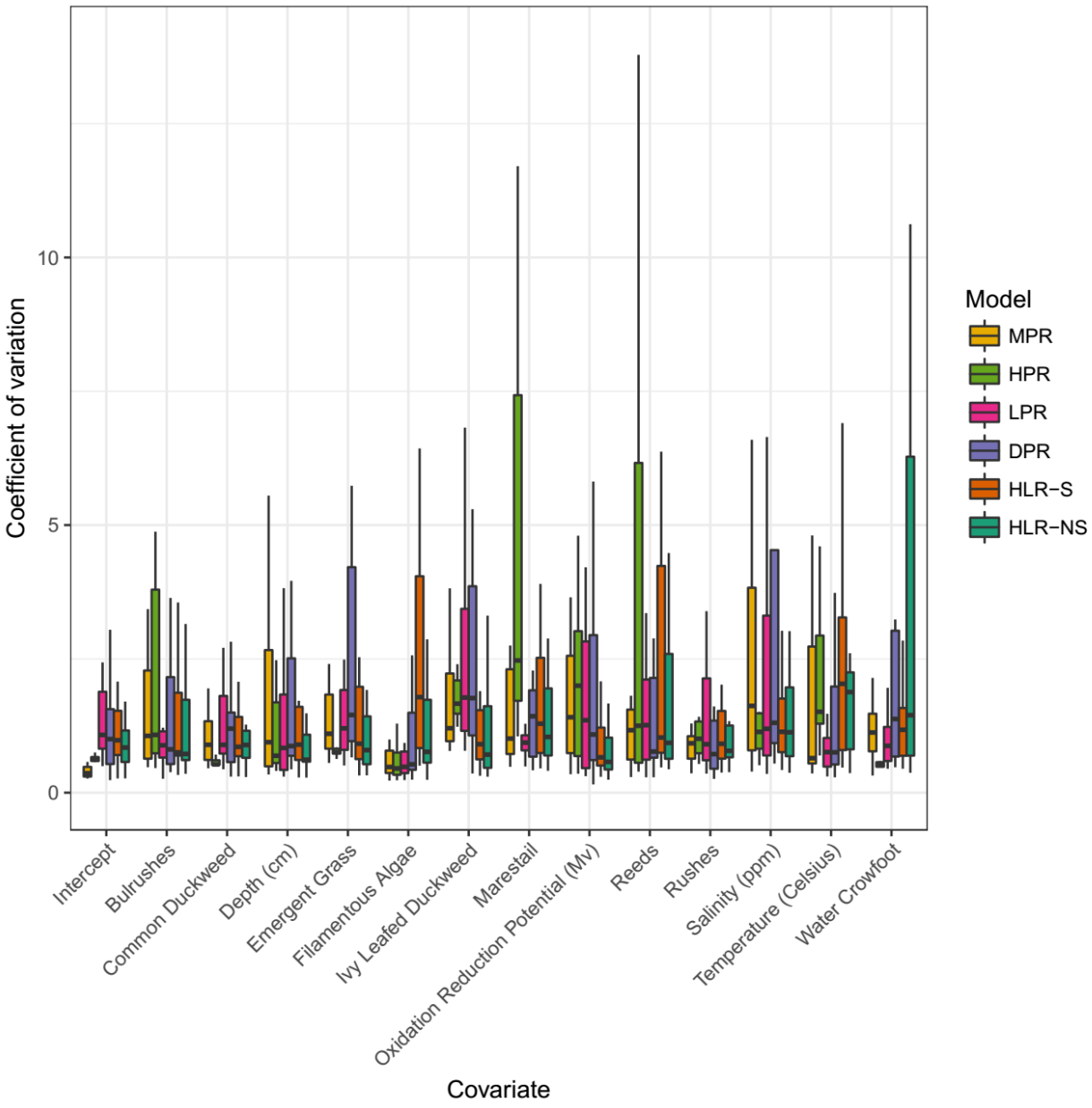


Figure 500: Coefficient of variation of the regression coefficient estimates of models on the mosquito dataset.

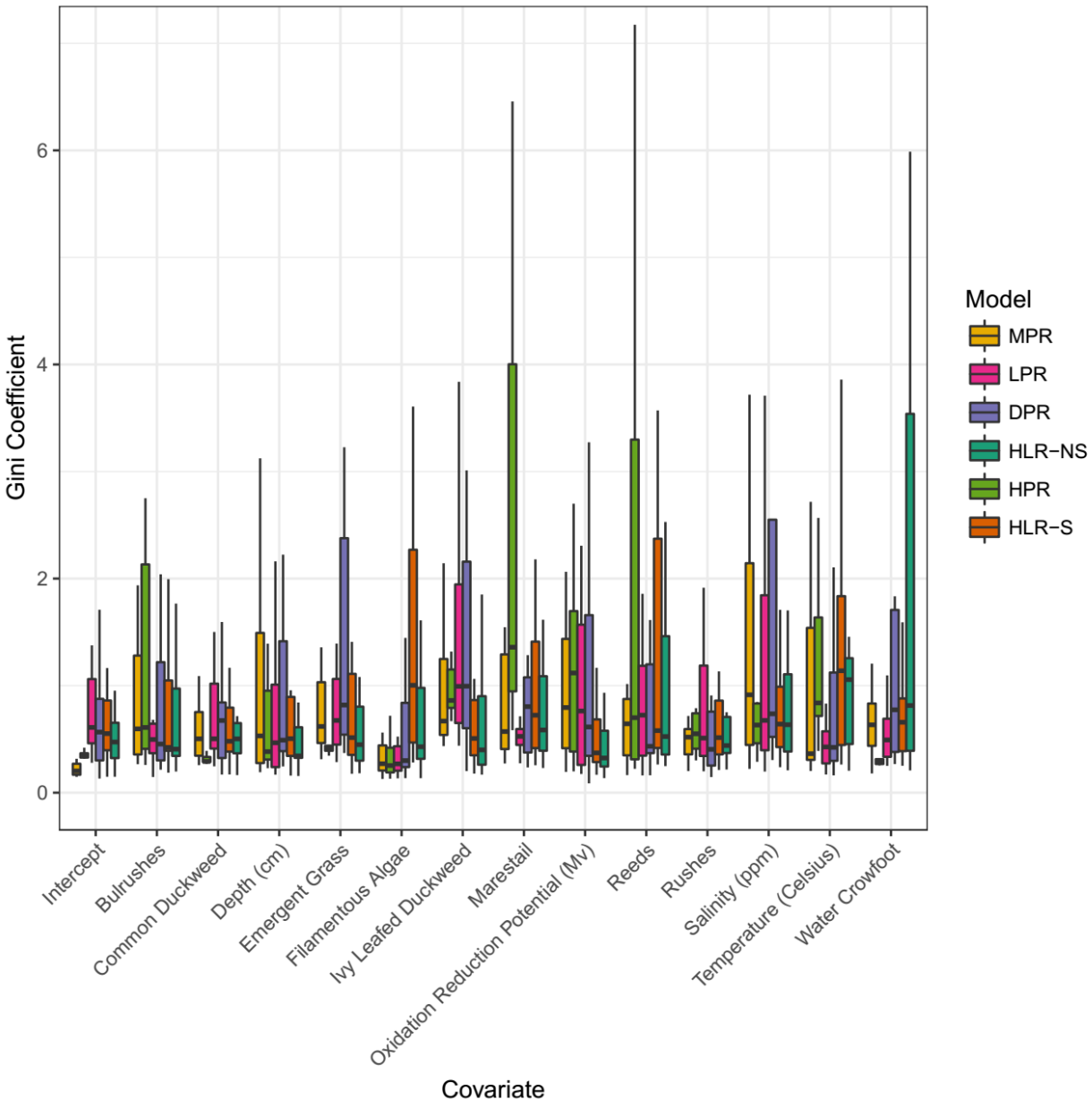


Figure 501: Gini coefficient of the regression coefficient estimates of models on the mosquito dataset.

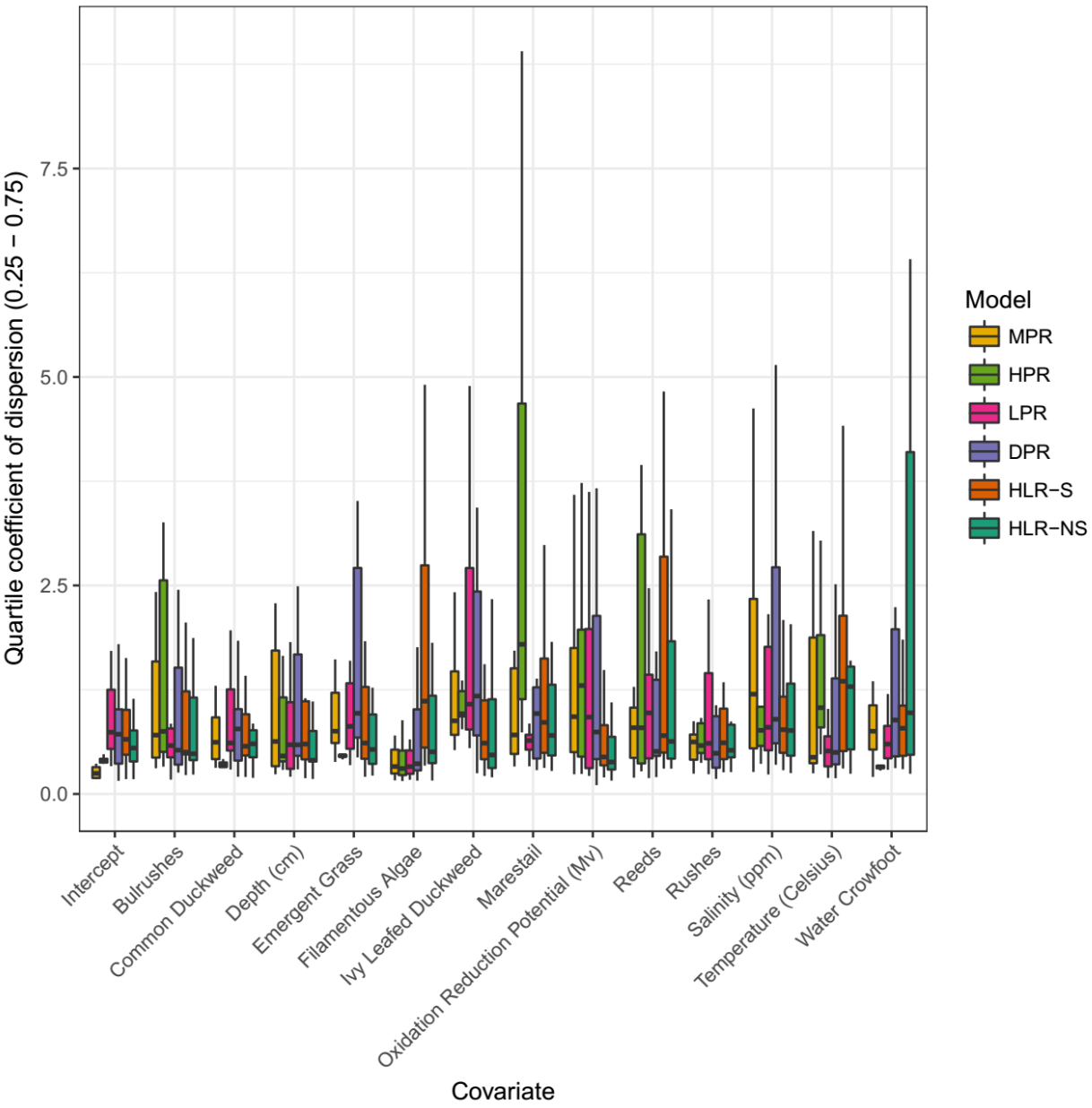


Figure 502: Quartile coefficient of dispersion of the regression coefficient estimates of models on the mosquito dataset.

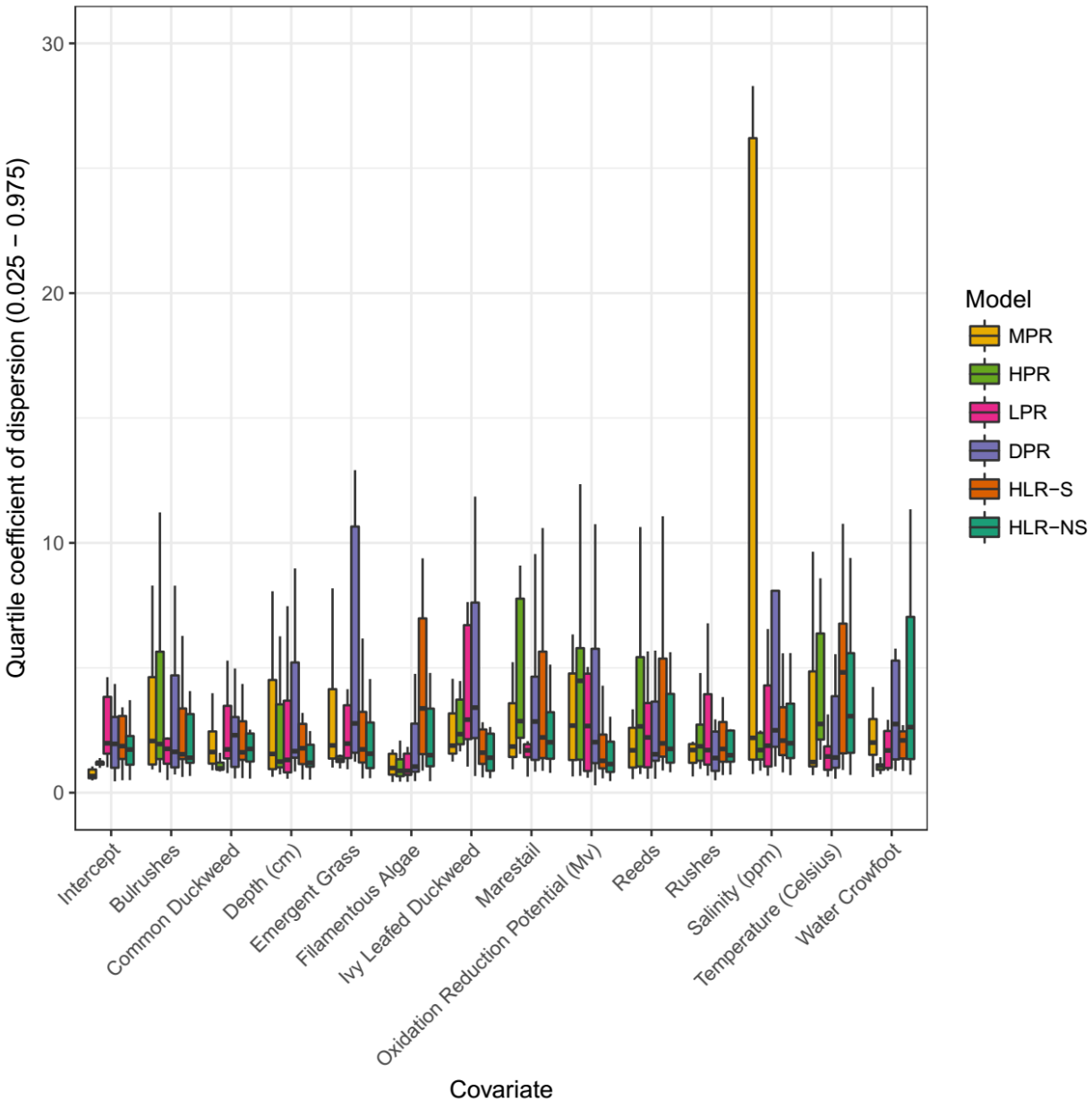


Figure 503: 90% quantile coefficient of dispersion of the regression coefficient estimates of models on the mosquito dataset.

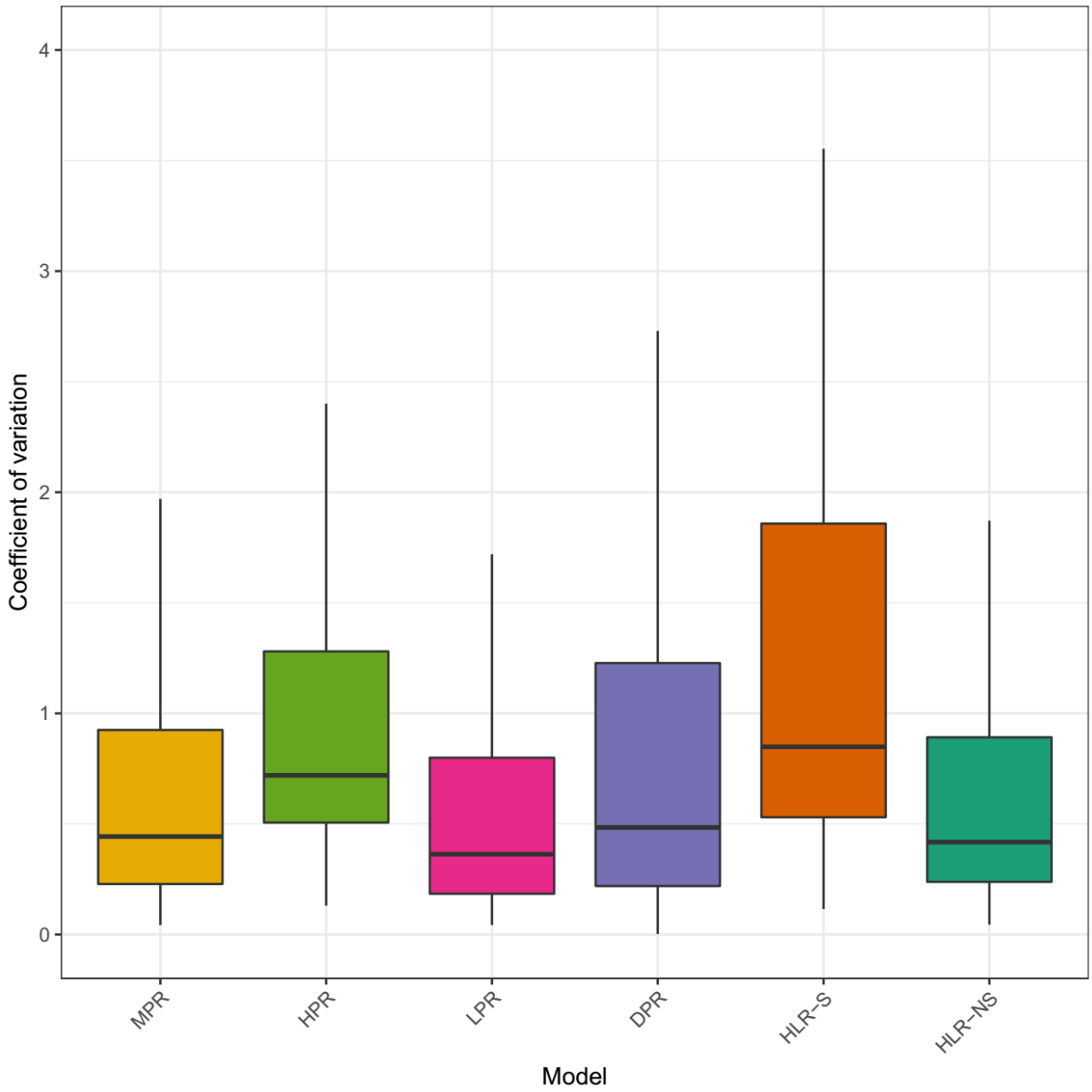


Figure 504: Coefficient of variation of the regression coefficient estimates of models across all datasets.

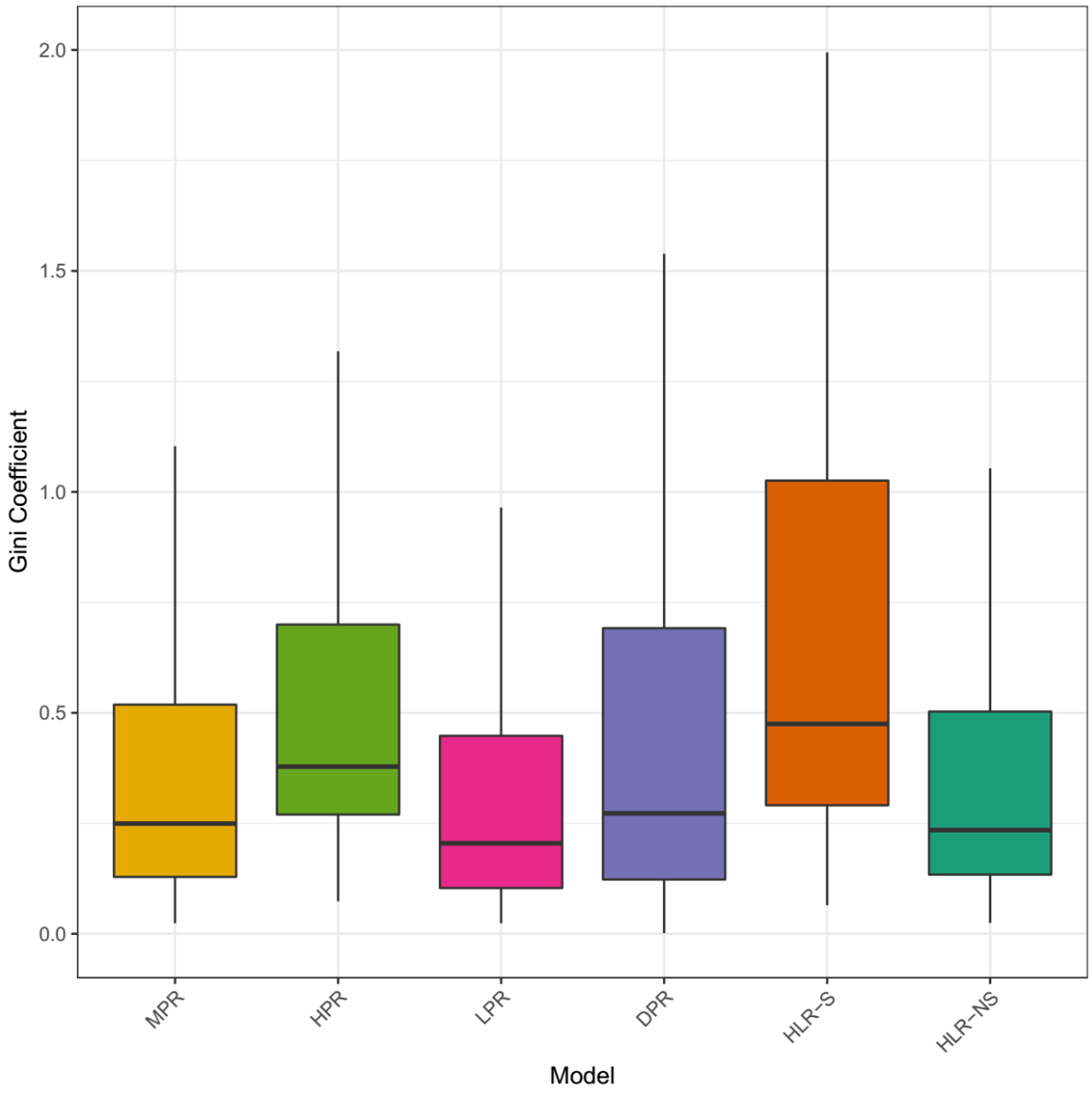


Figure 505: Gini coefficient of the regression coefficient estimates of models across all datasets.

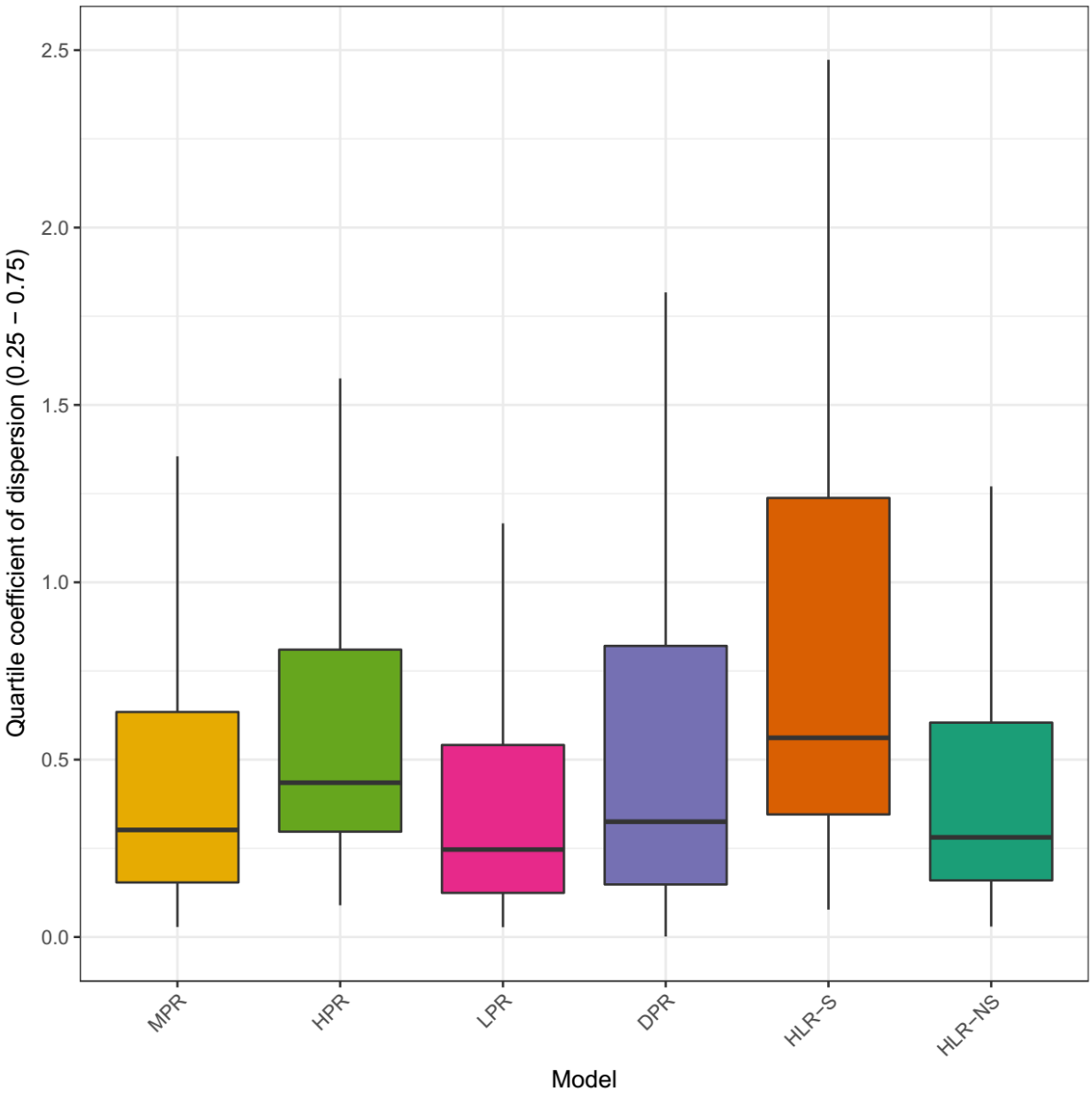


Figure 506: Quartile coefficient of dispersion of the regression coefficient estimates of models across all datasets.

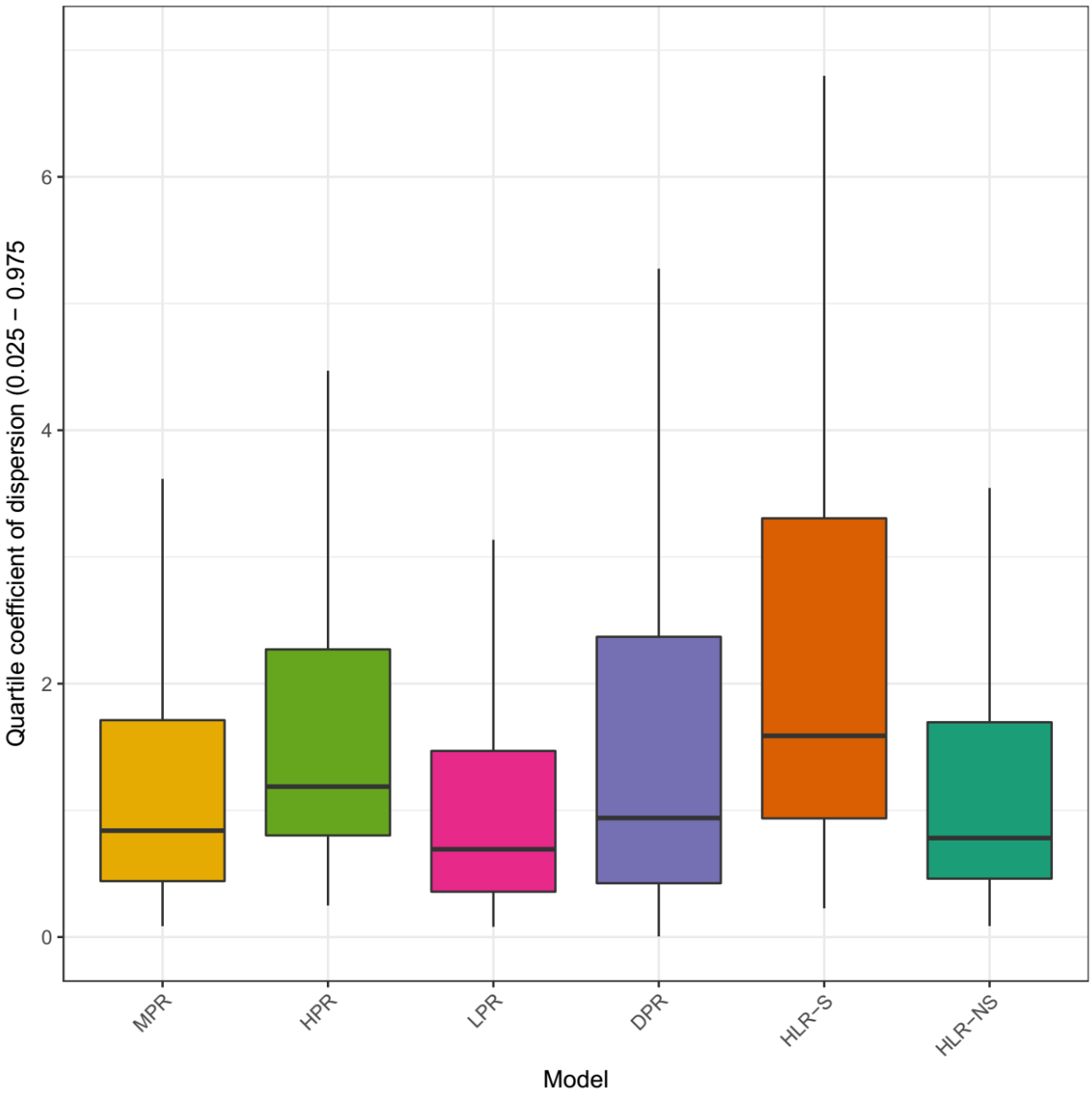


Figure 507: 90% quantile coefficient of dispersion of the regression coefficient estimates of models across all datasets.

APPENDIX E

CORRELATION IN REGRESSION AND CORRELATION COEFFICIENT ESTIMATES BETWEEN JSDMS

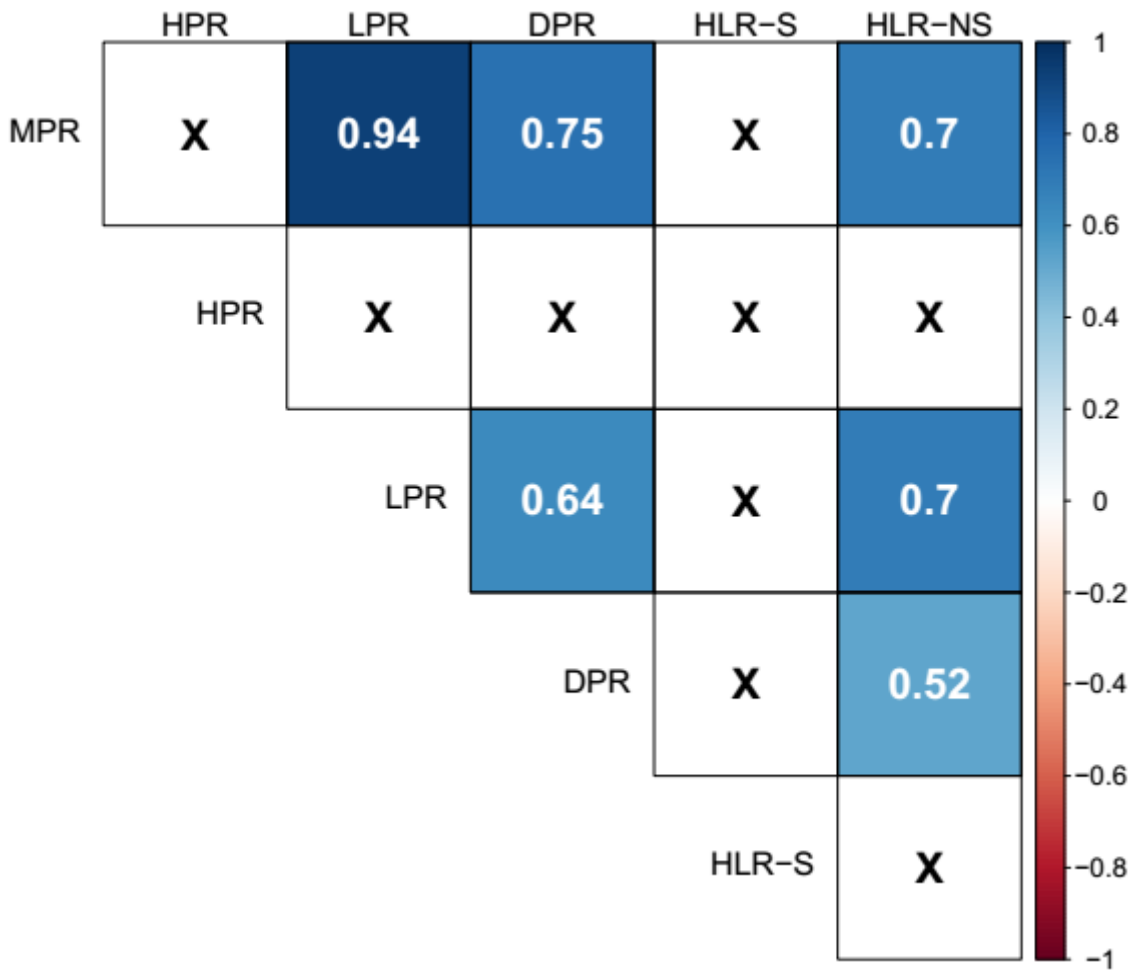


Figure 508: Average correlation of regression coefficient posterior mean estimates between different JSDMs on the bird dataset. 'X' marks pairs where one or more JSDMs did not run.

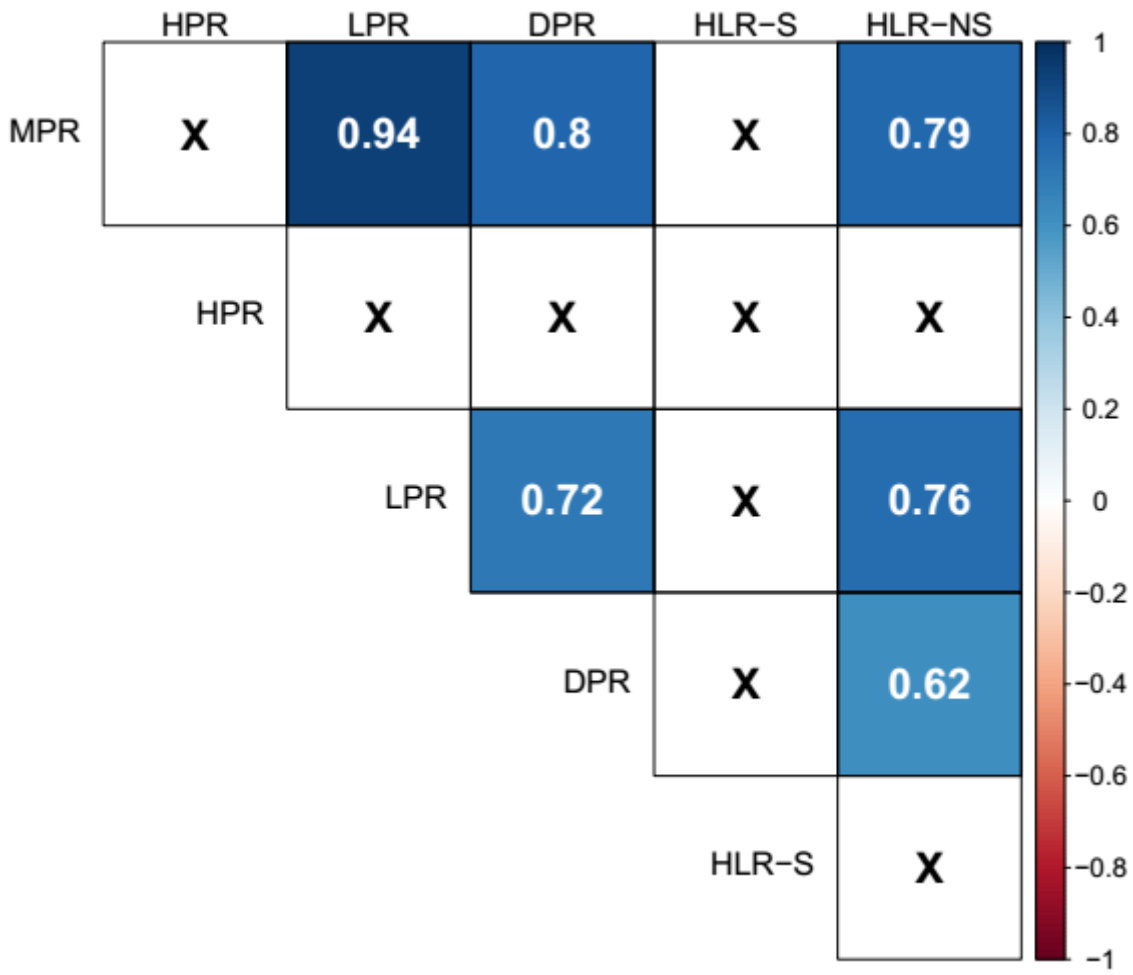


Figure 509: Average correlation of regression coefficient posterior mean estimates between different JSDMs on the butterfly dataset. 'X' marks pairs where one or more JSDMs did not run.

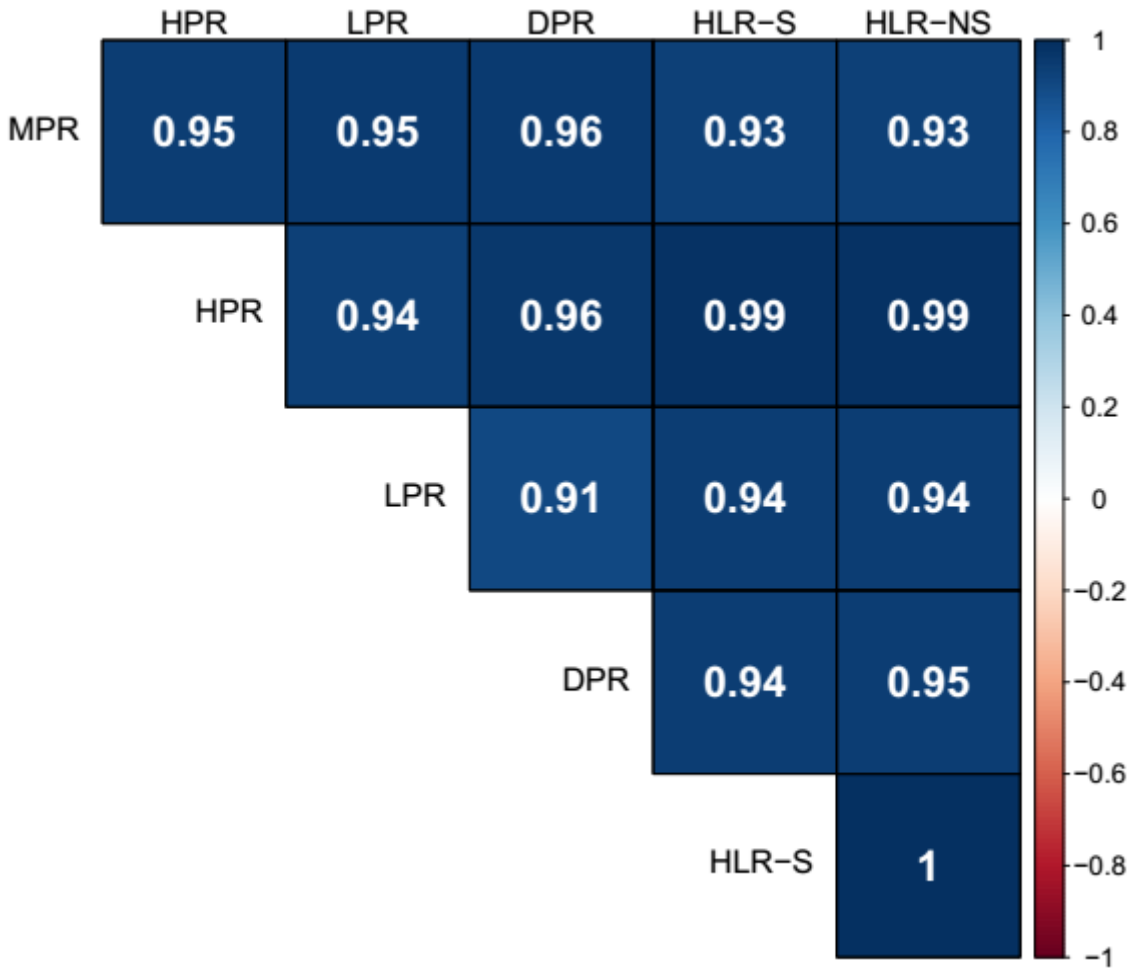


Figure 510: Average correlation of regression coefficient posterior mean estimates between different JSDMs on the eucalypt dataset.

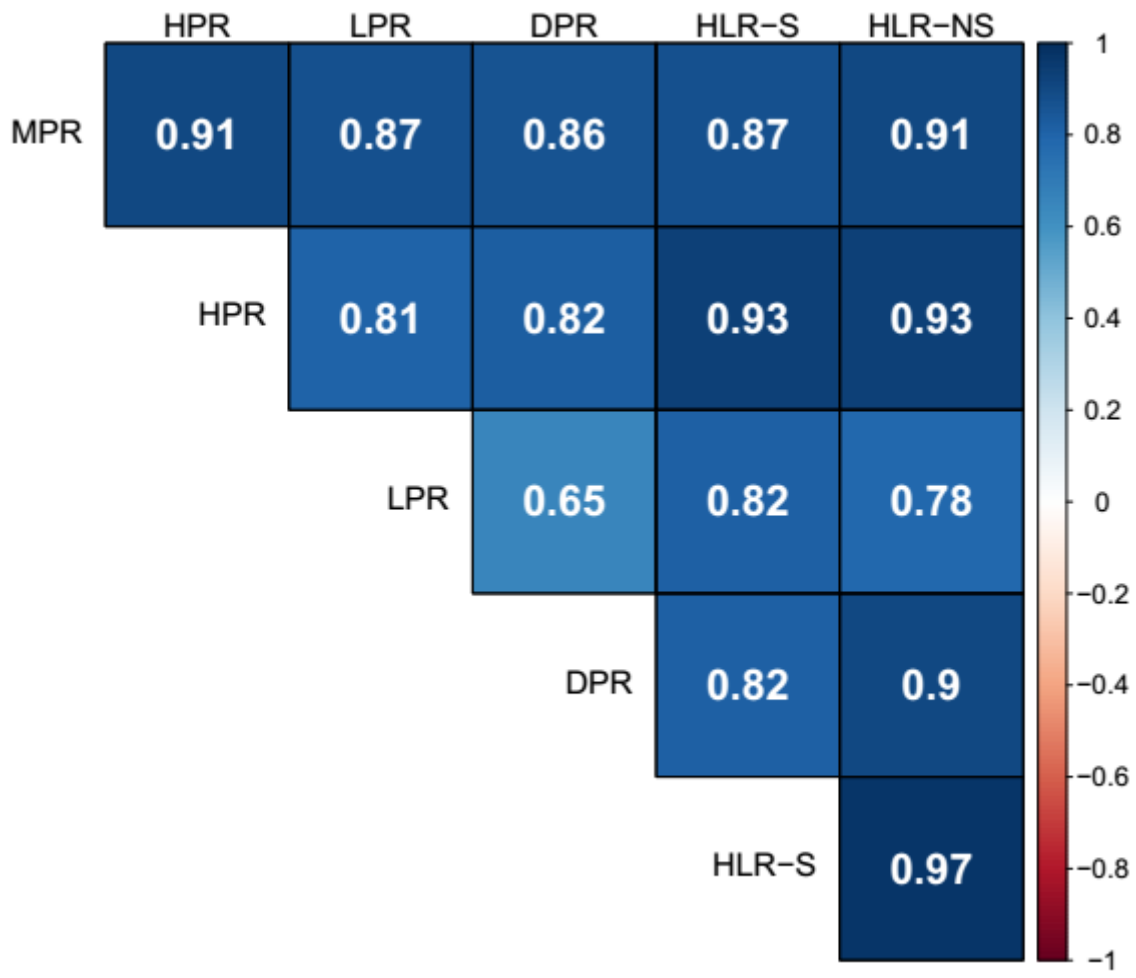


Figure 511: Average correlation of regression coefficient posterior mean estimates between different JSDMs on the frog dataset.

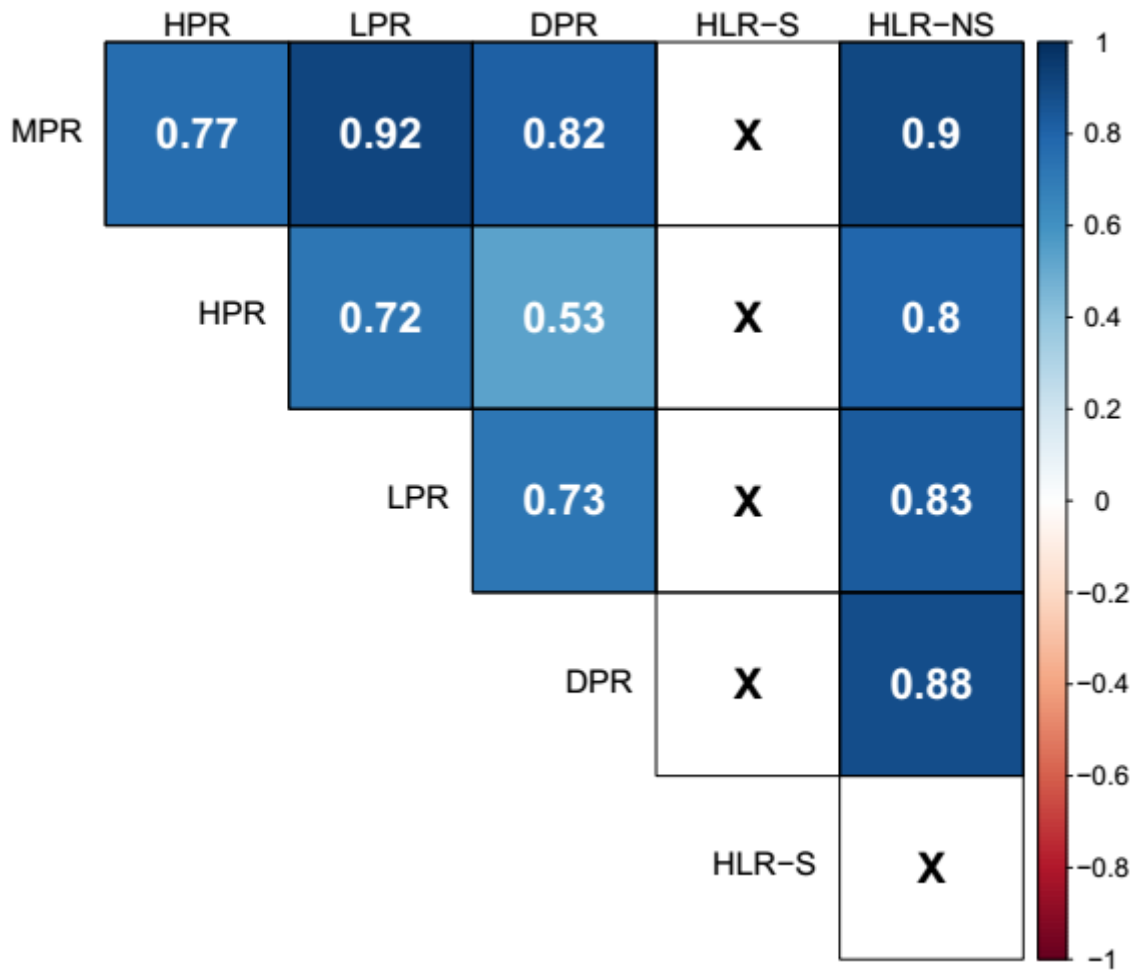


Figure 512: Average correlation of regression coefficient posterior mean estimates between different JSDMs on the fungi dataset. 'X' marks pairs where one or more JSDMs did not run.

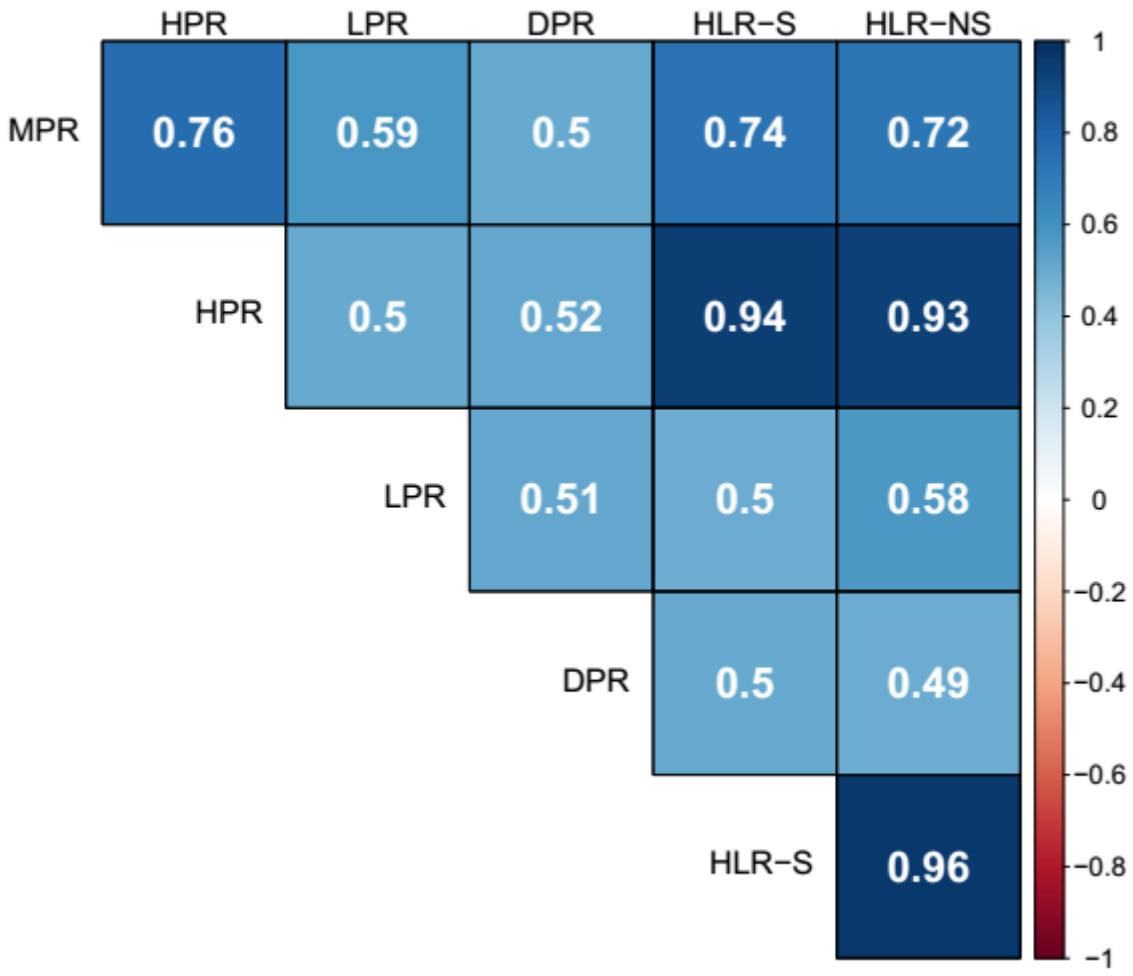


Figure 513: Average correlation of regression coefficient posterior mean estimates between different JSDMs on the mosquito dataset.

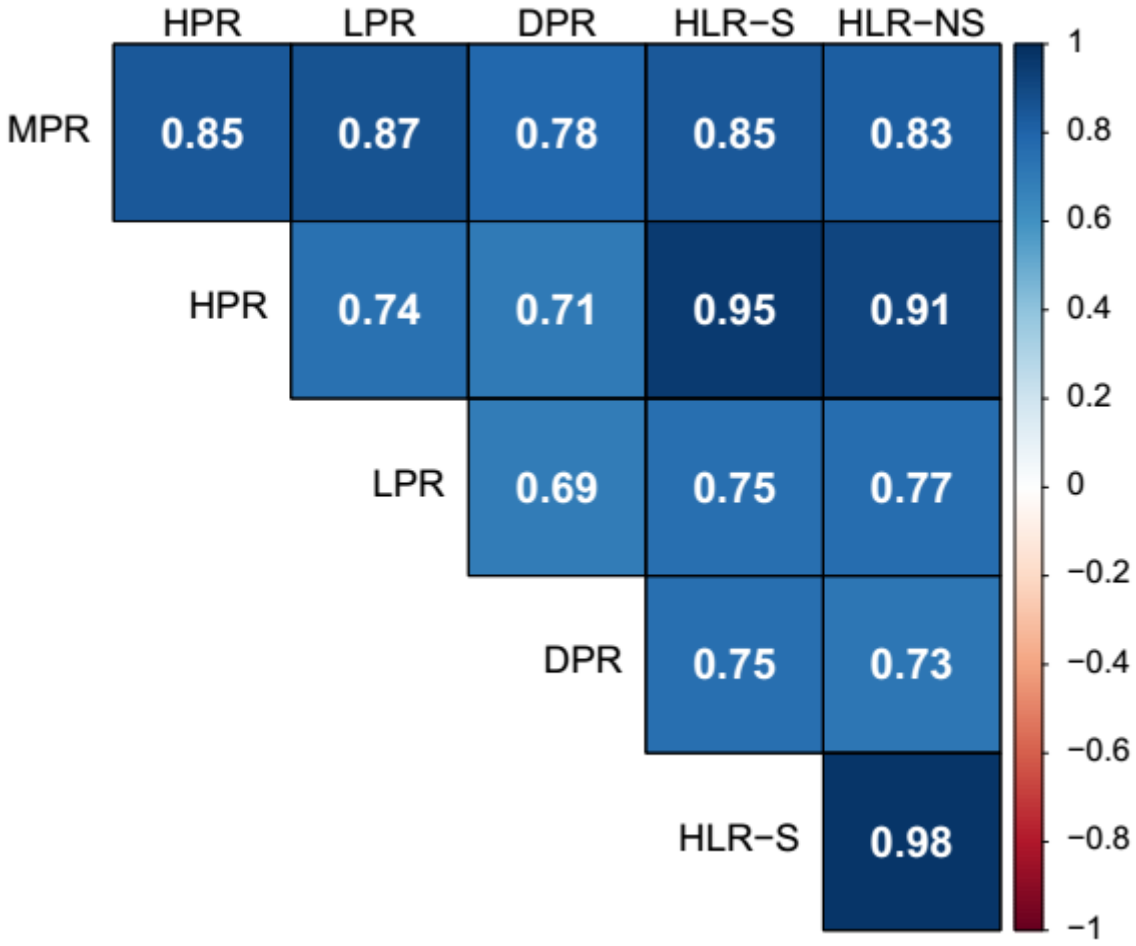


Figure 514: Average correlation of regression coefficient posterior mean estimates between different JSDMs for all datasets.

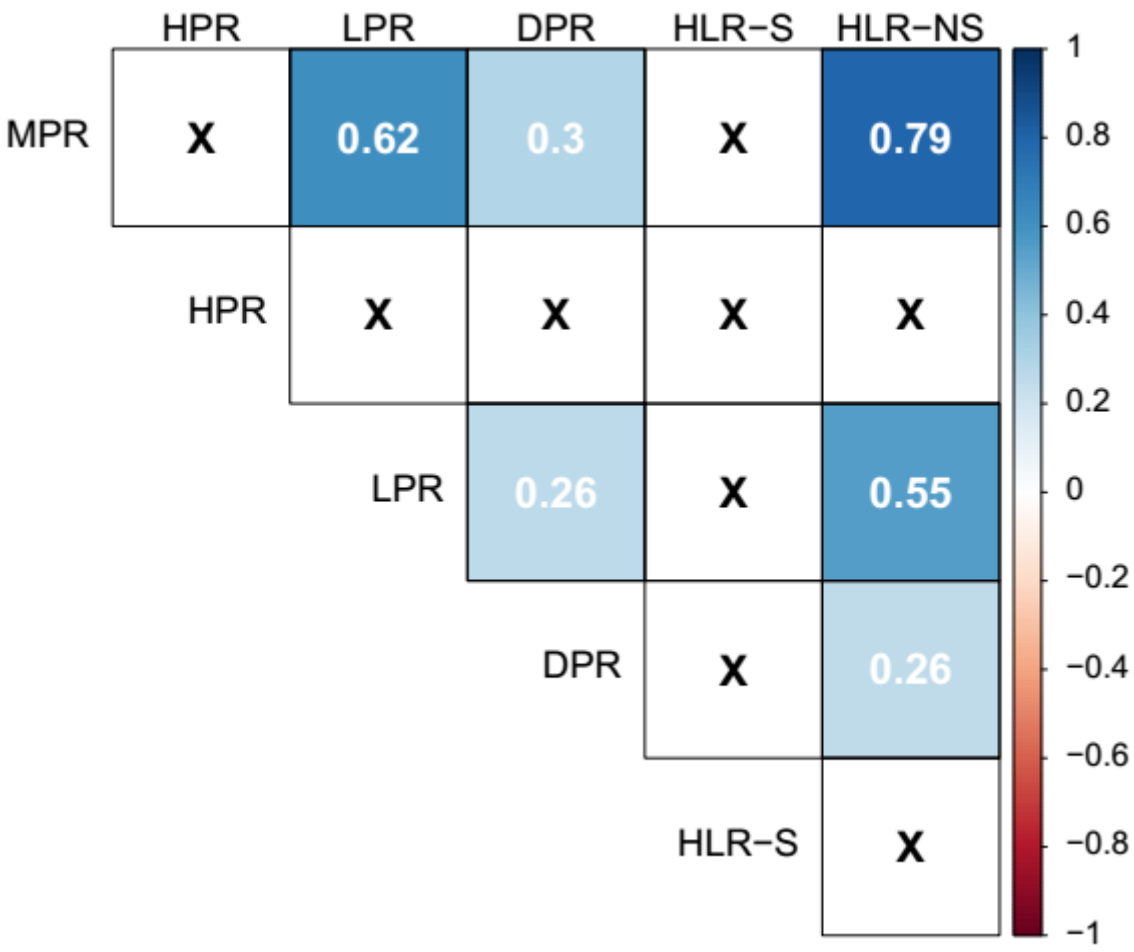


Figure 515: Average correlation of correlation coefficient posterior mean estimates between different JSDMs on the bird dataset. 'X' marks pairs where one or more JSDMs did not run.

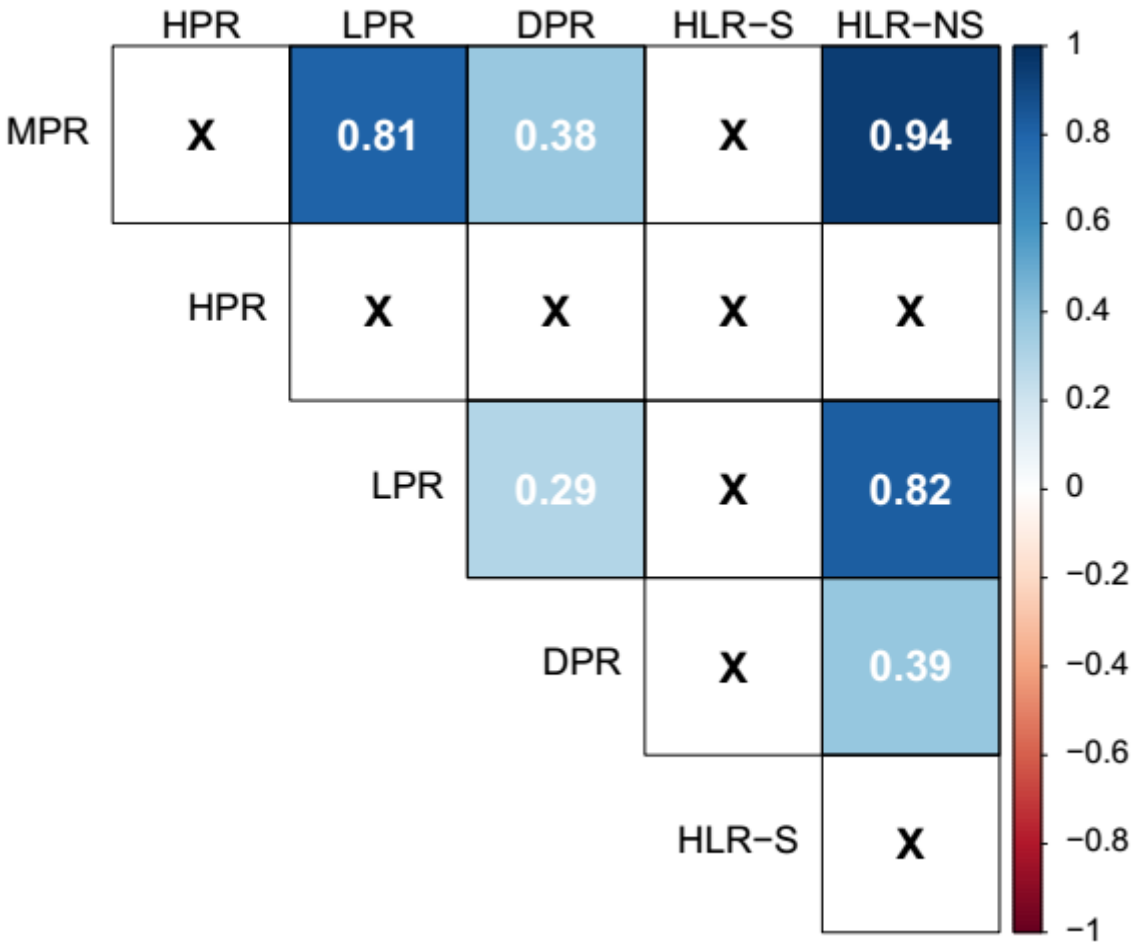


Figure 516: Average correlation of correlation coefficient posterior mean estimates between different JSDMs on the butterfly dataset. 'X' marks pairs where one or more JSDMs did not run.

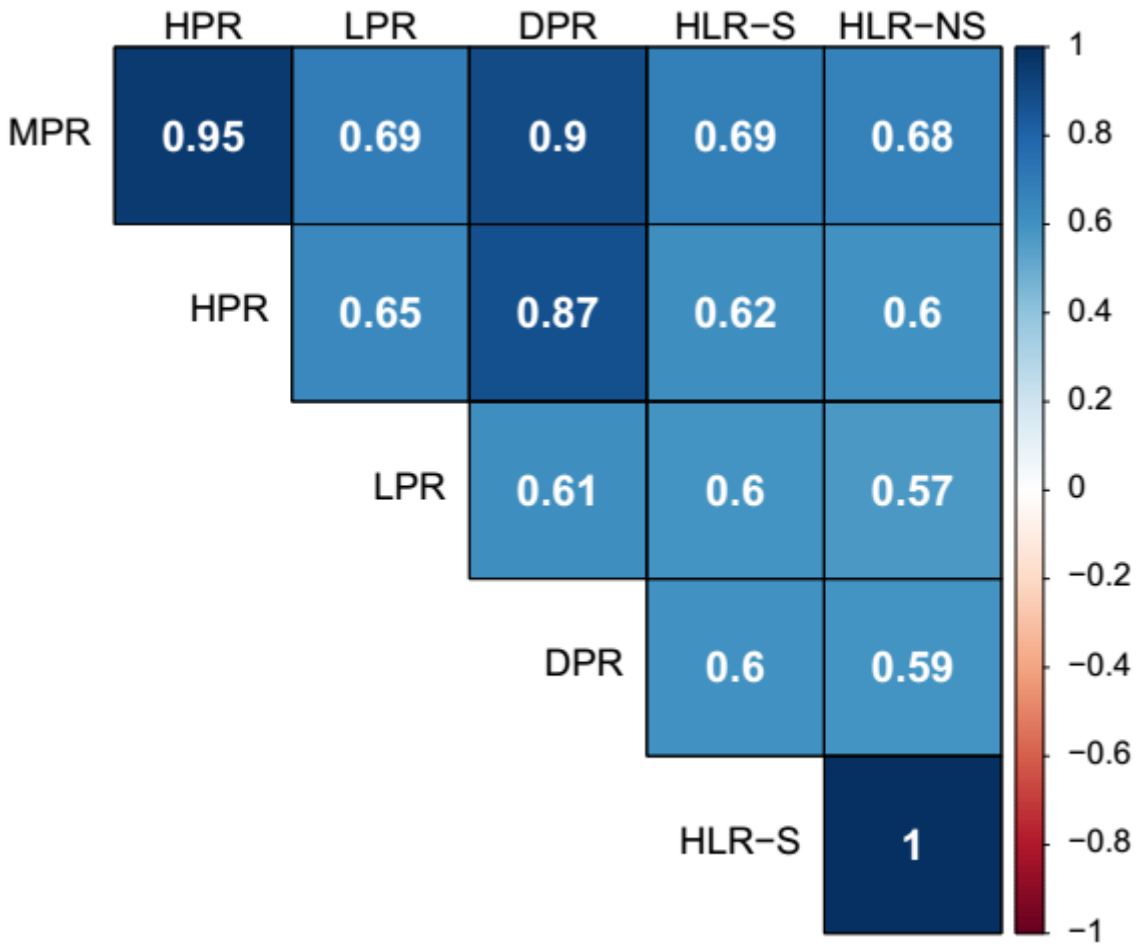


Figure 517: Average correlation of correlation coefficient posterior mean estimates between different JSDMs on the eucalypt dataset.

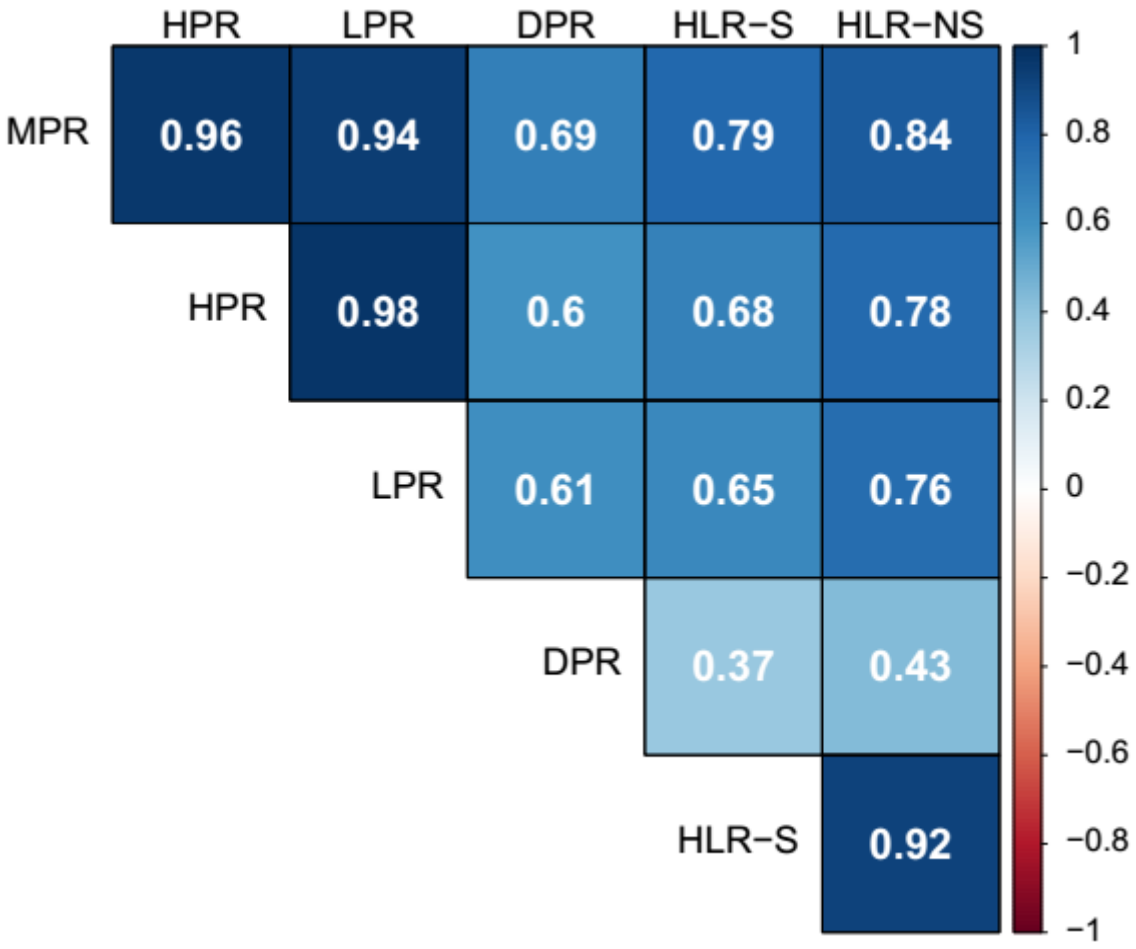


Figure 518: Average correlation of correlation coefficient posterior mean estimates between different JSDMs on the frog dataset.

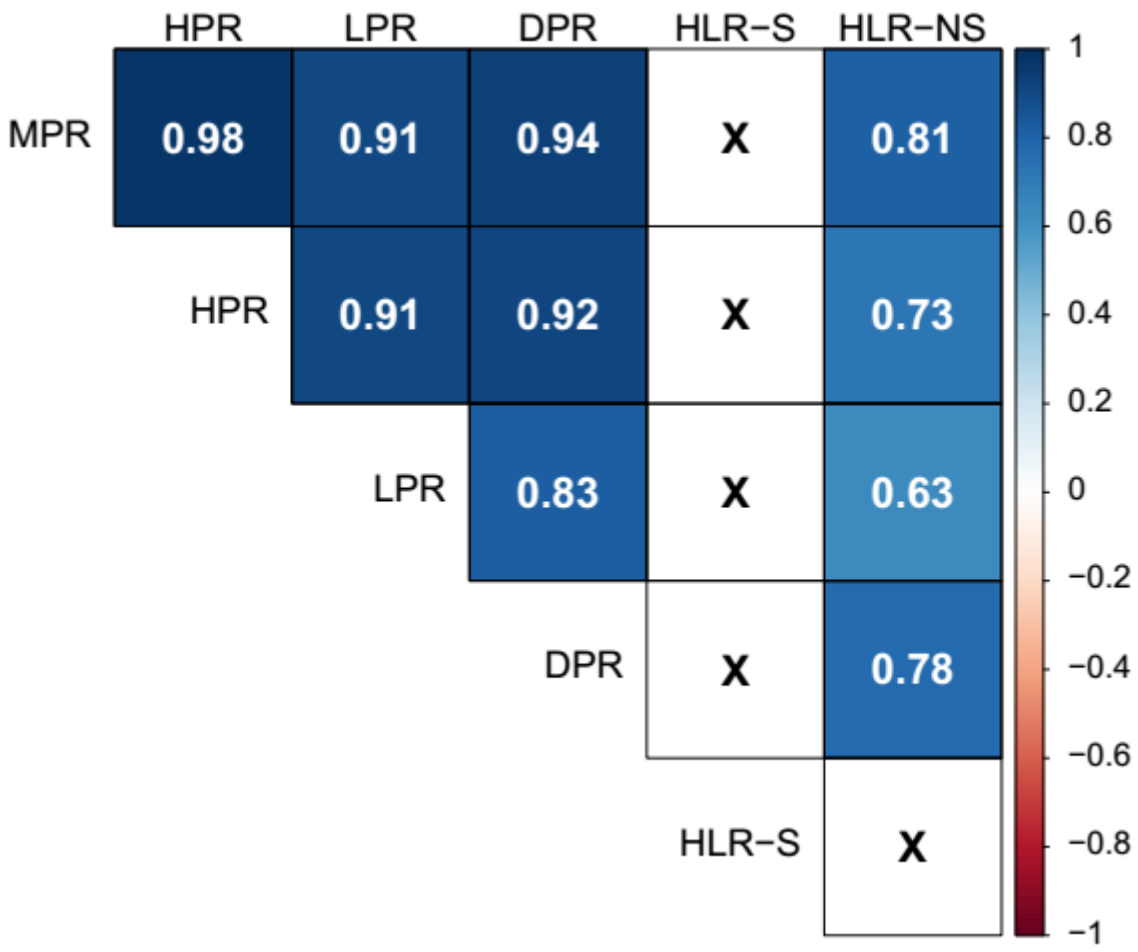


Figure 519: Average correlation of correlation coefficient posterior mean estimates between different JSDMs on the fungi dataset. 'X' marks pairs where one or more JSDMs did not run.

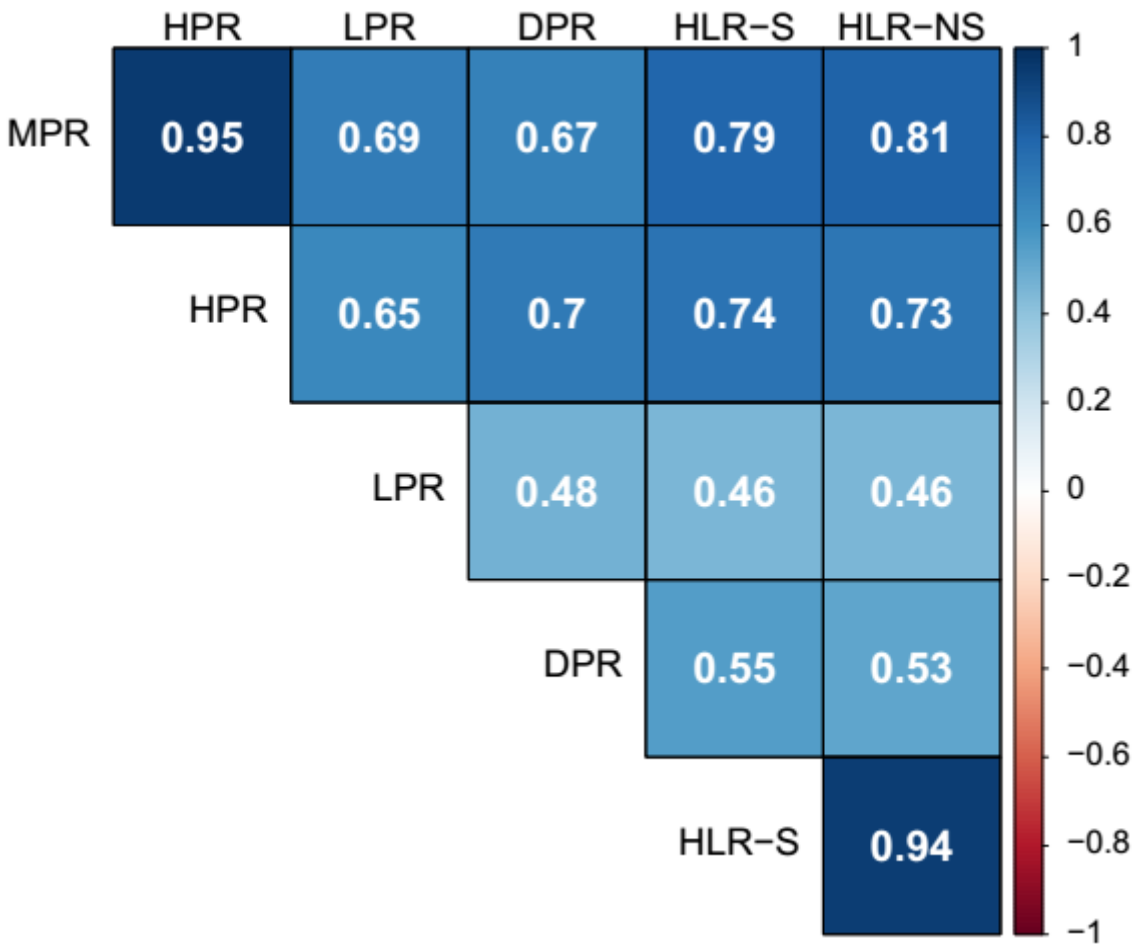


Figure 520: Average correlation of correlation coefficient posterior mean estimates between different JSDMs on the mosquito dataset.

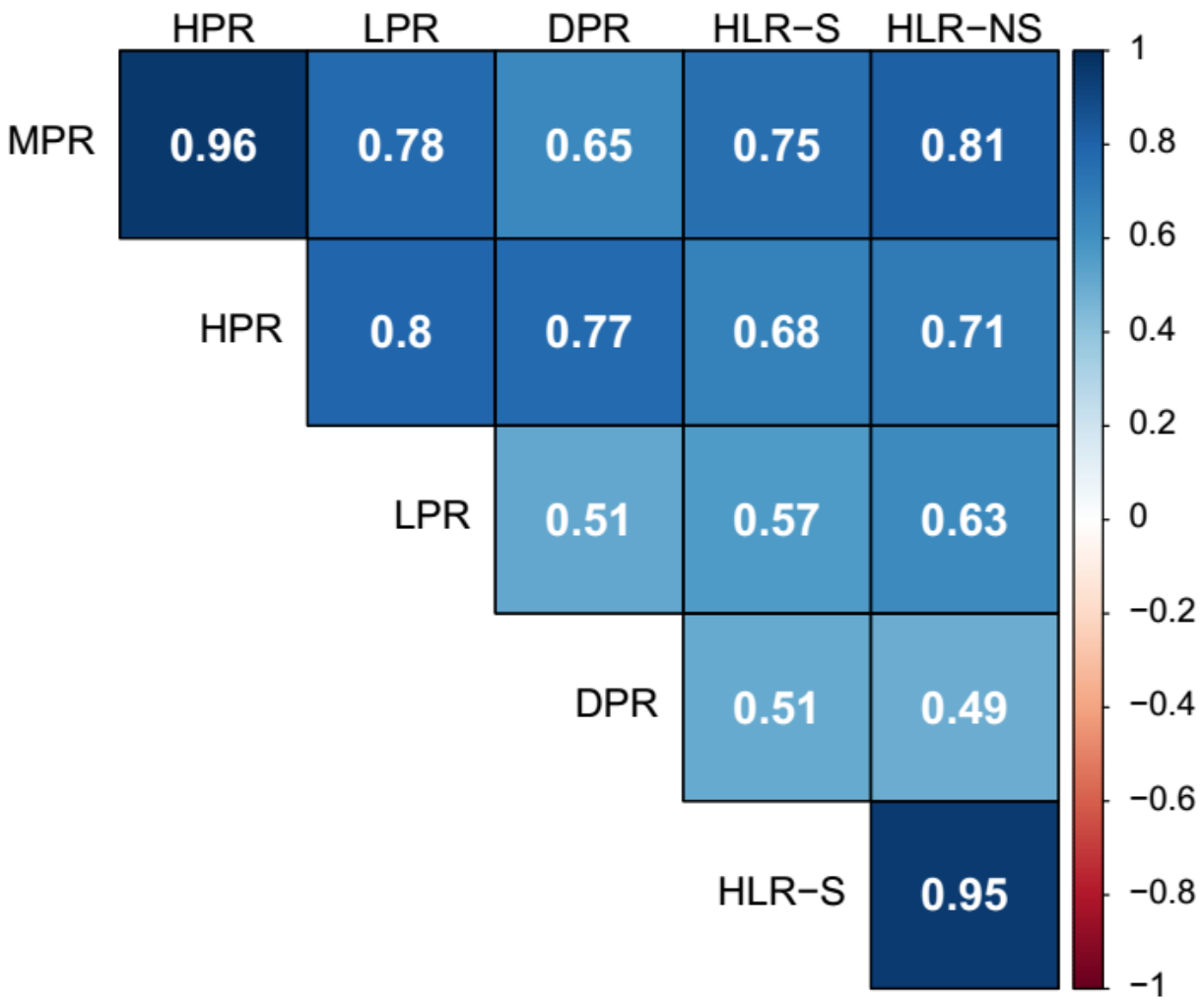


Figure 521: Average correlation of correlation coefficient posterior mean estimates between different JSDMs for all datasets.

APPENDIX F

SIMULATED DATASET COMPARISON

INTRODUCTION

The analysis in the main paper made use of real datasets, but we also explored the use of simulated datasets. By simulating the data according to different combinations of components corresponding to different model assumptions of the JSDMs under consideration, we can attempt to isolate if and when the models perform differently. These components correspond to the different model assumptions that make the JSDMs different to each other. In addition to being able to compare performance of JSDMs on the same dataset, we are able to compare their relative performance changes across datasets simulated under these different assumptions.

METHODS

Data simulation

We generated eight types of simulated datasets, with ten replications each, to cover the three main model assumptions in the JSDMs. These three assumptions are:

- Utilising a full-rank vs low-rank correlation matrix
- Hierarchical or non-hierarchical regression coefficients
- Only accounting for measured variables or the ability to estimate the effect of both measured and unmeasured variables

The datasets were generated for ten species at 100 sites. The different aspects of the datasets were generated as follows:

- Full-rank correlation matrix
    - The covariance matrix, $\boldsymbol{\Sigma}$, was generated using an inverse Wishart distribution with $2 * J$ degrees of freedom
    - $\boldsymbol{\Sigma}$ was rescaled to the correlation matrix $\boldsymbol{R}$
- Low-rank correlation matrix
    - The low-rank correlation matrix was simulated assuming three latent factors
    - The factor loadings matrix, λ, was drawn from a $N(0,1)$ distribution
    - $\boldsymbol{\Sigma}$ was calculated as $\boldsymbol{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \boldsymbol{I}$
    - $\boldsymbol{\Sigma}$ was rescaled to the correlation matrix $\boldsymbol{R}$
- Non-hierarchical regression coefficients

    - The species-specific coefficients were drawn from a $N(0,1)$ distribution
- Hierarchical regression coefficients
    - The covariate-specific distribution means, ω, were drawn from a $N(0,1)$ distribution
    - The species-specific coefficients were drawn from a $N(\omega, 1)$ distribution
- Measured variables
    - We generated four measured variables
    - The variables were drawn from a $N(0,1)$ distribution
    - An additional column of all 1s was supplied for models that require an intercept column in the data
- Measured and unmeasured variables
    - We generated four measured and two unmeasured variables using identical methods. The difference between the two is that only the measured variables were supplied to the models as data. The unmeasured variables are used to test the latent factor estimation
    - The variables were drawn from a $N(0,1)$ distribution
    - An additional column of all 1s was supplied for models that require an intercept column in the data
- Presence-absence data
    - Our presence-absence data was simulated using the latent variable parameterisation approach common to the JSDMs. We took random draws, z, from a multivariate normal distribution constructed using the using the simulated data from the above components depending on the dataset in question.
    - The mean values of the multivariate normal distribution were calculated as the regression coefficients multiplied by the values of the variables (dependent on dataset combination)
    - The covariance matrix of the multivariate normal distribution was either the full-rank or low-rank correlation matrix. The correlation matrix can be substituted for the covariance matrix here as the probit models constrain the variance to 1.
    - The latent variable draws were then converted to presence-absence records using a step function as follows$1(z > 0)$

Model fitting

For this simulation comparison we considered only a subset of the models covered in the main comparison: MPR, HPR, LPR, DPR, and HLR-NS. We did not include MLR as it was dropped in the main analysis for being too inefficient. We did not include HLR-S as we did not look at spatial components in residual co-occurrence in the simulation.

All models were fit using the same default settings as the main analysis (including prior selection and MCMC regimes). As we did not include a computational comparison for the simulated datasets all model/data combinations were instead run on high performance computing infrastructure (University of Melbourne 2017) instead of sequentially on a local machine.

Analysis

As per the main analysis, we conducted our analysis separately for the regression and correlation coefficients estimated by the models. We compared for each model the posterior mean of the regression coefficient estimates and the correlation estimates to their true values (separately for each dataset) using two metrics: mean difference of estimates from true values, and root mean square error (RMSE).

RESULTS

Across all of the simulated datasets there are no major trends suggesting results contradictory to the comparison of real datasets in the main paper. The different JSDMs all perform similarly to each other. The main difference is that the LPR model, like in the real data comparison, is symptomatic of poor fit to the data when using its default fitting values.

Across all datasets, the mean difference of parameter estimates from their true values for both regression (Table S1) and correlation (Table S2) coefficients showed no evidence of bias. In all cases the mean difference is near zero and much smaller than their respective standard deviations.

The RMSE values for the regression coefficients (Table S3) show no significant difference between the different models except for LPR. LPR returns relatively high RMSE values, regularly double or more than the RMSE of other models, symptomatic of the same poor convergence using default settings issue we explored in Appendix S2. The RMSE values for

the correlation coefficients (Table S4) show no major differences between the models. The RMSE values for HPR and LPR are slightly higher than those returned by the other models.

If we look at the effect of data component choice we can start to identify some trends that occur across all models. First, we do not observe any effect of simulating the data with a full or low-rank correlation matrix on the estimation performance across models for either the regression or correlation coefficients. Second, simulating the data with either hierarchical or non-hierarchical regression coefficients has an effect on the estimation of regression coefficients but no impact on correlation coefficients. Models that account for hierarchical regression coefficients have better performance than other models where the data has been generated hierarchically, but only if they are not also impacted by other model assumptions. HLR-NS, which accounts for unmeasured variables, always has better performance (lower RMSE) than non-hierarchical models, while HPR does not show better performance for any dataset where latent factor related assumptions (low-rank correlation matrix and/or unmeasured variables) were included. Third, we observe trends in models responding to the datasets simulated either with or without unmeasured variables driving species occurrence. Ignoring LPR and its convergence issues, all models perform better for regression coefficients on datasets with only measured variables except for MPR which performs similarly for both. For correlation coefficients, all models perform better for datasets with only measured variables except for HPR which performs similarly for both. HLR-NS, the latent factor model without convergence issues, showed a much smaller difference in RMSE between datasets that did and did not include unmeasured variables.

DISCUSSION

We observed no evidence of any apparent bias in parameter estimation for any JSDM, and that all JSDMs performed similarly for the simulated datasets just as they did for the real datasets in the main paper. The LPR model, however, showed relatively high RMSE values for the regression coefficients compared to the other models. We believe this can be explained by the fact that, in the implementation of the model we tested (*boral* v1.3.1), there are no constraints applied to the variance of the normal distribution. The variance is not constrained to equal 1, unlike in a standard probit model, and the RMSE for regression coefficients increases as more latent factors are included in the model. This is related to the issue of lack of identifiability of the variance in the probit model. Tests not reported here suggest that, if regression coefficients are rescaled to correct for this lack of identifiability,

then LPR would return values more in line with the other models. The trends we observed were related to the components of the simulated data and not between models.

Two of the datasets correspond to aspects relevant to the latent factor approach and we found that only one of them had an observable effect on parameter estimation. We observed no difference in model estimates when considering data simulated with either a full or low-rank correlation matrix approach. In contrast, we did observe differences in parameter estimates when comparing datasets that were simulated with only measured variables or with both measured and unmeasured variables. All models performed better (smaller RMSE) for both regression and correlation coefficients when fit to datasets that did not include unmeasured variables. This is not surprising as the models are trying to explain the variation in species occurrence due to environmental effects without information on all of the driving factors. When fit to datasets with unmeasured variables, the models are trying to explain the variation resulting from six covariates with information on only four of them, with the other variables effectively acting as noise, hence making estimation more uncertain. This was observed in HLR-NS as well, despite using the latent factor approach to account for unmeasured variables, but the difference in performance between these two dataset types was much smaller than the other models. This suggests the method can successfully attribute some of the residual variation as unmeasured variables.

The third dataset component was whether the regression coefficients were simulated hierarchically or non-hierarchically. Models that account for hierarchical regression coefficients performed better than other models for datasets generated hierarchically, but only where they also matched assumptions related to latent factors. This suggests that there are benefits to accounting for hierarchical regression coefficients, but that the effect of latent factor assumptions, especially unmeasured variables, has a much greater impact on parameter estimation.

Here we have shown that the different JSDMs perform similarly across the different simulated datasets in the same way as they do to the real datasets in the main paper. We do find, however, that all models respond to certain assumptions in how they data are structured. Models that account for these assumptions show less of an impact on parameter estimation than those that don’t but are not able to entirely explain the additional variation.

Datasets	Dataset Components			Models				
	Correlation Matrix	Regression Coefficients	Variables	MPR	DPR	HPR	LPR	HLR-NS
1	Full-rank	Non-hierarchical	Measured only	**0.004 ± 0.48**	**-0.015 ± 0.25**	-0.013 ± 0.34	0.011 ± 1.56	-0.010 ± 0.26
2	Full-rank	Non-hierarchical	Measured + unmeasured	0.029 ± 0.41	0.045 ± 0.47	0.031 ± 0.54	-0.052 ± 1.13	0.023 ± 0.42
3	Full-rank	Hierarchical	Measured only	0.029 ± 0.59	-0.006 ± 0.36	**-0.074 ± 0.43**	0.186 ± 1.72	-0.004 ± 0.30
4	Full-rank	Hierarchical	Measured + unmeasured	-0.029 ± 0.56	-0.023 ± 0.69	-0.058 ± 0.74	-0.100 ± 1.21	-0.033 ± 0.55
5	Low-rank	Non-hierarchical	Measured only	-0.005 ± 0.44	-0.000 ± 0.24	0.006 ± 0.29	**0.103 ± 1.69**	-0.006 ± 0.24
6	Low-rank	Non-hierarchical	Measured + unmeasured	0.030 ± 0.40	0.034 ± 0.44	0.030 ± 0.50	-0.077 ± 1.24	0.028 ± 0.40
7	Low-rank	Hierarchical	Measured only	0.038 ± 0.59	-0.020 ± 0.34	-0.065 ± 0.66	0.176 ± 1.82	**0.010 ± 0.33**
8	Low-rank	Hierarchical	Measured + unmeasured	-0.020 ± 0.59	-0.021 ± 0.70	-0.013 ± 0.70	0.068 ± 1.18	**-0.015 ± 0.55**

Table 1: Mean $\pm$ standard deviation for the difference of regression coefficient estimates from their true values for each model/dataset combination. The dataset components correspond to the method used to simulate the datasets. The bold boxes denote where the model assumptions of a JSDM match those used to simulate a dataset.

Datasets	Dataset Components			Models				
	Correlation Matrix	Regression Coefficients	Variables	MPR	DPR	HPR	LPR	HLR-NS
1	Full-rank	Non-hierarchical	Measured only	**0.014 ± 0.21**	**0.016 ± 0.22**	0.025 ± 0.38	0.015 ± 0.32	0.017 ± 0.26
2	Full-rank	Non-hierarchical	Measured + unmeasured	0.014 ± 0.36	0.011 ± 0.31	0.029 ± 0.38	0.015 ± 0.50	0.015 ± 0.31
3	Full-rank	Hierarchical	Measured only	0.010 ± 0.23	0.014 ± 0.25	**0.032 ± 0.36**	-0.003 ± 0.31	0.018 ± 0.27
4	Full-rank	Hierarchical	Measured + unmeasured	0.068 ± 0.34	0.047 ± 0.30	0.010 ± 0.36	0.076 ± 0.47	0.056 ± 0.31
5	Low-rank	Non-hierarchical	Measured only	-0.023 ± 0.24	-0.021 ± 0.24	-0.042 ± 0.45	**0.002 ± 0.29**	-0.035 ± 0.32
6	Low-rank	Non-hierarchical	Measured + unmeasured	-0.033 ± 0.38	-0.033 ± 0.34	-0.038 ± 0.44	-0.027 ± 0.51	-0.030 ± 0.36
7	Low-rank	Hierarchical	Measured only	-0.027 ± 0.25	-0.017 ± 0.24	-0.032 ± 0.44	-0.017 ± 0.29	**-0.034 ± 0.32**
8	Low-rank	Hierarchical	Measured + unmeasured	0.046 ± 0.38	0.024 ± 0.34	-0.060 ± 0.42	0.078 ± 0.50	**0.016 ± 0.36**

Table 2: Mean $\pm$ standard deviation for the difference of correlation coefficient estimates from their true values for each model/dataset combination. The dataset components correspond to the method used to simulate the datasets. The bold boxes denote where the model assumptions of a JSDM match those used to simulate a dataset.

Dataset	Dataset Components			Models				
	Correlation Matrix	Regression Coefficients	Variables	MPR	DPR	HPR	LPR	HLR-NS
1	Full-rank	Non-hierarchical	Measured only	0.481	0.255	0.342	1.560	0.261
2	Full-rank	Non-hierarchical	Measured + unmeasured	0.413	0.473	0.543	1.128	0.421
3	Full-rank	Hierarchical	Measured only	0.591	0.363	0.436	1.726	0.304
4	Full-rank	Hierarchical	Measured + unmeasured	0.562	0.690	0.746	1.217	0.555
5	Low-rank	Non-hierarchical	Measured only	0.435	0.236	0.295	1.688	0.236
6	Low-rank	Non-hierarchical	Measured + unmeasured	0.400	0.442	0.505	1.246	0.405
7	Low-rank	Hierarchical	Measured only	0.590	0.340	0.666	1.826	0.326
8	Low-rank	Hierarchical	Measured + unmeasured	0.586	0.701	0.702	1.180	0.546

Table S3: RMSE calculated for the estimated vs true regression coefficients for each model/dataset combination. The dataset components correspond to the method used to simulate the datasets. The bold boxes denote where the model assumptions of a JSDM match those used to simulate a dataset.

Dataset	Dataset Components			Models				
	Correlation Matrix	Regression Coefficients	Variables	MPR	DPR	HPR	LPR	HLR-NS
1	Full-rank	Non-hierarchical	Measured only	0.210	0.225	0.378	0.321	0.256
2	Full-rank	Non-hierarchical	Measured + unmeasured	0.356	0.311	0.376	0.499	0.311
3	Full-rank	Hierarchical	Measured only	0.228	0.249	0.363	0.314	0.271
4	Full-rank	Hierarchical	Measured + unmeasured	0.344	0.305	0.363	0.475	0.314
5	Low-rank	Non-hierarchical	Measured only	0.244	0.241	0.454	0.286	0.323
6	Low-rank	Non-hierarchical	Measured + unmeasured	0.386	0.343	0.442	0.514	0.358
7	Low-rank	Hierarchical	Measured only	0.250	0.243	0.443	0.292	0.320
8	Low-rank	Hierarchical	Measured + unmeasured	0.383	0.345	0.427	0.511	0.365

Table S4: RMSE calculated for the estimated vs true correlation coefficients for each model/dataset combination. The dataset components correspond to the method used to simulate the datasets. The bold boxes denote where the model assumptions of a JSDM match those used to simulate a dataset.

APPENDIX G

EVALUATION METRICS

CHAPTER TWO

Name	Formula	Range	Binary predictions	Probabilistic predictions	Threshold-dependent	Threshold-independent	Community dissimilarity metrics	Species Richness	Likelihood	Interpretation
Accuracy / True Skill Statistic	$\frac{TP+TN}{TP+FP+TN+FN}$	0-1	X	X	X					Higher = more correct classifications
Area under the Receiver Operating Characteristic curve (AUC)		0-1		X		X				Higher = more correct classifications
Bray-Curtis dissimilarity	$\frac{A+B-2J}{A+B}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X			X			0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Canberra Index	$\frac{A+B-2J}{A+B-J}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X			X			0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Cohen's Kappa	$\frac{(TP+TN)-\frac{(TP+FN)*(TP+FP)+(TN+FP)*(TN+FN)}{TP+FP+TN+FN}}{(TP+FP+TN+FN)-\frac{(TP+FN)*(TP+FP)+(TN+FP)*(TN+FN)}{TP+FP+TN+FN}}$	0-1	X		X					Higher = more correct classifications
Diagnostic odds ratio	$\frac{TP/FP}{FN/TN}$ or $\frac{Positive\ likelihood\ ratio}{Negative\ likelihood\ ratio}$	0-∞	X	X	X					Higher = more correct classifications

F_1 score	$\frac{2}{\frac{1}{Sensitivity}+\frac{1}{Precision}}$	0-1	X	X	X		Higher = more correct classifications
False discovery ratio	$\frac{FP}{TP+FP}$	0-1	X	X	X		Higher = more predicted presences are incorrect
False negative rate	$\frac{FN}{TP+FN}$	0-1	X	X	X		Higher = more observed presences are predicted incorrect
False omission rate	$\frac{FN}{TN+FN}$	0-1	X	X	X		Higher = more predicted absences are incorrect
False positive rate	$\frac{FP}{FP+TN}$	0-1	X	X	X		Higher = more observed absences are predicted incorrect
Gower Index	$\frac{A+B-2J}{M}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences M = Total number of species	0-1	X	X		X	0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Gower Index (alternative)	$\frac{A+B-2J}{A+B-J}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X		X	0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Jaccard distance	$\frac{2D}{1+D}$ Where: D = Bray-Curtis	0-1	X	X		X	0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Kendall rank correlation coefficient	$\frac{(\#\ concordant\ pairs)-(\#\ discordant\ pairs)}{n(n-1)/2}$	-1 – 1		X	X		>0 = more correct predictions <0 = more incorrect predictions
Kulczynski Index	$1-\frac{\frac{J}{A}+\frac{J}{B}}{2}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X		X	0 = same composition 1 = opposite composition Lower = more correct assemblage prediction

Log Likelihood – Independent		-∞ - ∞		X			X	Higher = Model fit to data better
Log Likelihood – Joint		-∞ - ∞		X			X	Higher = model fit to data better
Mountford Index	Positive root of the equation: $e^{A*E} + e^{B*E} = 1 + e^{(A+B-J)*E}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences E = Mountford Index	0-1	X	X		X		0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Mean error (bias)	$\frac{1}{n}\sum_{i=1}^{n}(observed - predicted)$	-∞ - ∞		X		X		Increasing magnitude = more bias Positive value = more false absences (underprediction) Negative value = more false presences (overprediction)
Mean square error	$\frac{1}{n}\sum_{i=1}^{n}(observed - predicted)^2$	0-∞		X		X		Higher = increased error/bias Can't differentiate between over/underprediction
Negative likelihood ratio	$\frac{False\ negative\ rate}{True\ negative\ rate}$	0-∞	X	X	X			Higher = more predicted absences are incorrect
Negative predictive performance	$\frac{TN}{TN + FN}$	0-1	X	X	X			Higher = more observed absences are correctly identified
Pearson correlation coefficient	$\frac{cov(observed, predicted)}{\sigma_{observed}\sigma_{predicted}}$	-1 – 1		X		X		>0 = more correct predictions <0 = more incorrect predictions
Positive likelihood ratio	$\frac{True\ positive\ rate}{False\ positive\ rate}$	0-∞	X	X	X			Higher = more predicted presences are correct
Positive predictive performance / Precision	$\frac{TP}{TP + FP}$	0-1	X	X	X			Higher = more predicted presences are correct
R^2 / Coefficient of determination	$1 - \frac{\sum(observed - predicted)^2}{\sum(observed - \overline{observed})^2}$	0-1		X		X		Higher = predictions more closely match the observed values

Raup-Crick dissimilarity	$1 - \int_0^A hypergeometric(A, B, M, 1)$ Where: A = Number of observed presences B = Number of predicted presences M = Total number of species	0-1	X	X	X			0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Root mean square error	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(observed - predicted)^2}$	0-∞		X		X		Higher = increased error/bias Can't differentiate between over/underprediction
Spearman rank correlation coefficient	$\frac{cov(r_{obs}, r_{pred})}{\sigma_{r_{obs}}\sigma_{r_{pred}}}$	-1 – 1		X		X		>0 = more correct predictions <0 = more incorrect predictions
Species richness difference	$Predicted\ richness - observed\ richness$	-∞ - ∞	X	X			X	>0 = overpredicts species richness <0 = underpredicts species richness
Sum of squared errors	$\sum_{i=1}^{n}(observed - predicted)^2$	0-∞		X		X		Higher = increased error/bias Can't differentiate between over/underprediction
True negative rate / Specificity	$\frac{TN}{FP + TN}$	0-1	X	X	X			Higher = more correctly identified absences
True positive rate / Sensitivity	$\frac{TP}{TP + FN}$	0-1	X	X	X			Higher = more correctly identified presences
Youden's J statistic	$Sensitivity + Specificity - 1$	0-1	X	X	X			Higher = more correct classifications

APPENDIX H

FOREST PLOTS FOR EVALUATION METRIC ABSOLUTE VALUES

This appendix contains forest plots for each evaluation metric/prediction type combination. For convenience of navigation the figures are presented in subsection corresponding to prediction type and then ordered alphabetically by evaluation metric name.

1. **Marginal (probabilistic) predictions**

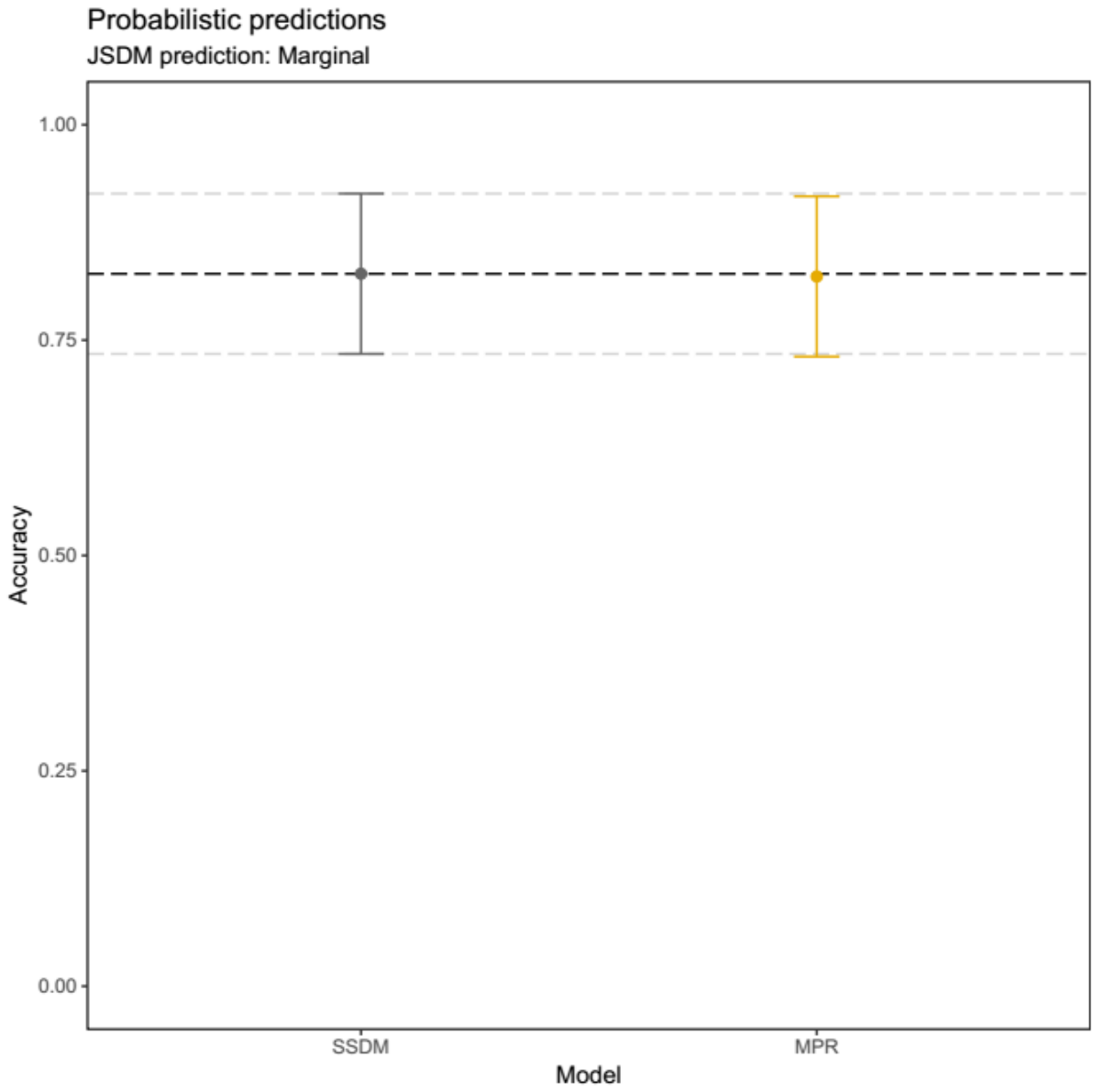


Figure 1: Performance of the SSDM and JSDM for the accuracy evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

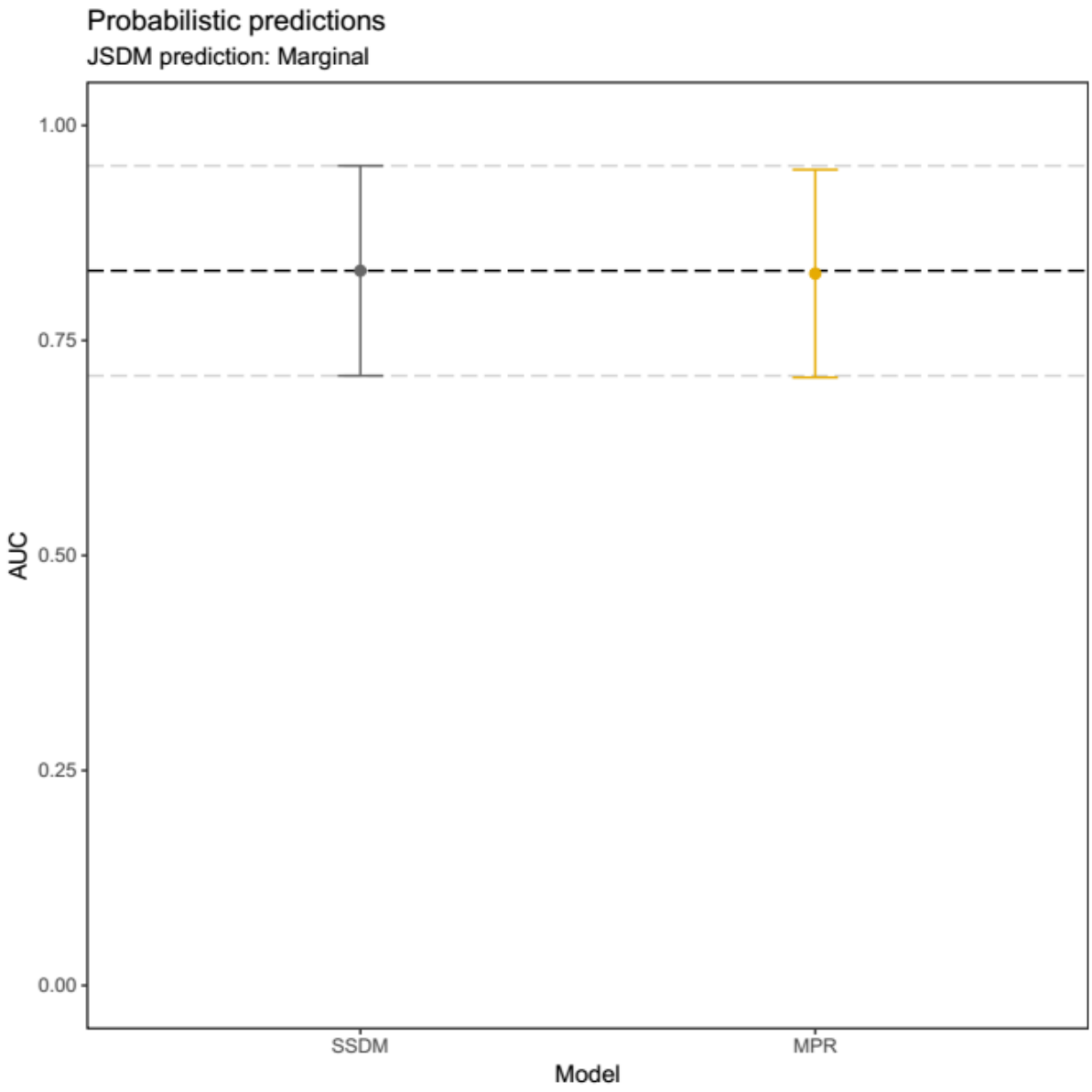


Figure 2: Performance of the SSDM and JSDM for the AUC evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

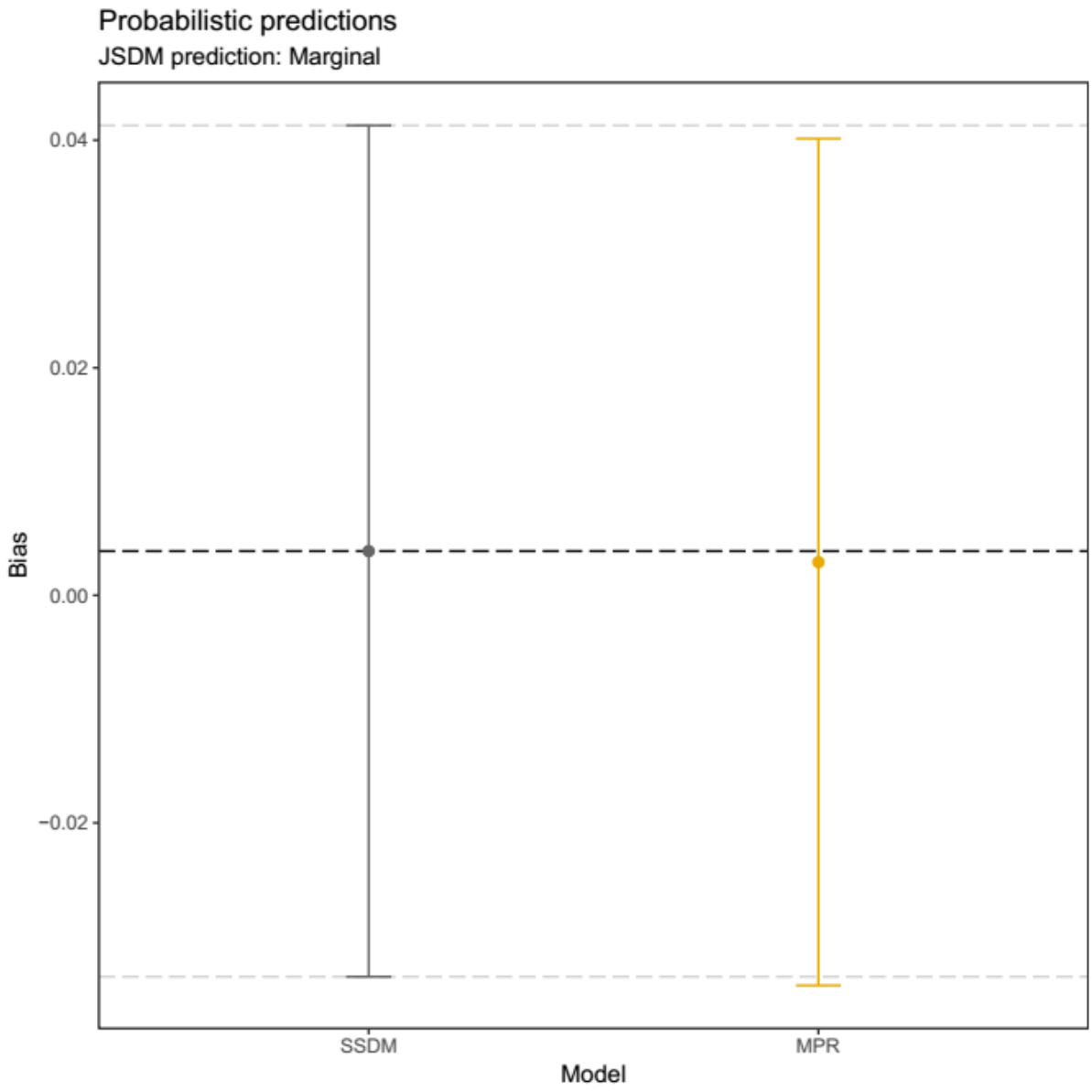


Figure 3: Performance of the SSDM and JSDM for the bias/mean error evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

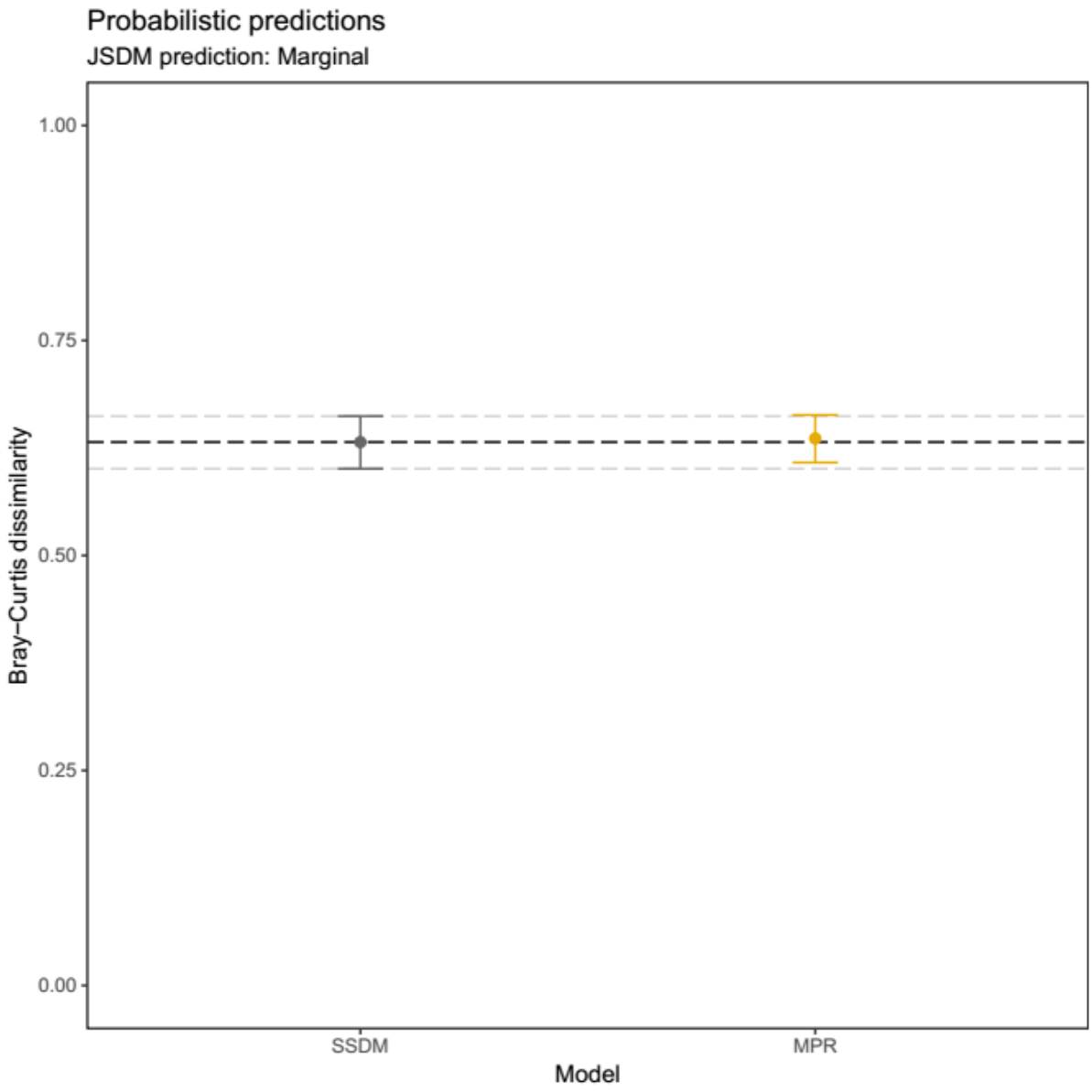


Figure 4: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

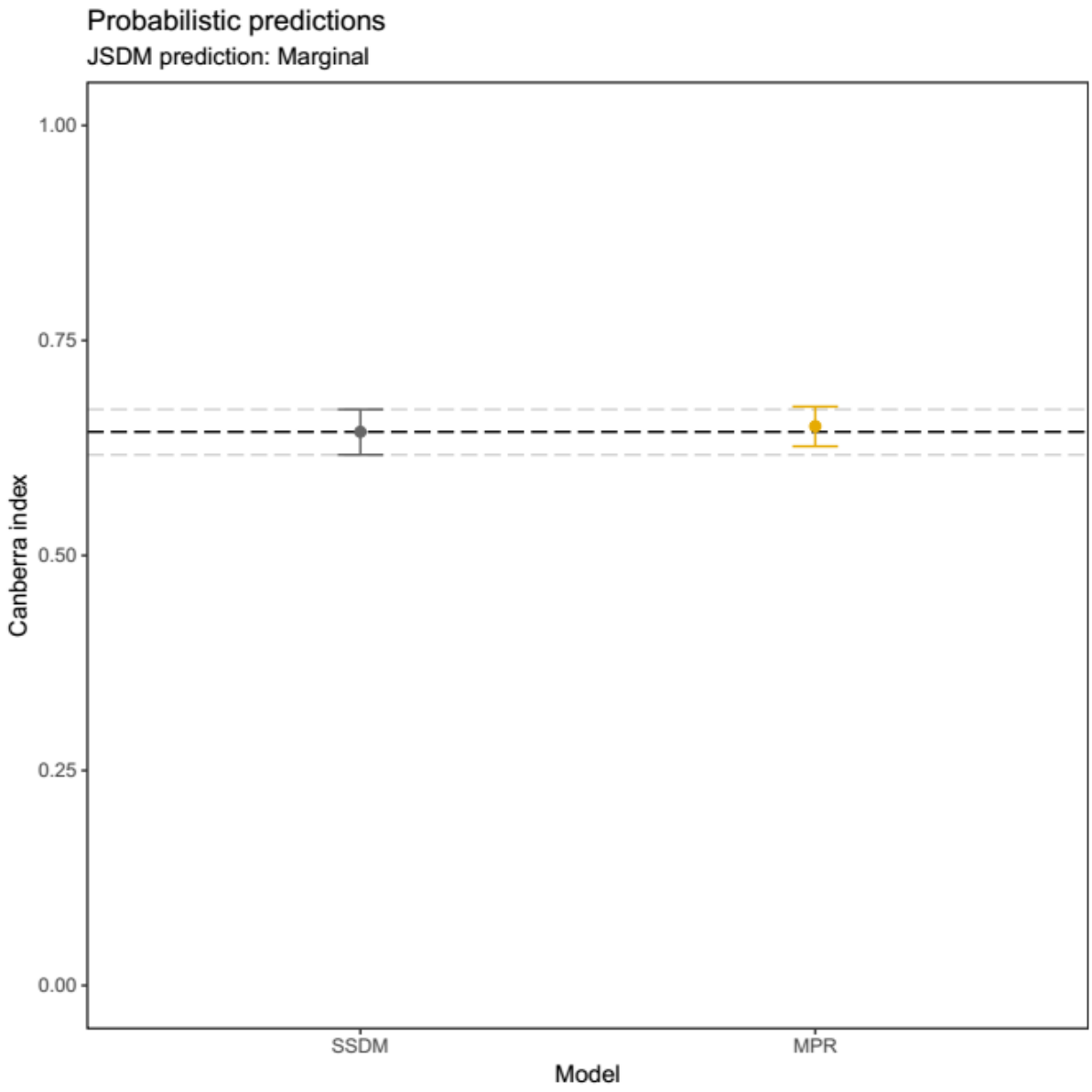


Figure 5: Performance of the SSDM and JSDM for the Canberra index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

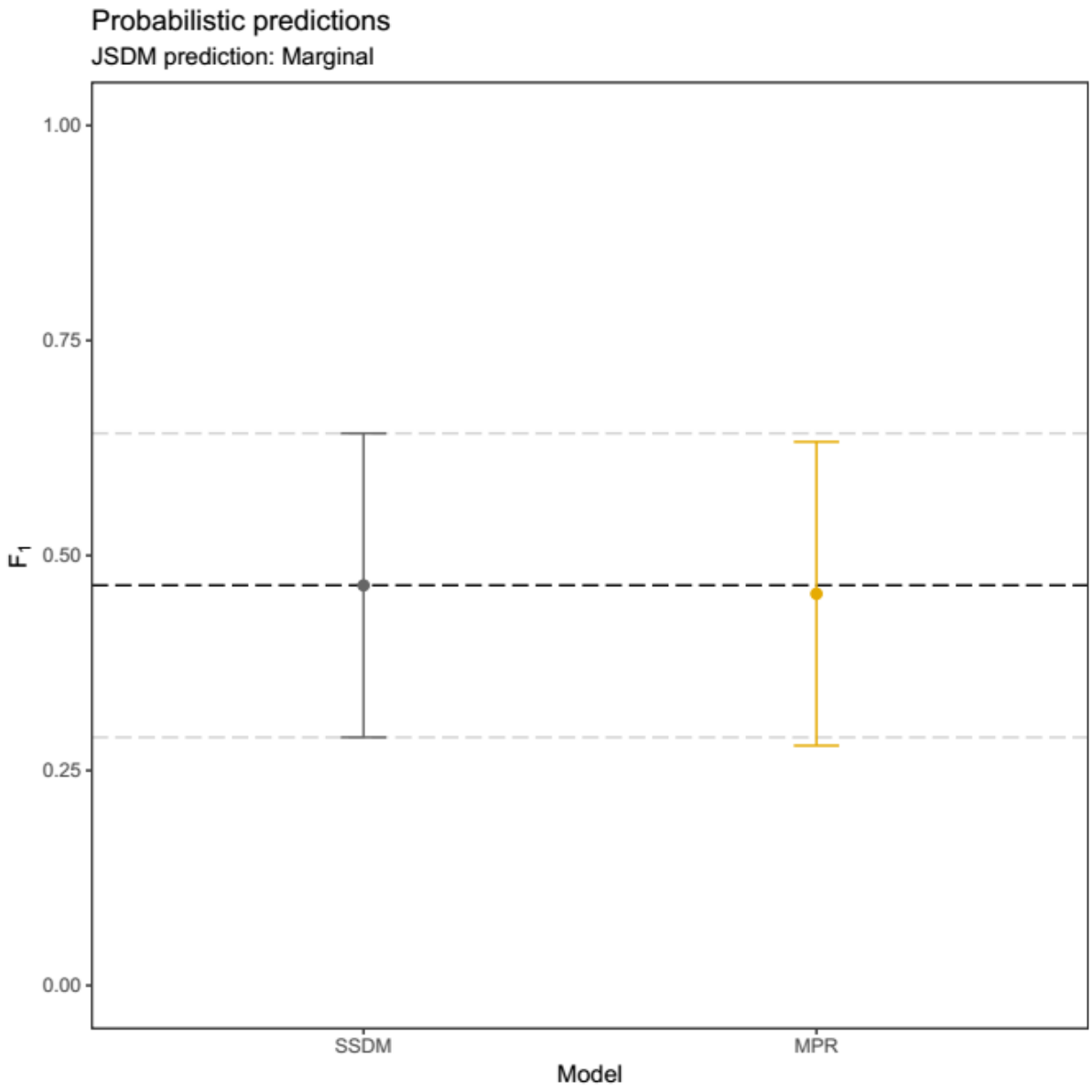


Figure 6: Performance of the SSDM and JSDM for the F_1evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

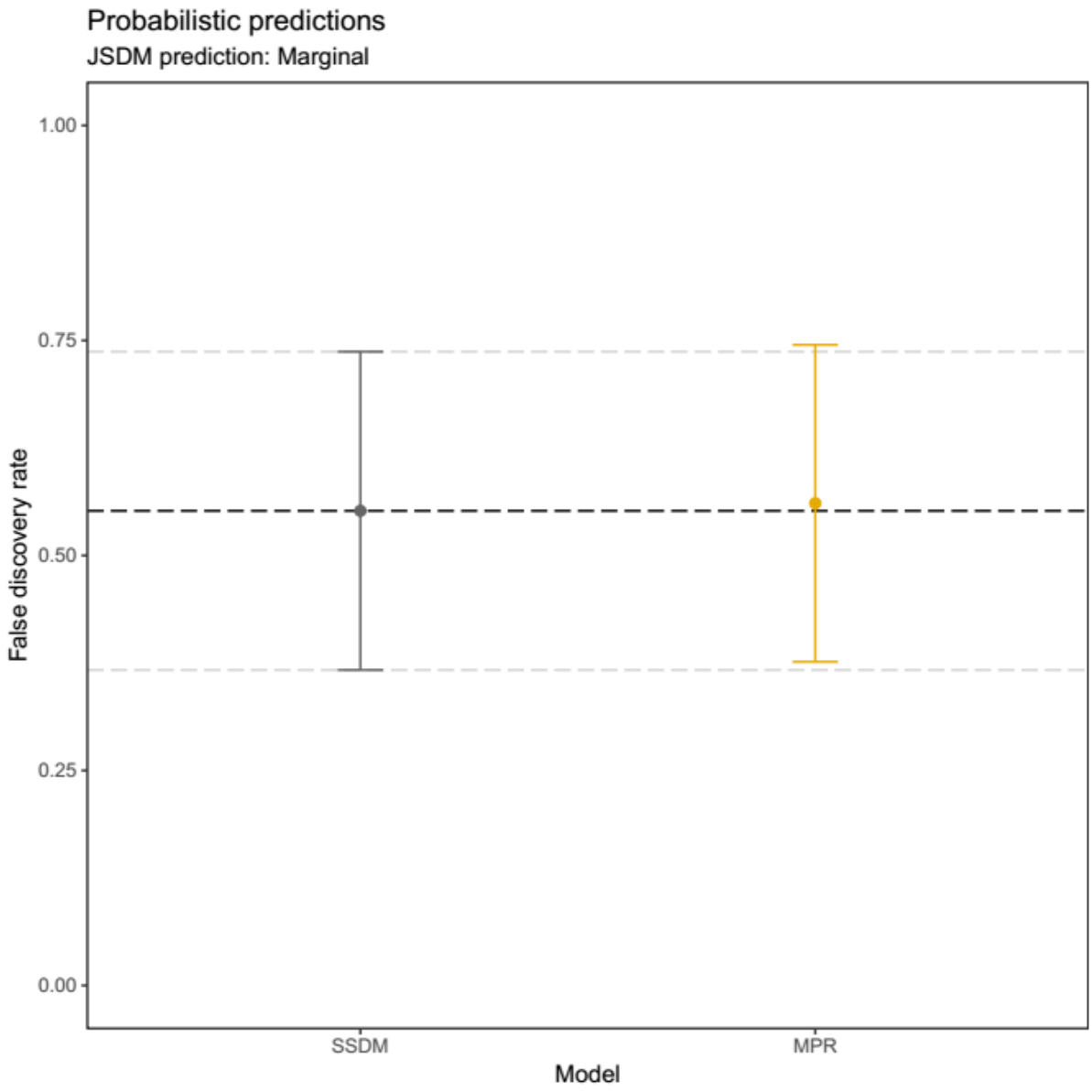


Figure 7: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

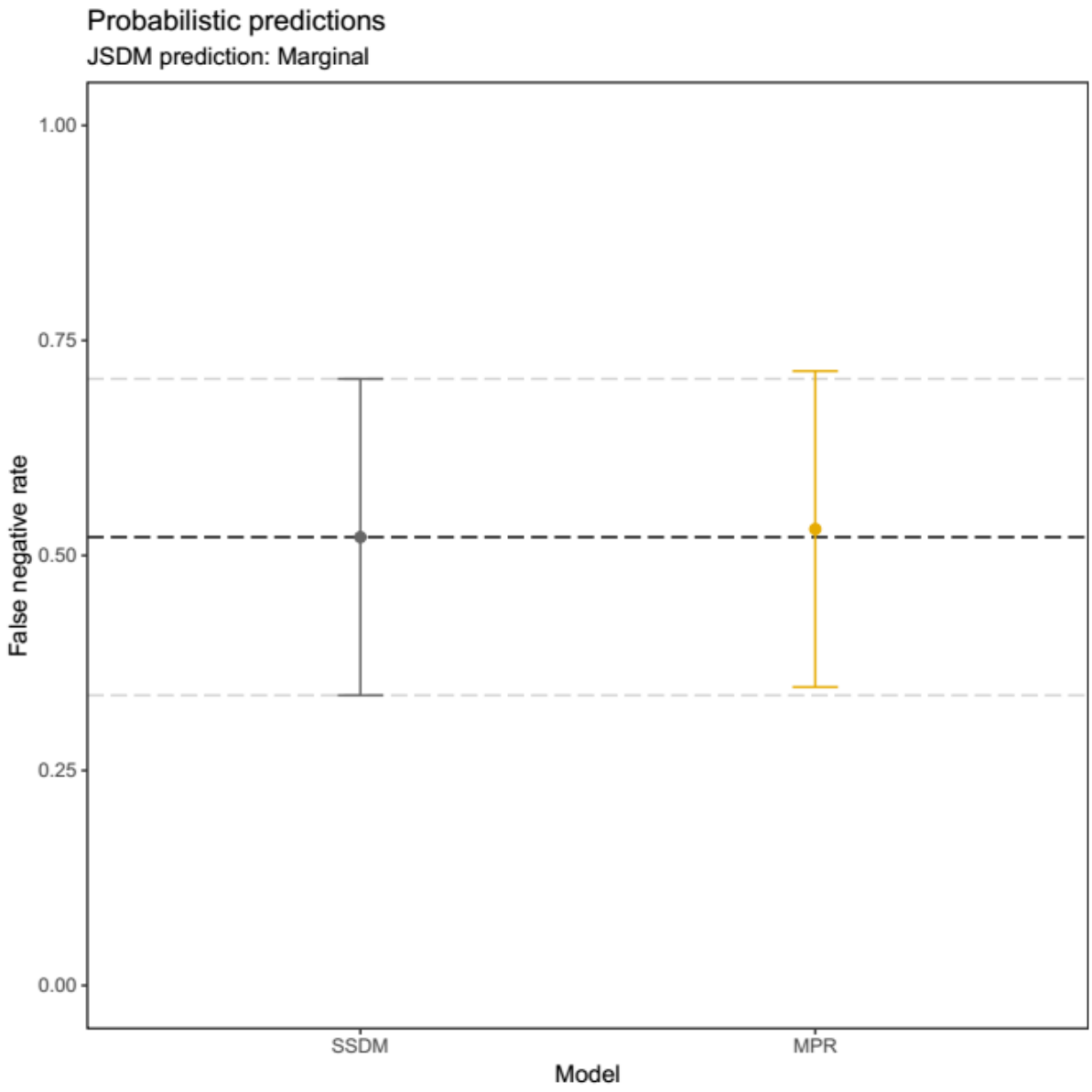


Figure 8: Performance of the SSDM and JSDM for the false negative rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

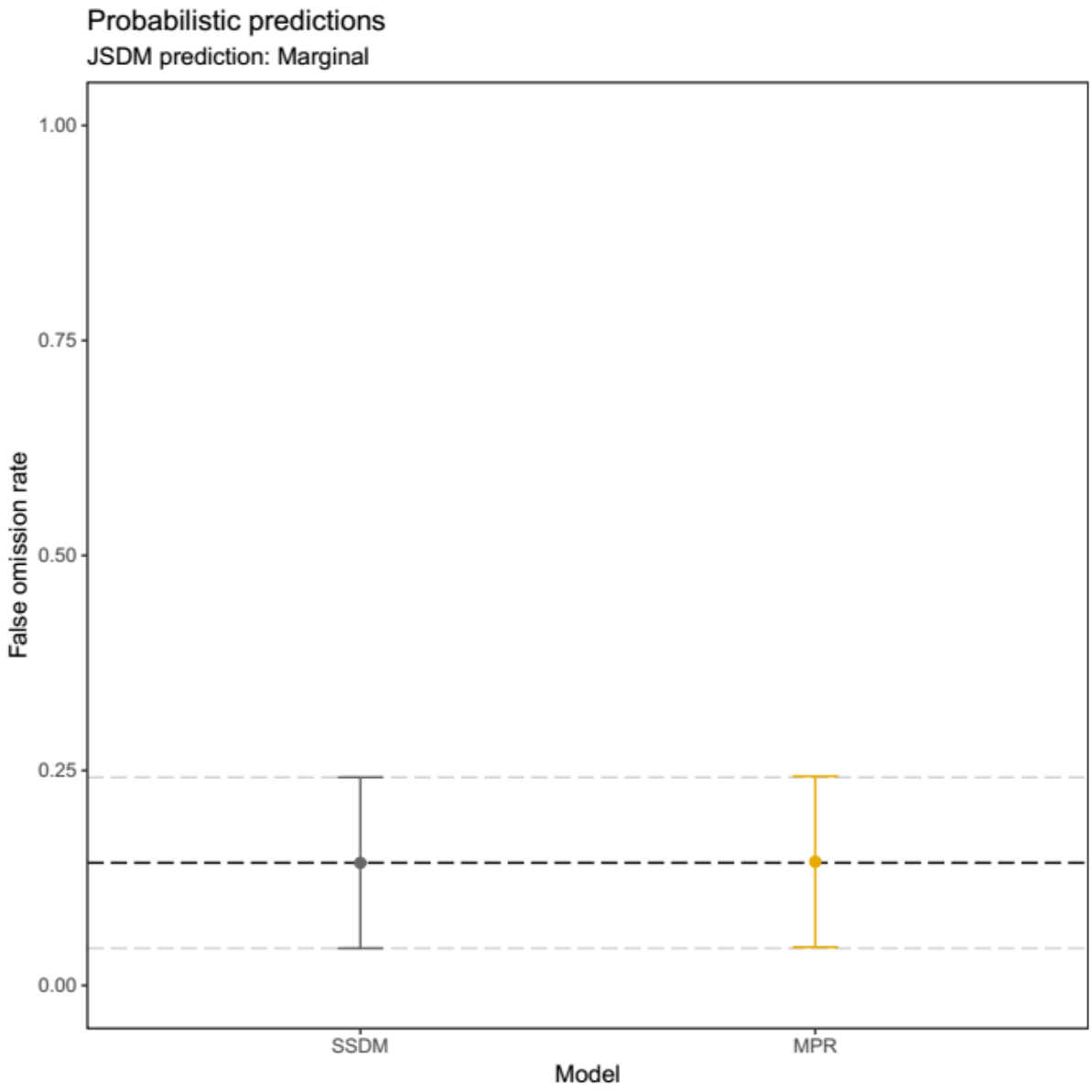


Figure 9: Performance of the SSDM and JSDM for the false omission rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

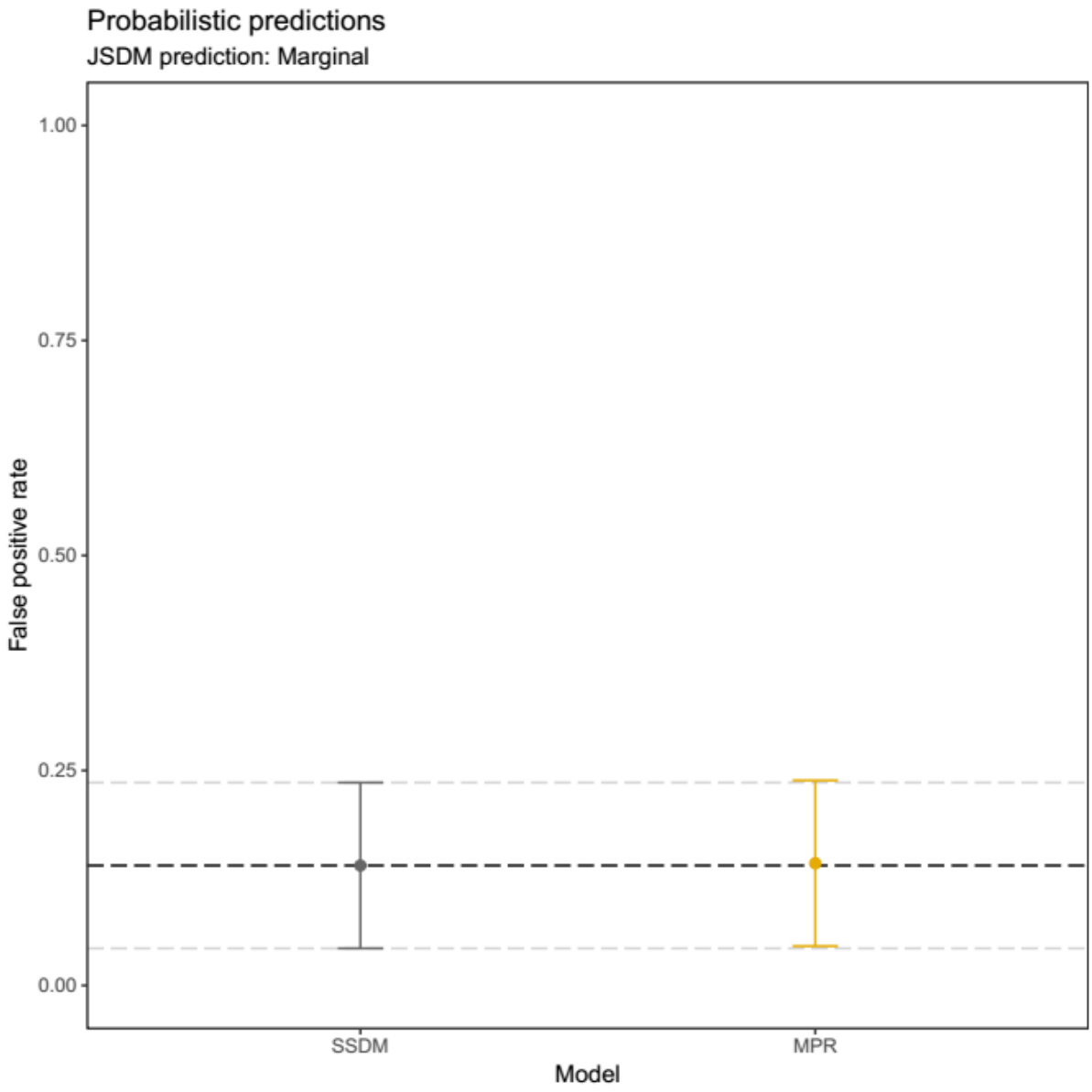


Figure 10: Performance of the SSDM and JSDM for the false positive rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

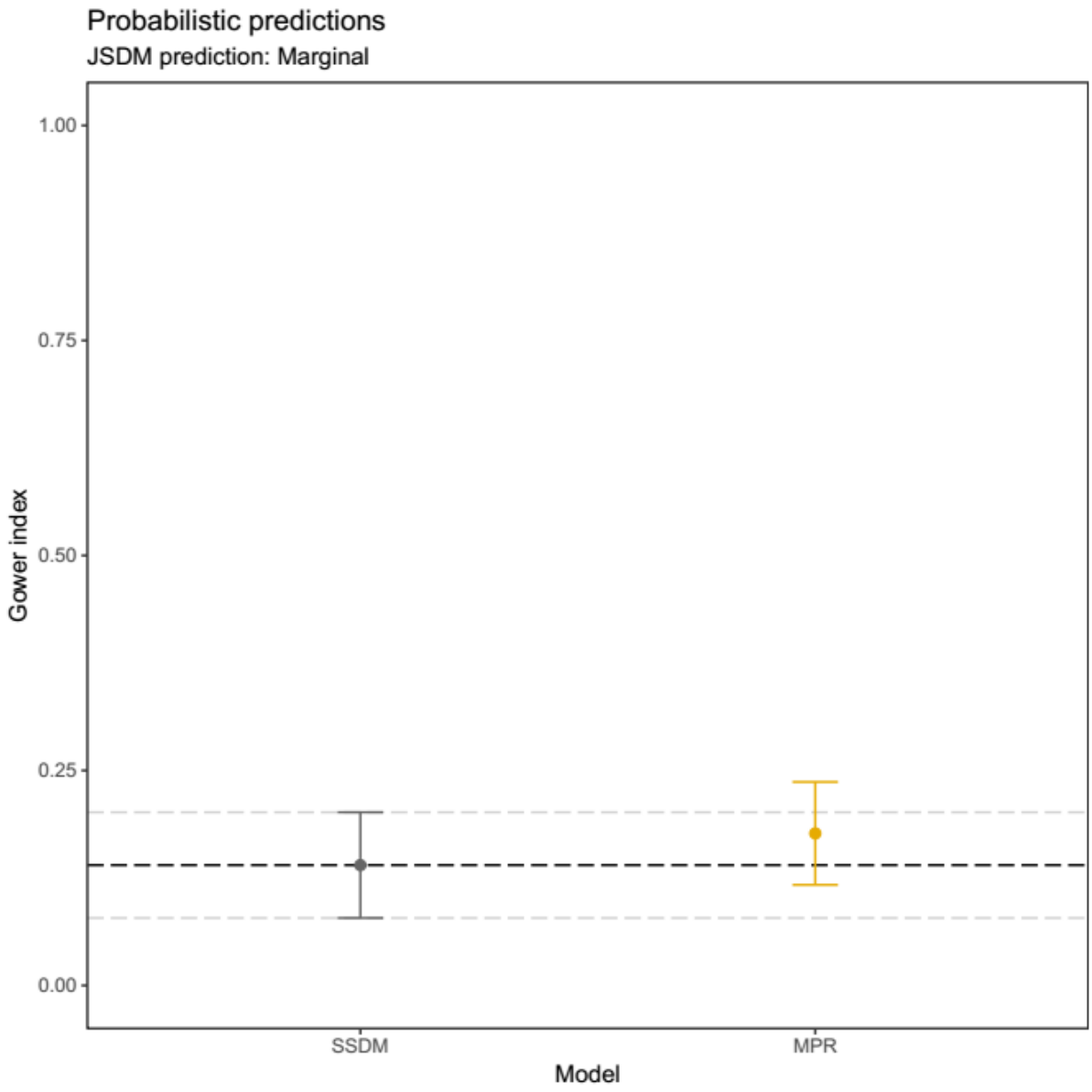


Figure 11: Performance of the SSDM and JSDM for the Gower index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

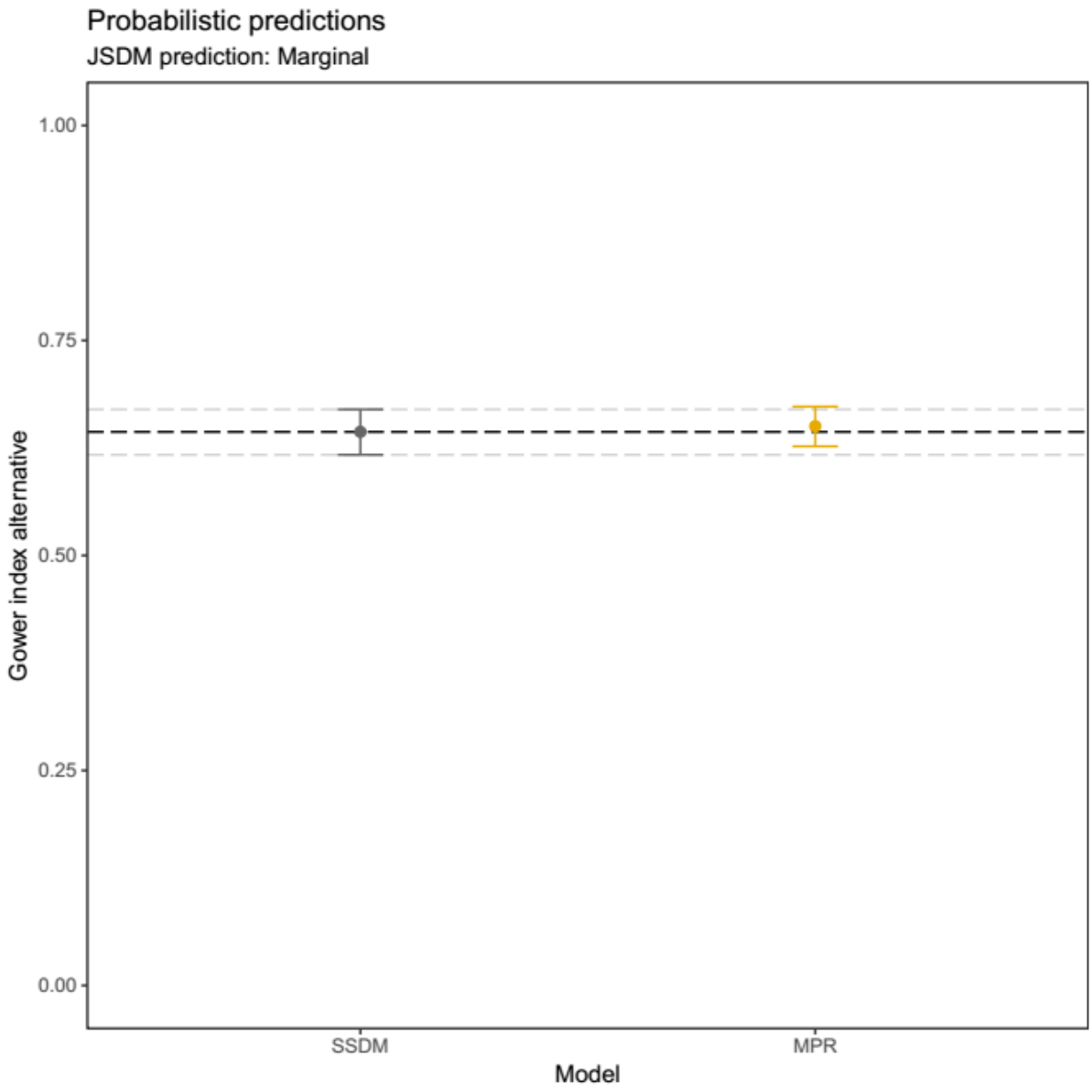


Figure 12: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

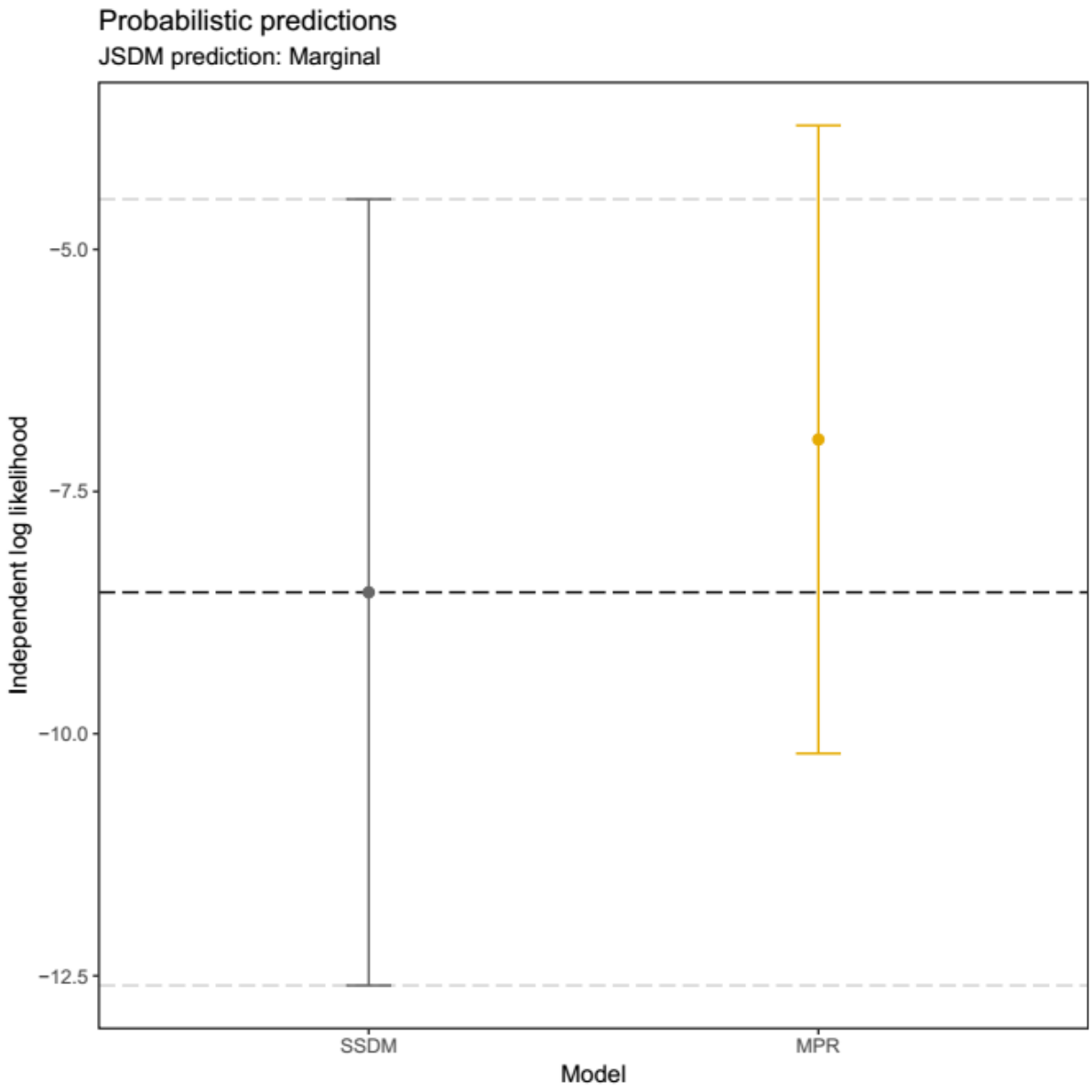


Figure 13: Performance of the SSDM and JSDM for the independent log likielihood evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

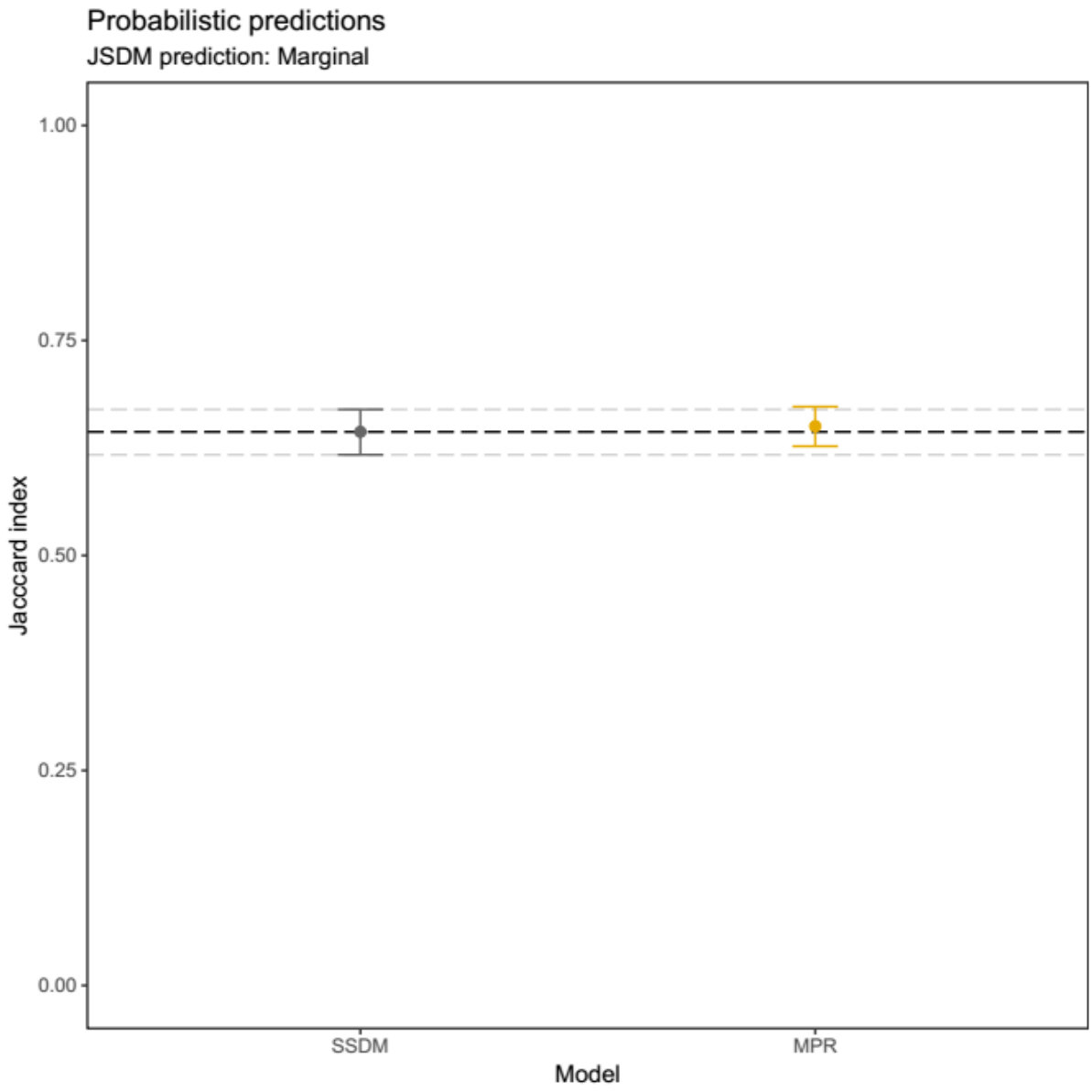


Figure 14: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

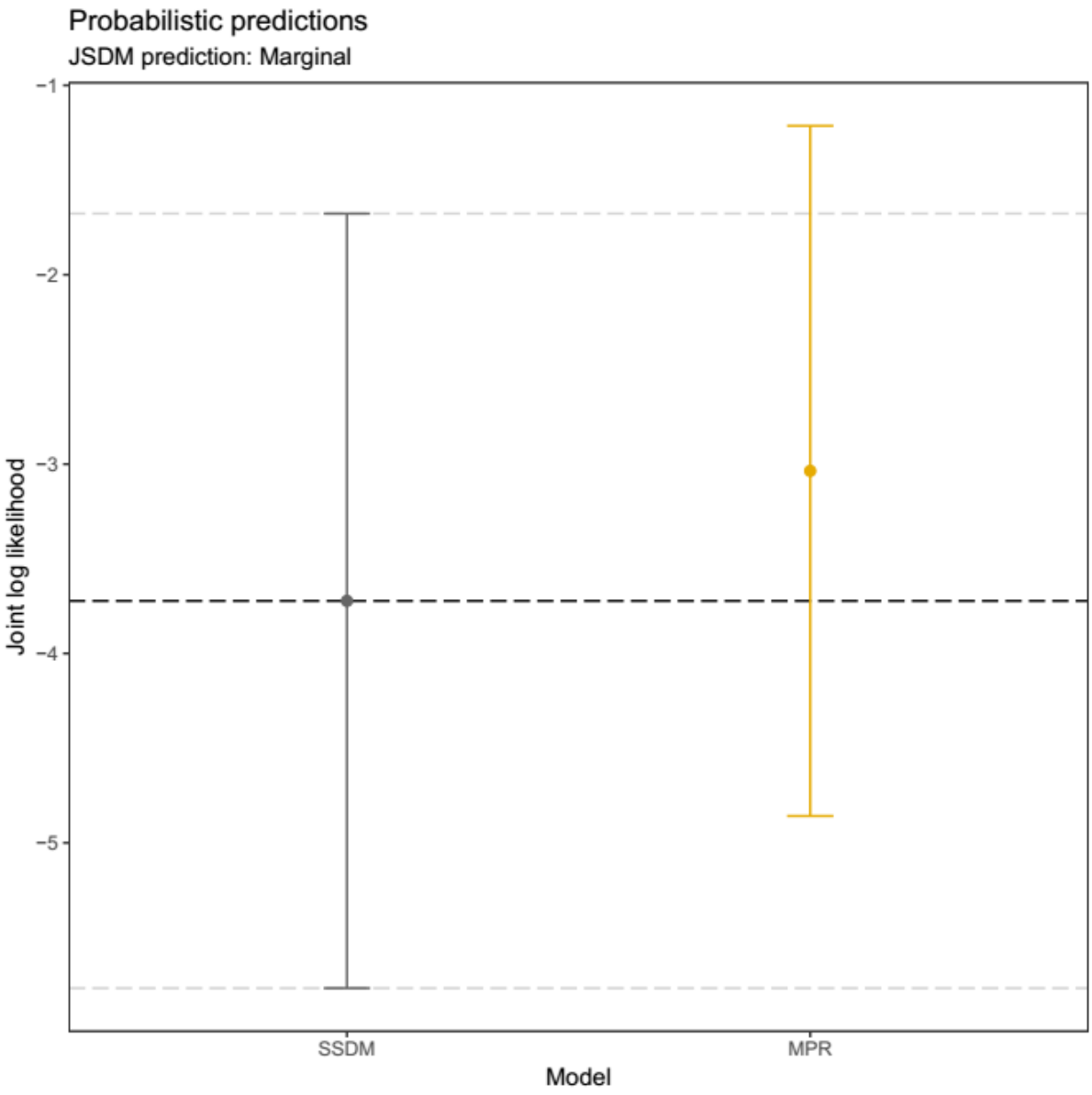


Figure 15: Performance of the SSDM and JSDM for the joint log likelihood evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

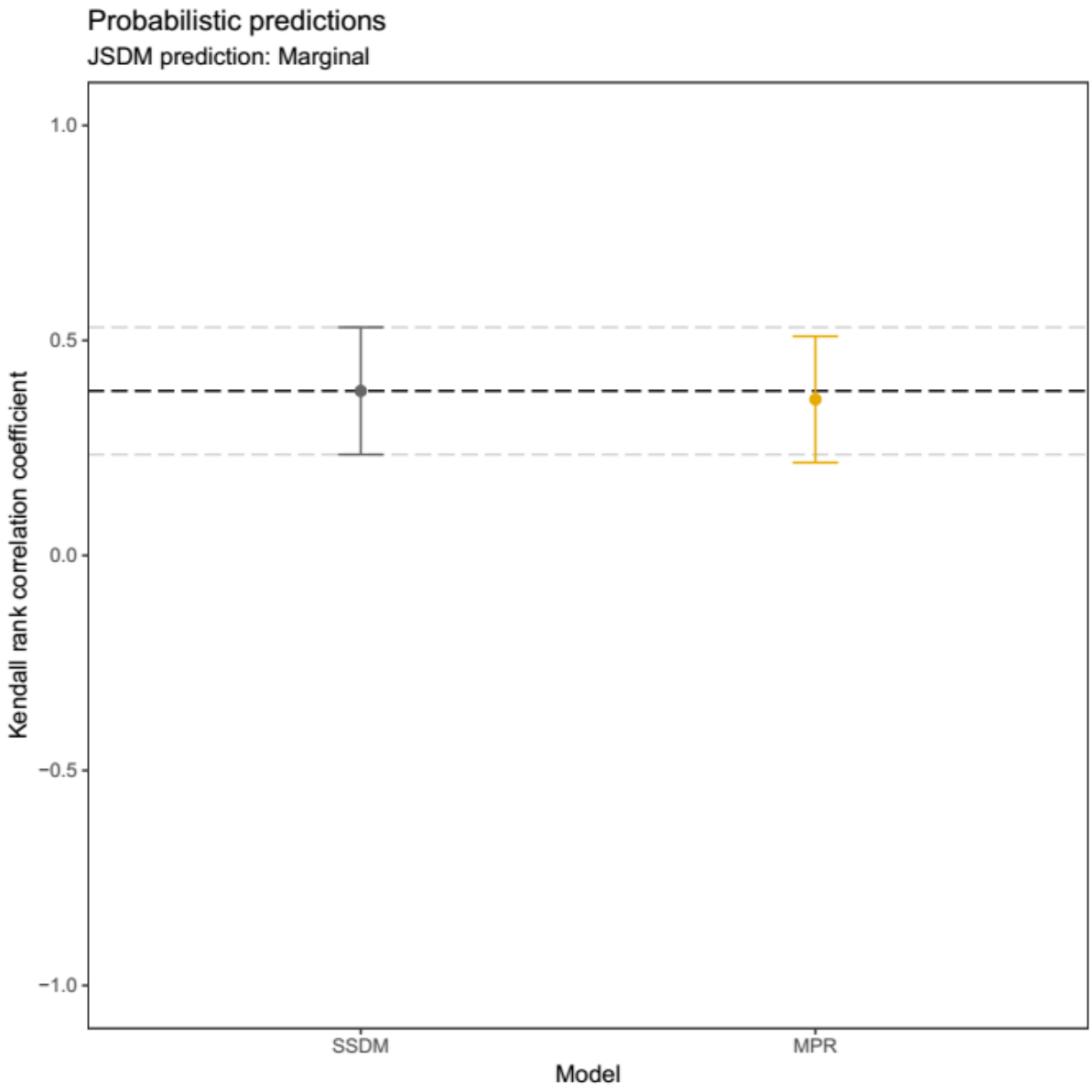


Figure 16: Performance of the SSDM and JSDM for the Kendall rank correlation coefficient evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

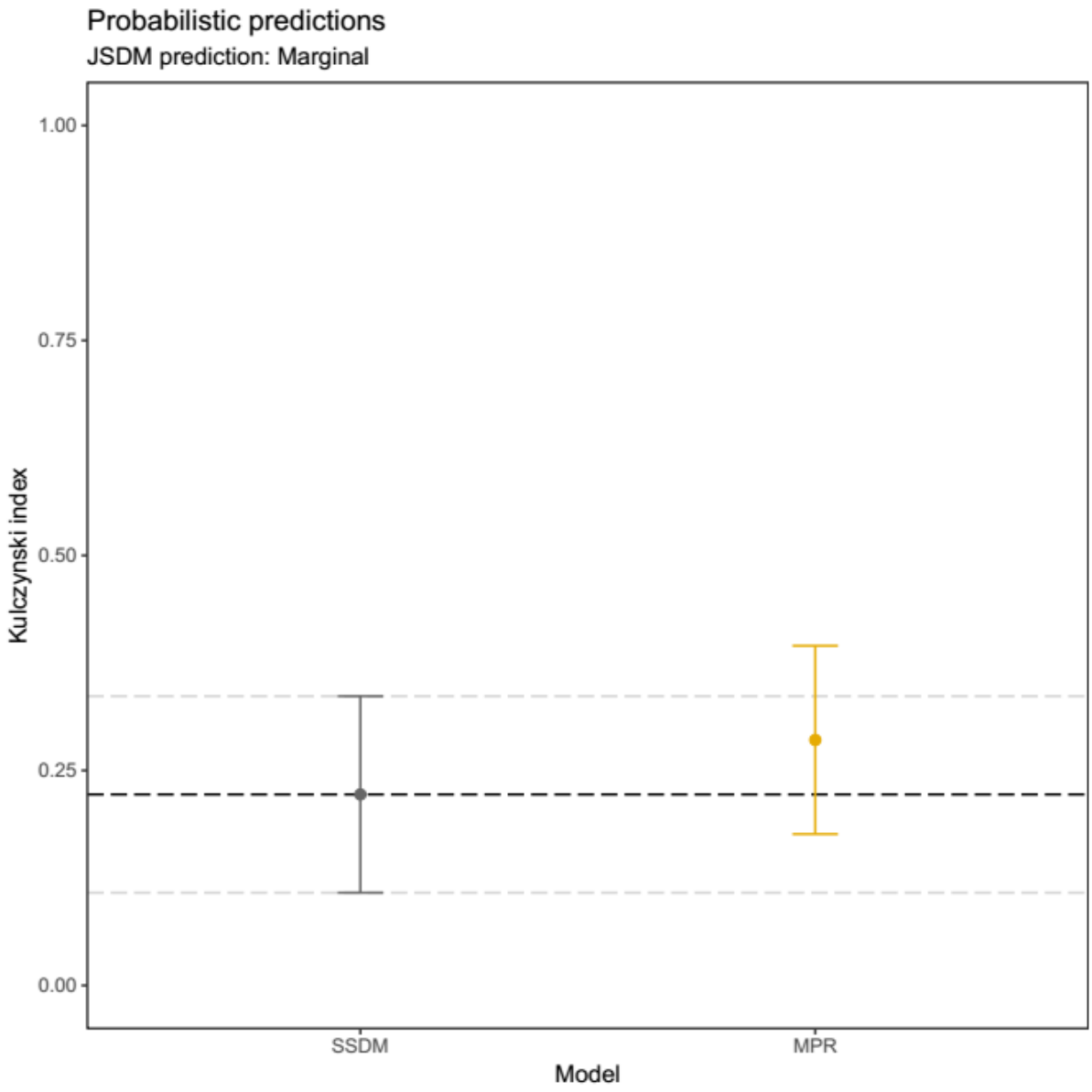


Figure 17: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

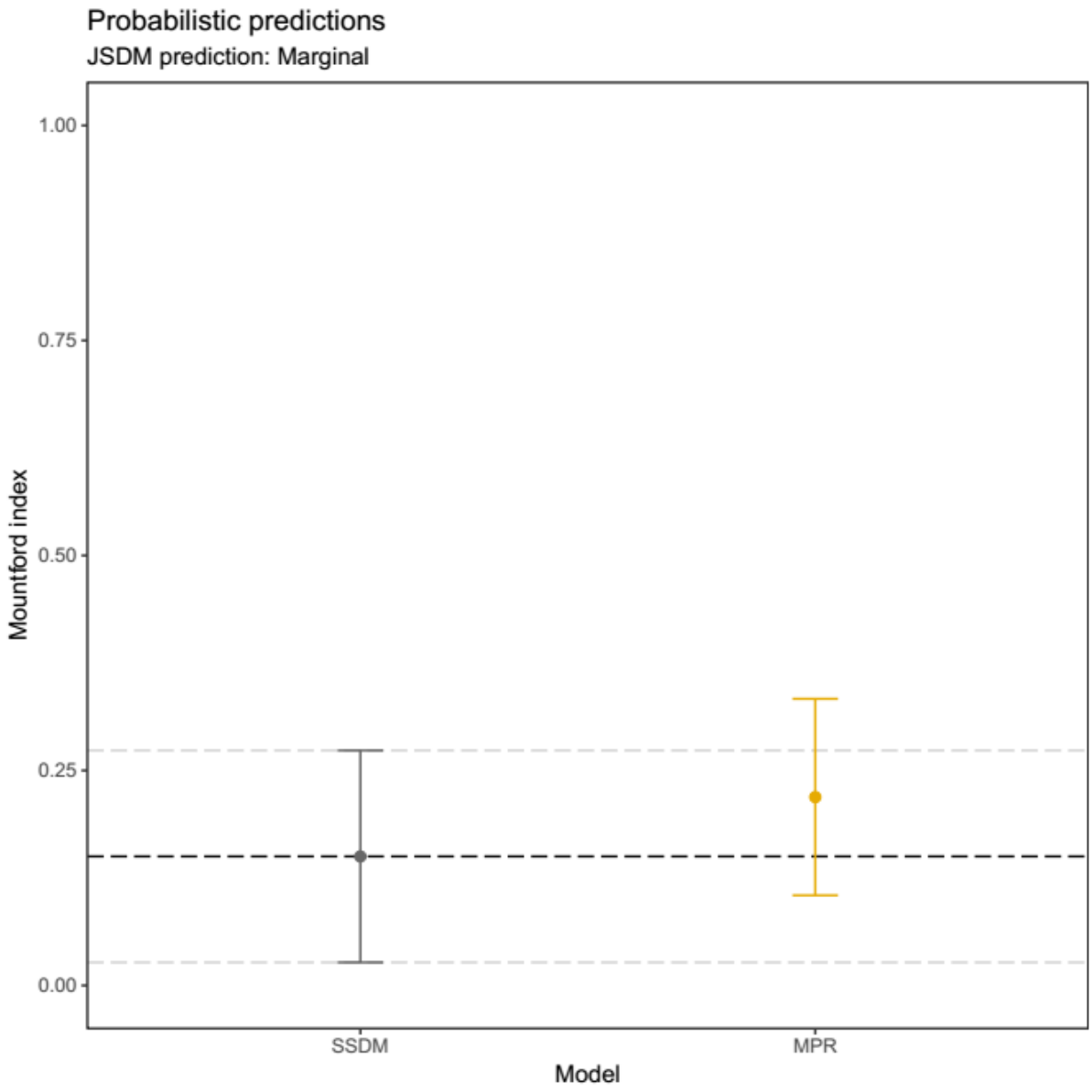


Figure 18: Performance of the SSDM and JSDM for the Mountford index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

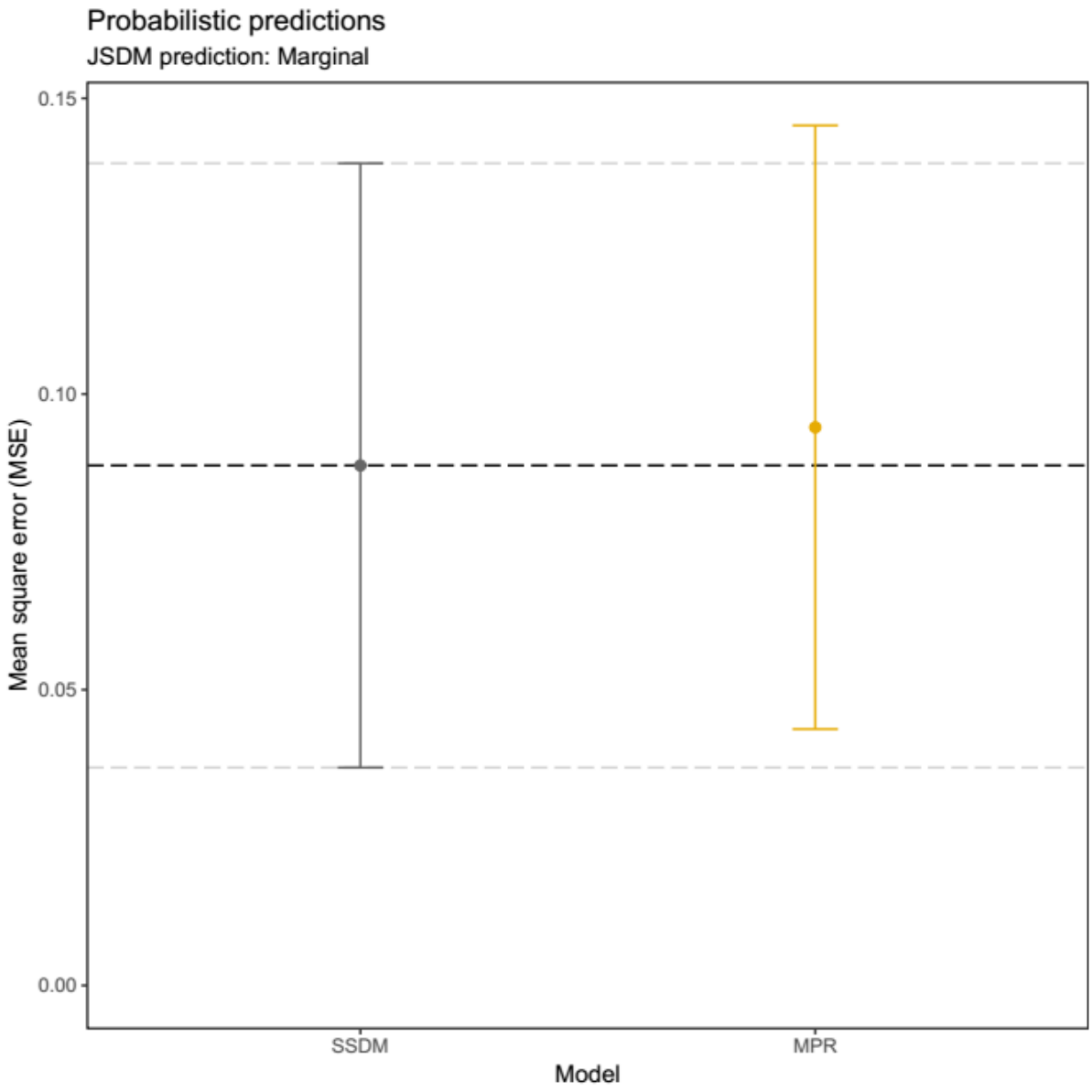


Figure 19: Performance of the SSDM and JSDM for the mean square error evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

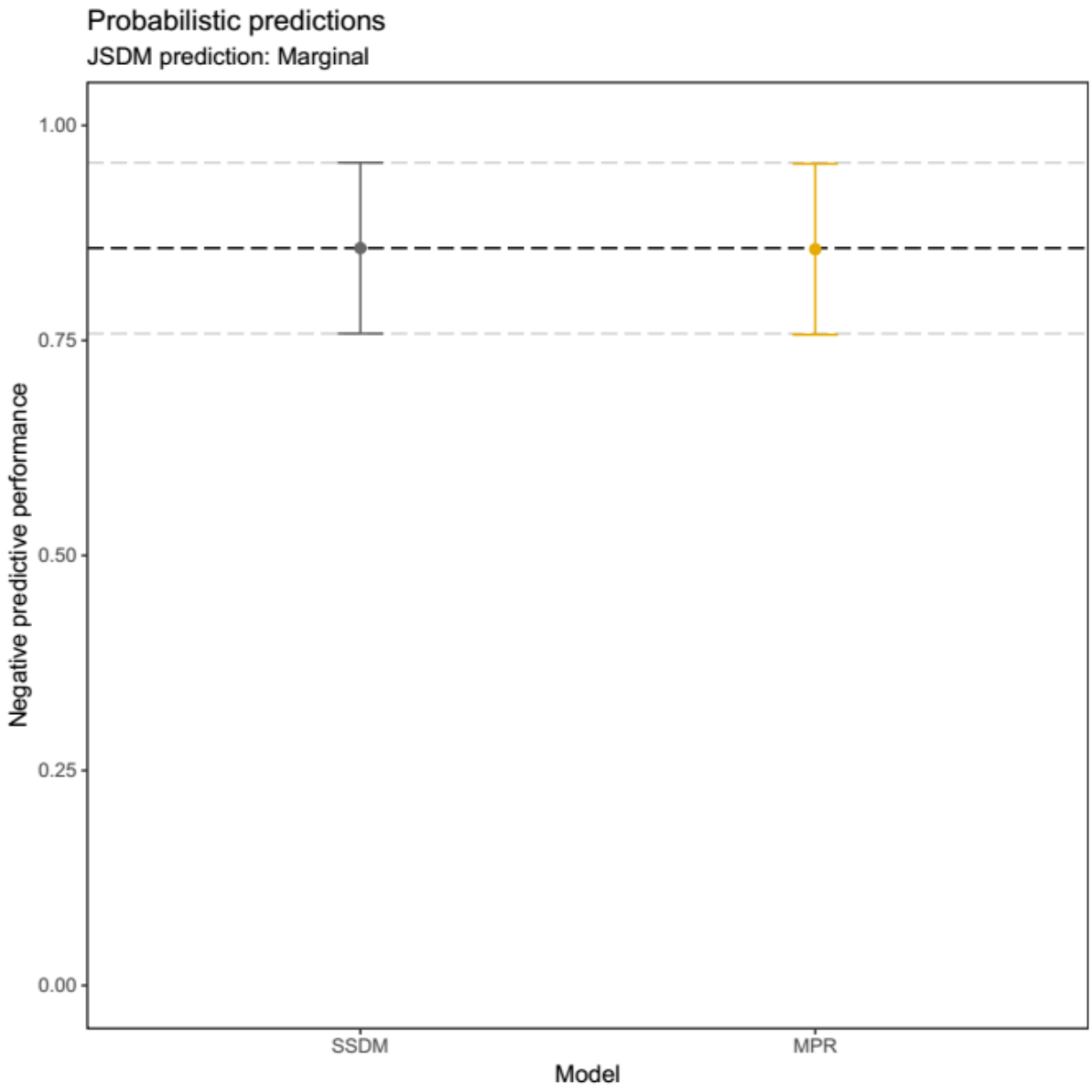


Figure 20: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

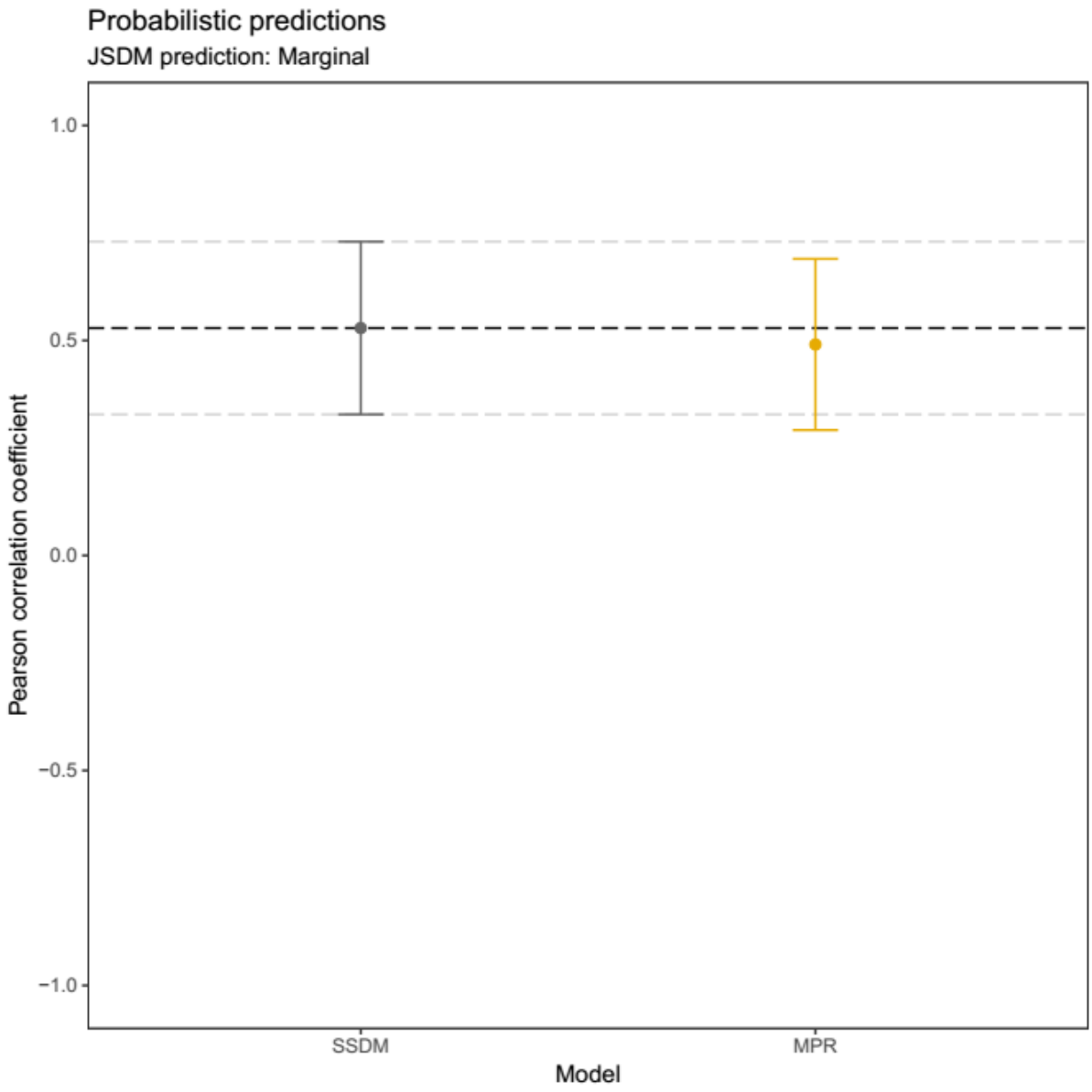


Figure 21: Performance of the SSDM and JSDM for the Pearson correlation coefficient evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

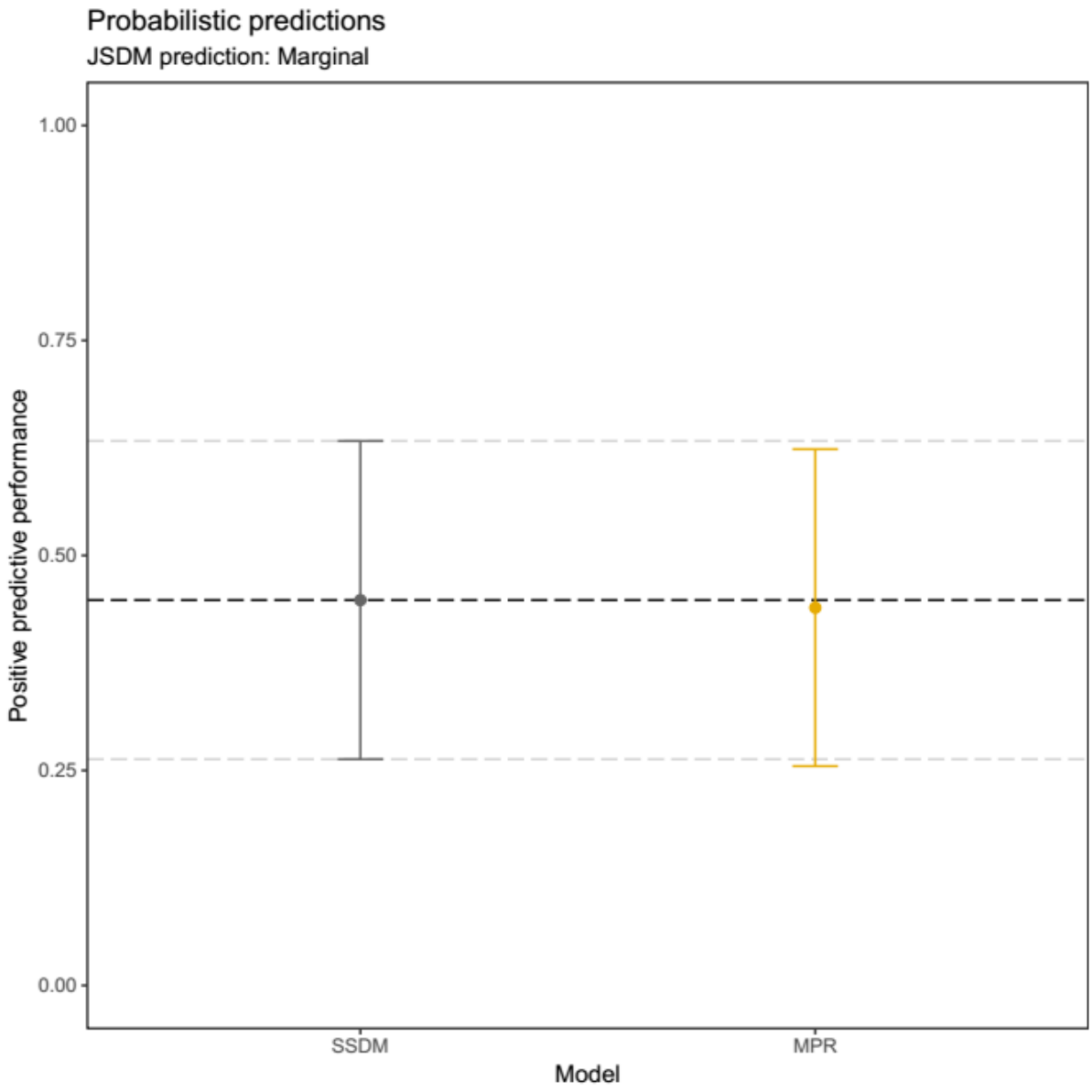


Figure 22: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

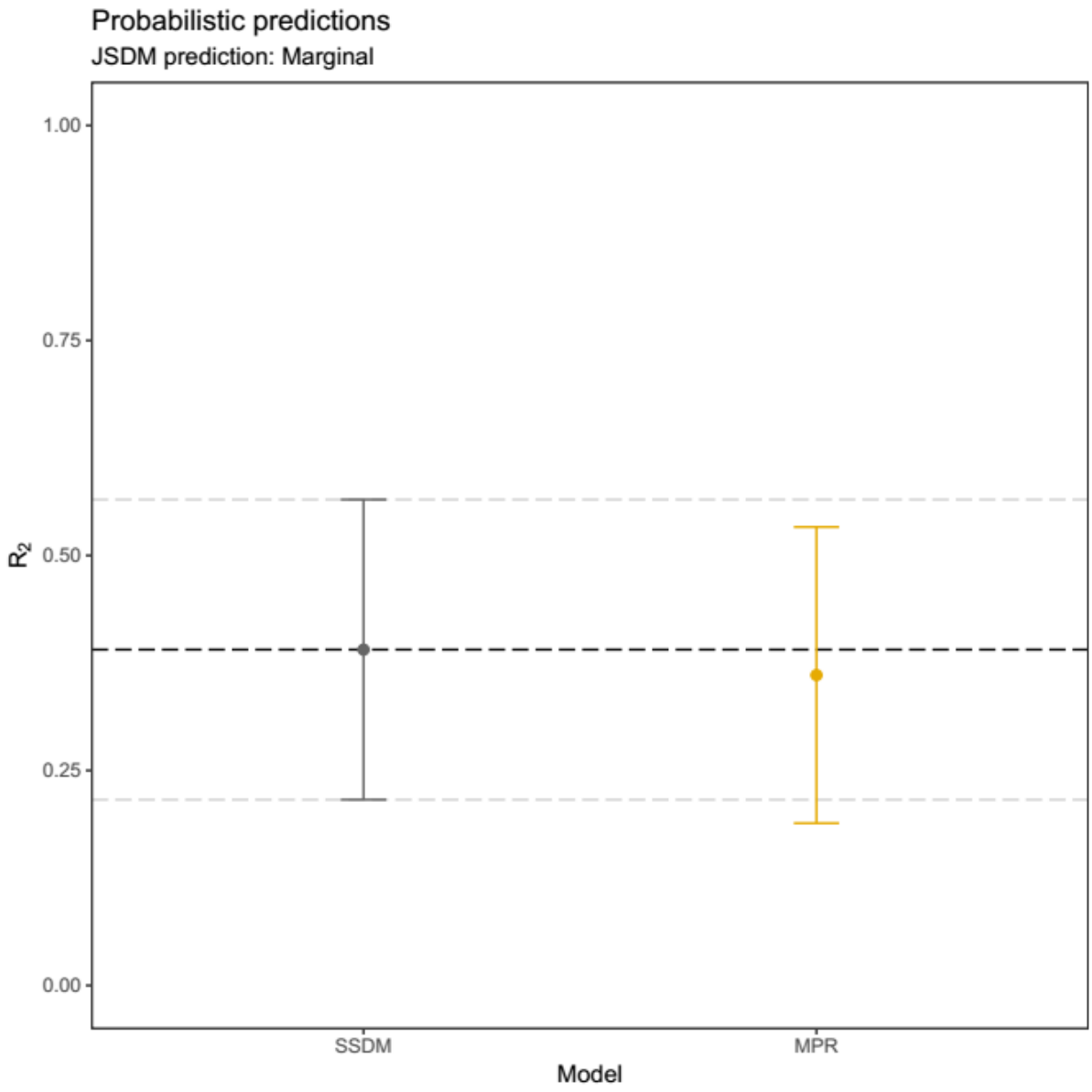


Figure 23: Performance of the SSDM and JSDM for the R^2 evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

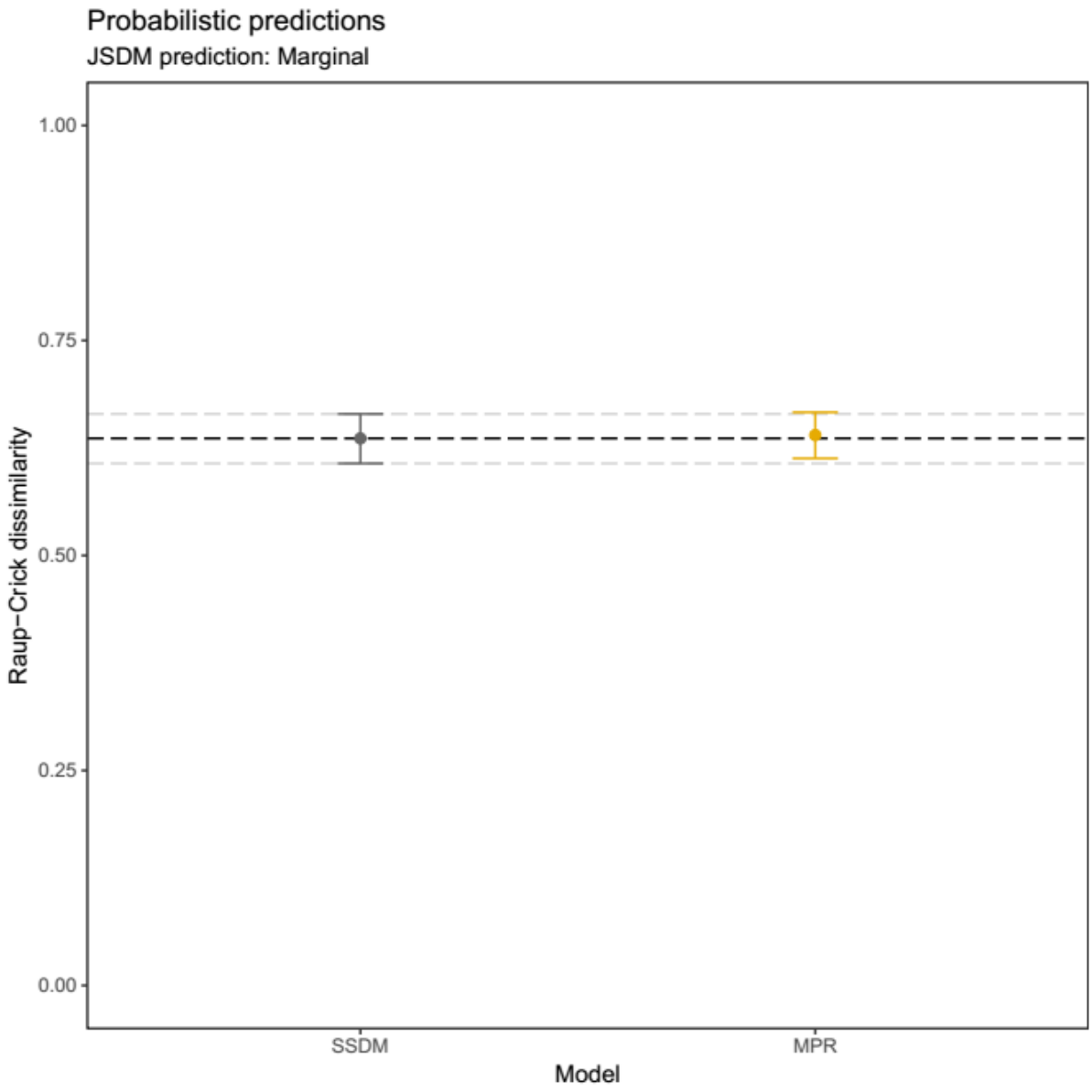


Figure 24: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

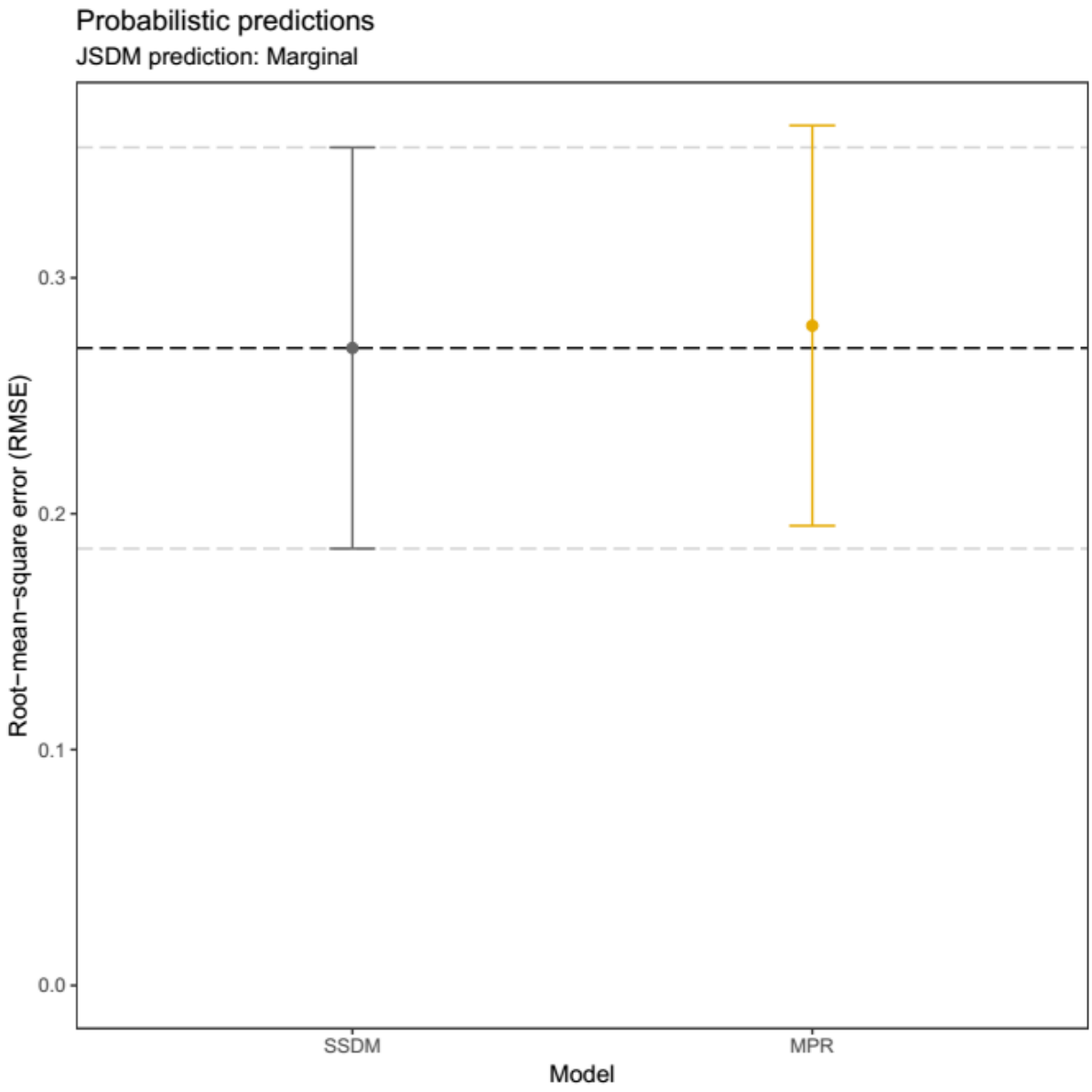


Figure 25: Performance of the SSDM and JSDM for the root-mean-square error evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

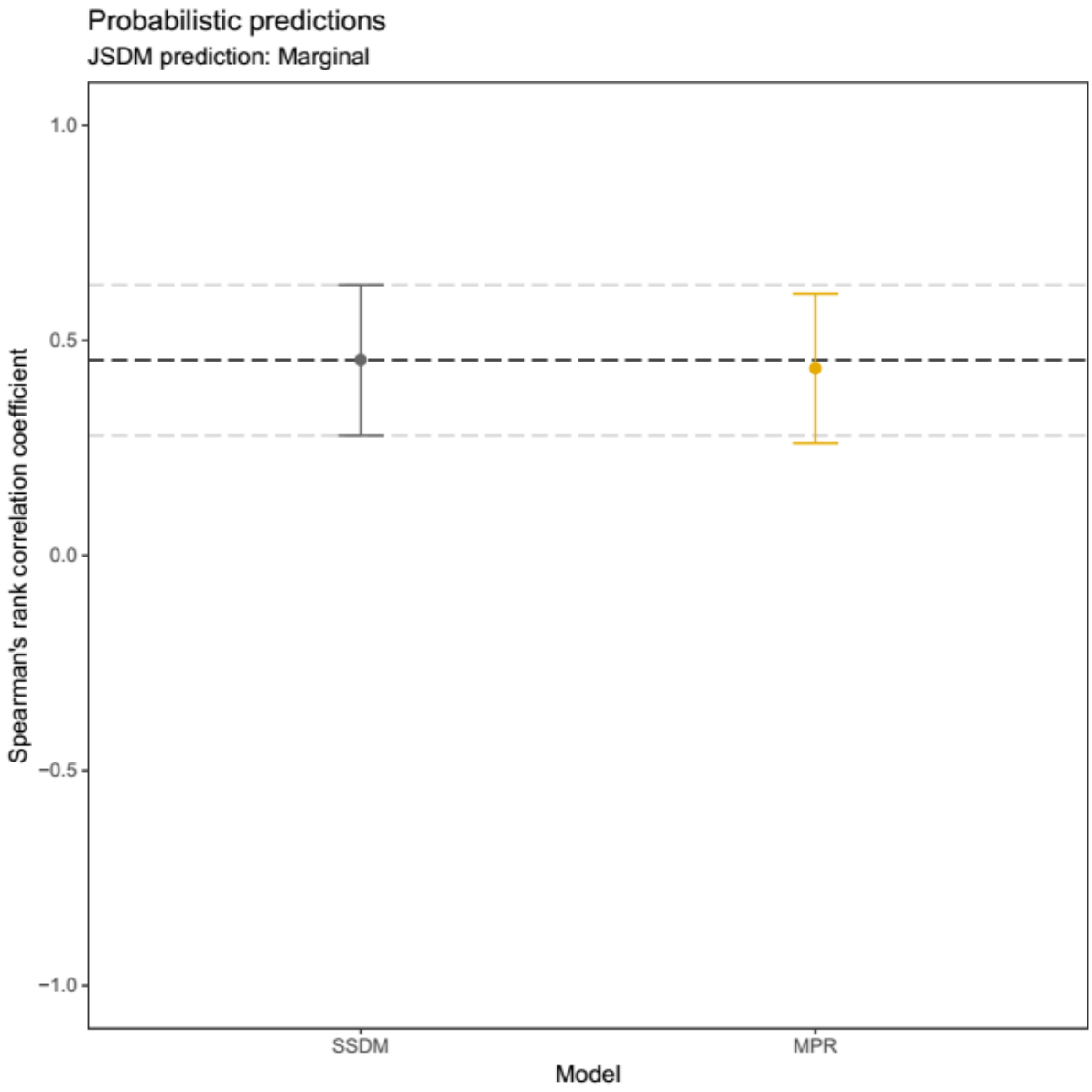


Figure 26: Performance of the SSDM and JSDM for the Spearman's rank correlation coefficient evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

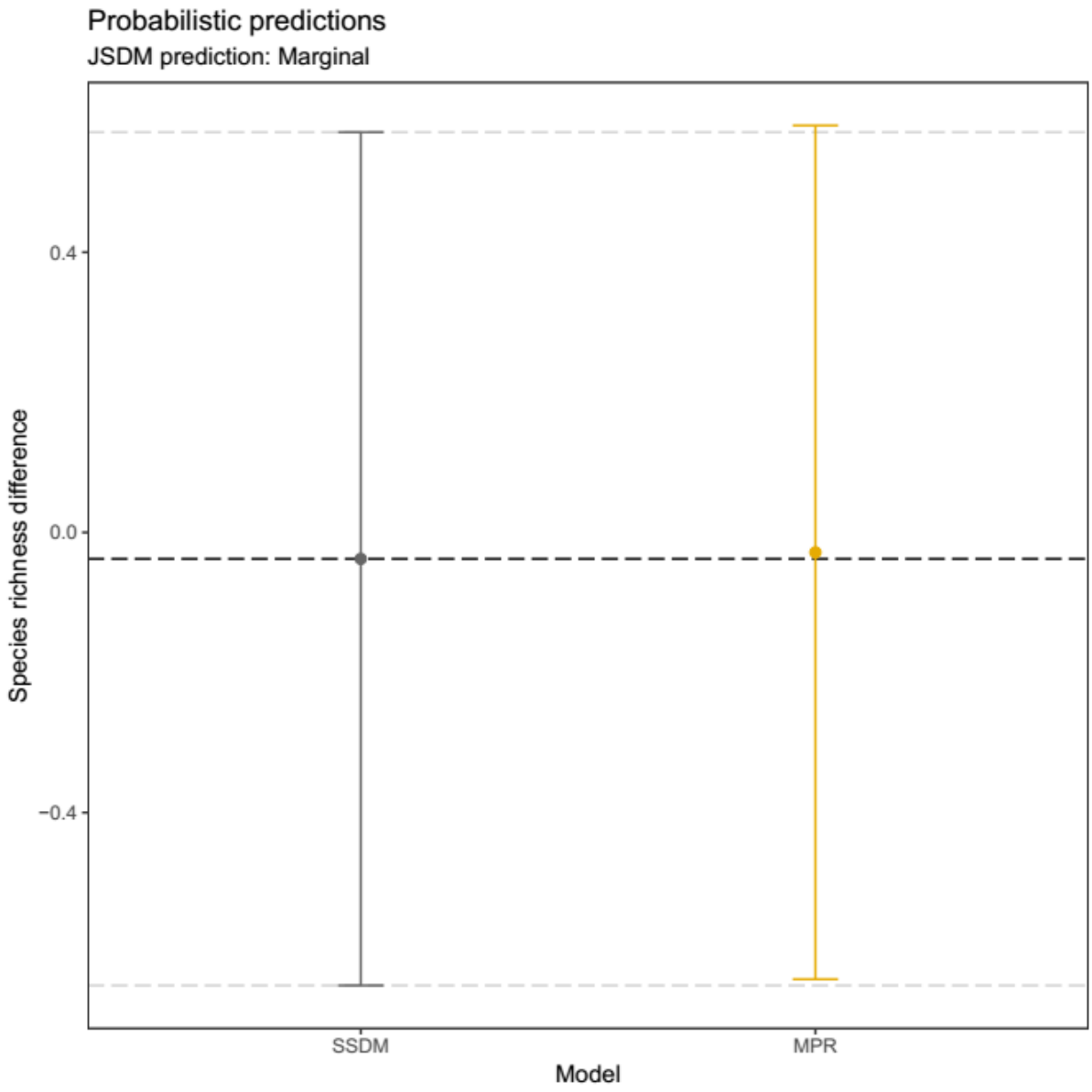


Figure 27: Performance of the SSDM and JSDM for the species richness difference evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

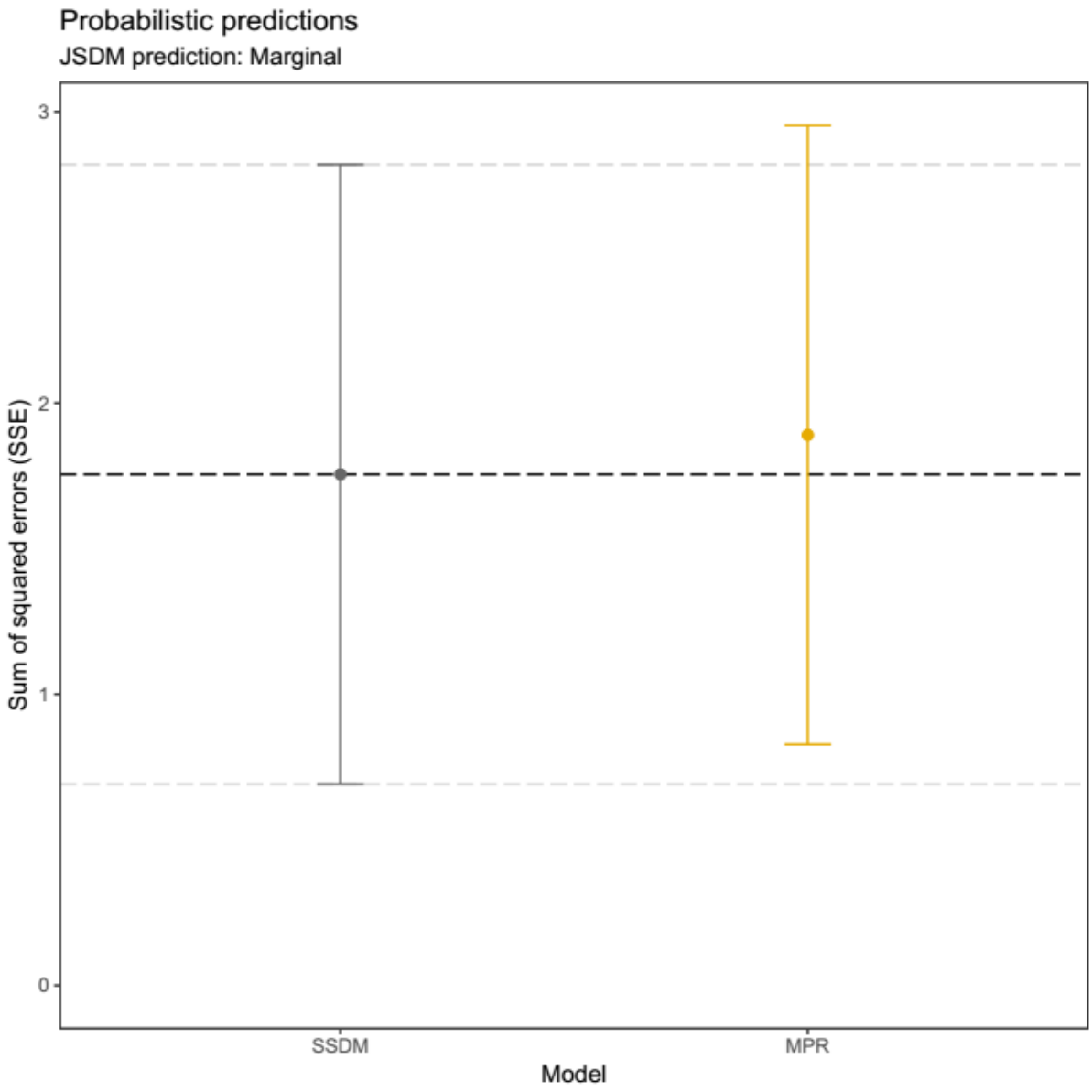


Figure 28: Performance of the SSDM and JSDM for the sum of squared errors evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

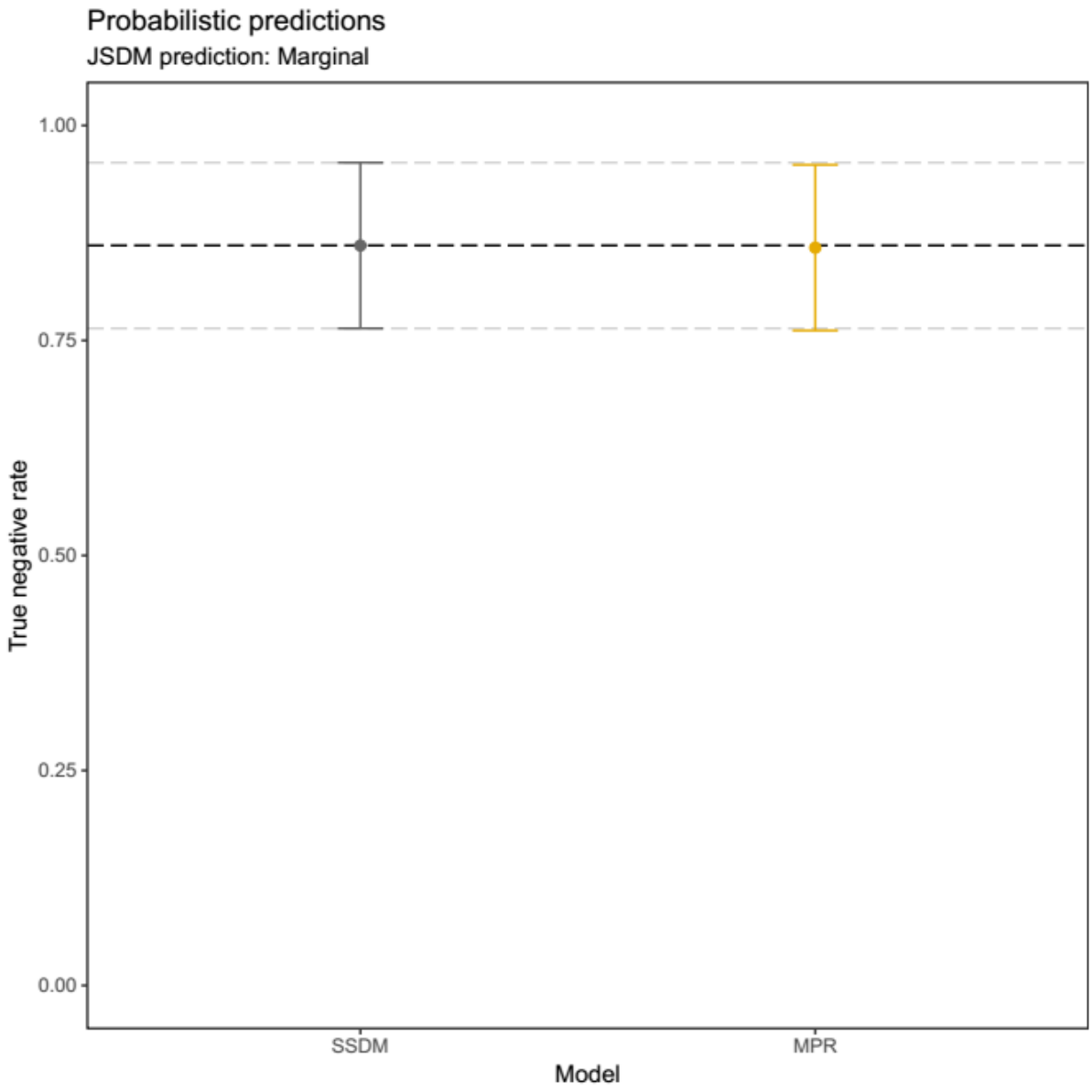


Figure 29: Performance of the SSDM and JSDM for the true negative rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

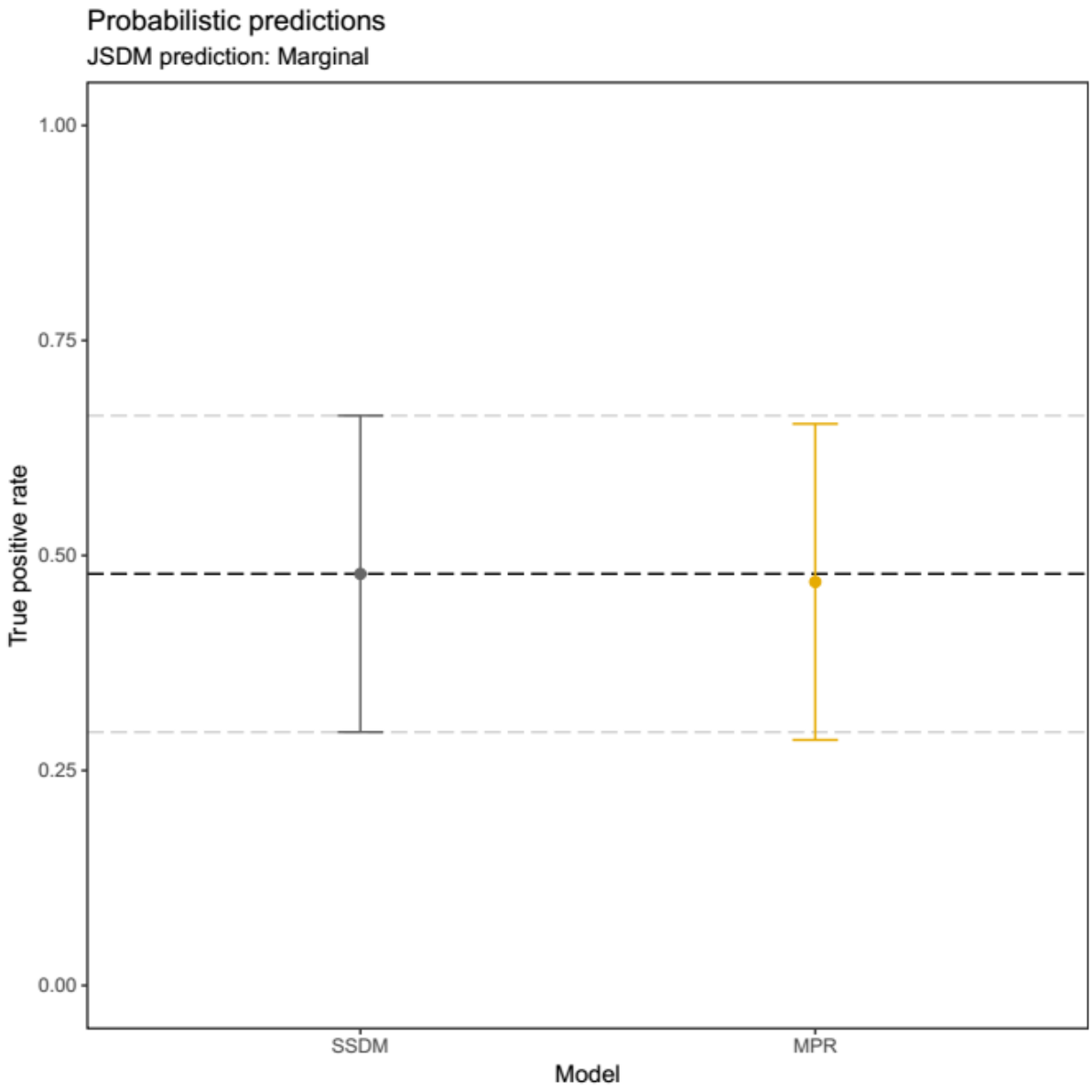


Figure 30: Performance of the SSDM and JSDM for the true positive rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

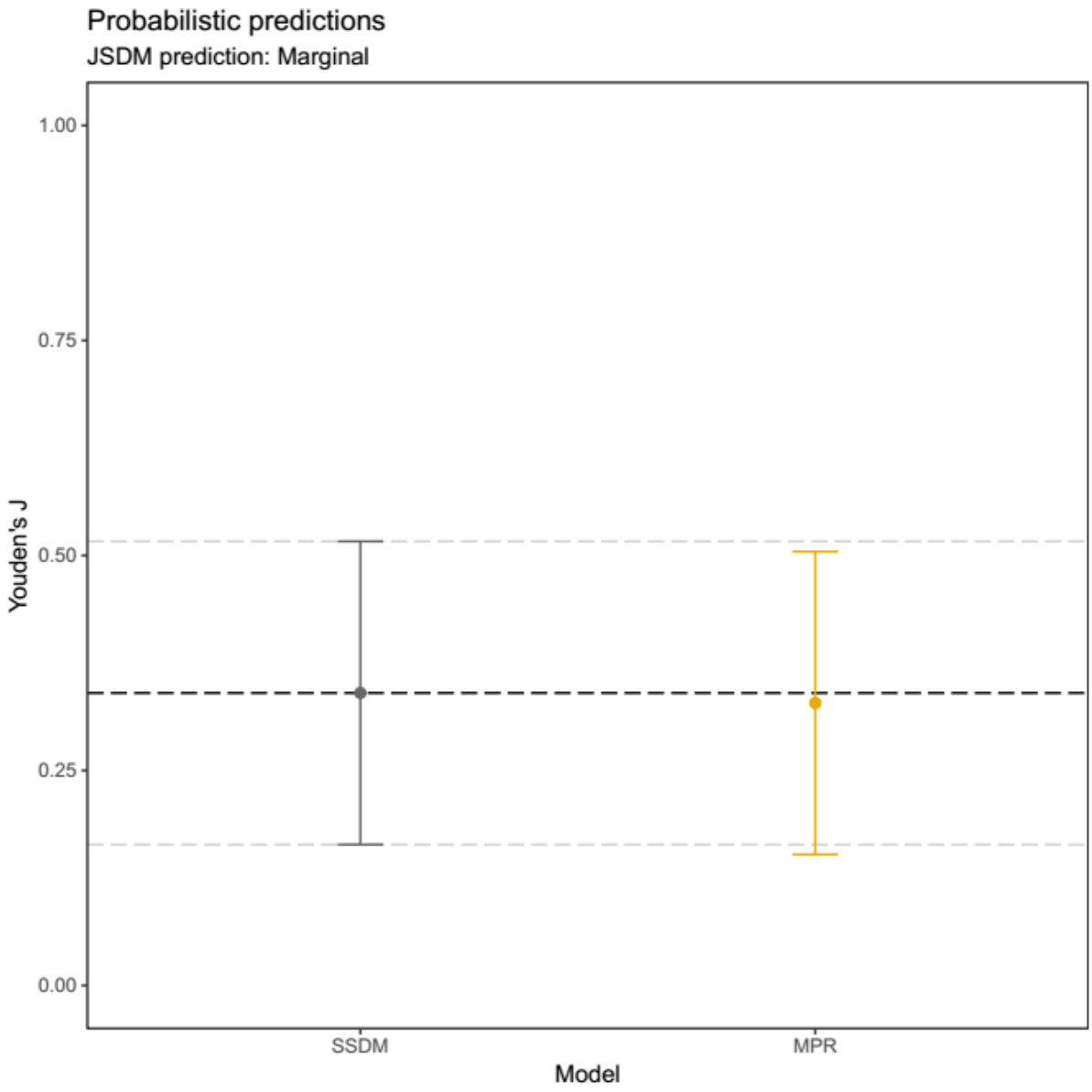


Figure 31: Performance of the SSDM and JSDM for the Youden's J evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

2. Marginal (binary) predictions

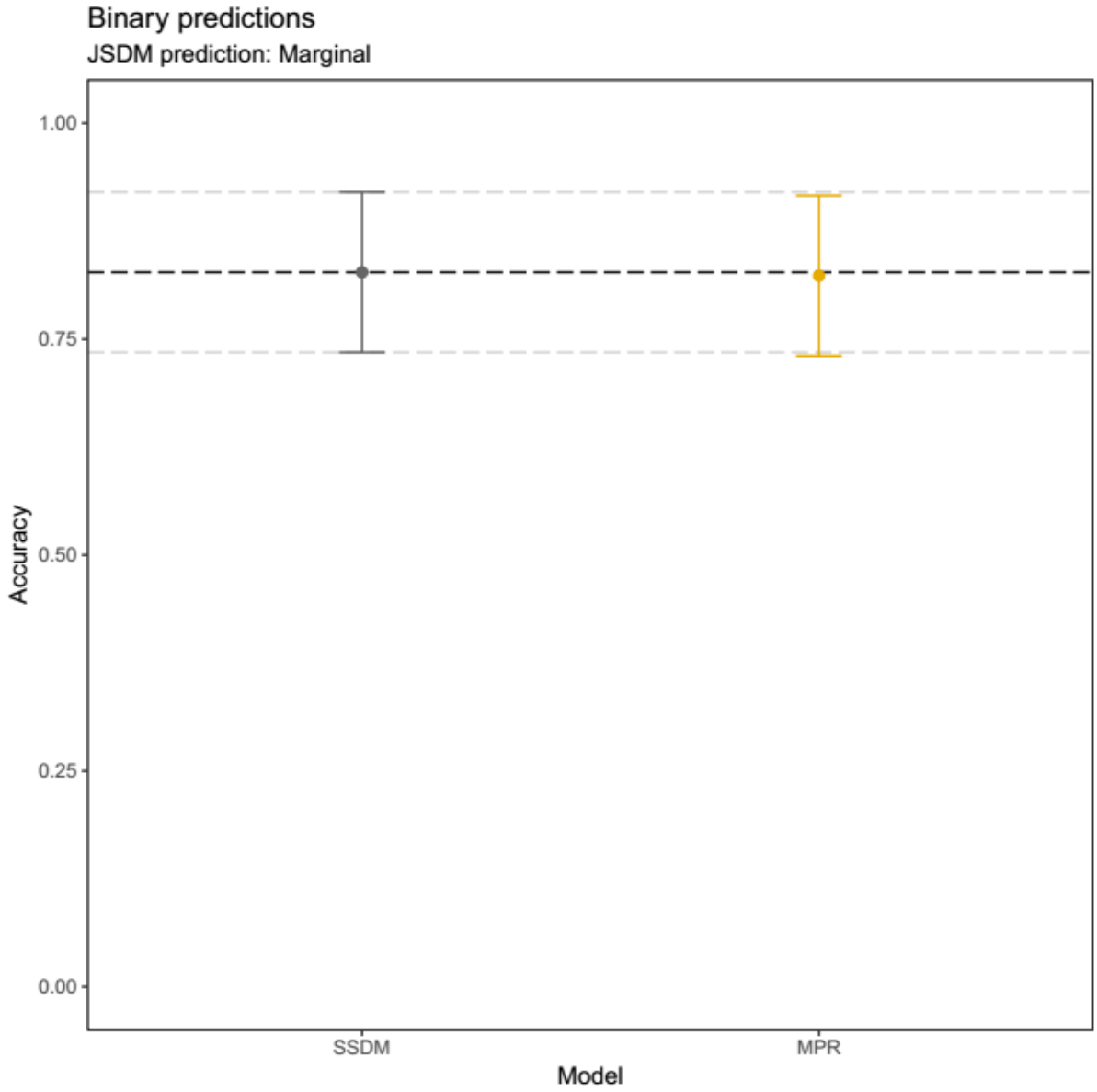


Figure 32: Performance of the SSDM and JSDM for the accuracy evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

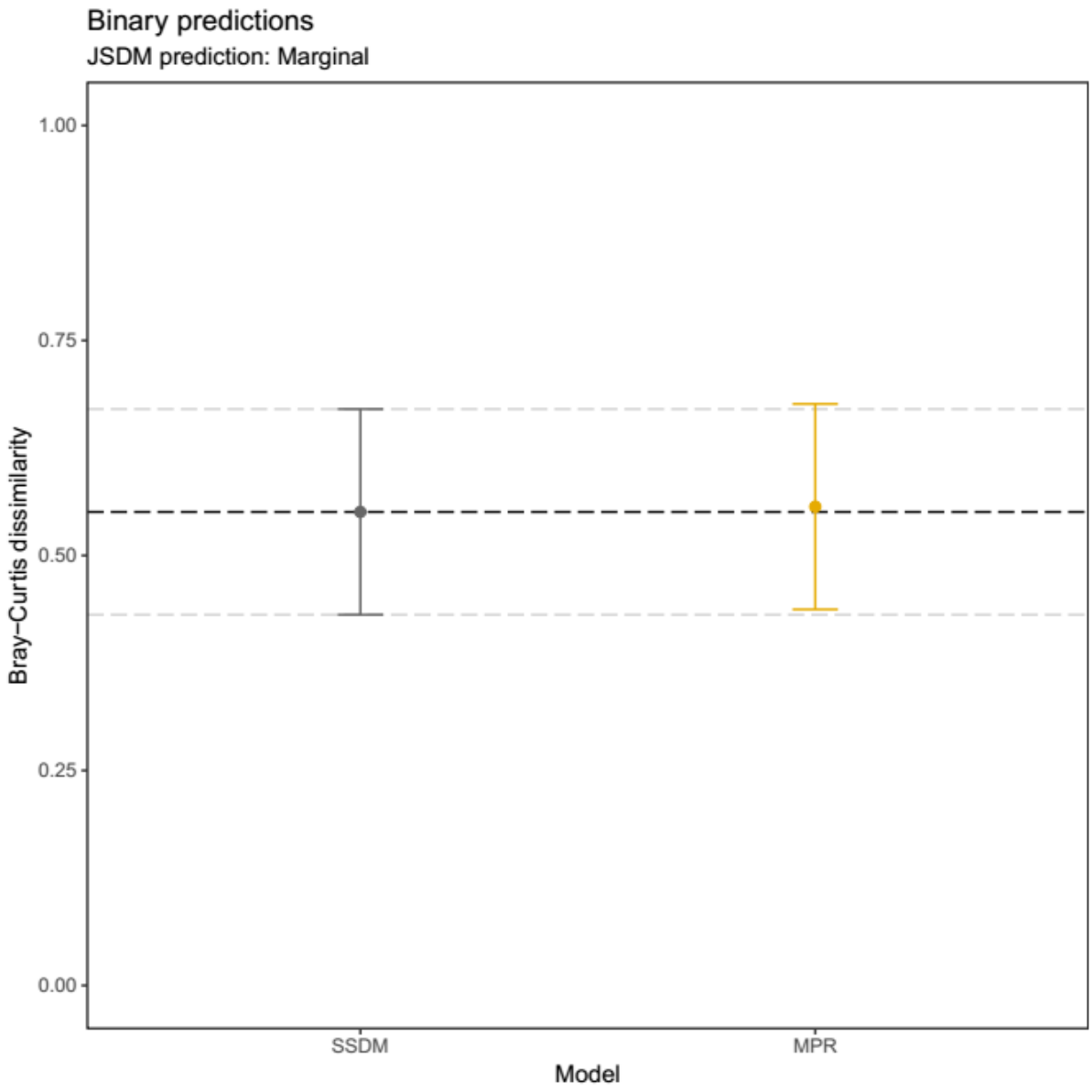


Figure 33: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

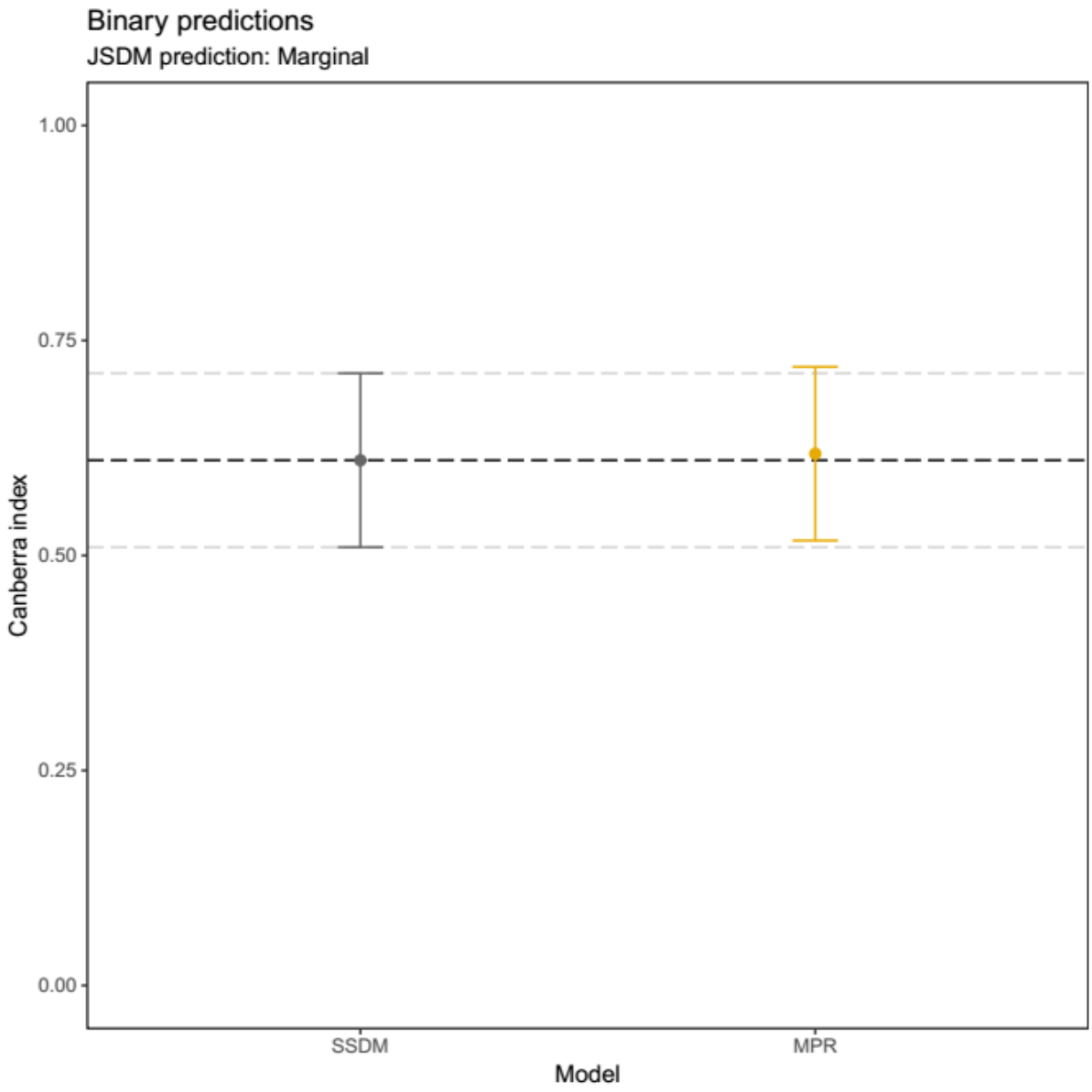


Figure 34: Performance of the SSDM and JSDM for the Canberra index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

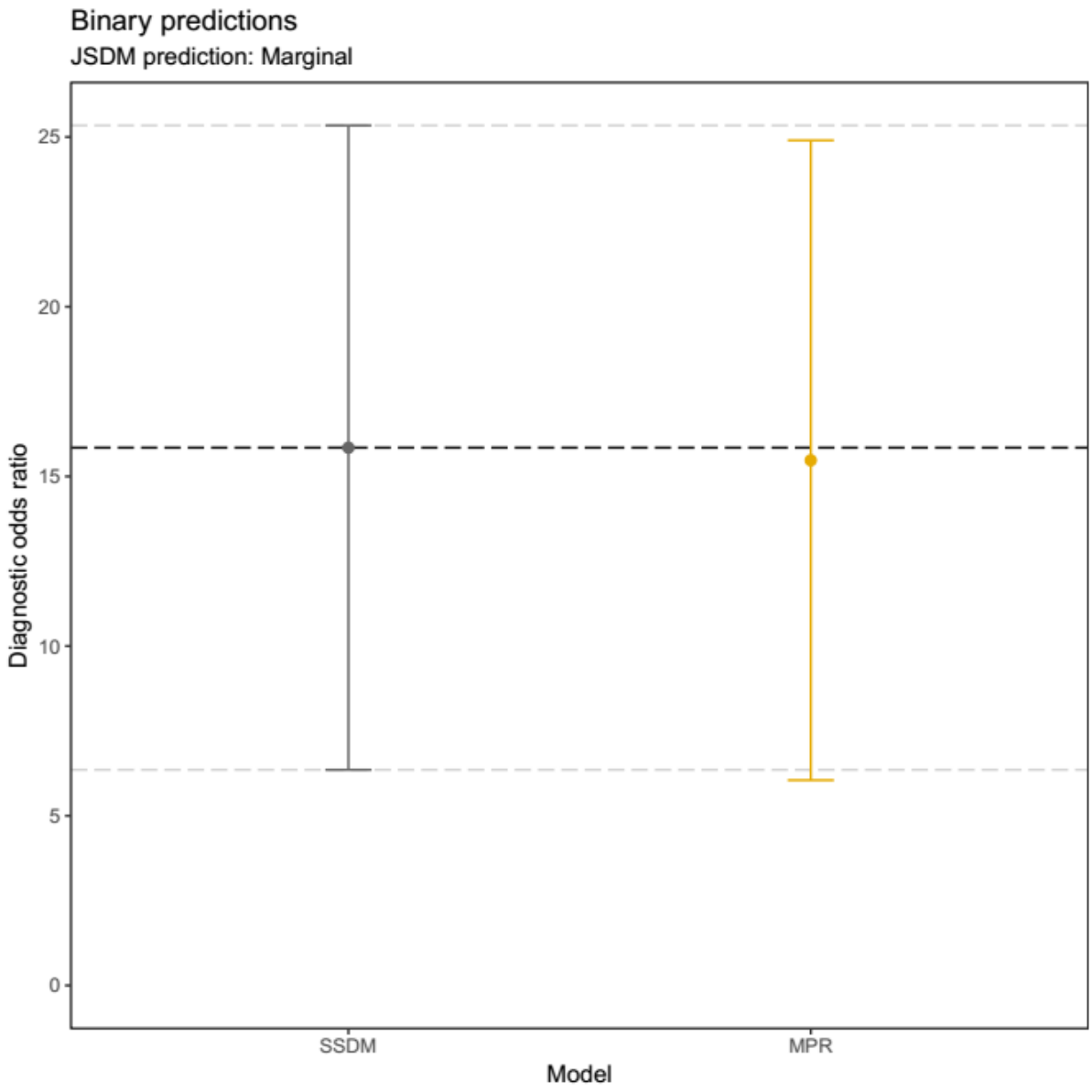


Figure 35: Performance of the SSDM and JSDM for the diagnostic odds ratio evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

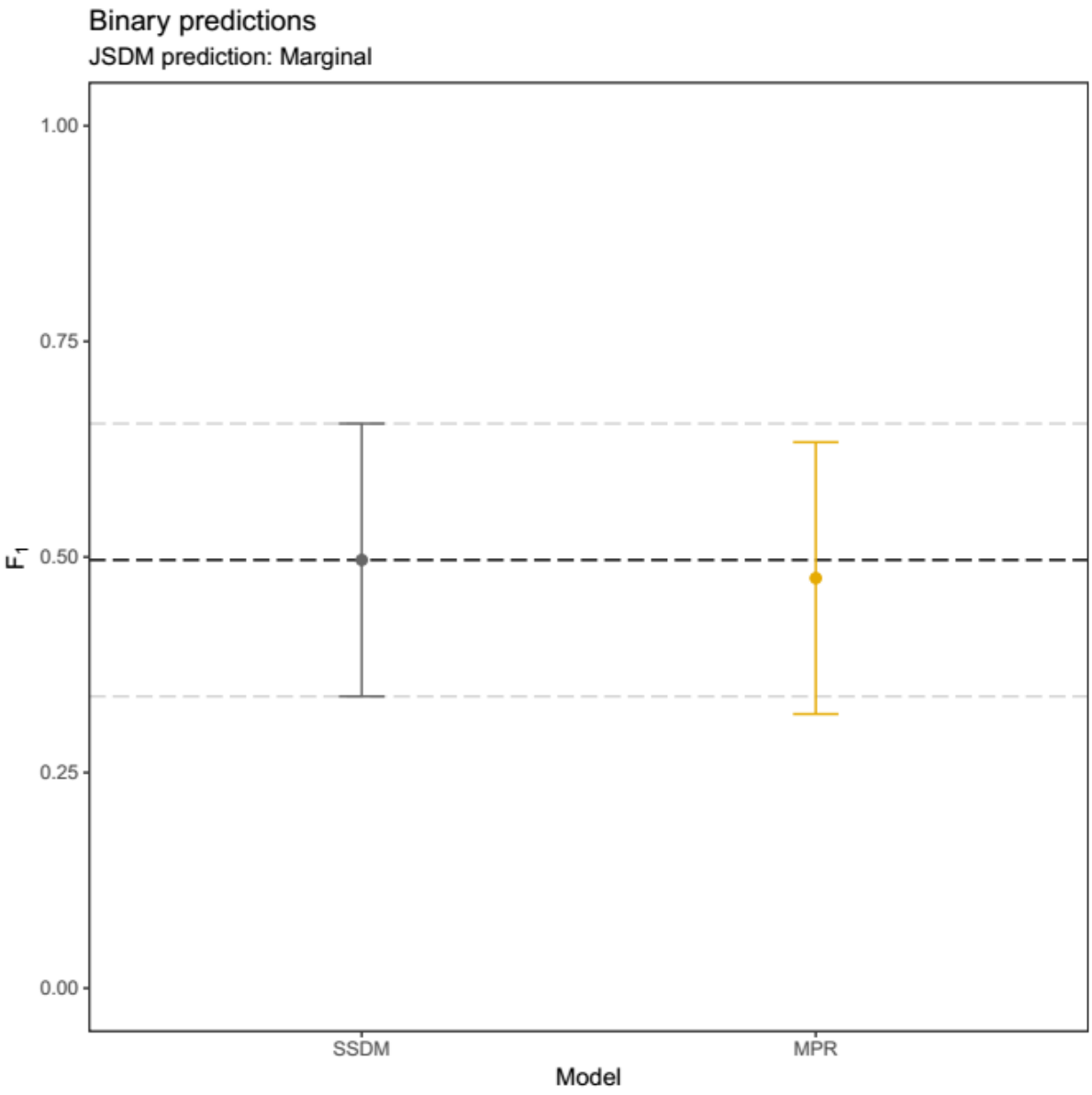


Figure 36: Performance of the SSDM and JSDM for the F_1 evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

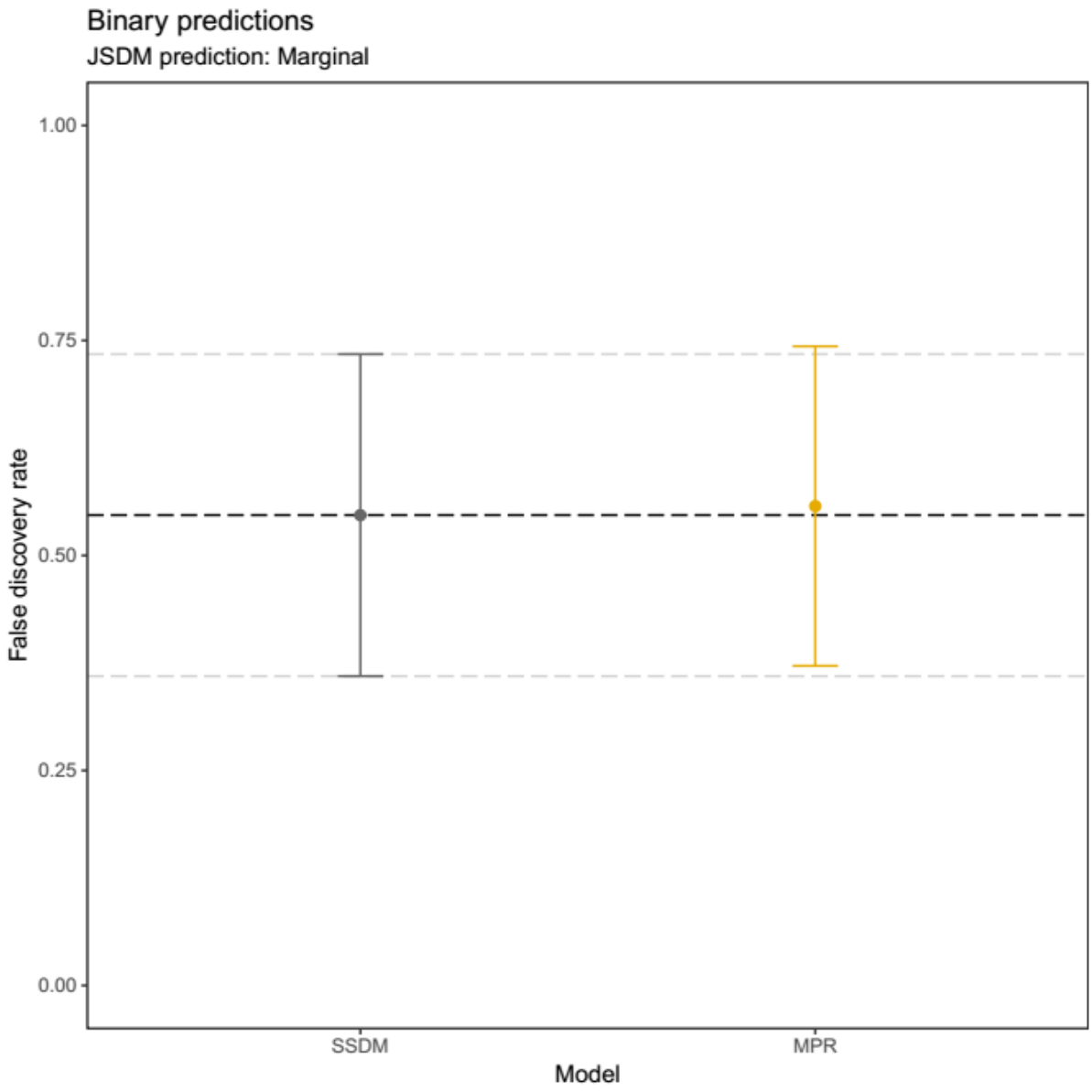


Figure 37: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

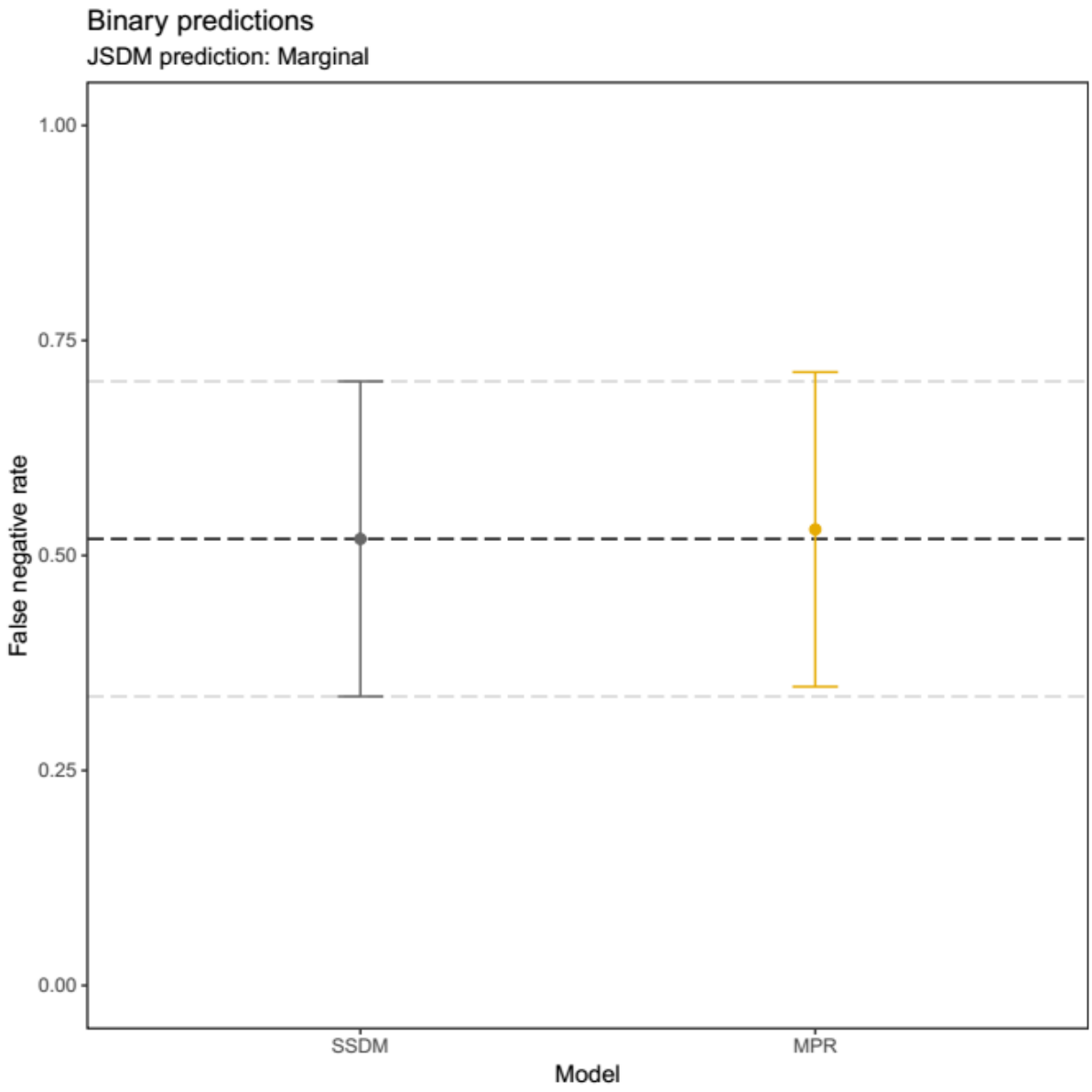


Figure 38: Performance of the SSDM and JSDM for the false negative rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

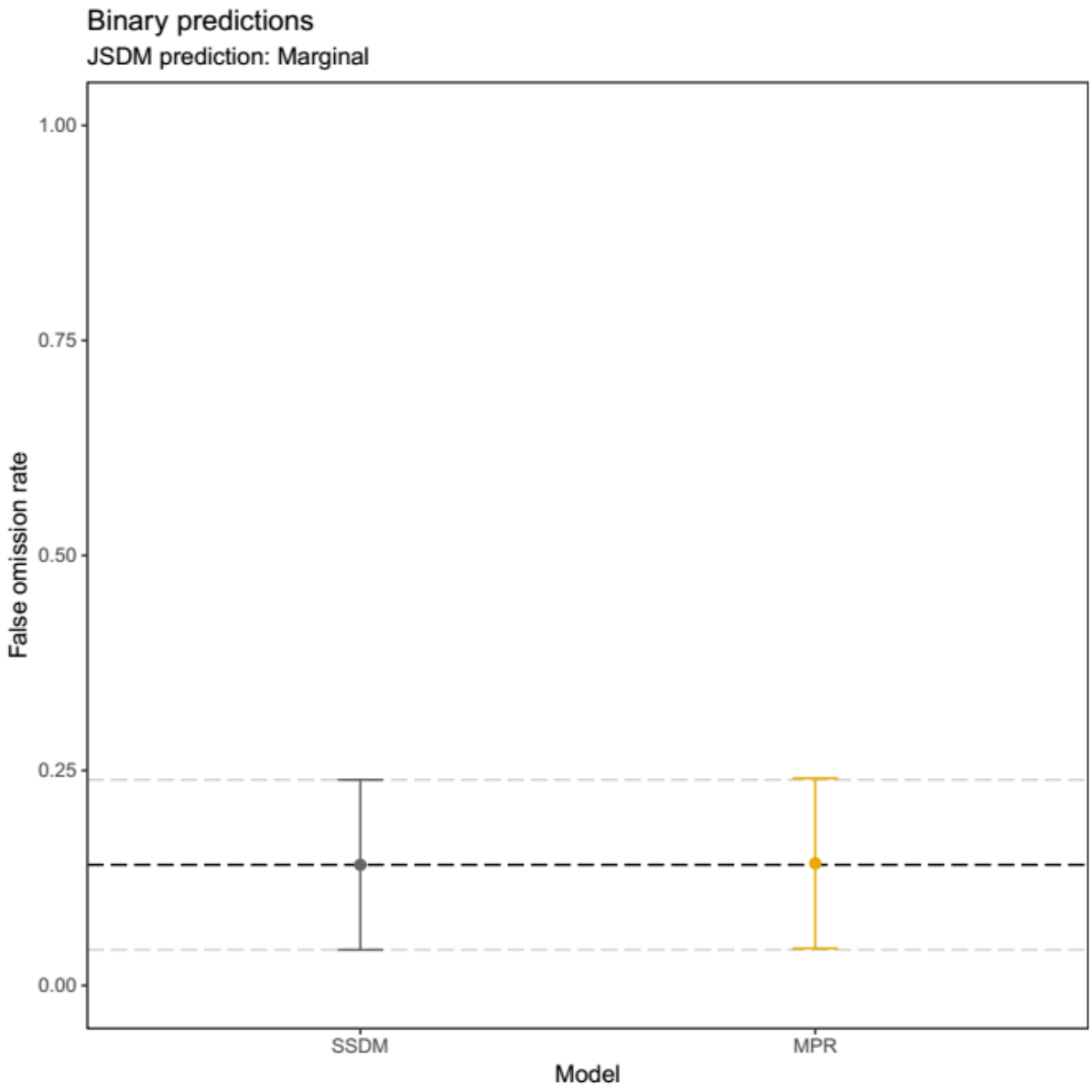


Figure 39: Performance of the SSDM and JSDM for the false omission rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

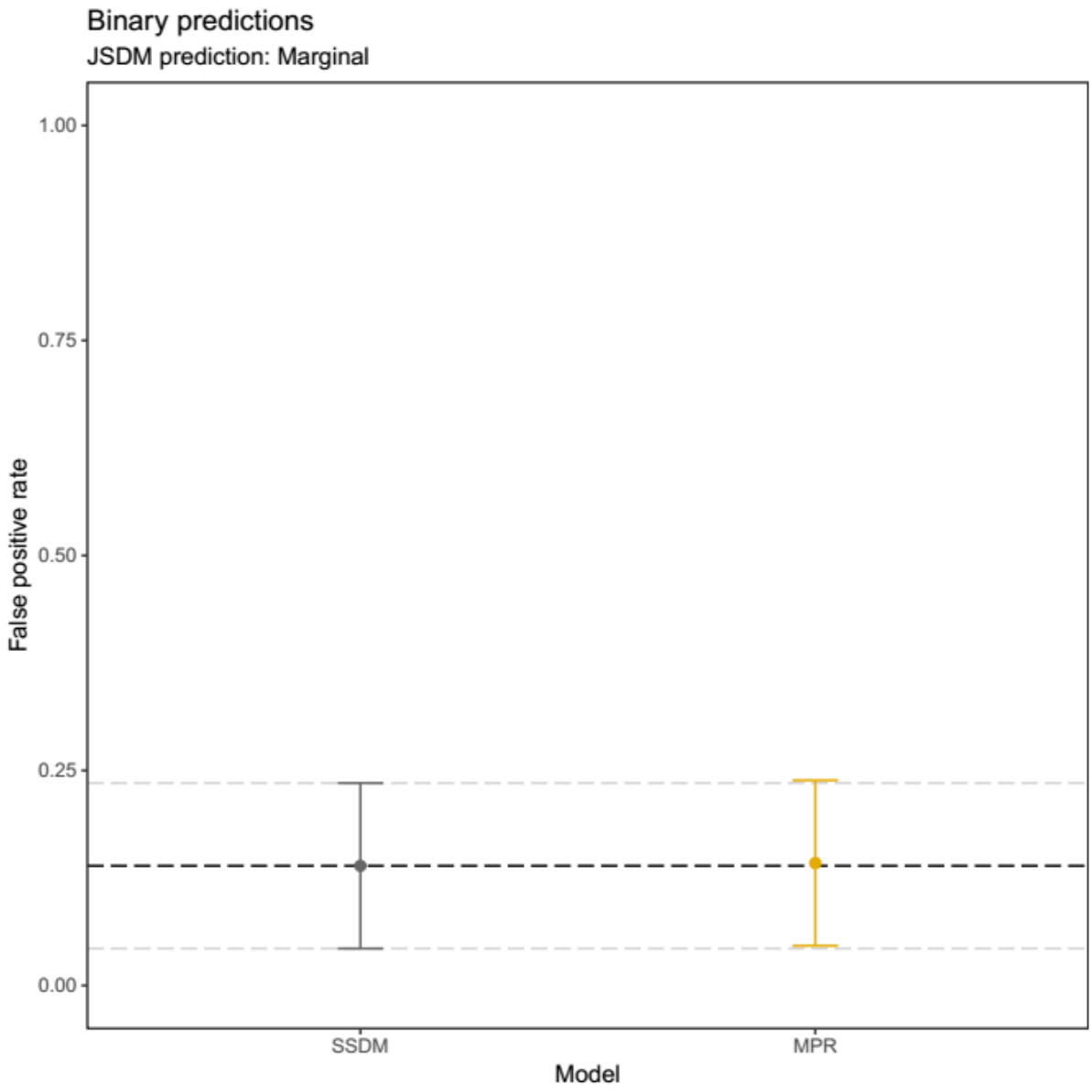


Figure 40: Performance of the SSDM and JSDM for the false positive rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

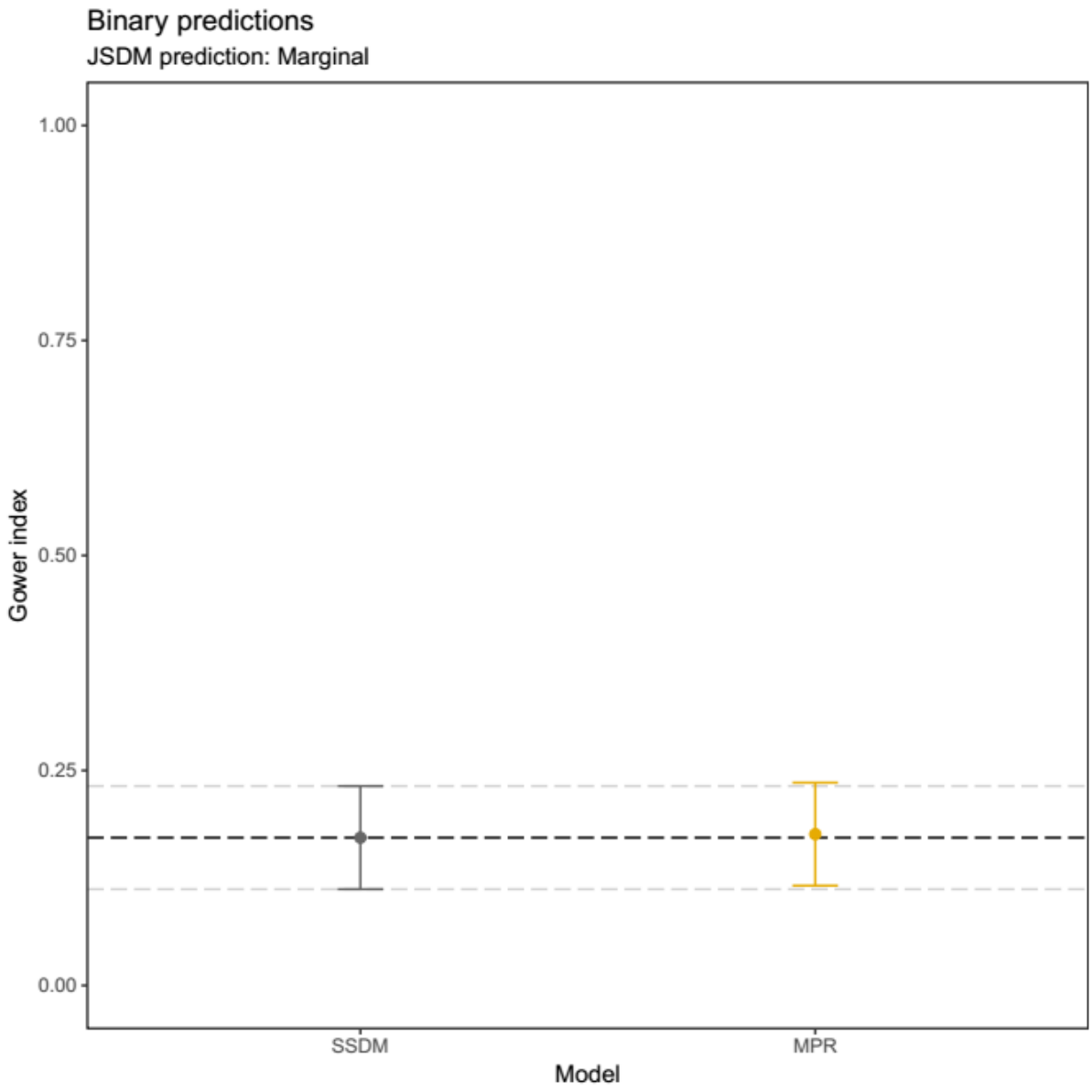


Figure 41: Performance of the SSDM and JSDM for the Gower index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

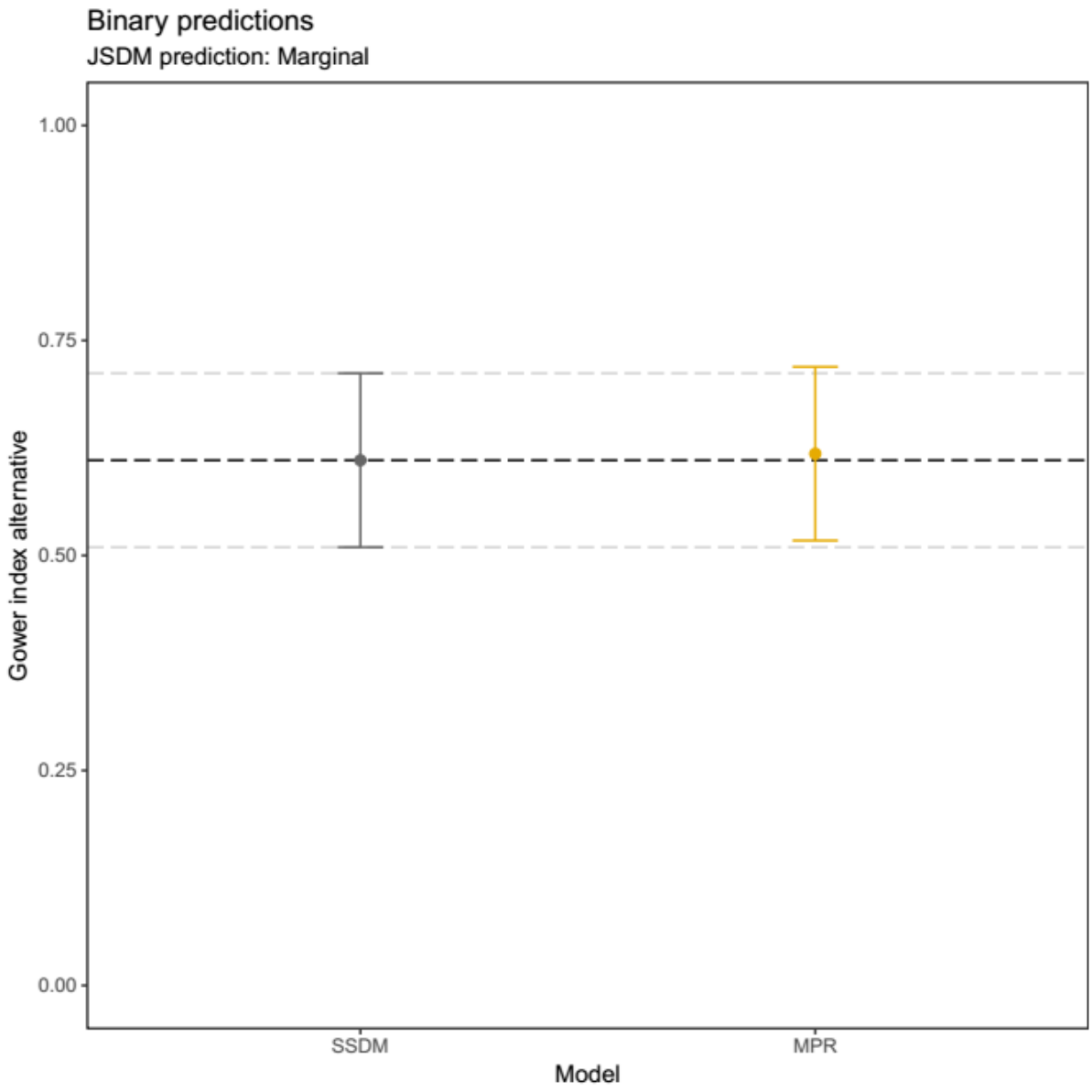


Figure 42: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

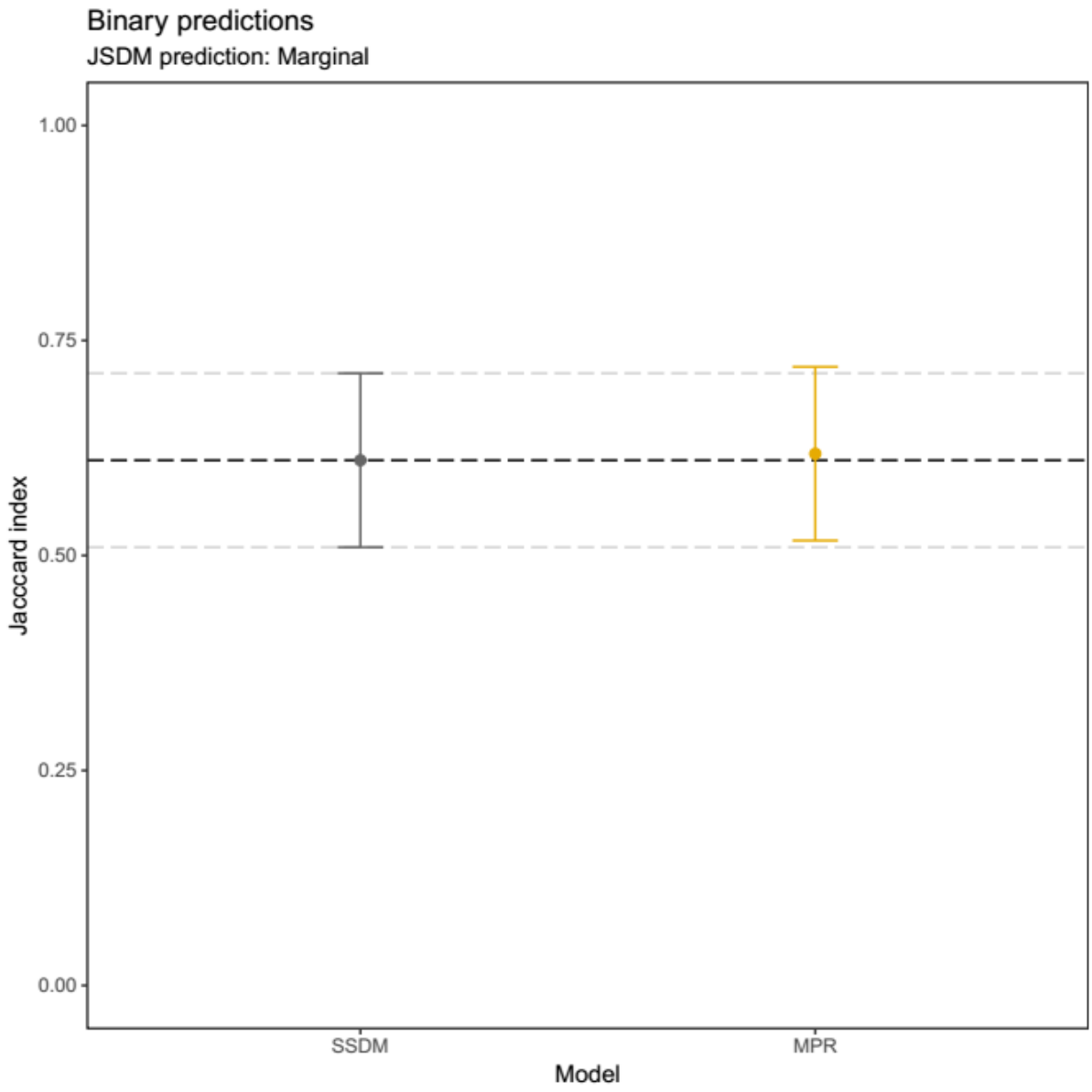


Figure 43: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

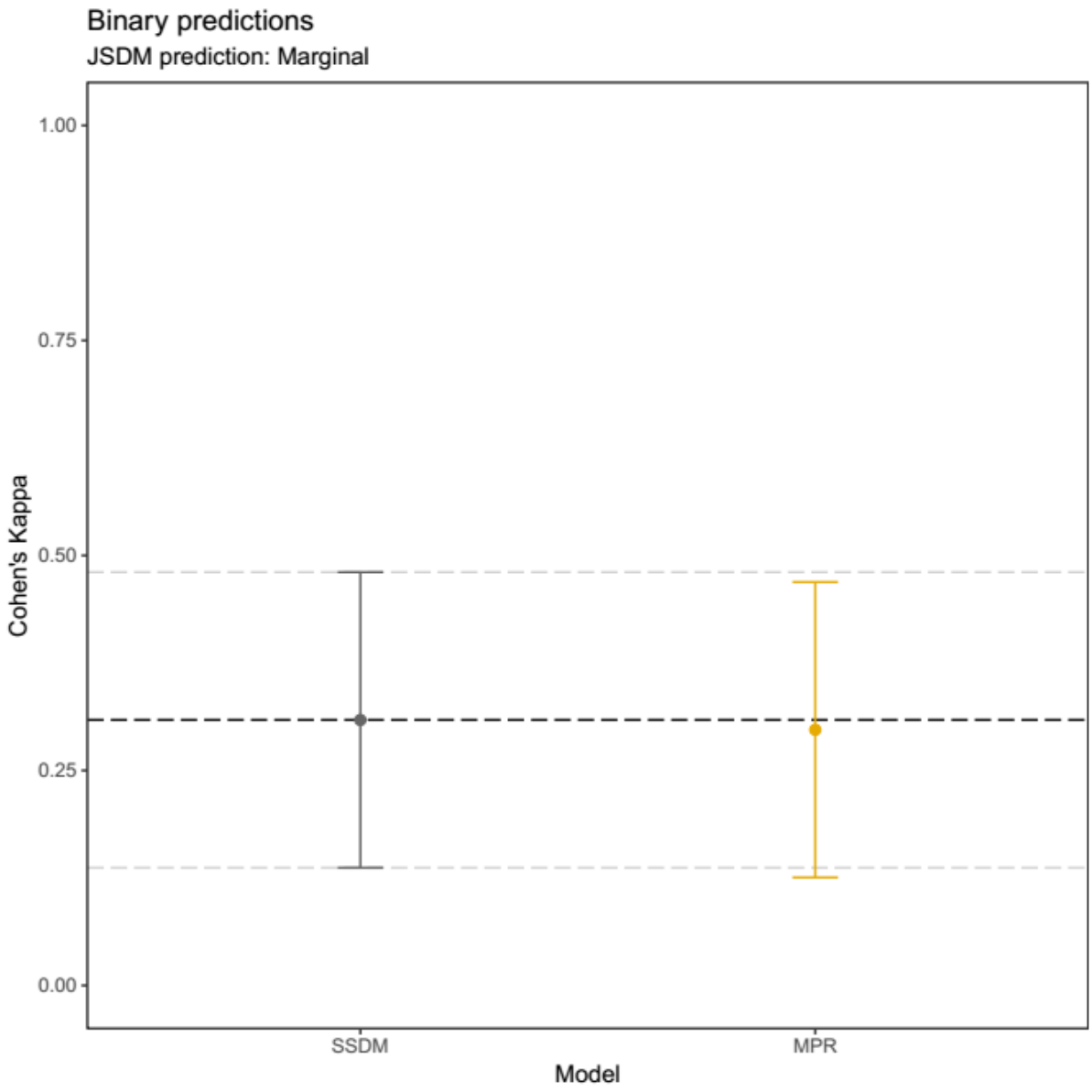


Figure 44: Performance of the SSDM and JSDM for the Cohen's Kappa evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

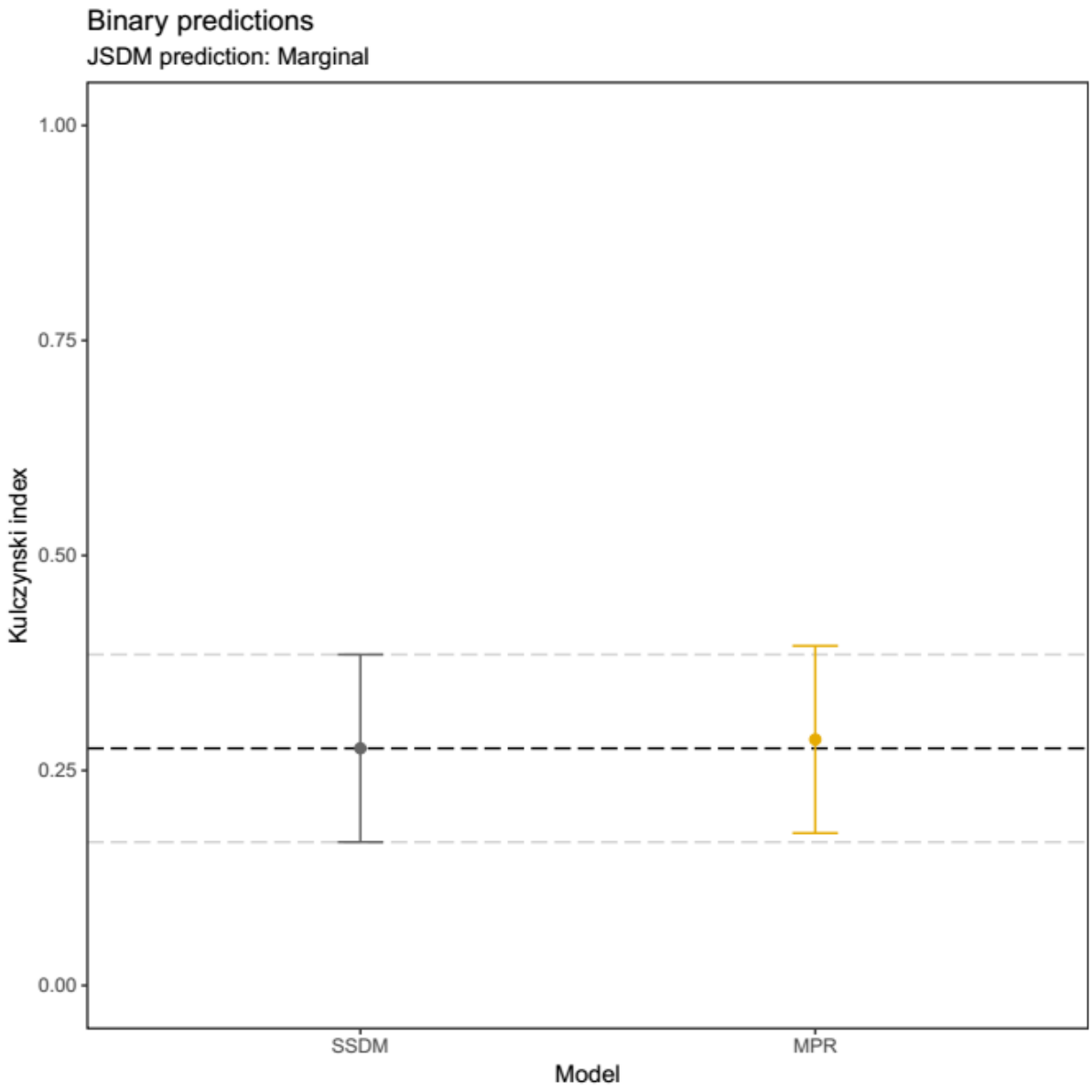


Figure 45: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

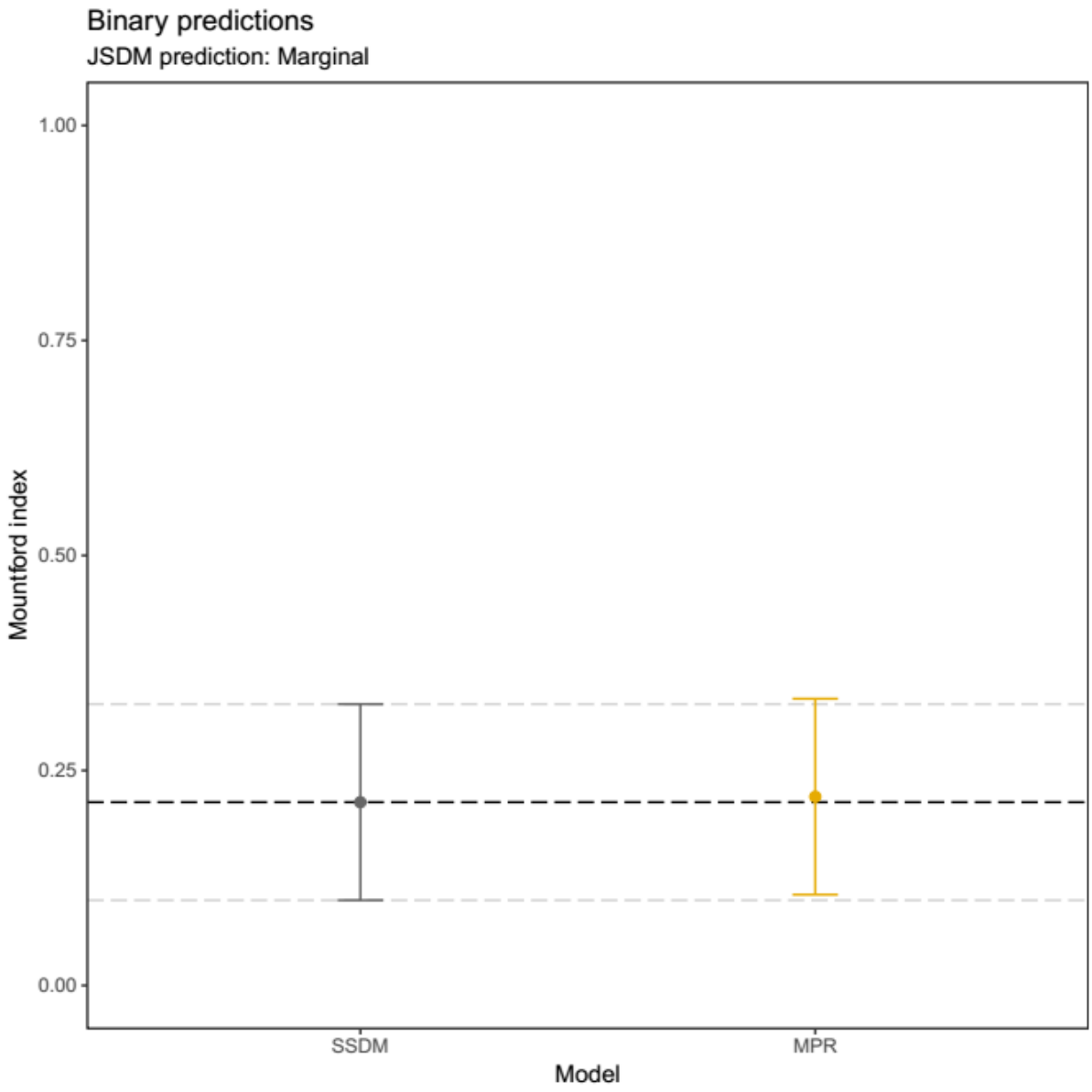


Figure 46: Performance of the SSDM and JSDM for the Mountford index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

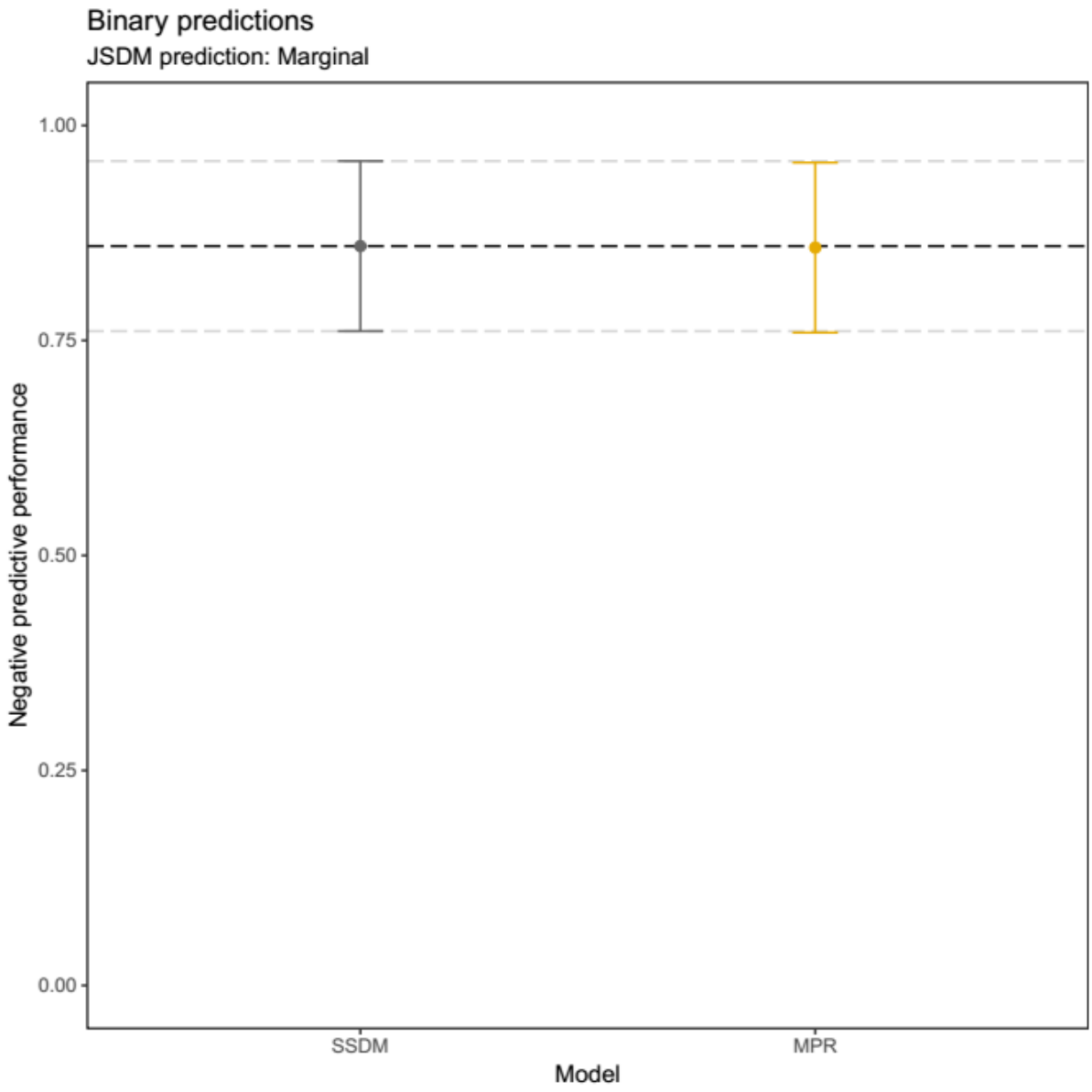


Figure 47: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

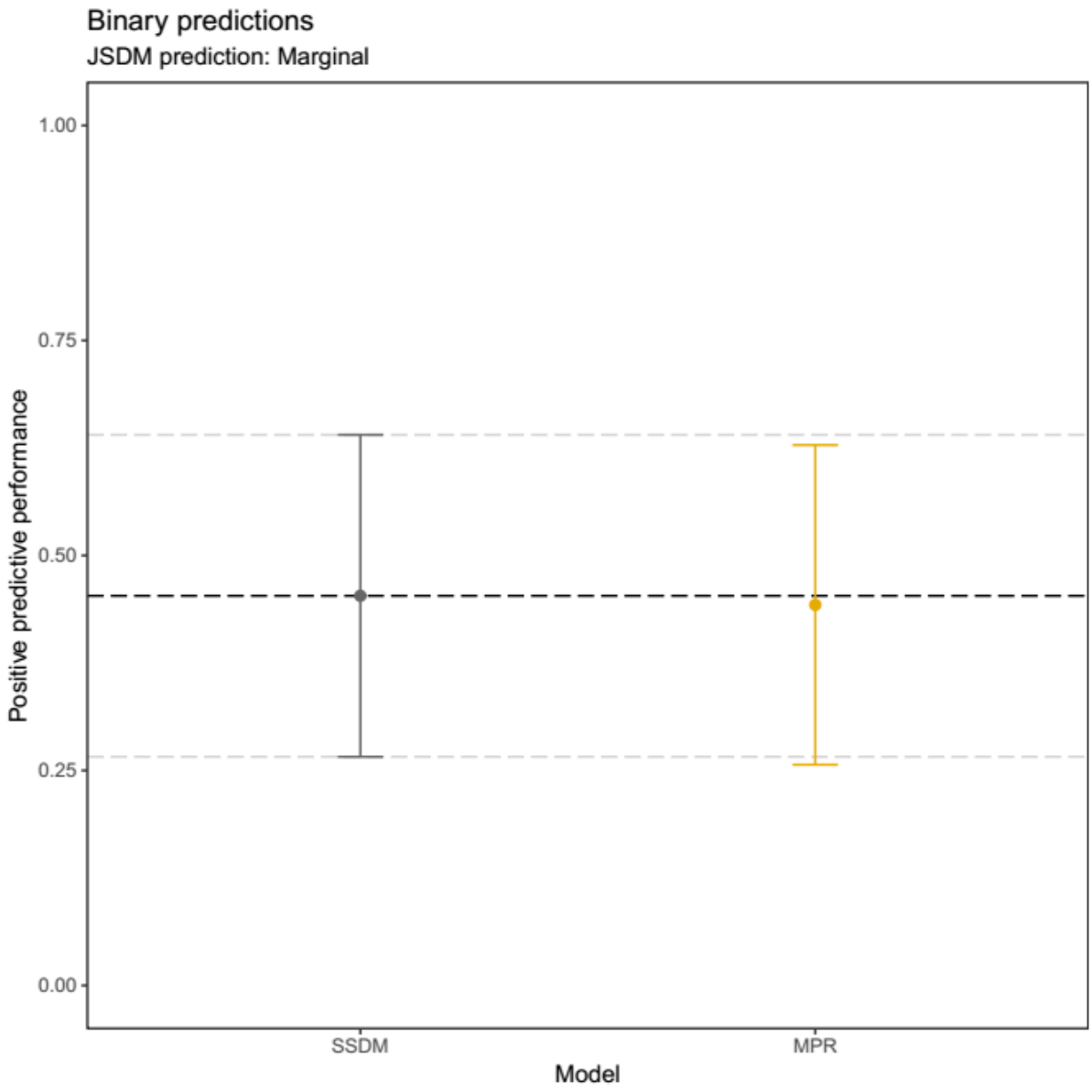


Figure 48: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

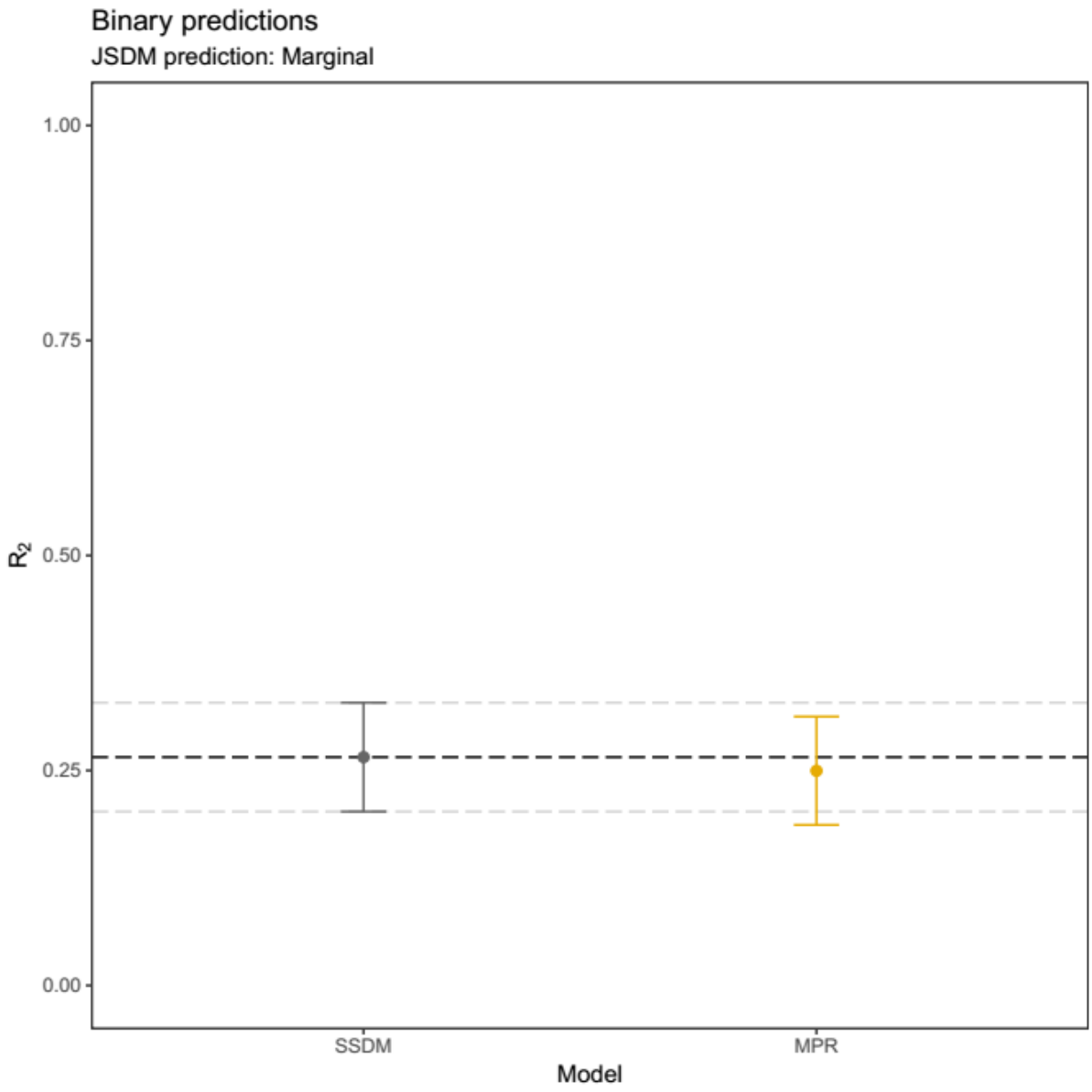


Figure 49: Performance of the SSDM and JSDM for the R^2 evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

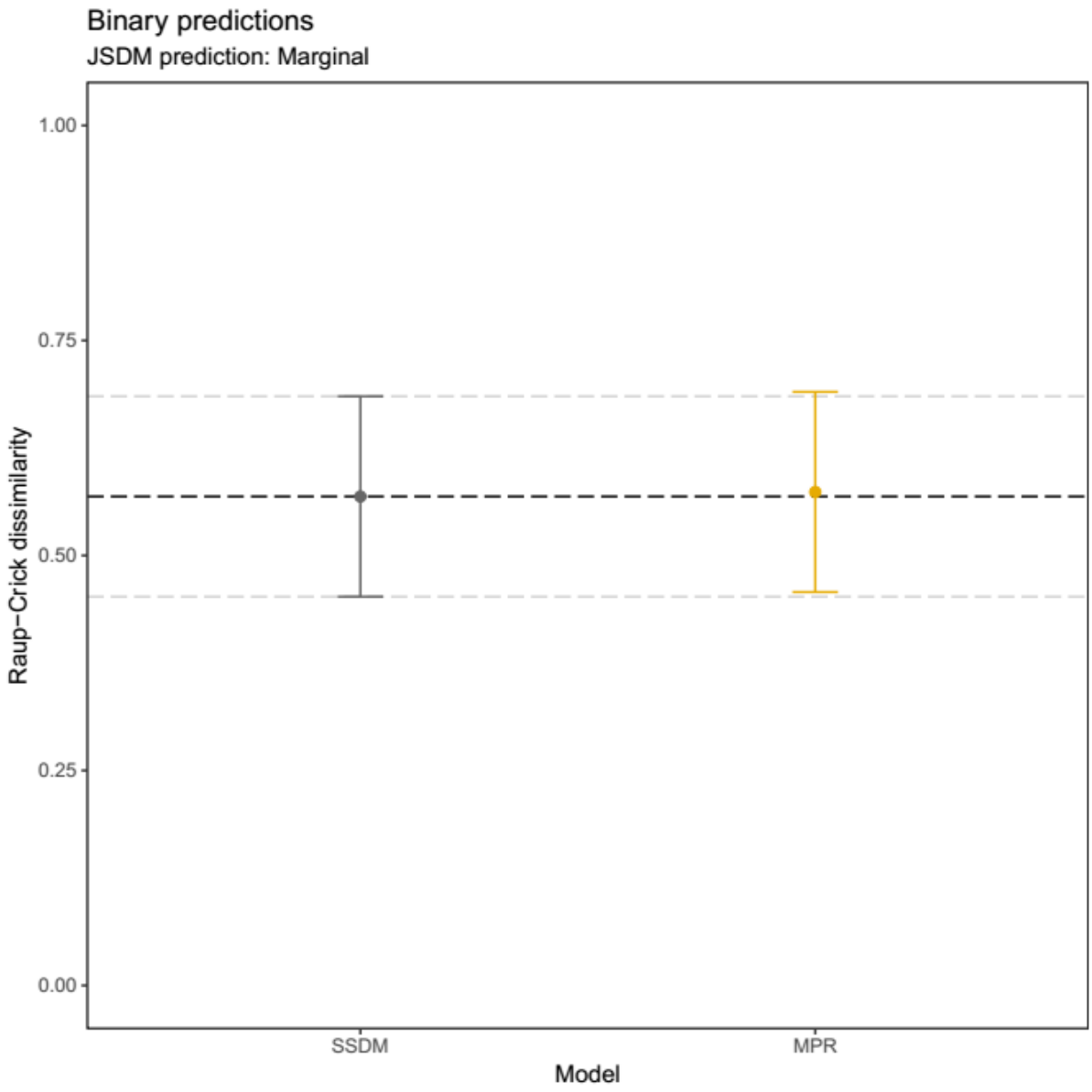


Figure 50: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

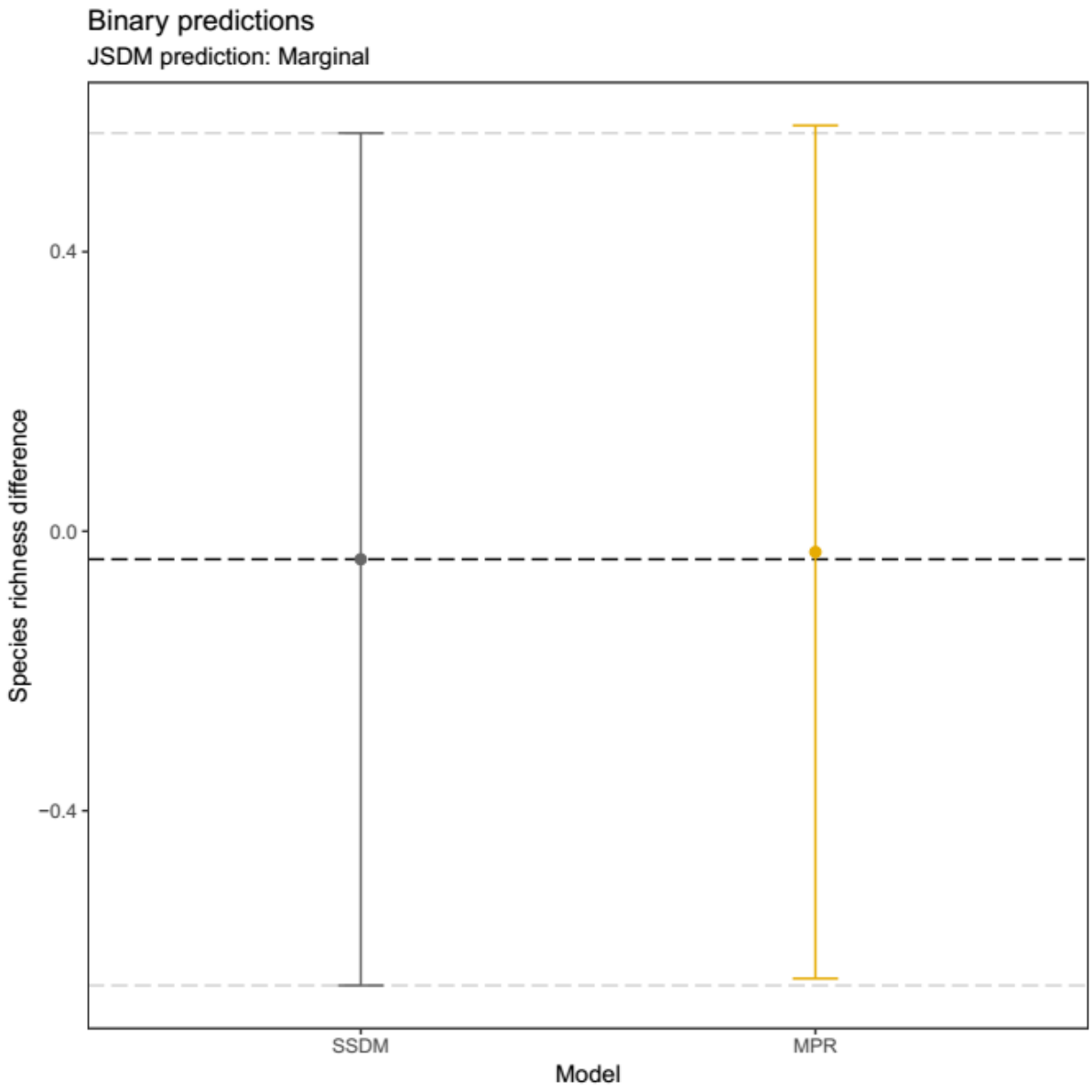


Figure 51: Performance of the SSDM and JSDM for the species richness difference evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

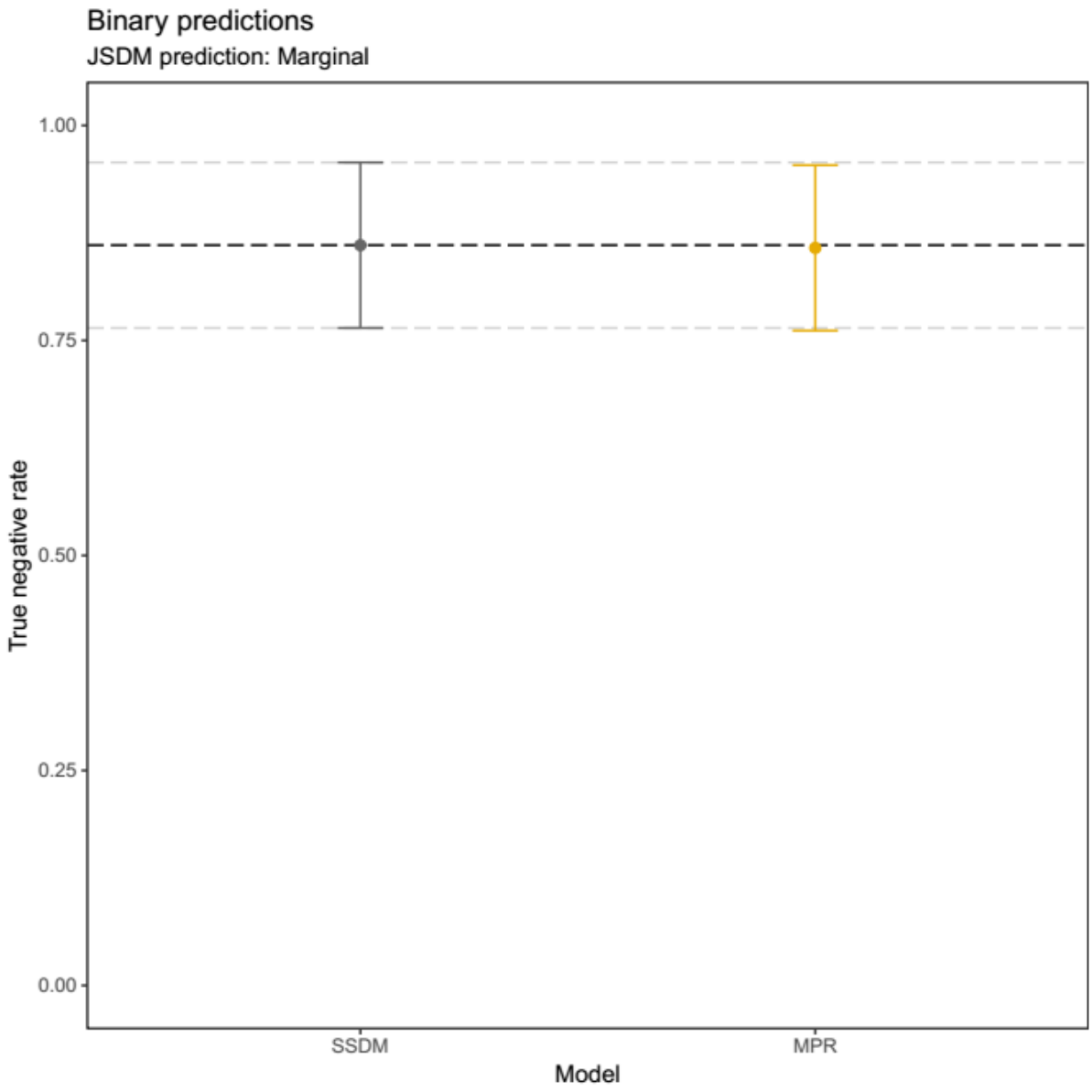


Figure 52: Performance of the SSDM and JSDM for the true negative rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

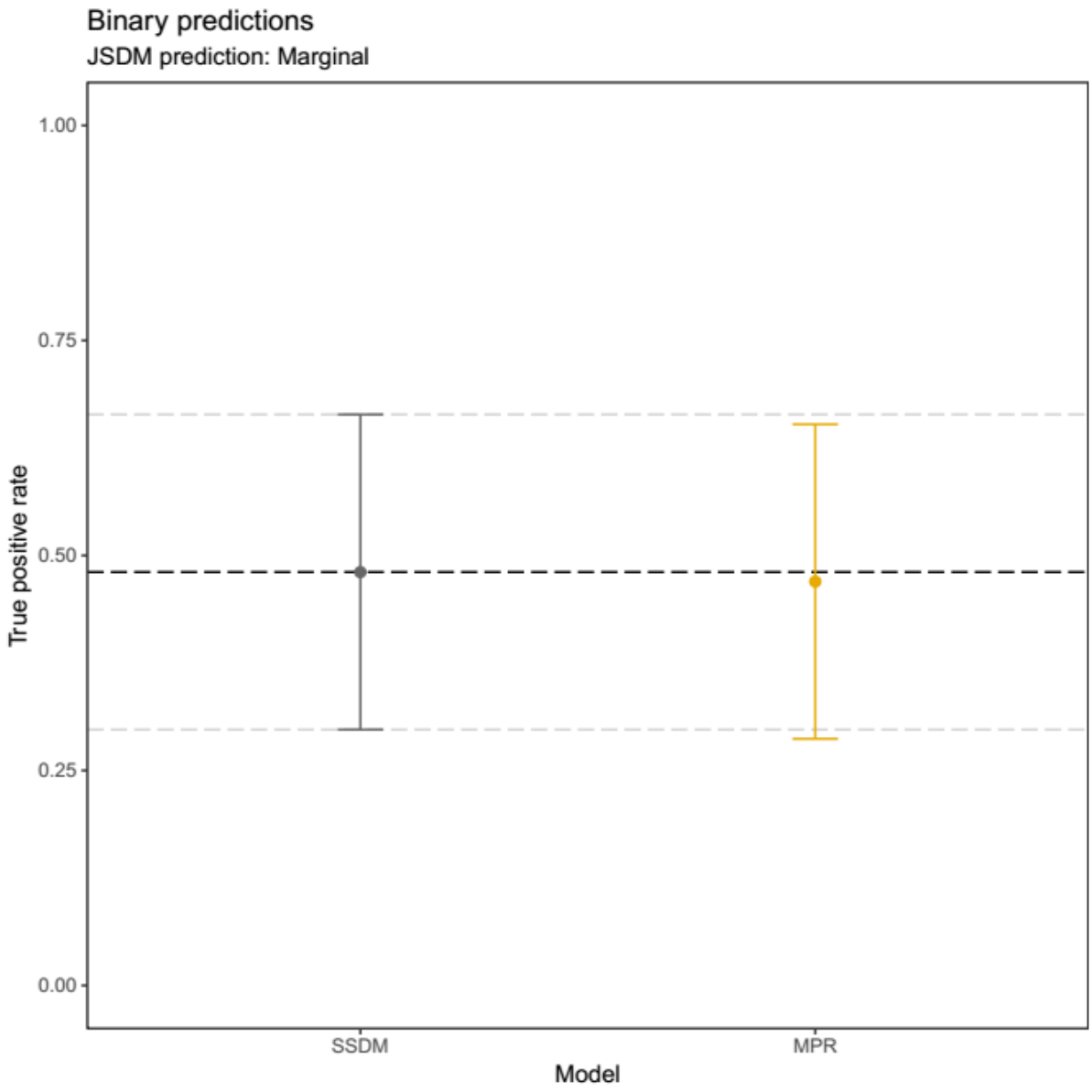


Figure 53: Performance of the SSDM and JSDM for the true positive rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

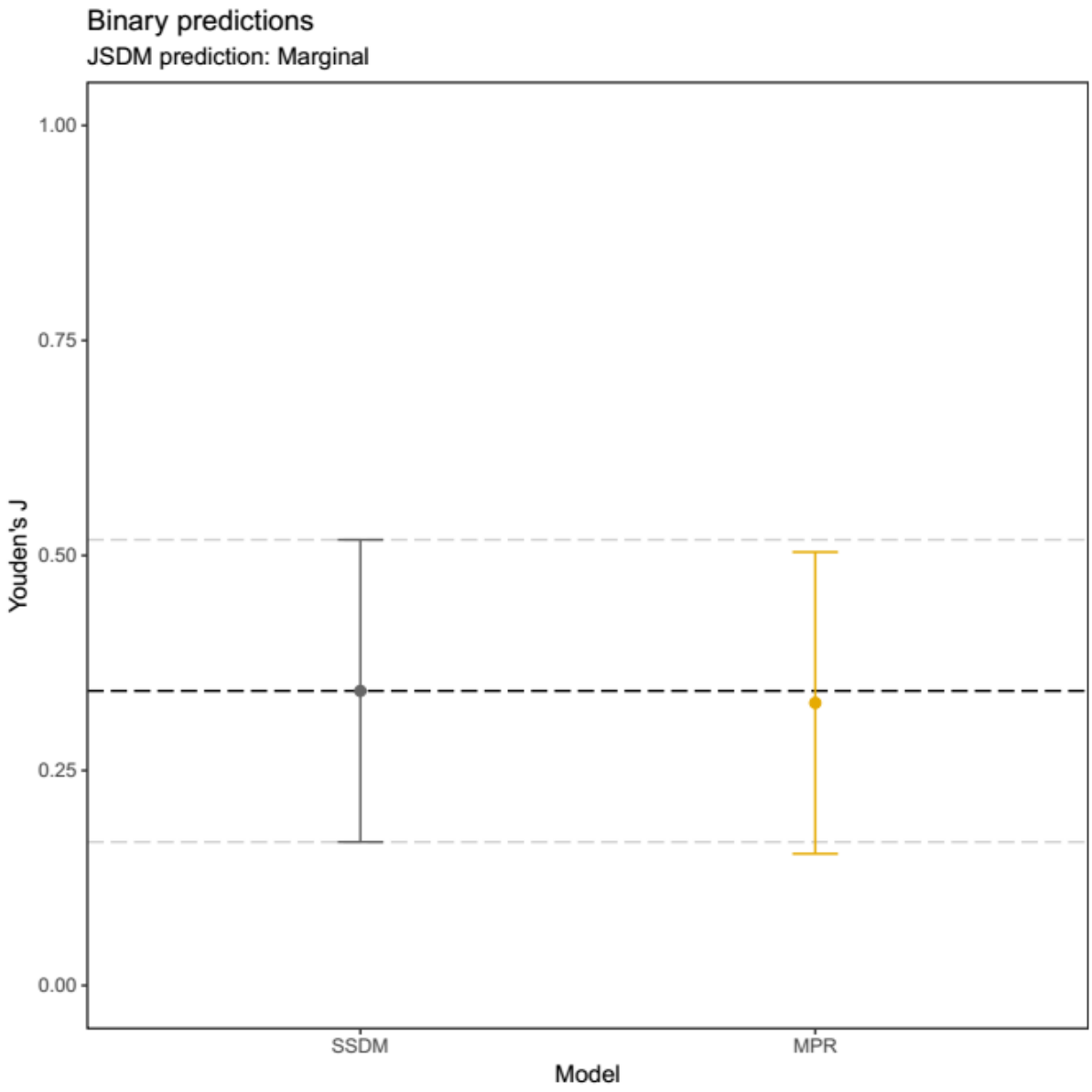


Figure 54: Performance of the SSDM and JSDM for the Youden's J evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

3. Joint predictions

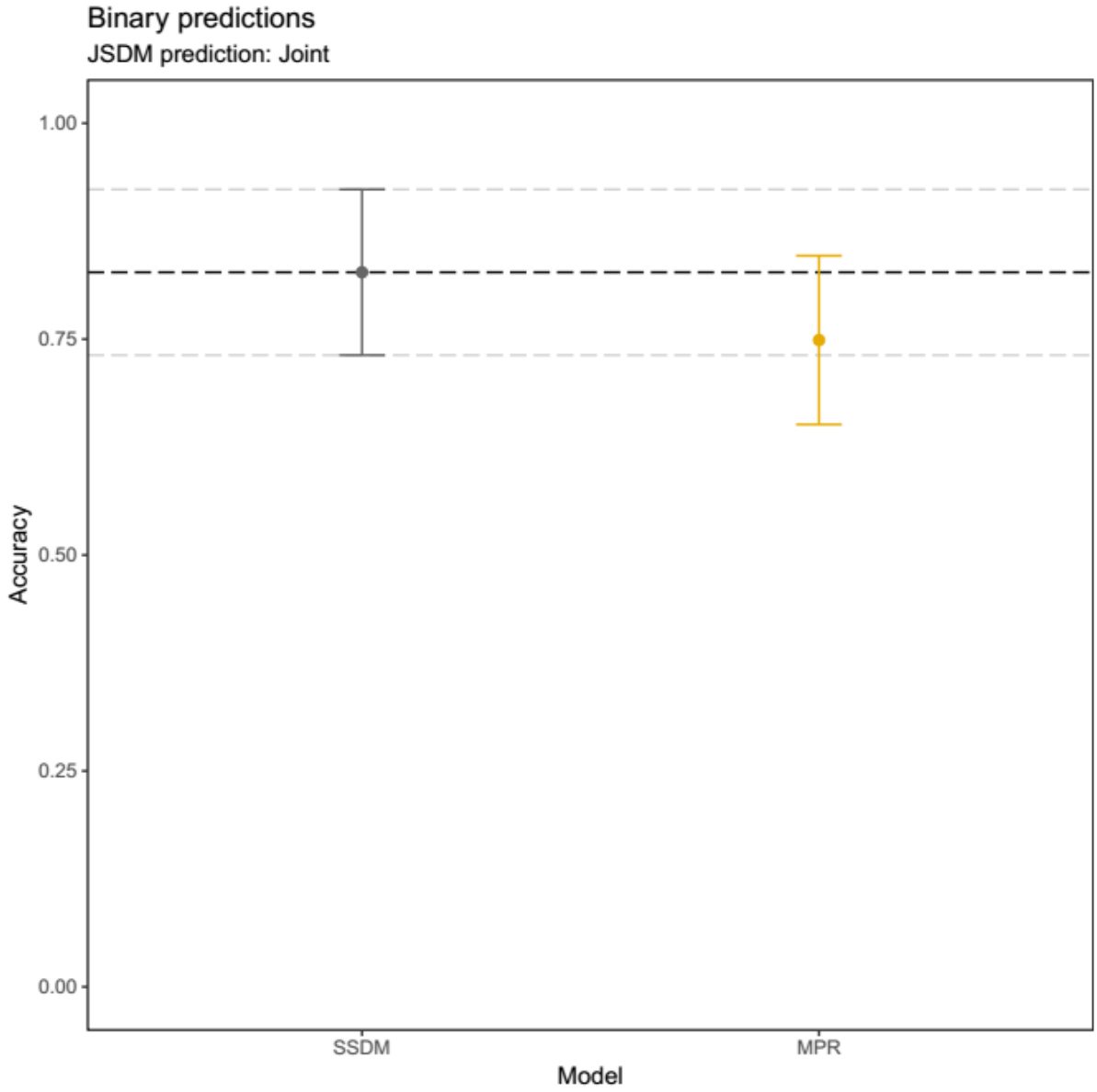


Figure 55: Performance of the SSDM and JSDM for the accuracy evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

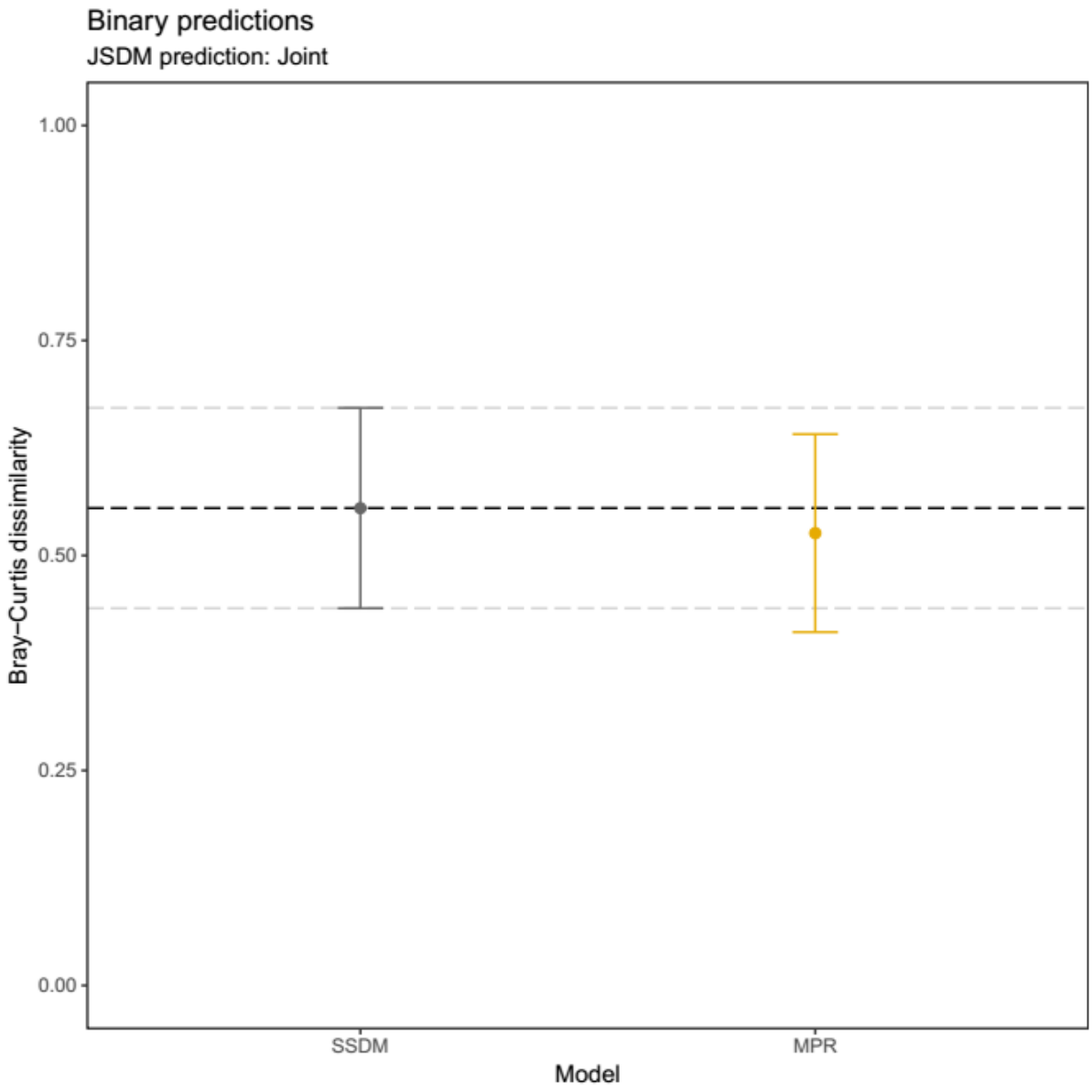


Figure 56: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

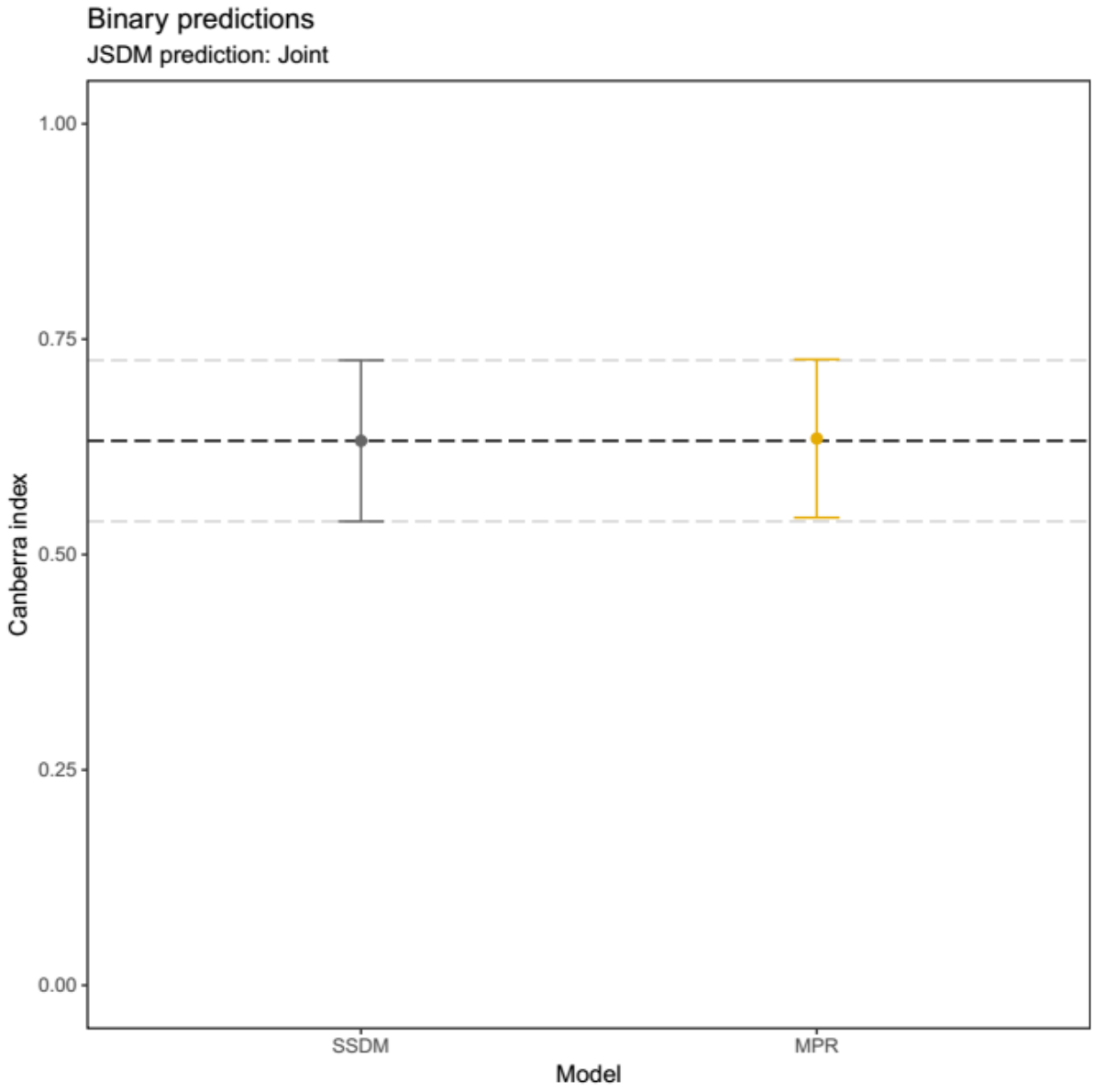


Figure 57: Performance of the SSDM and JSDM for the Canberra index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

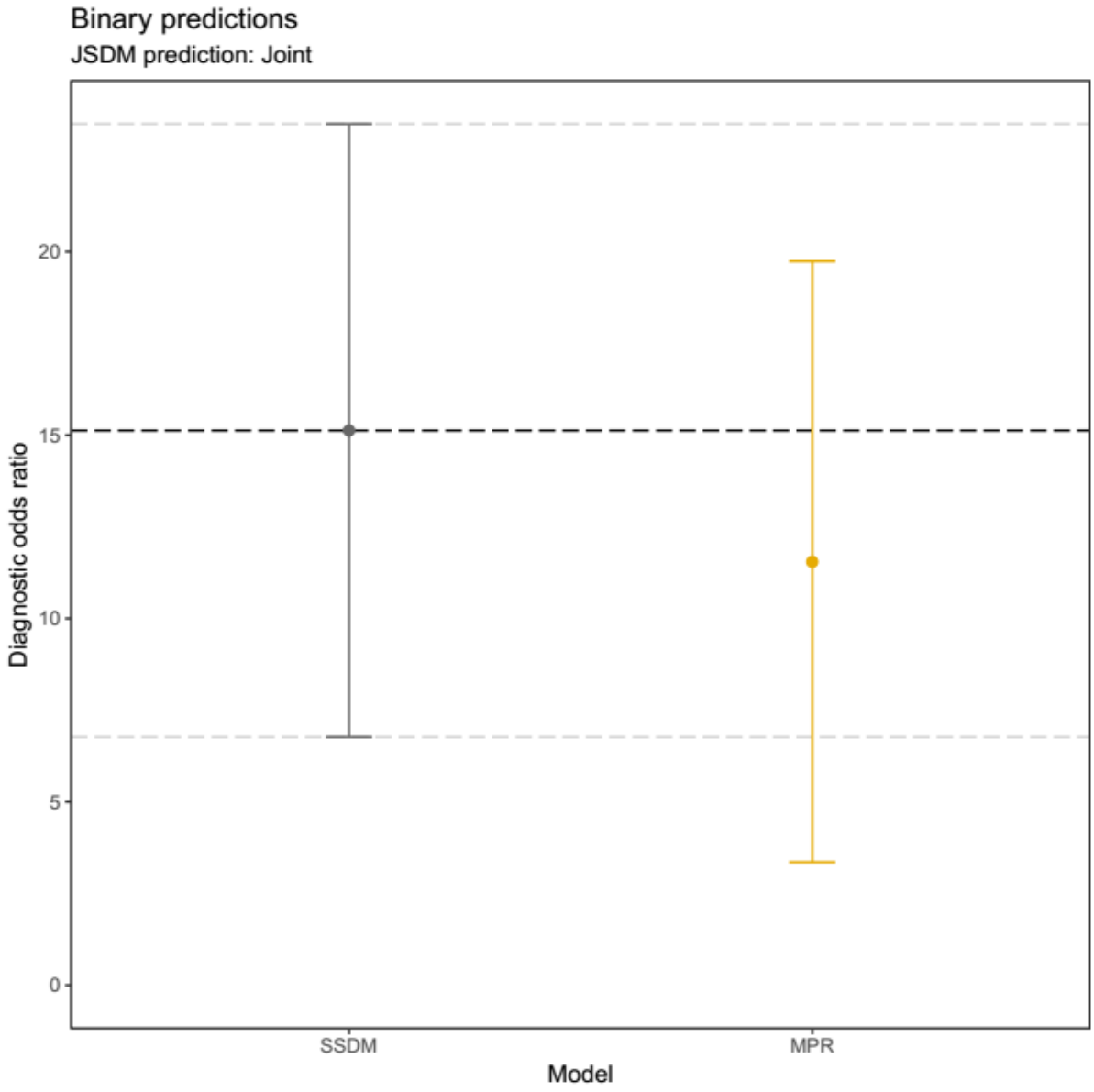


Figure 58: Performance of the SSDM and JSDM for the diagnostic odds ratio evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

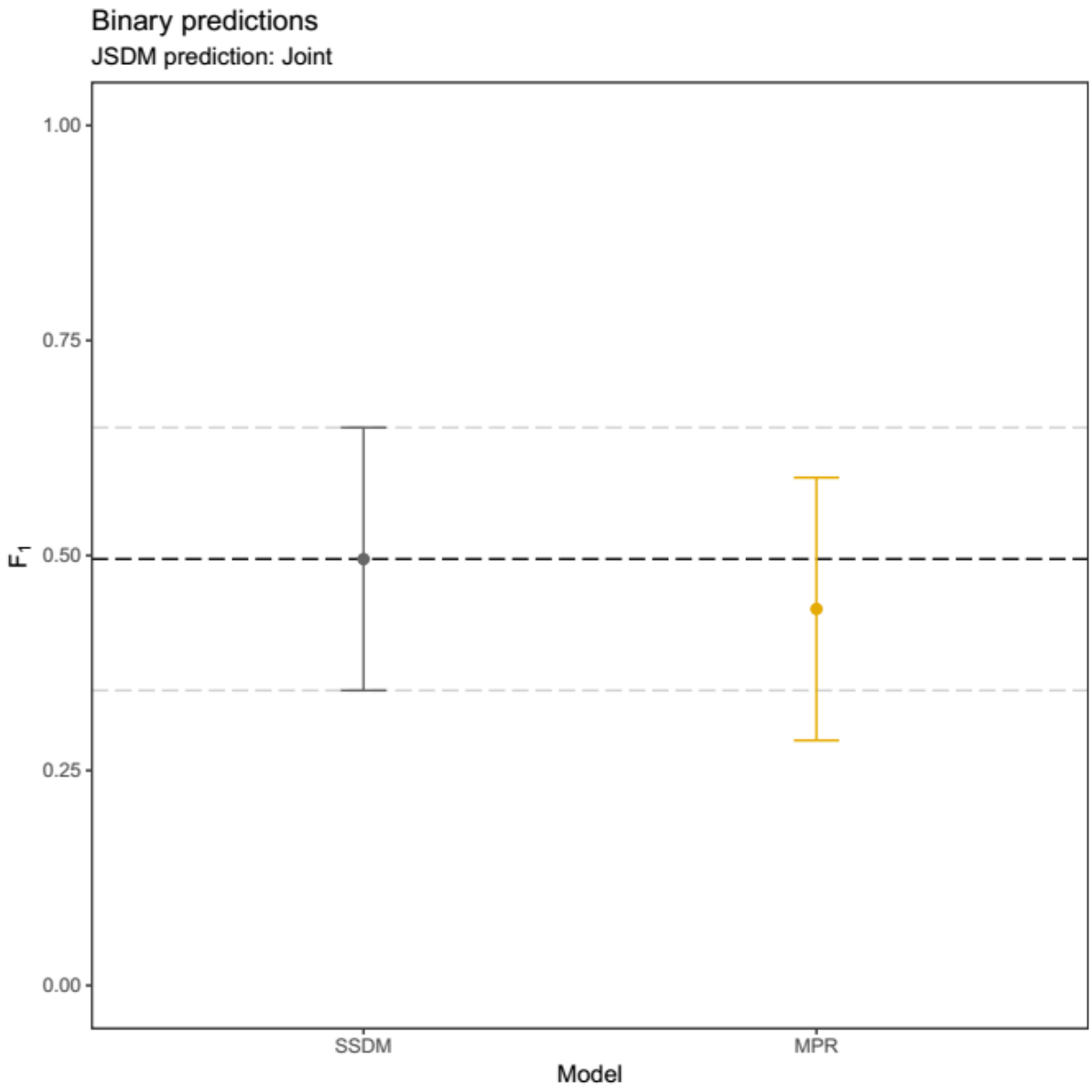


Figure 59: Performance of the SSDM and JSDM for the F_1 evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

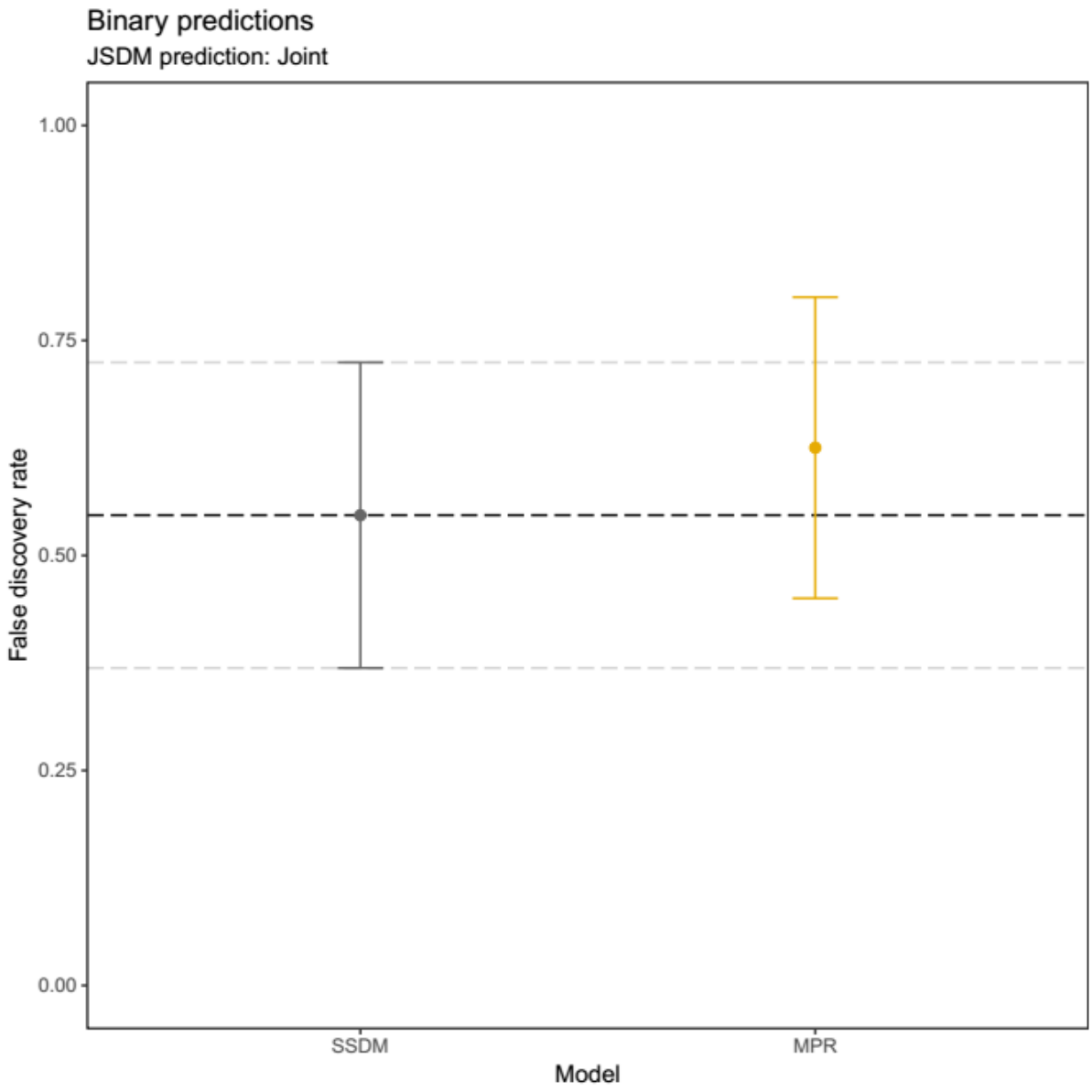


Figure 60: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

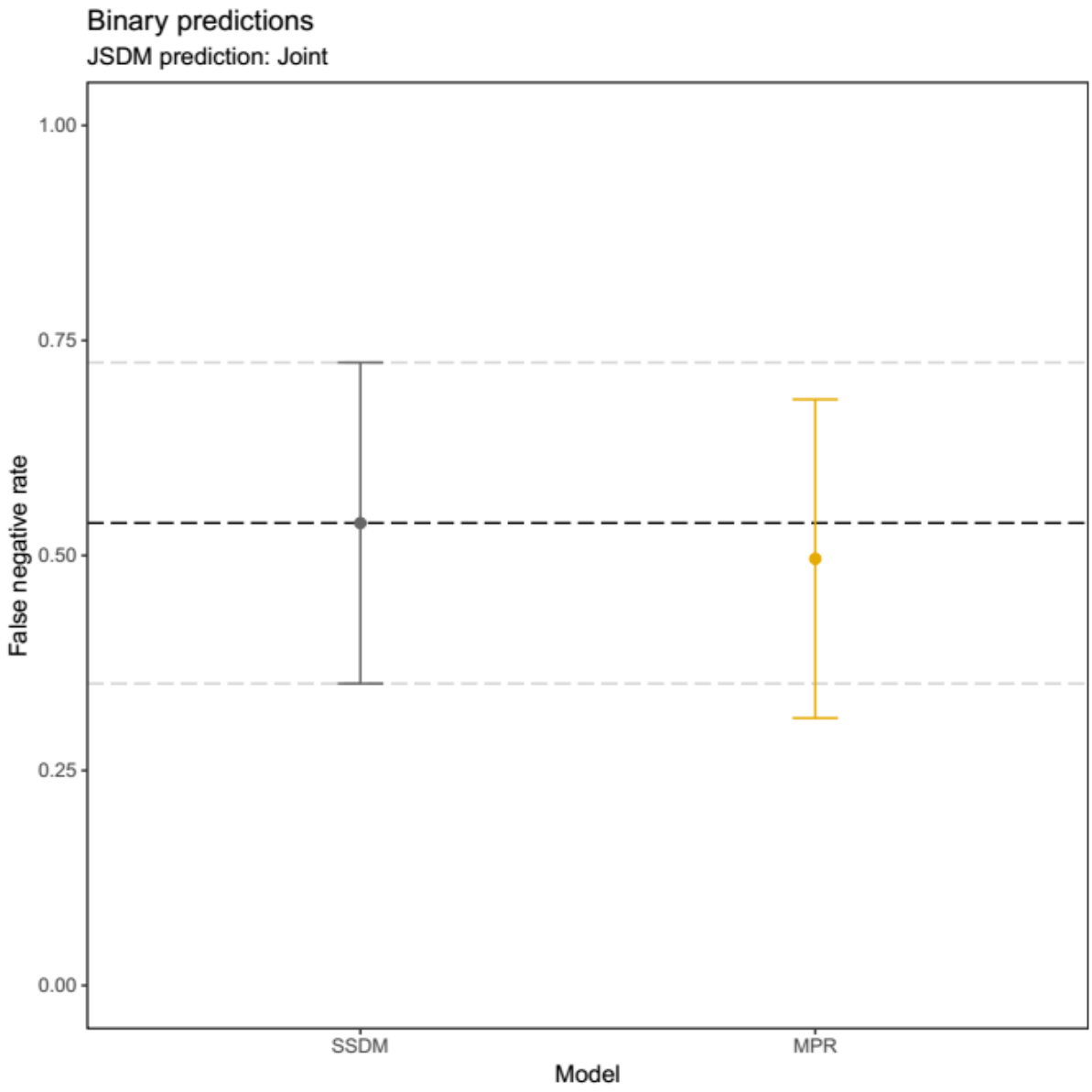


Figure 61: Performance of the SSDM and JSDM for the false negative rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

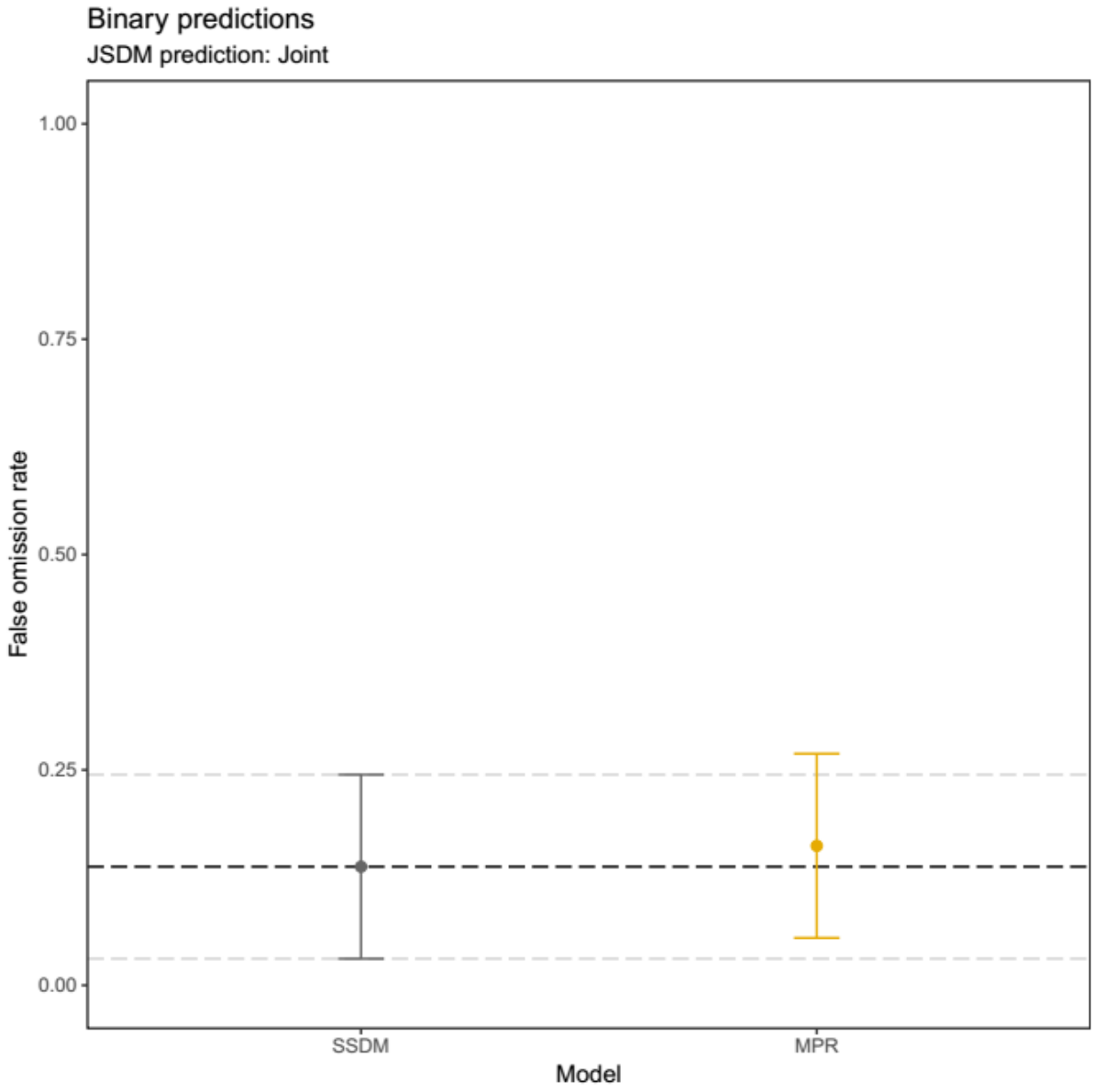


Figure 62: Performance of the SSDM and JSDM for the false omission rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

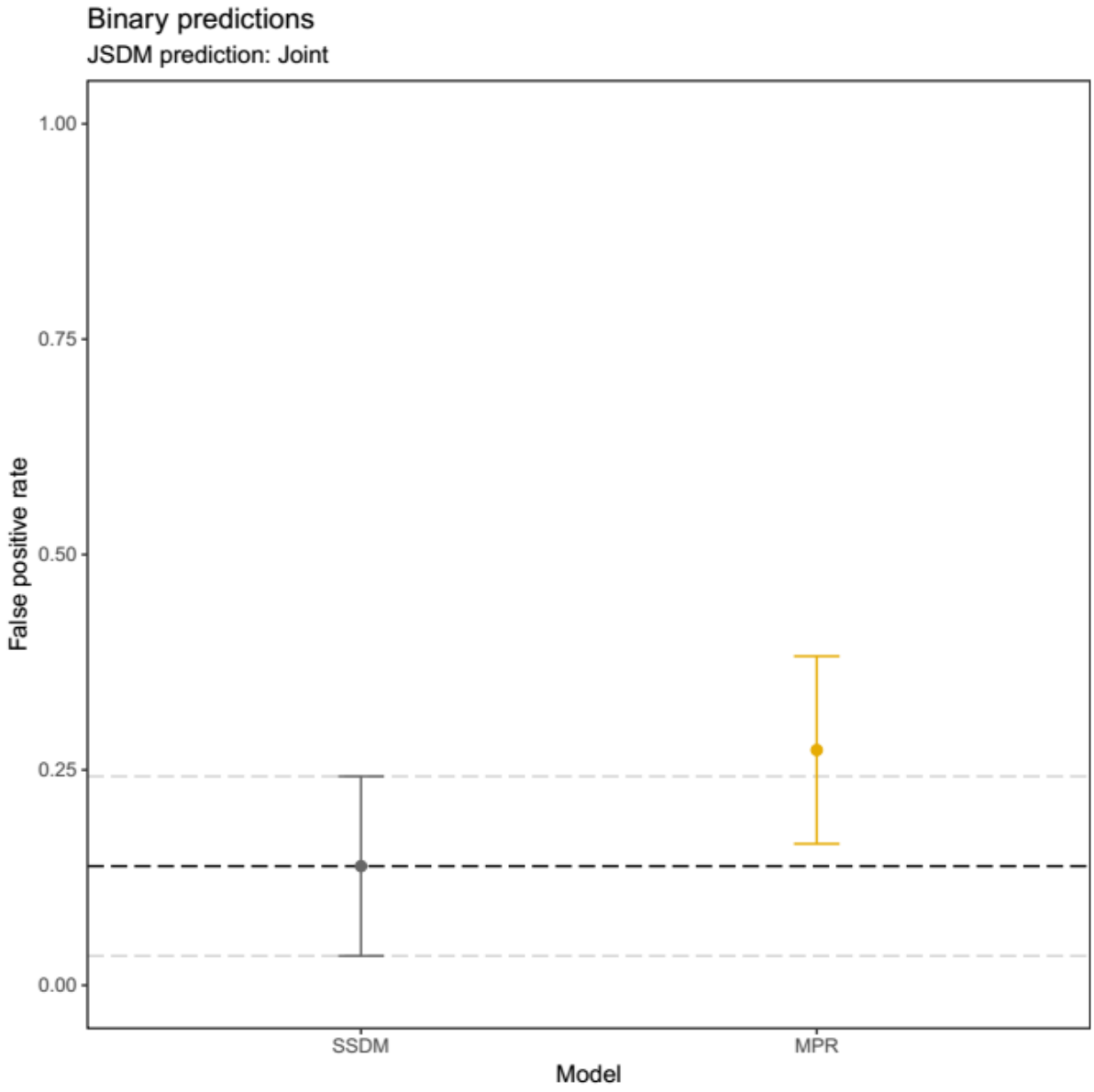


Figure 63: Performance of the SSDM and JSDM for the false positive rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

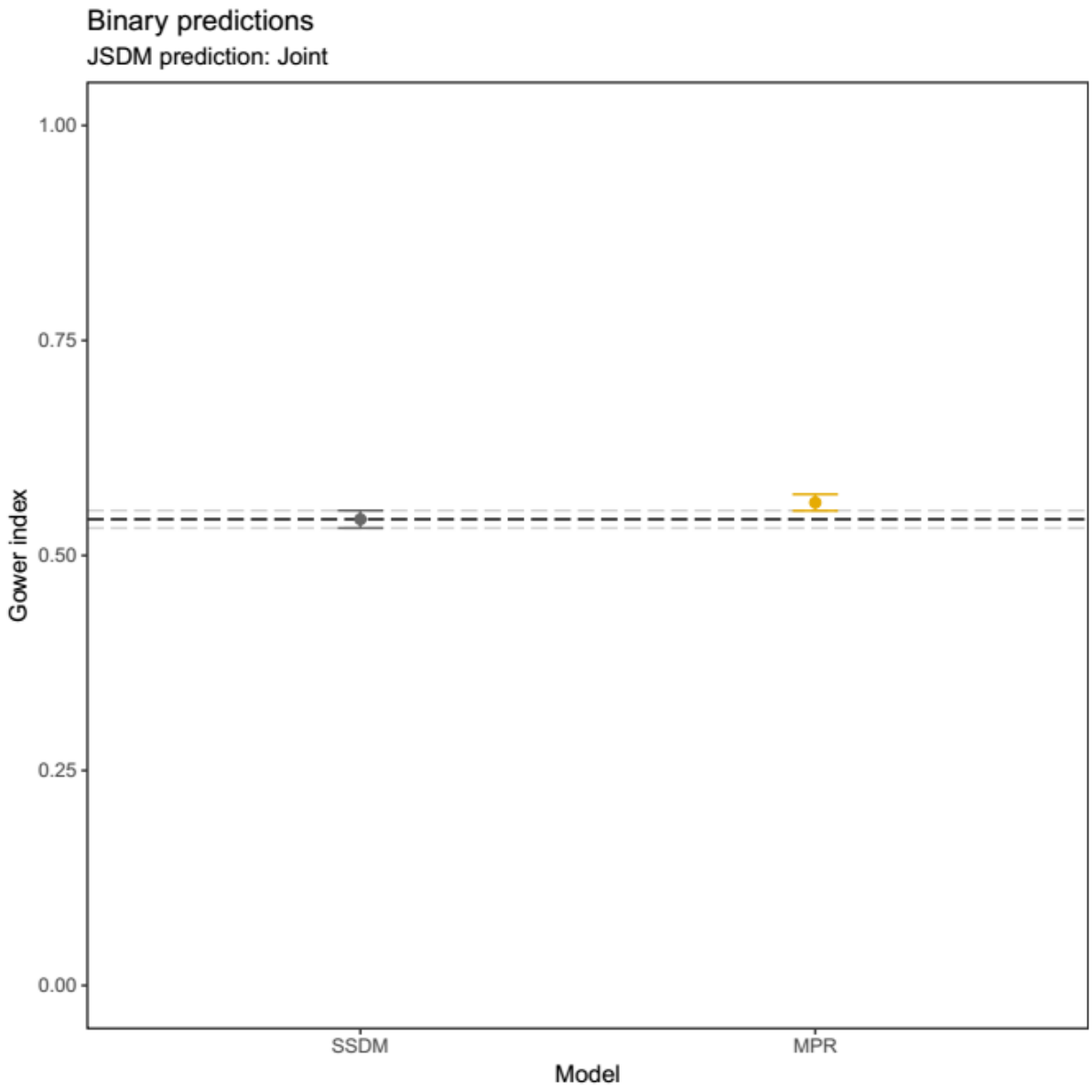


Figure 64: Performance of the SSDM and JSDM for the Gower index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

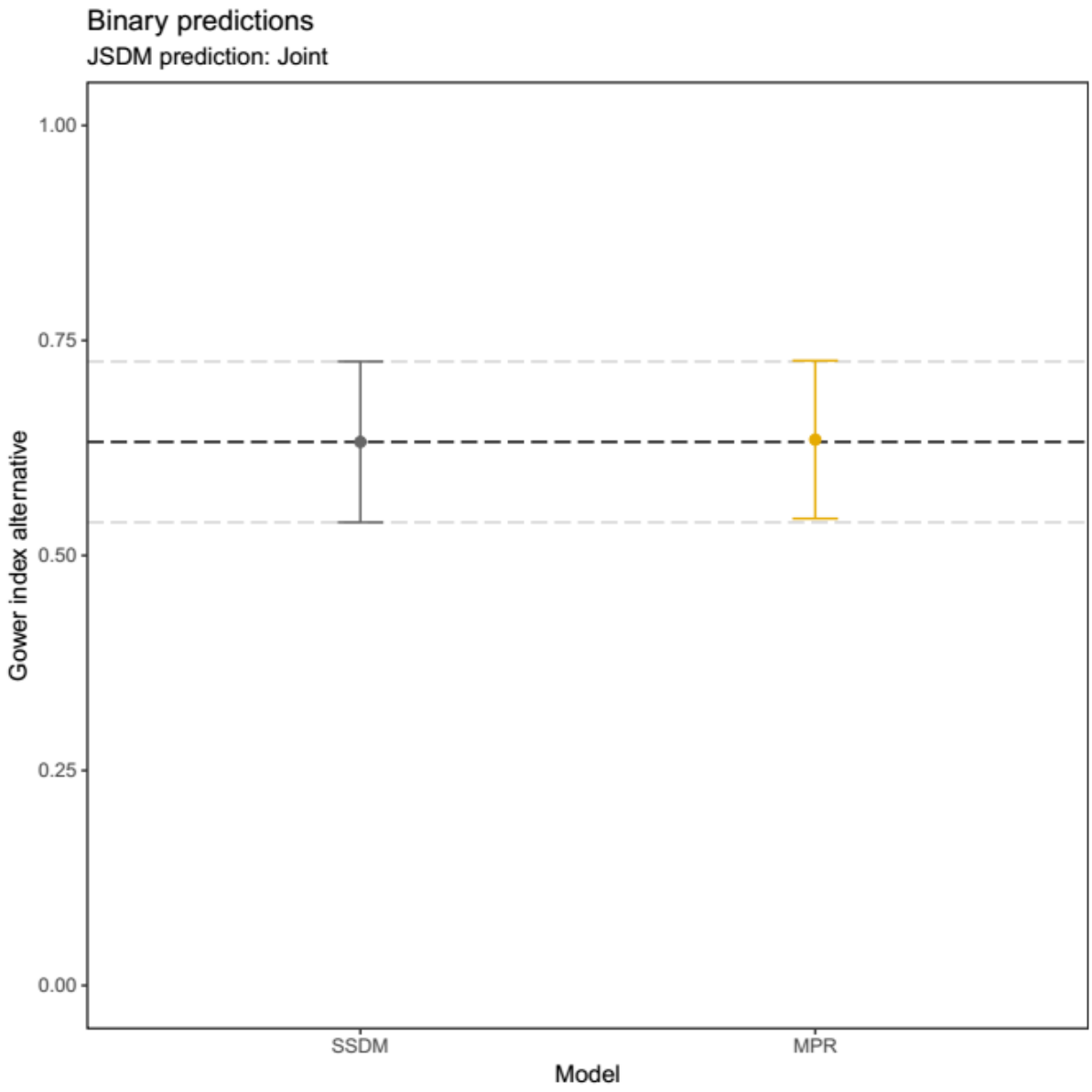


Figure 65: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

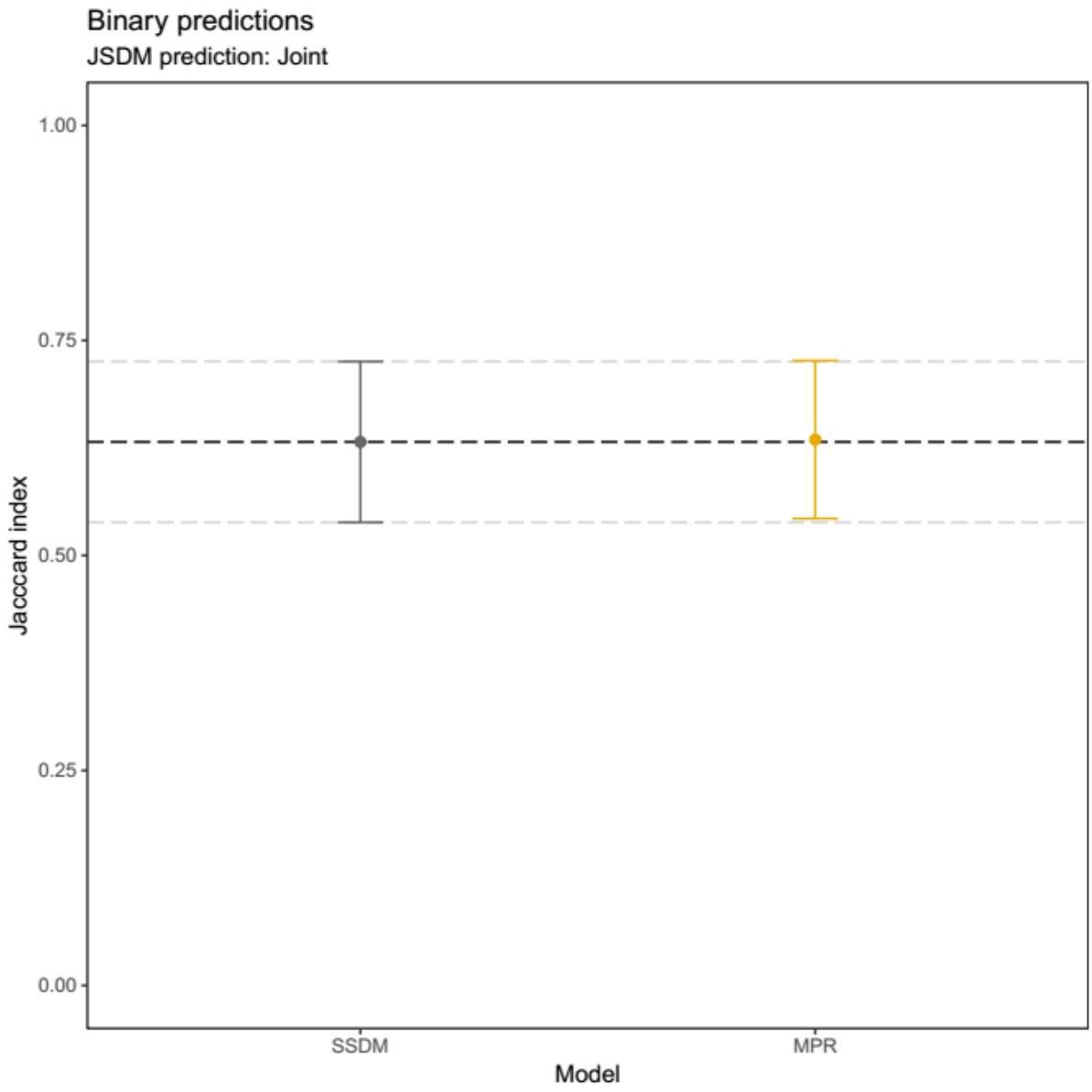


Figure 66: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

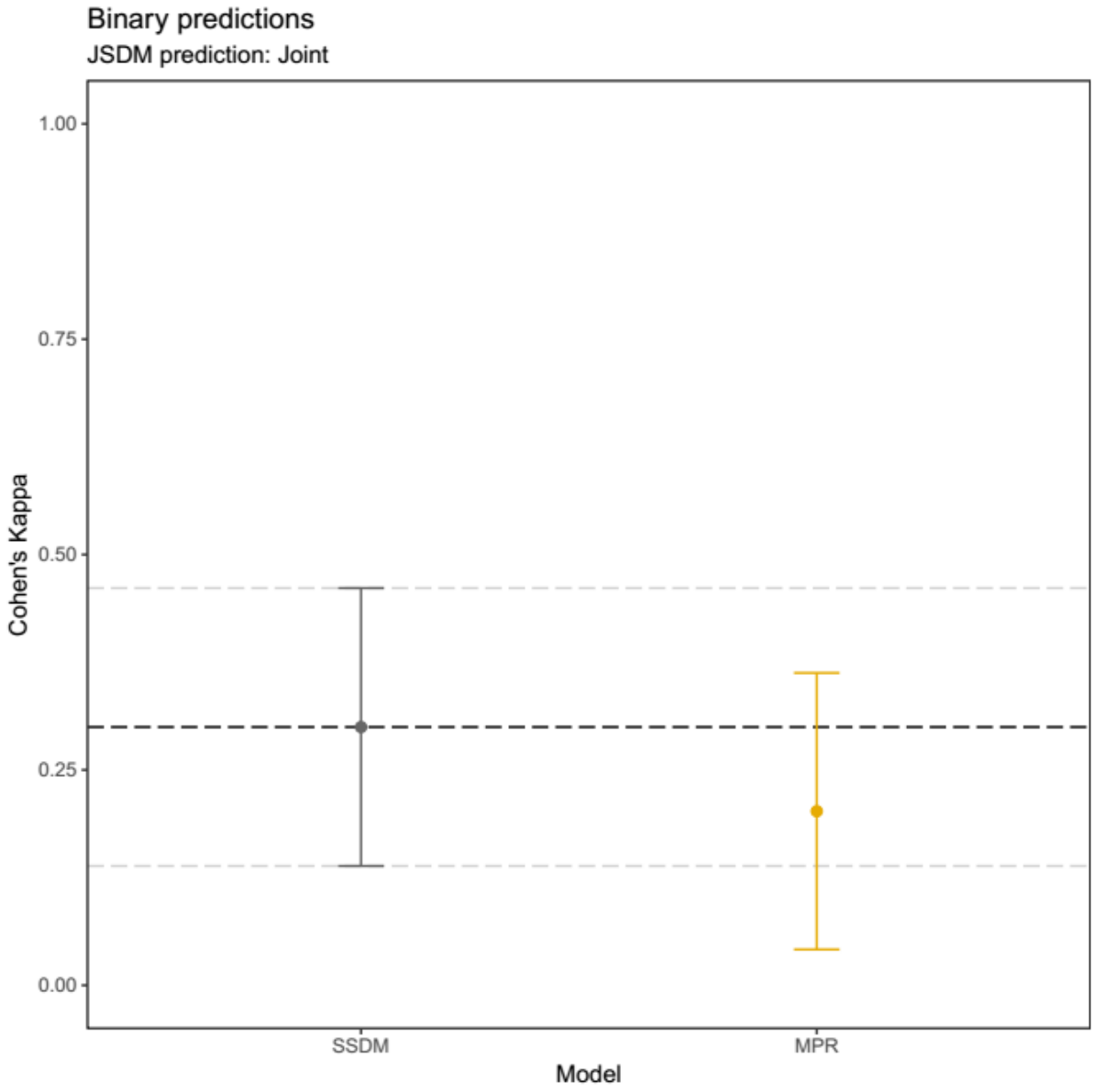


Figure 67: Performance of the SSDM and JSDM for the Cohen's Kappa evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

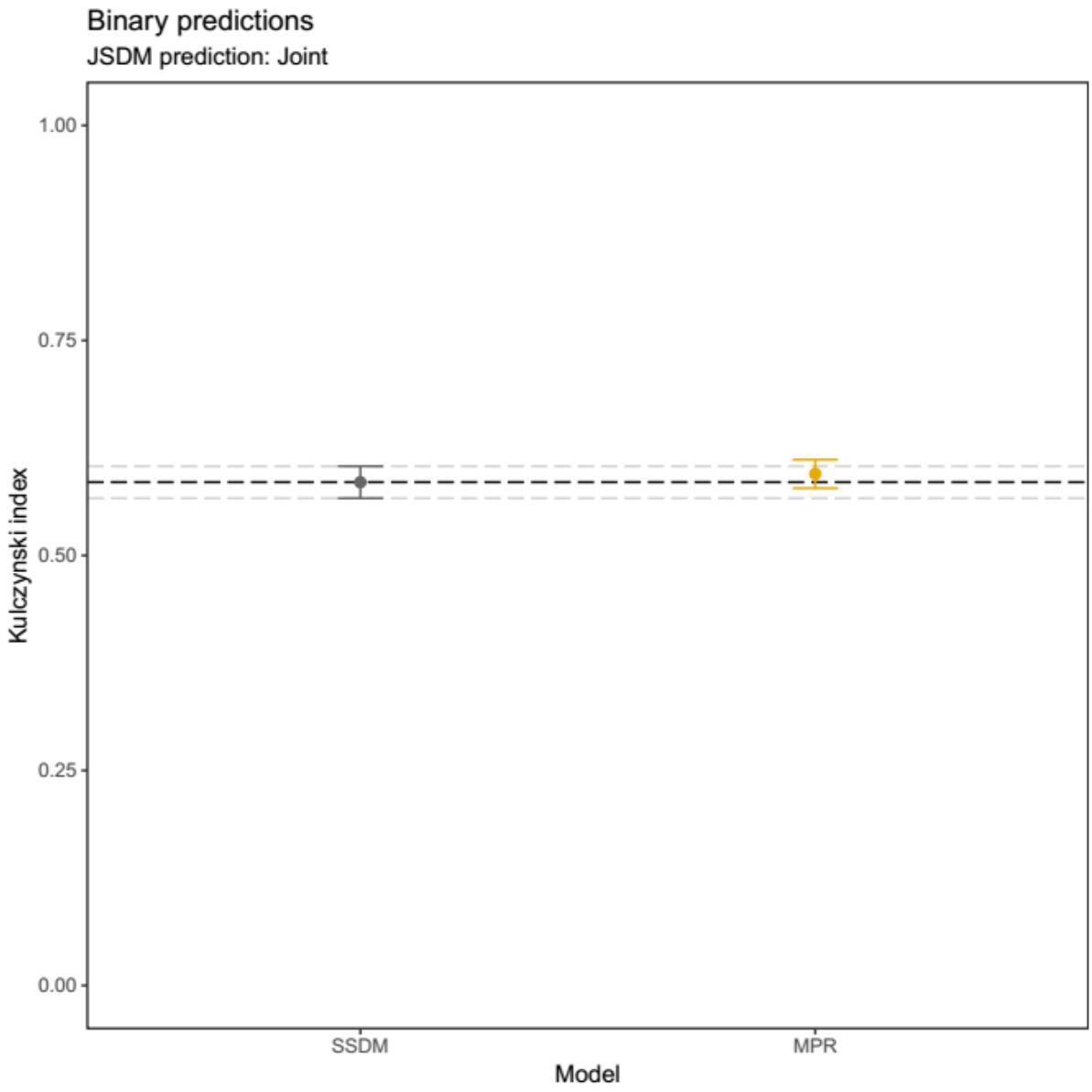


Figure 68: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

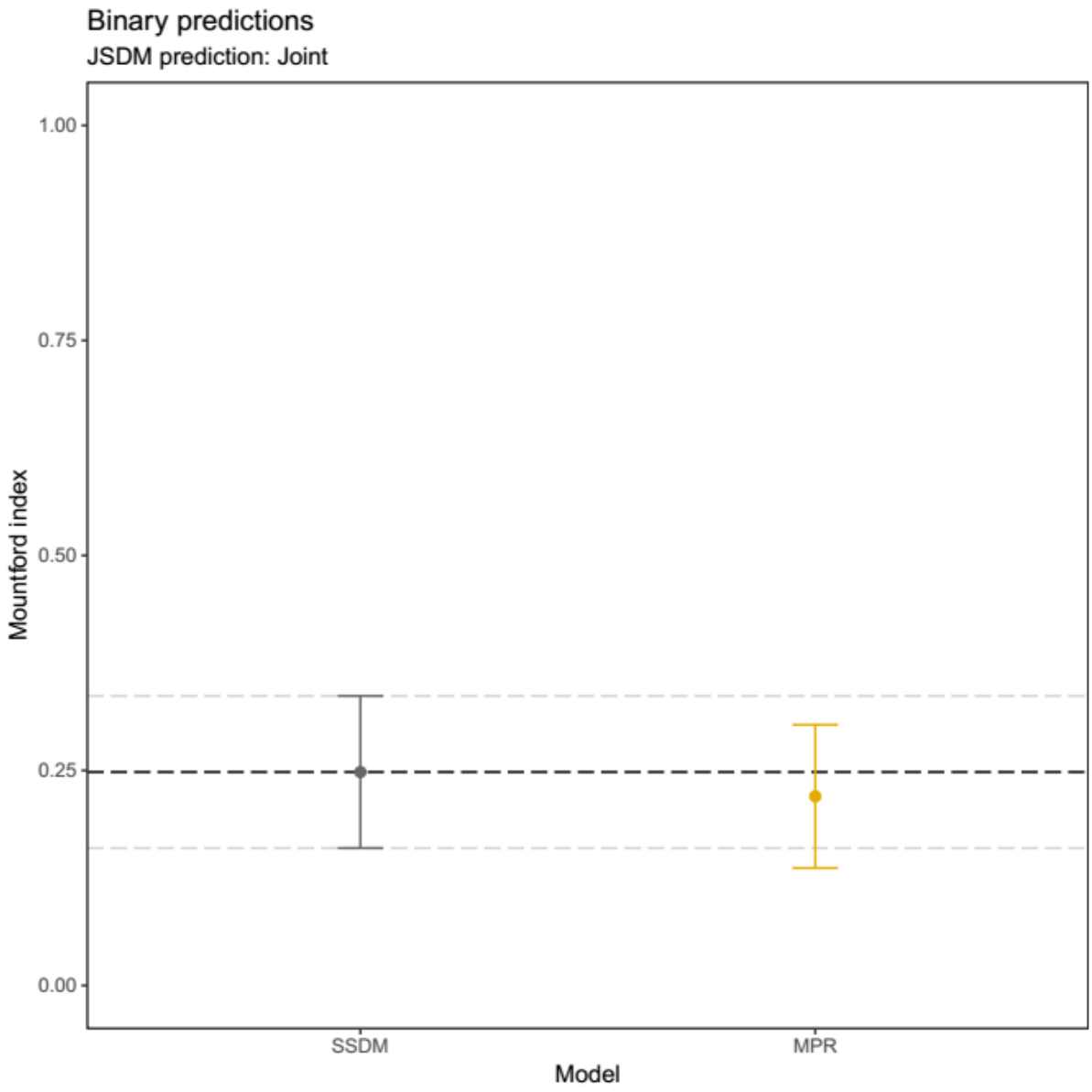


Figure 69: Performance of the SSDM and JSDM for the Mountford index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

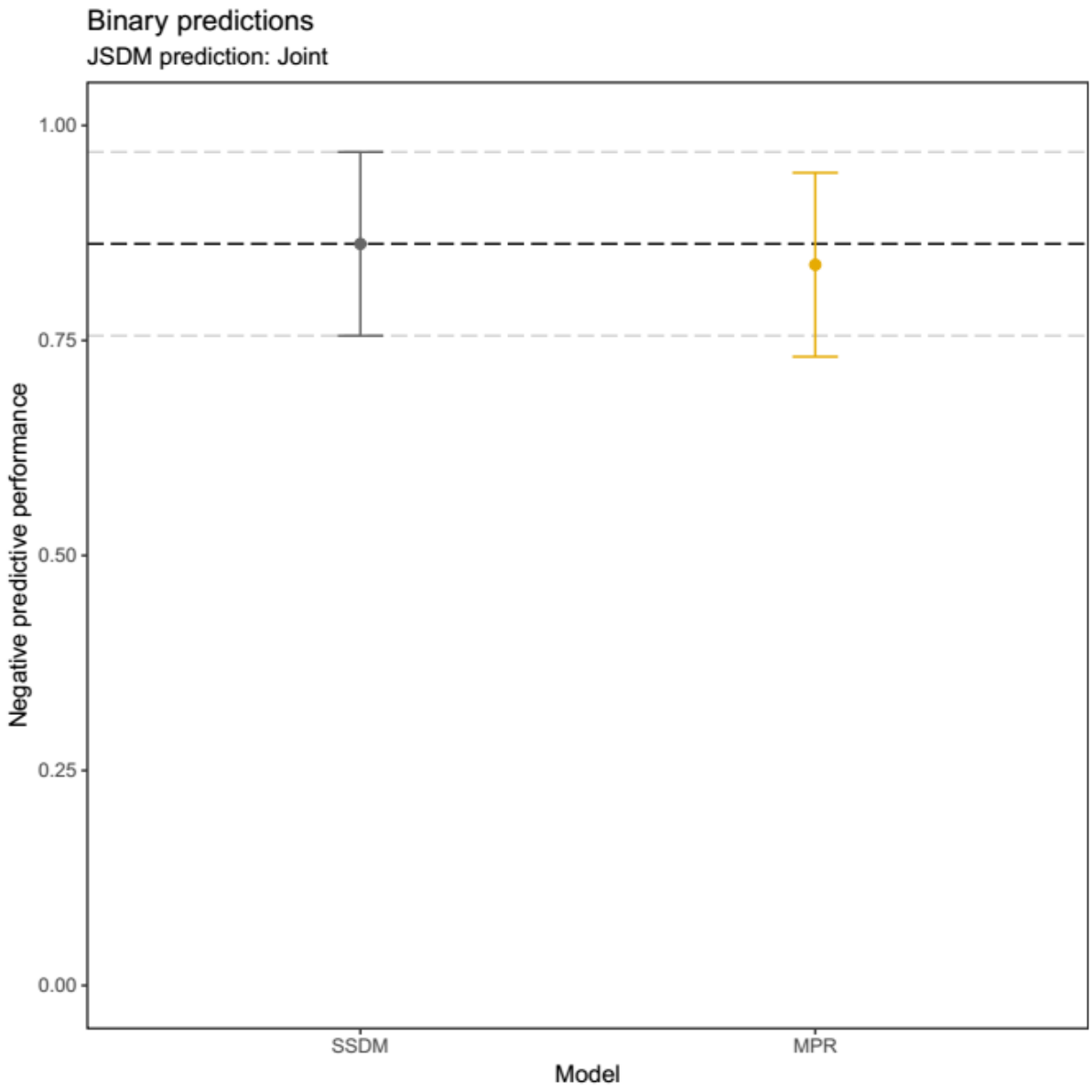


Figure 70: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

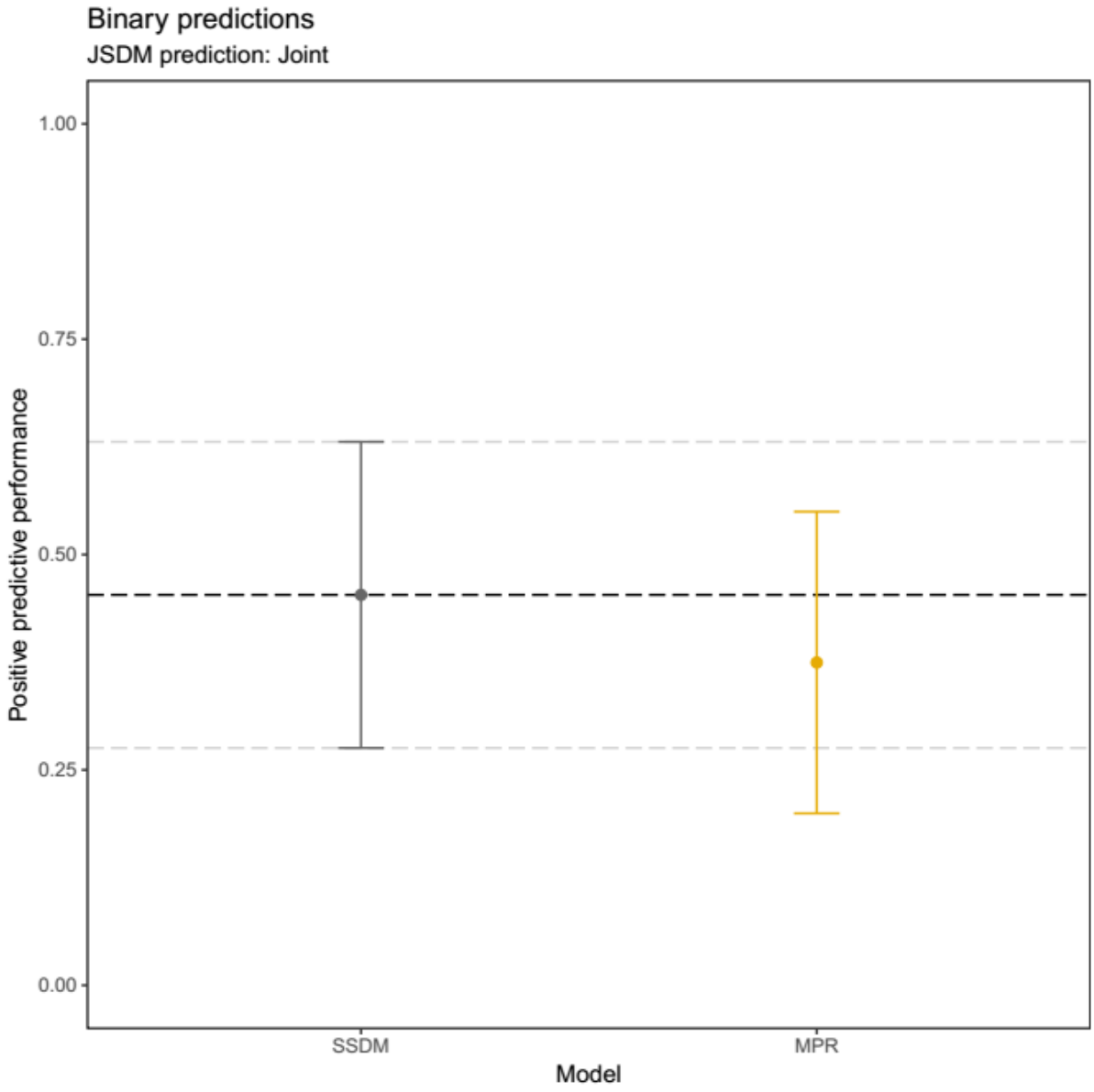


Figure 71: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

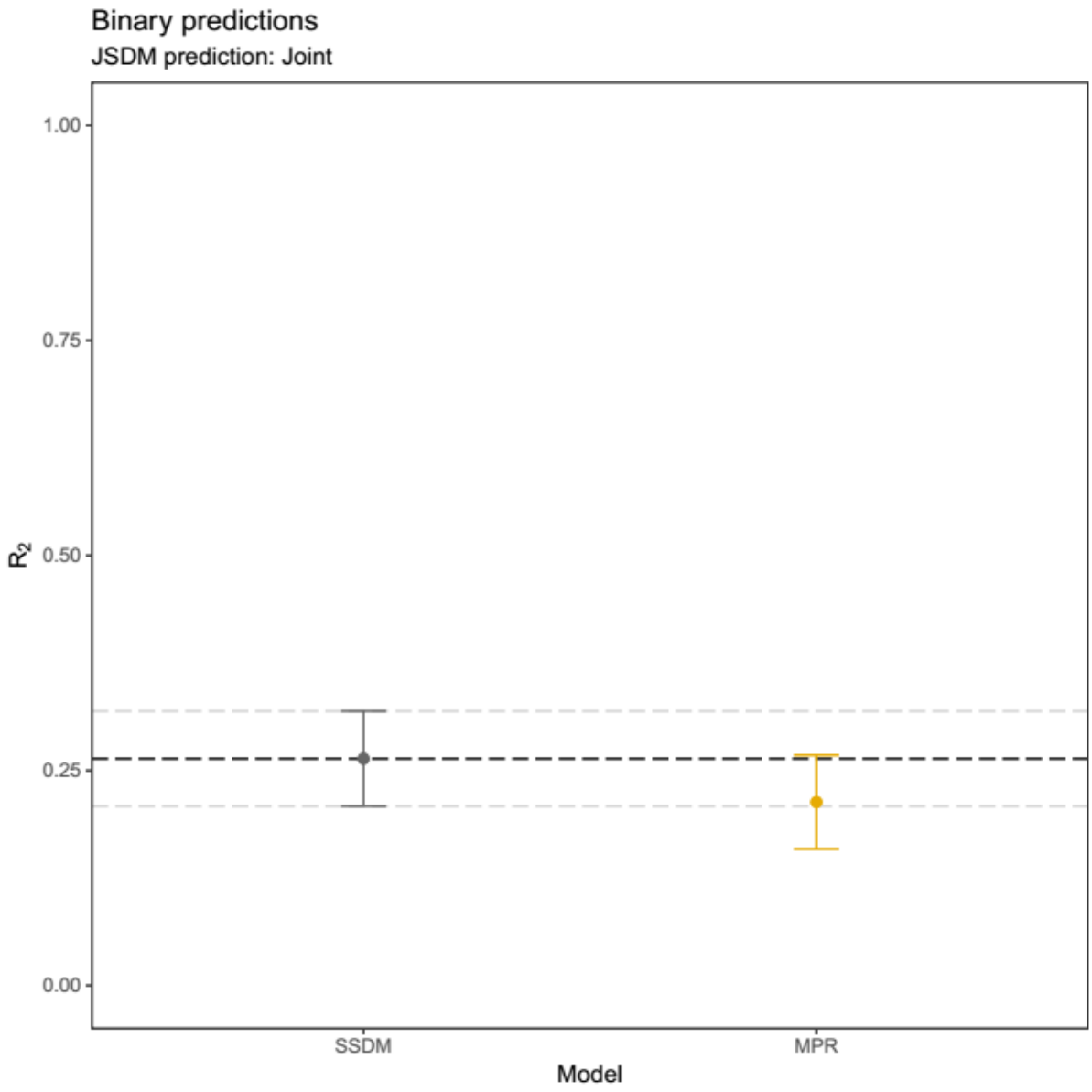


Figure 72: Performance of the SSDM and JSDM for the R^2 evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

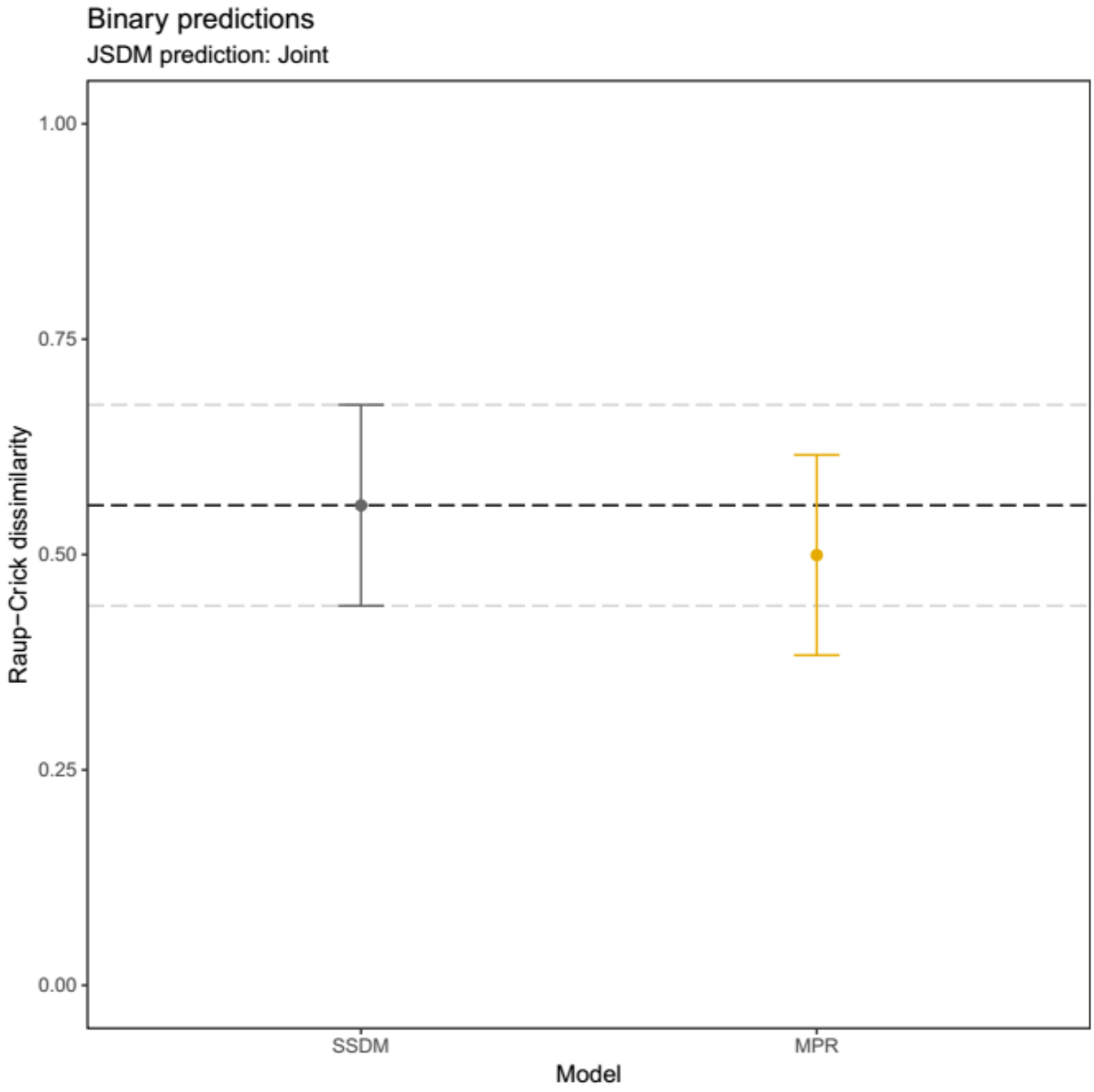


Figure 73: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

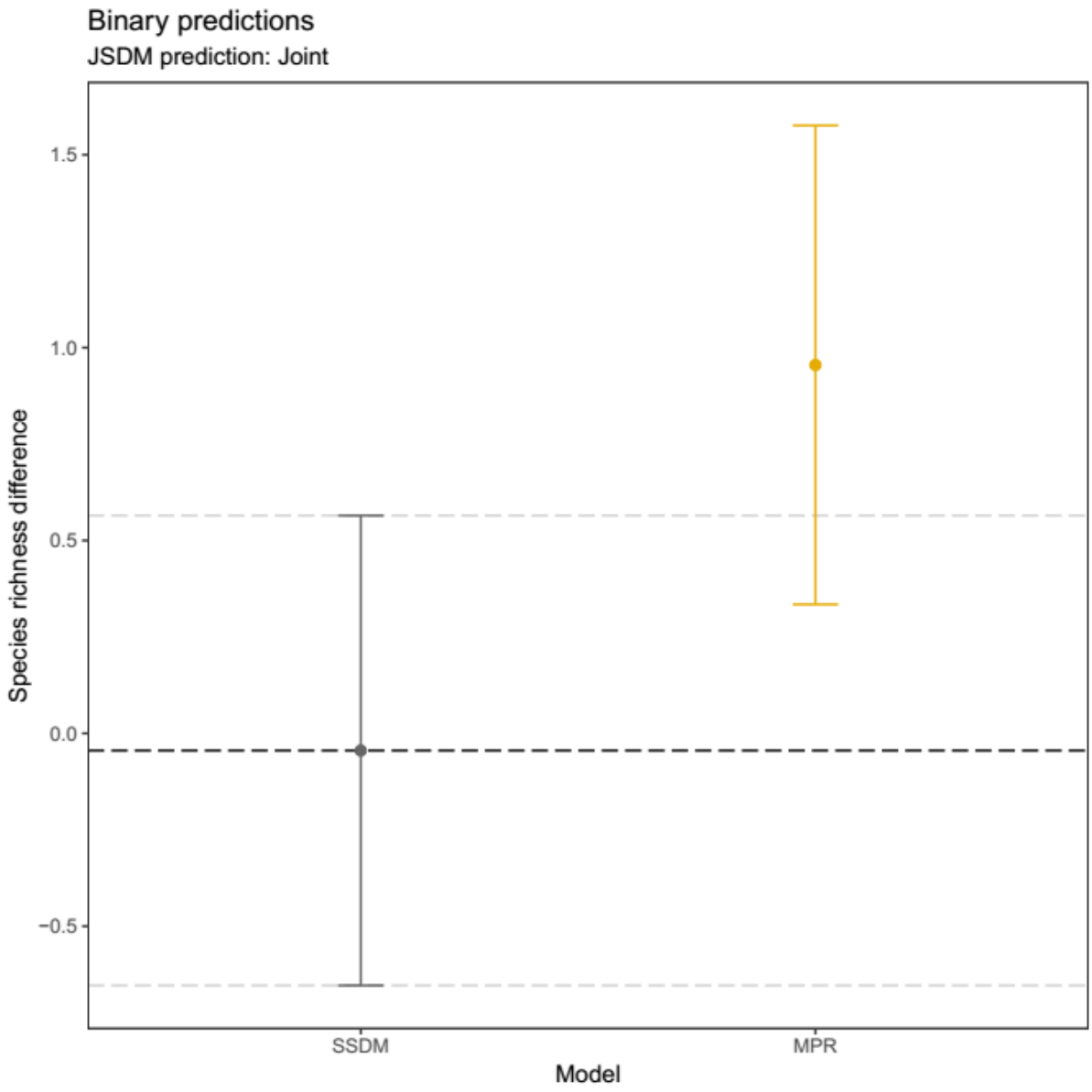


Figure 74: Performance of the SSDM and JSDM for the species richness difference evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

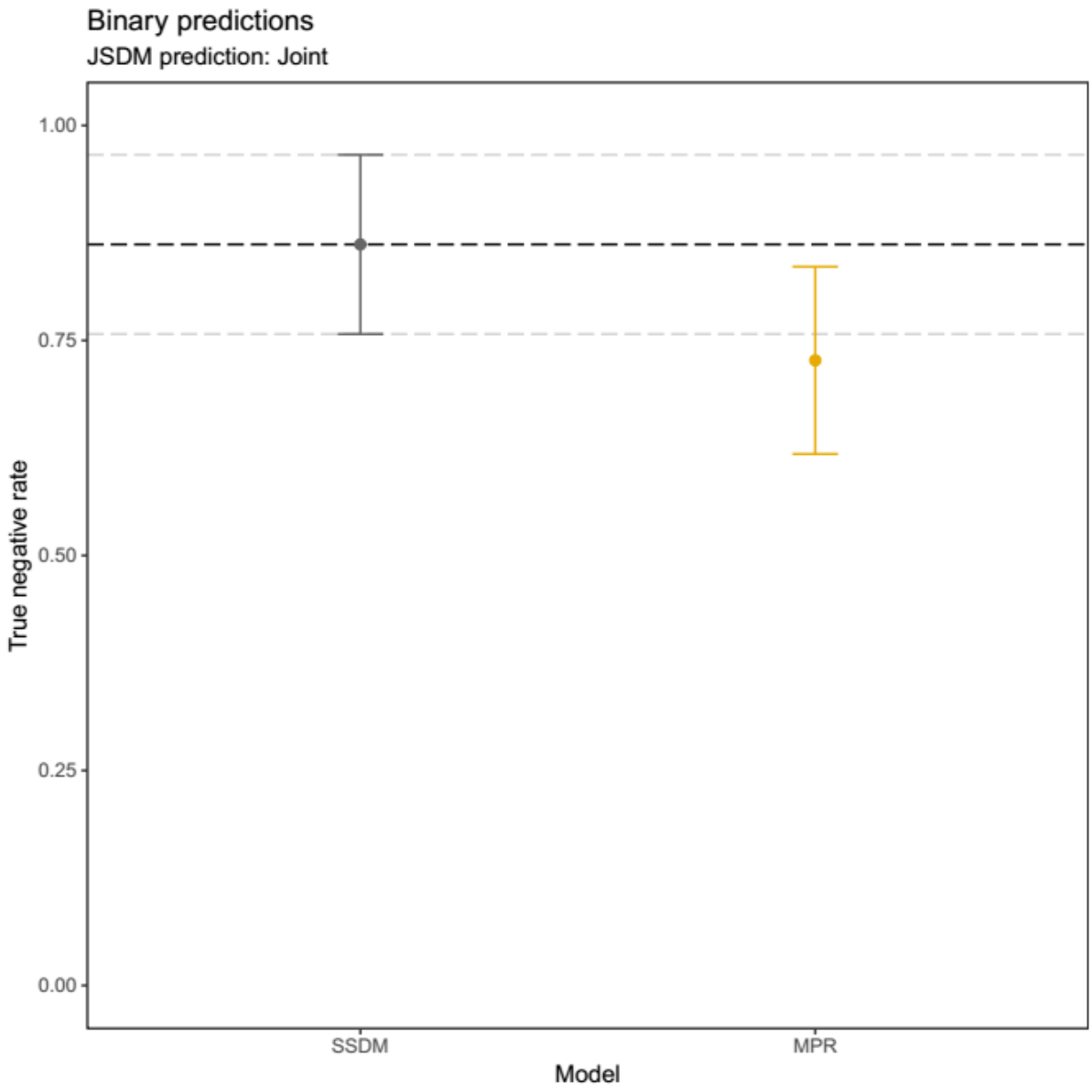


Figure 75: Performance of the SSDM and JSDM for the true negative rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

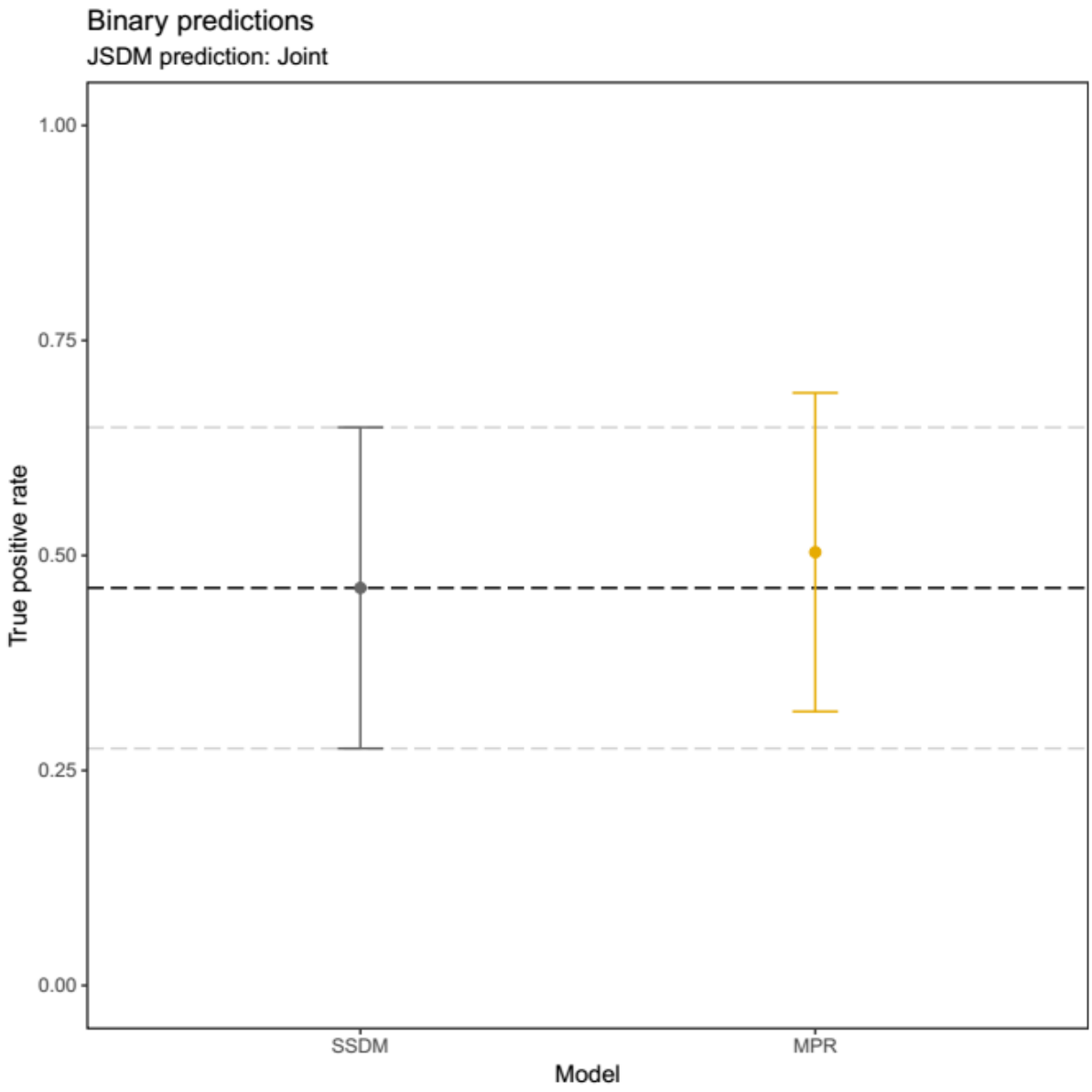


Figure 76: Performance of the SSDM and JSDM for the true positive rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

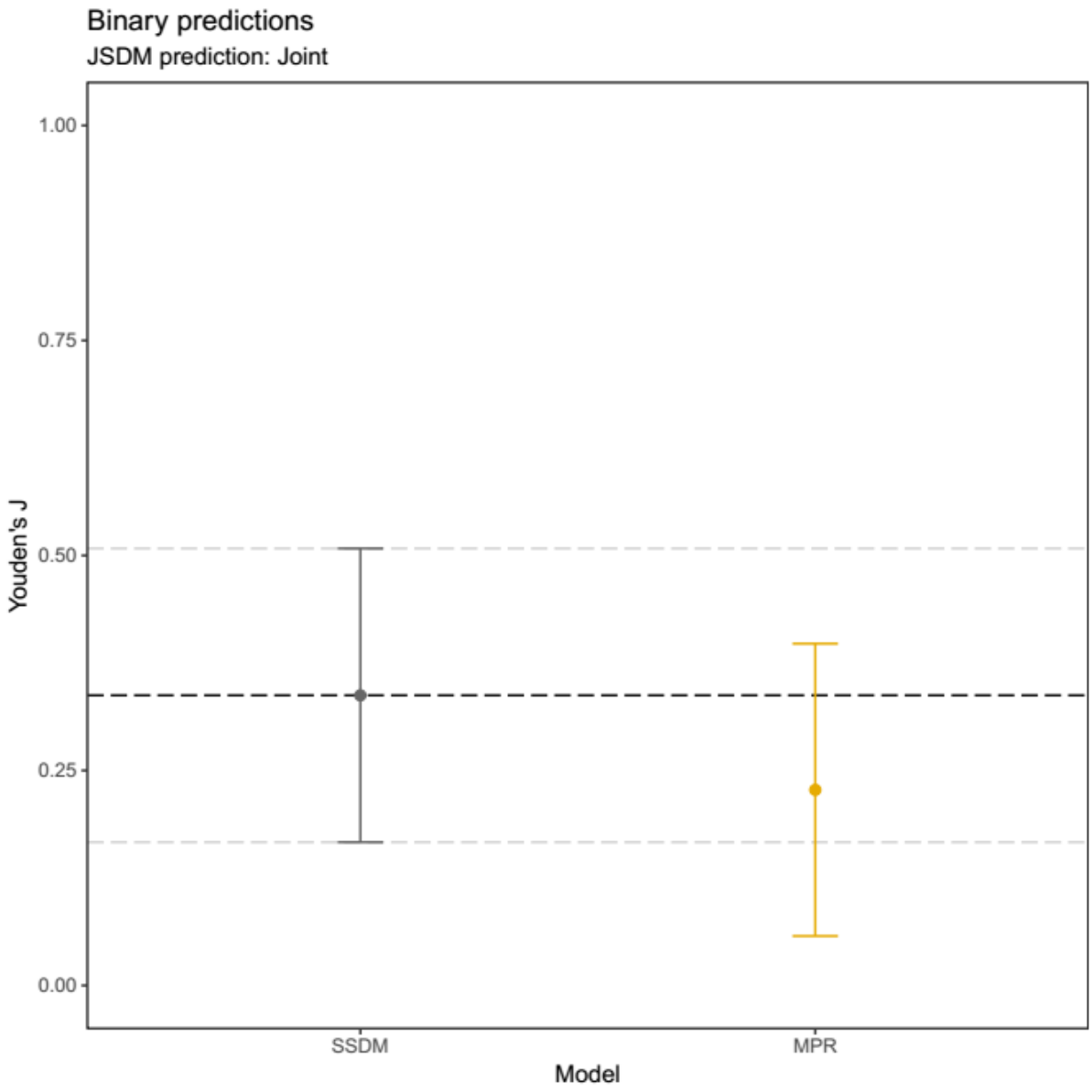


Figure 77: Performance of the SSDM and JSDM for the Youden's J evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

4. Conditional (low-prevalence) predictions

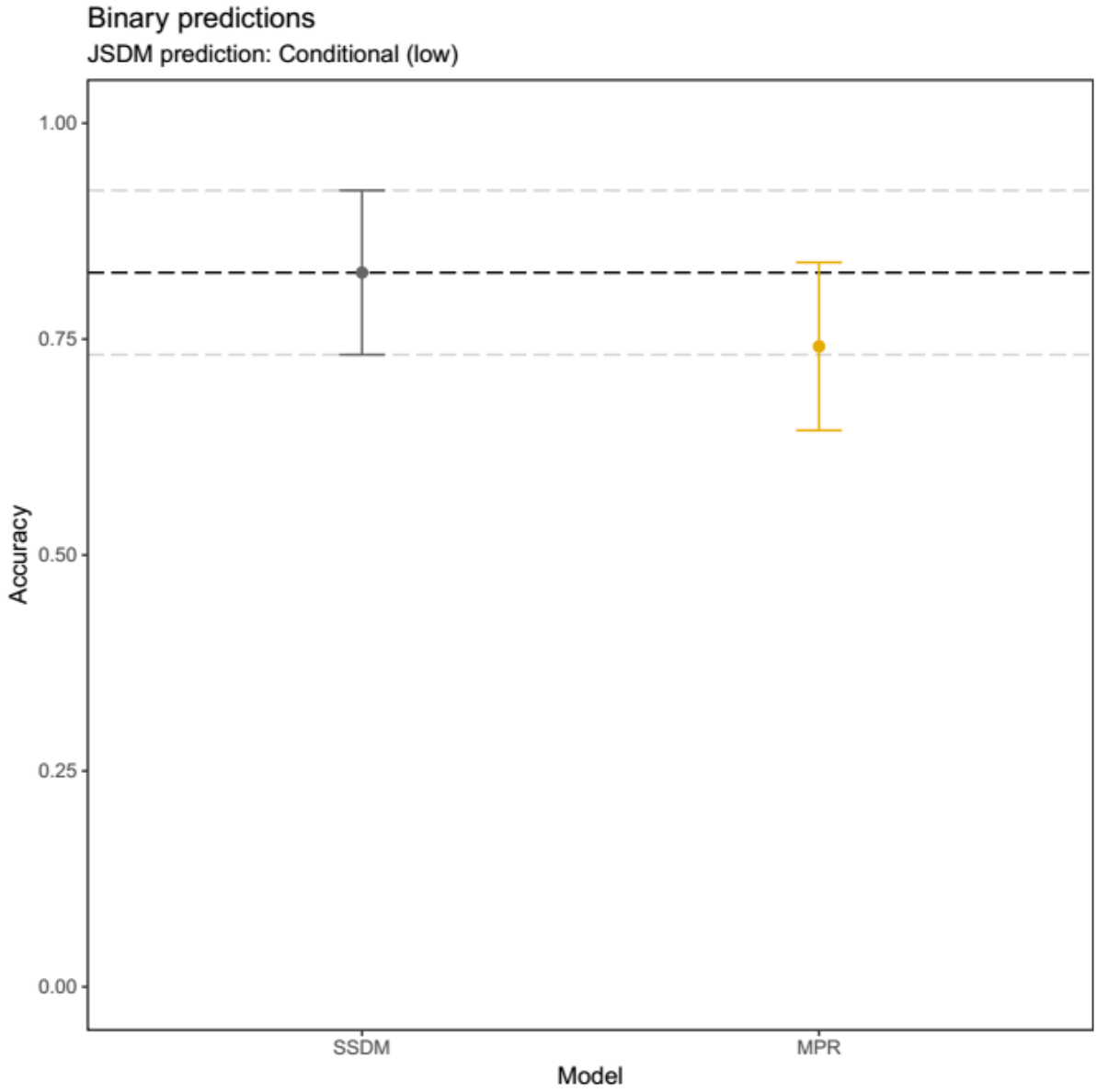


Figure 78: Performance of the SSDM and JSDM for the accuracy evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

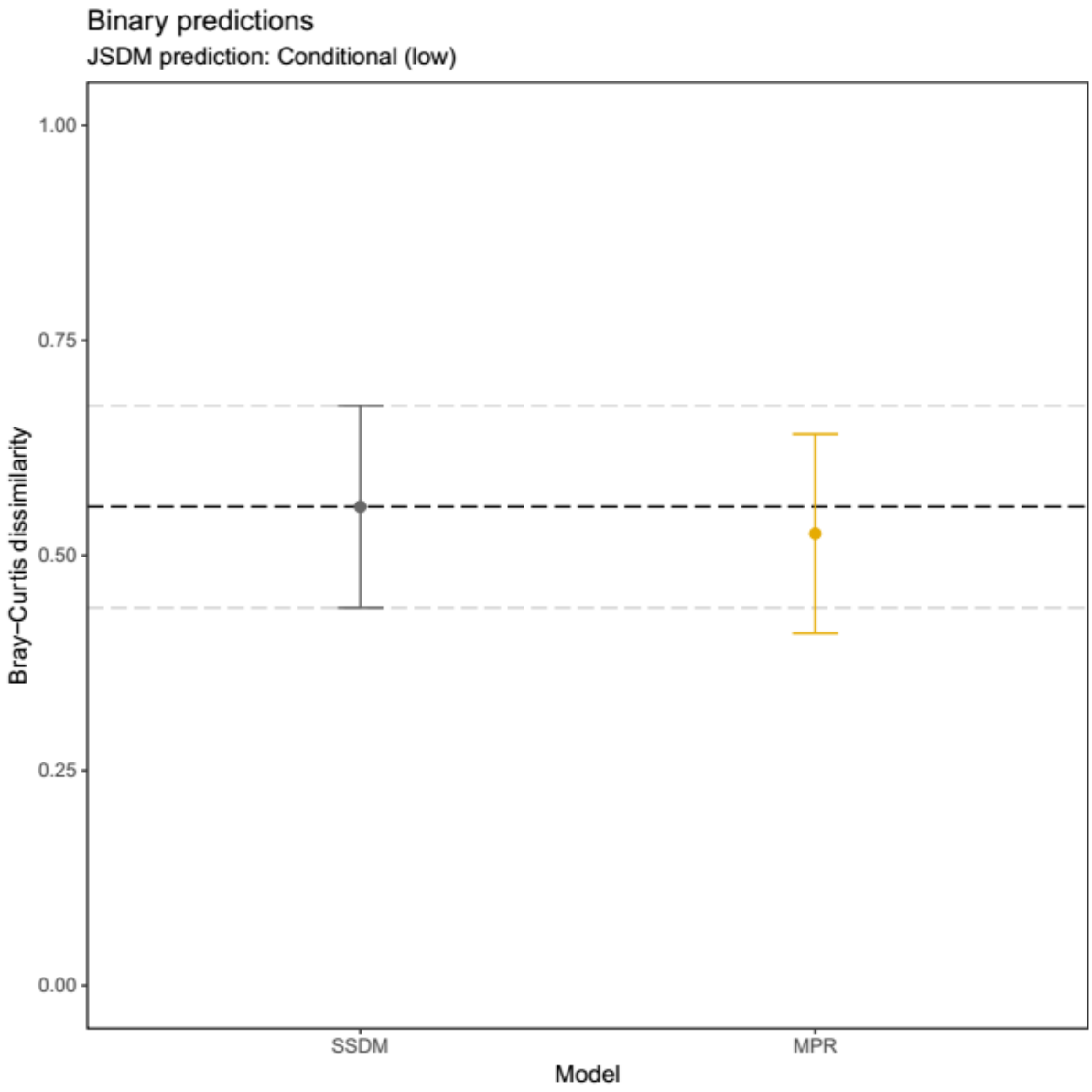


Figure 79: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

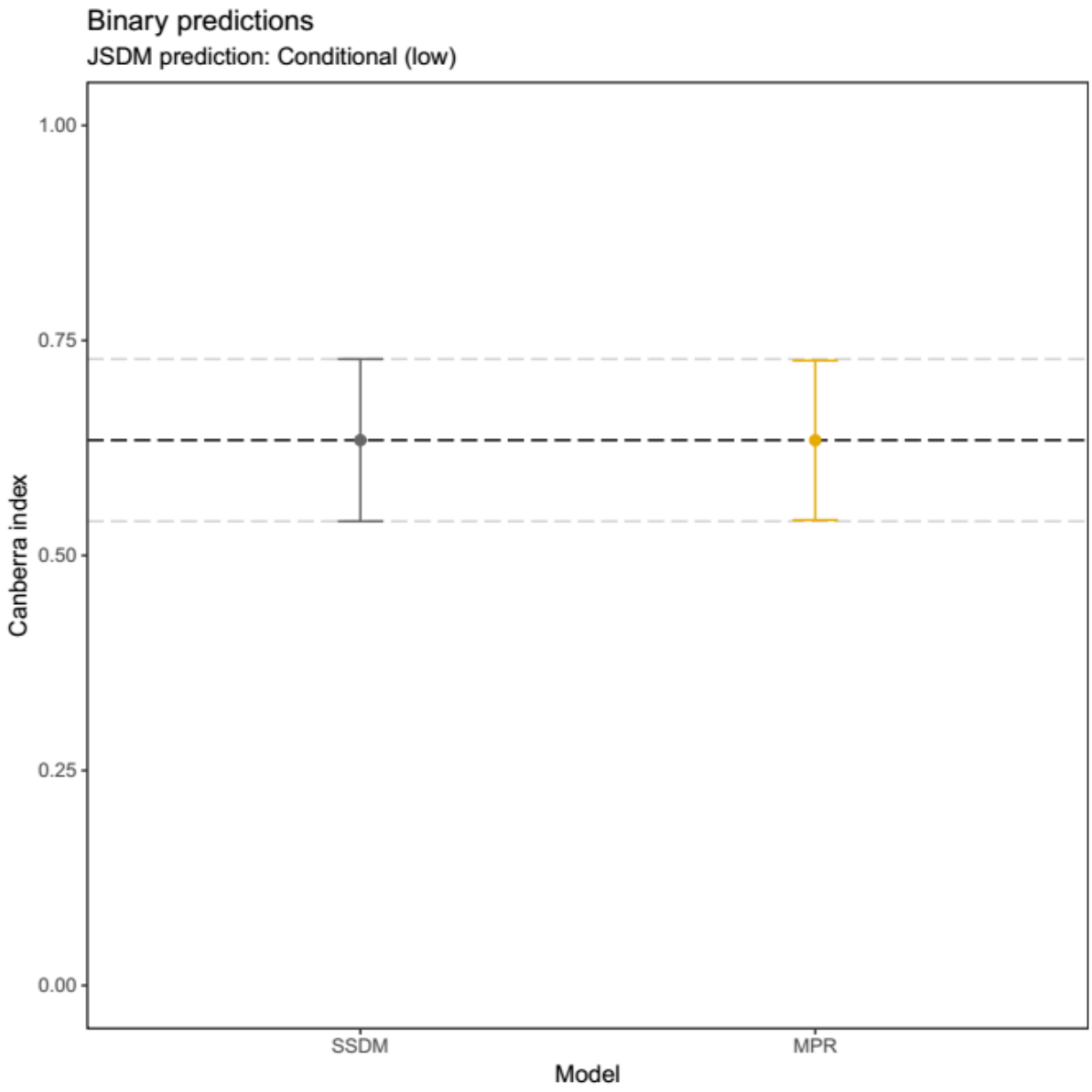


Figure 80: Performance of the SSDM and JSDM for the Canberra index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

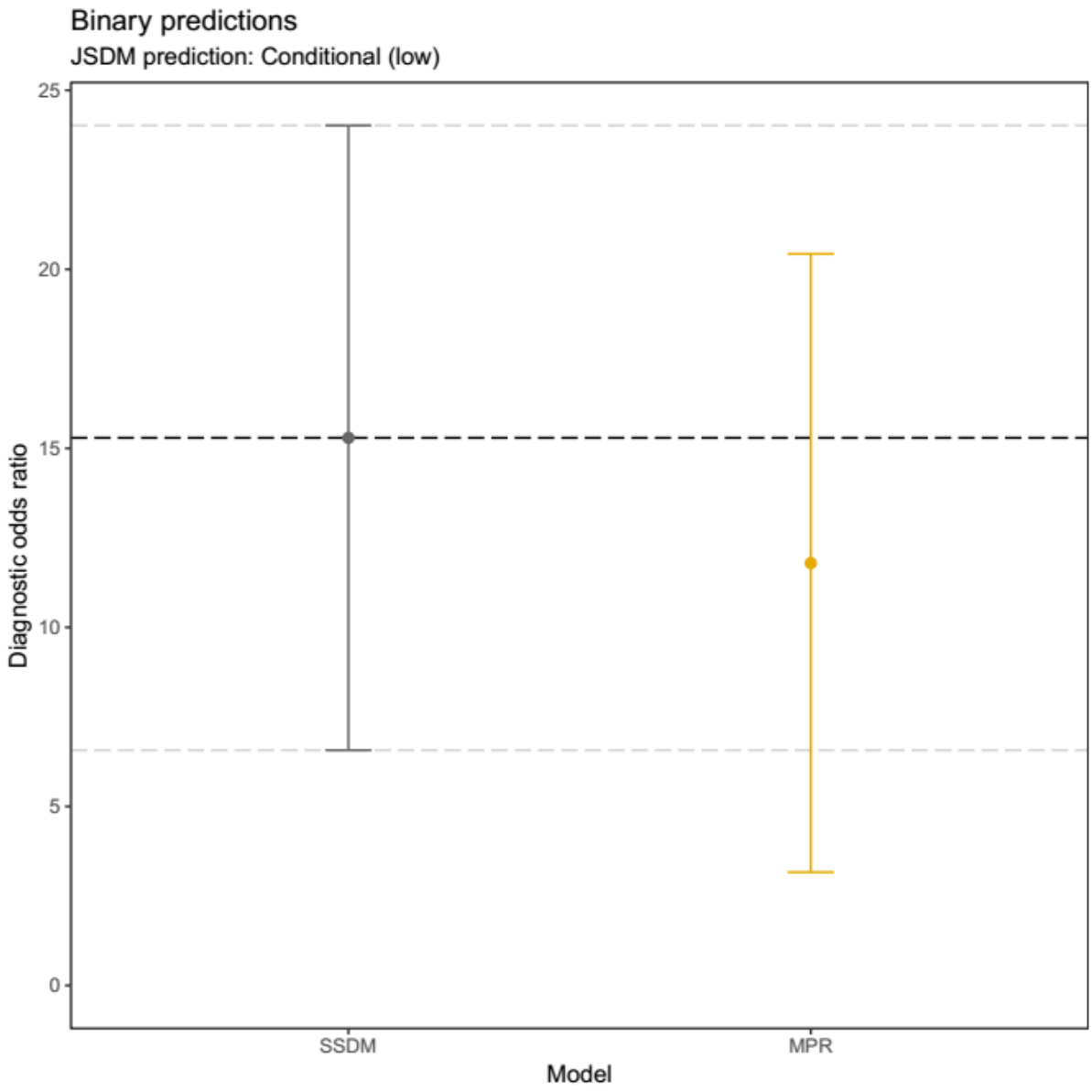


Figure 81: Performance of the SSDM and JSDM for the diagnostic odds ratio evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

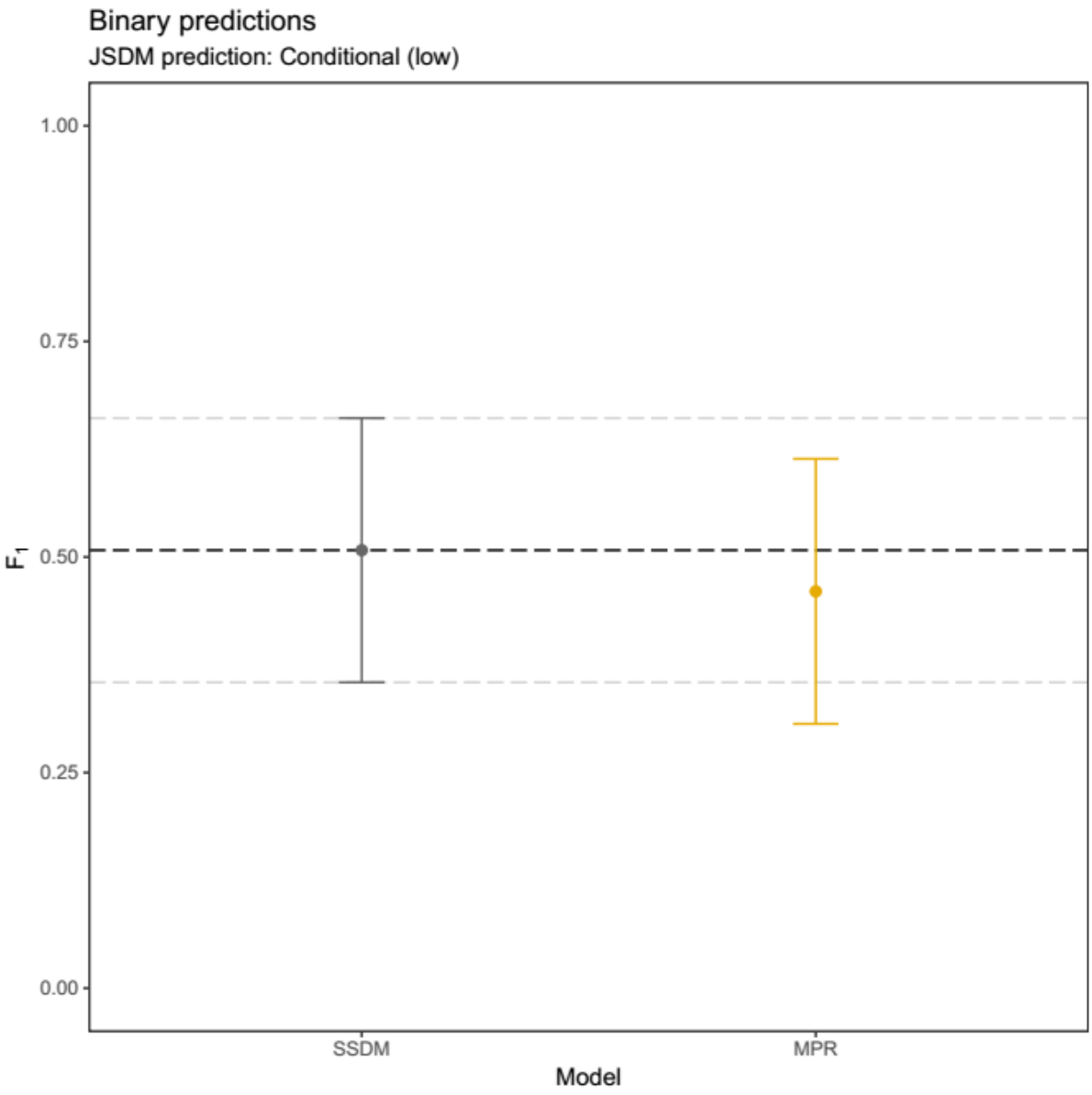


Figure 82: Performance of the SSDM and JSDM for the F_1 evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

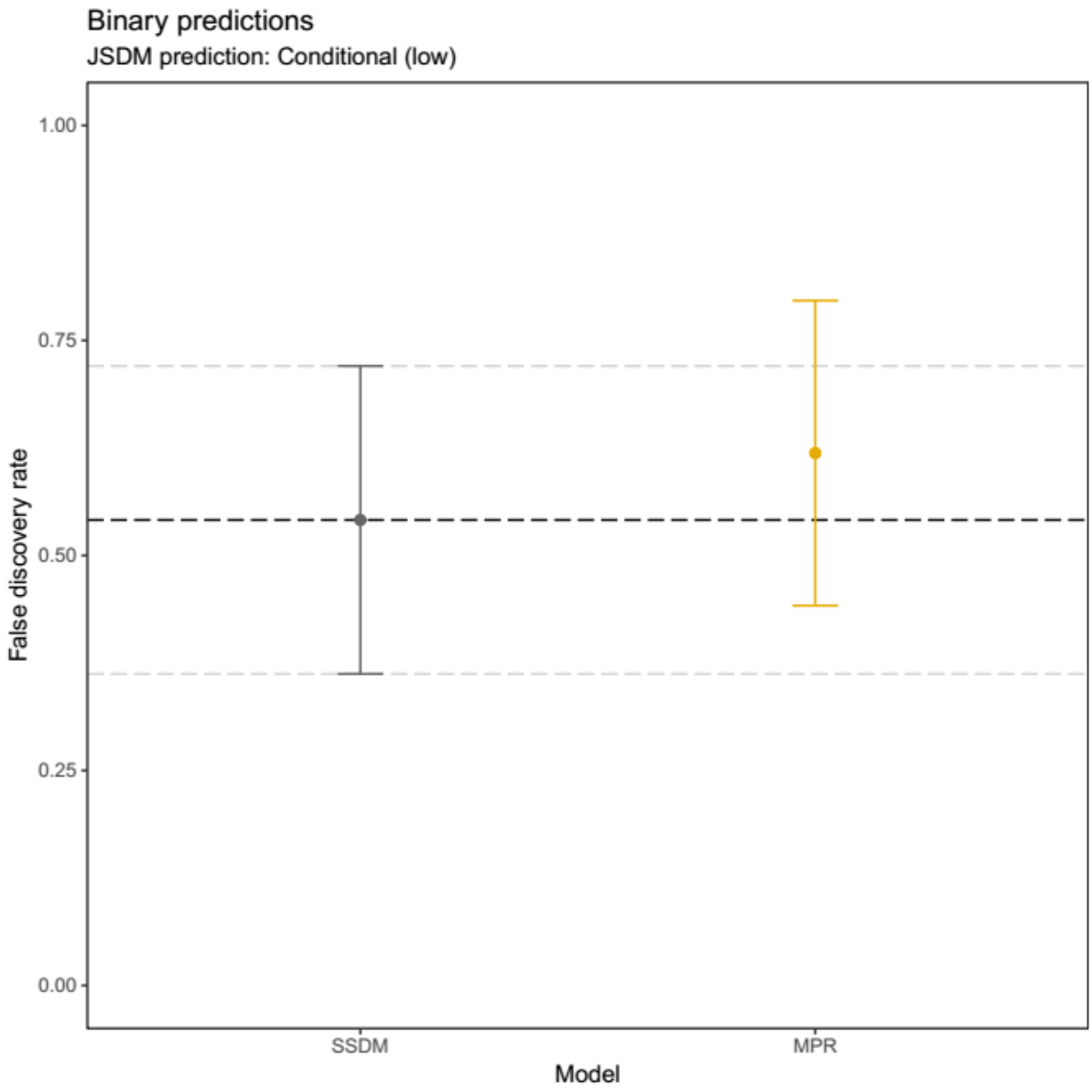


Figure 83: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

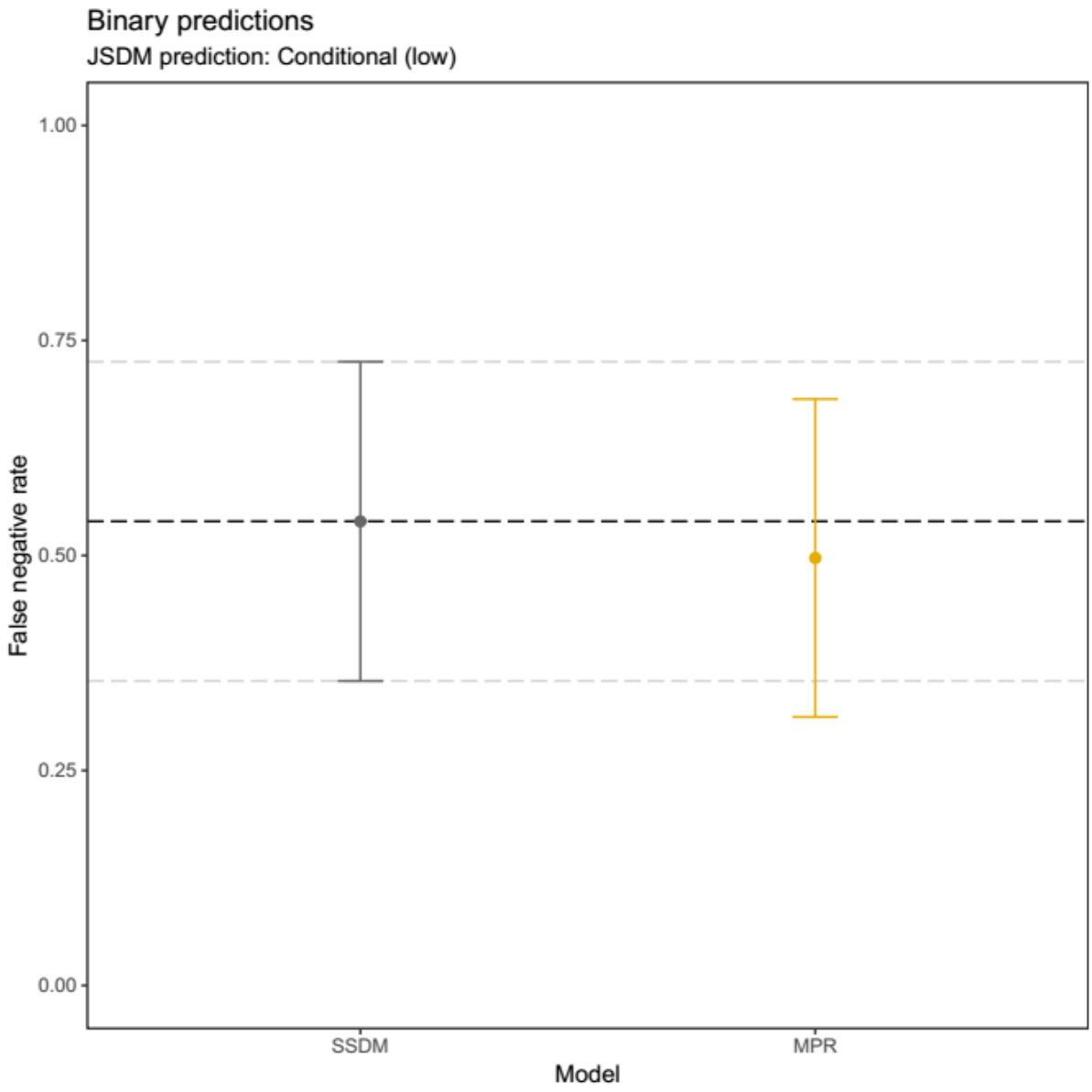


Figure 84: Performance of the SSDM and JSDM for the false negative rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

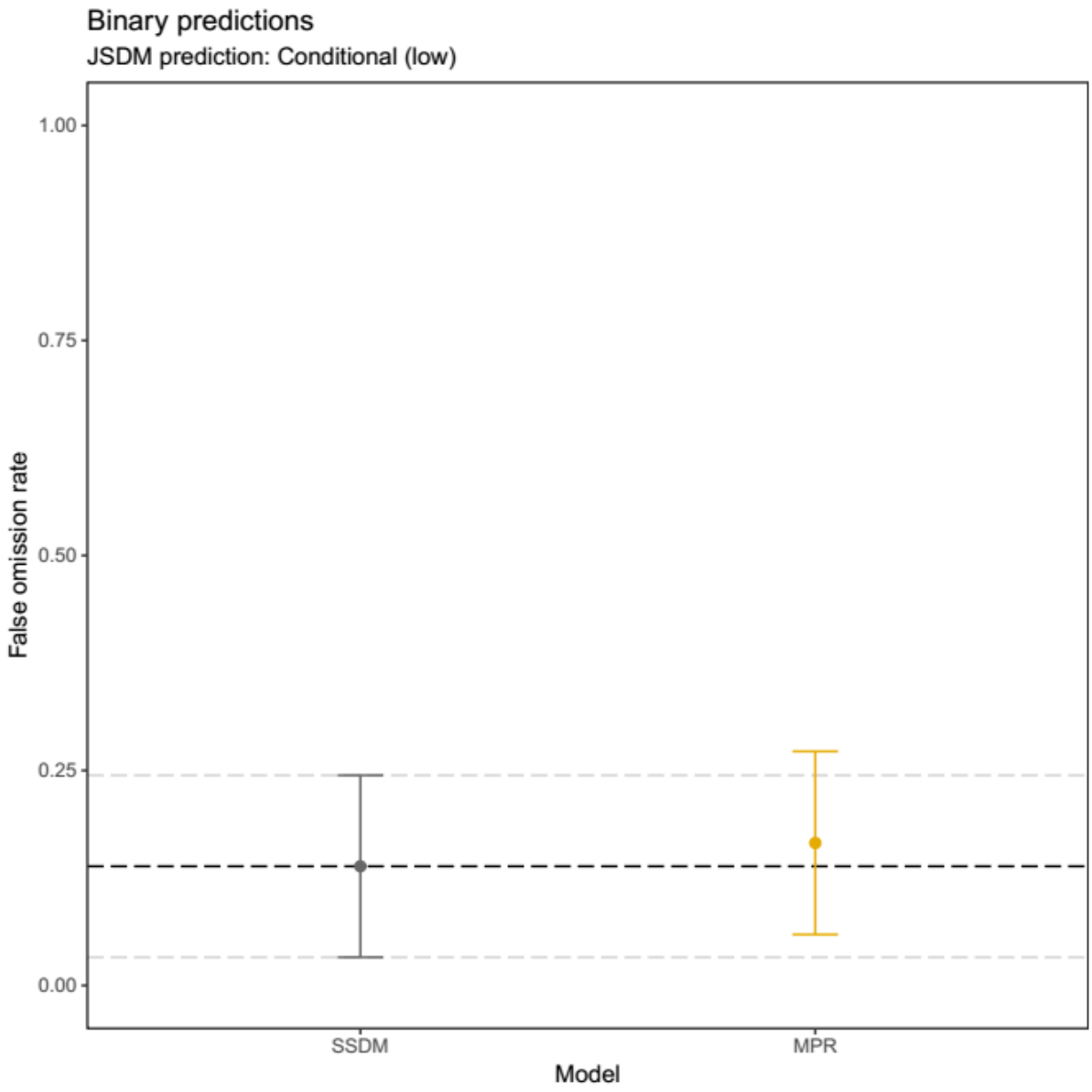


Figure 85: Performance of the SSDM and JSDM for the false omission rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

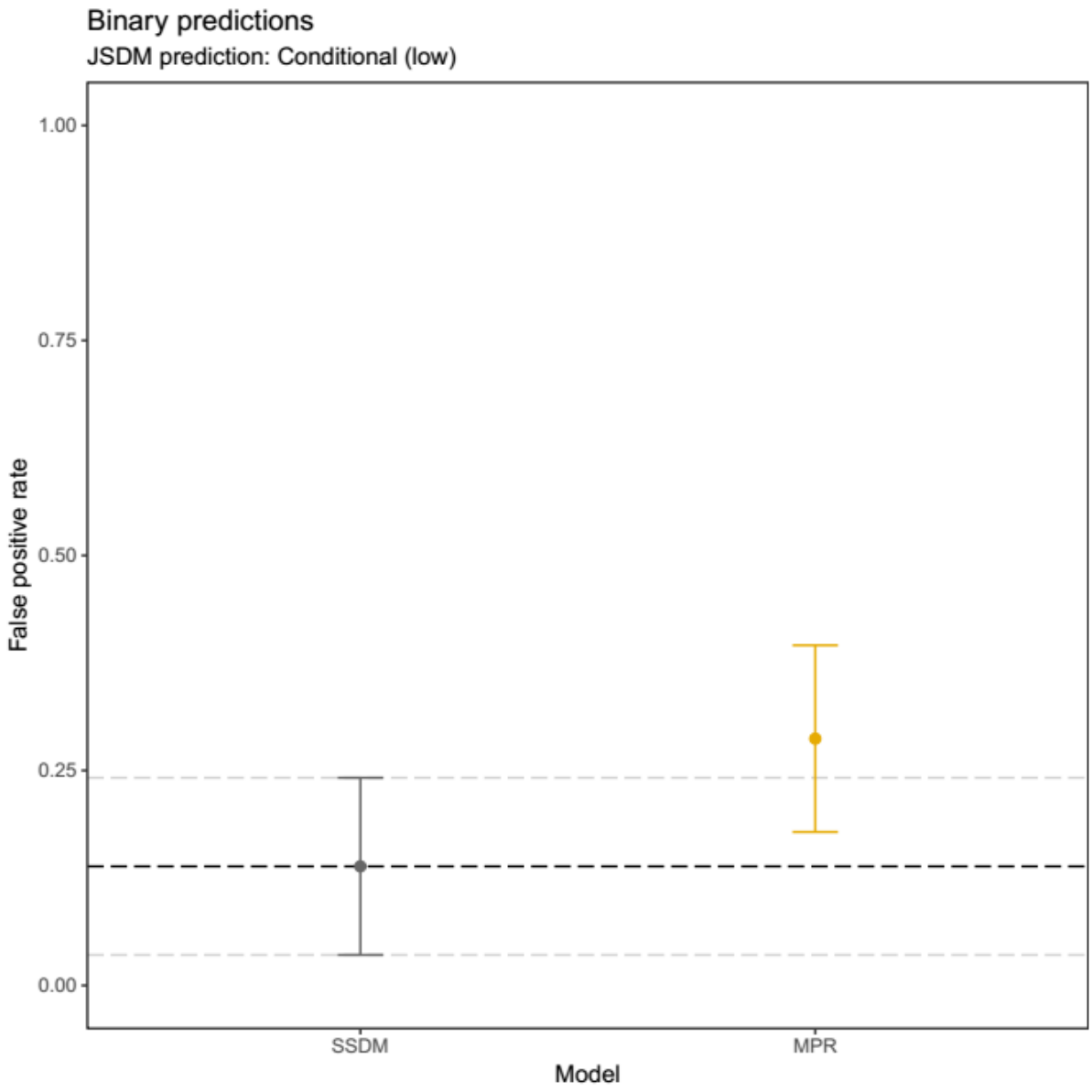


Figure 86: Performance of the SSDM and JSDM for the false positive rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

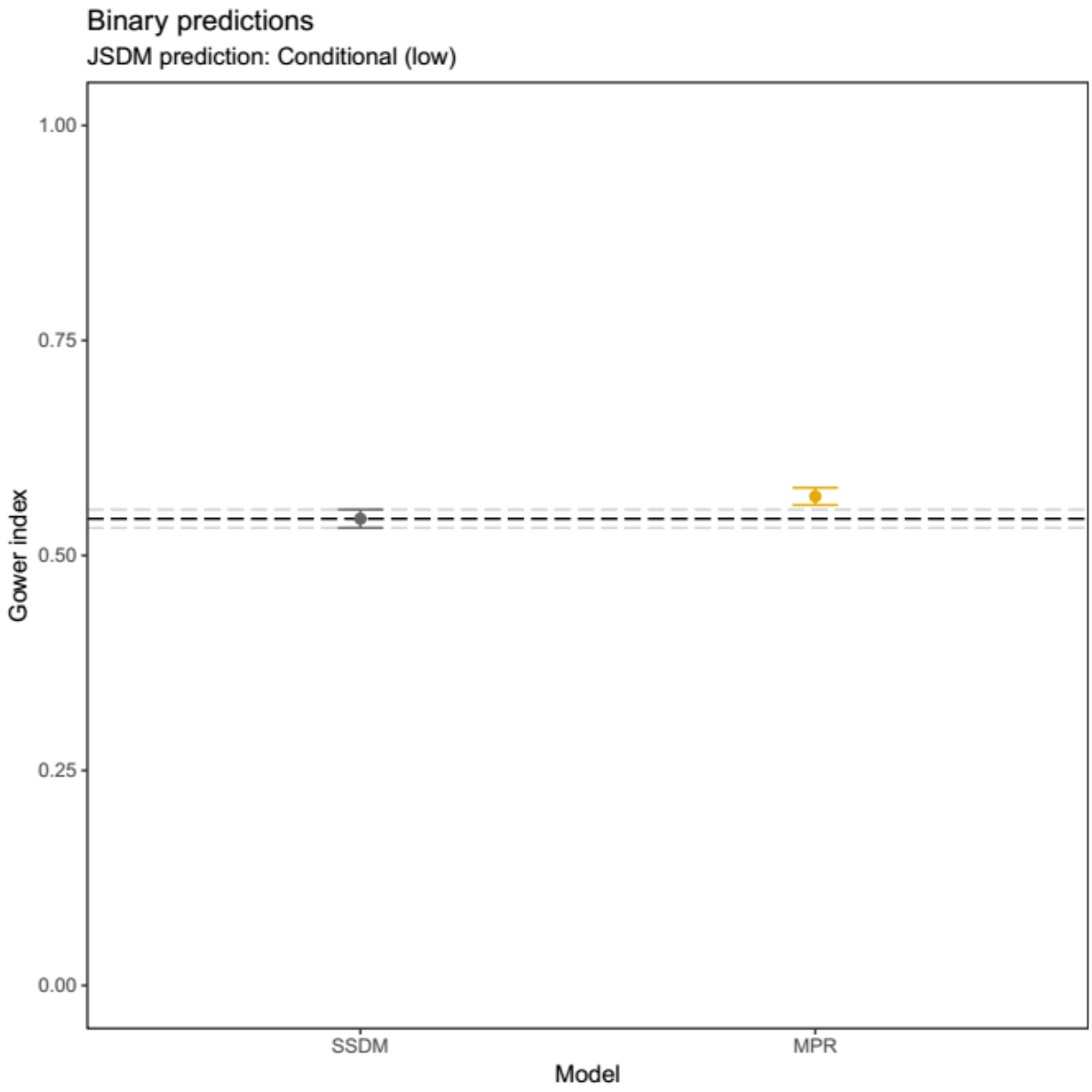


Figure 87: Performance of the SSDM and JSDM for the Gower index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

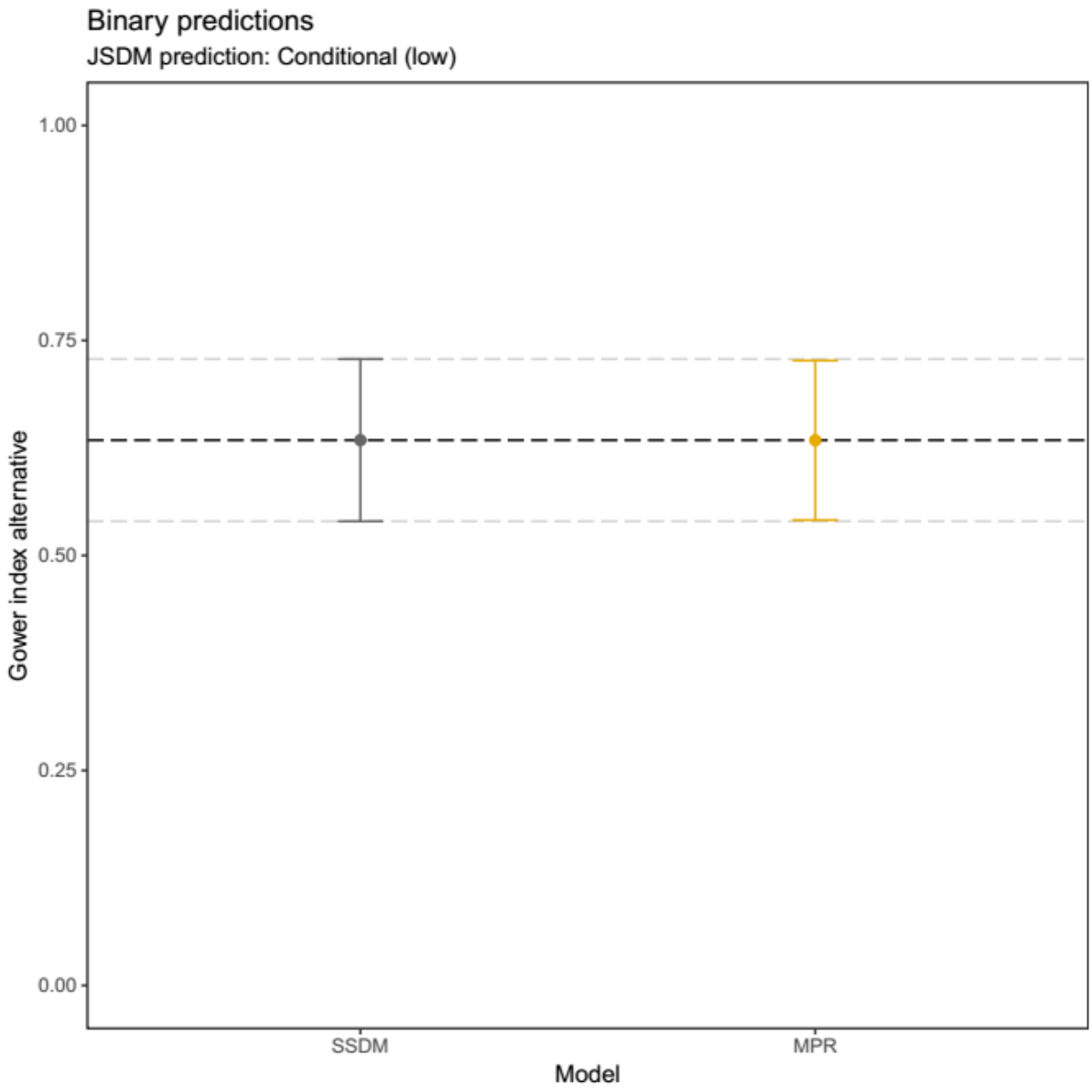


Figure 88: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

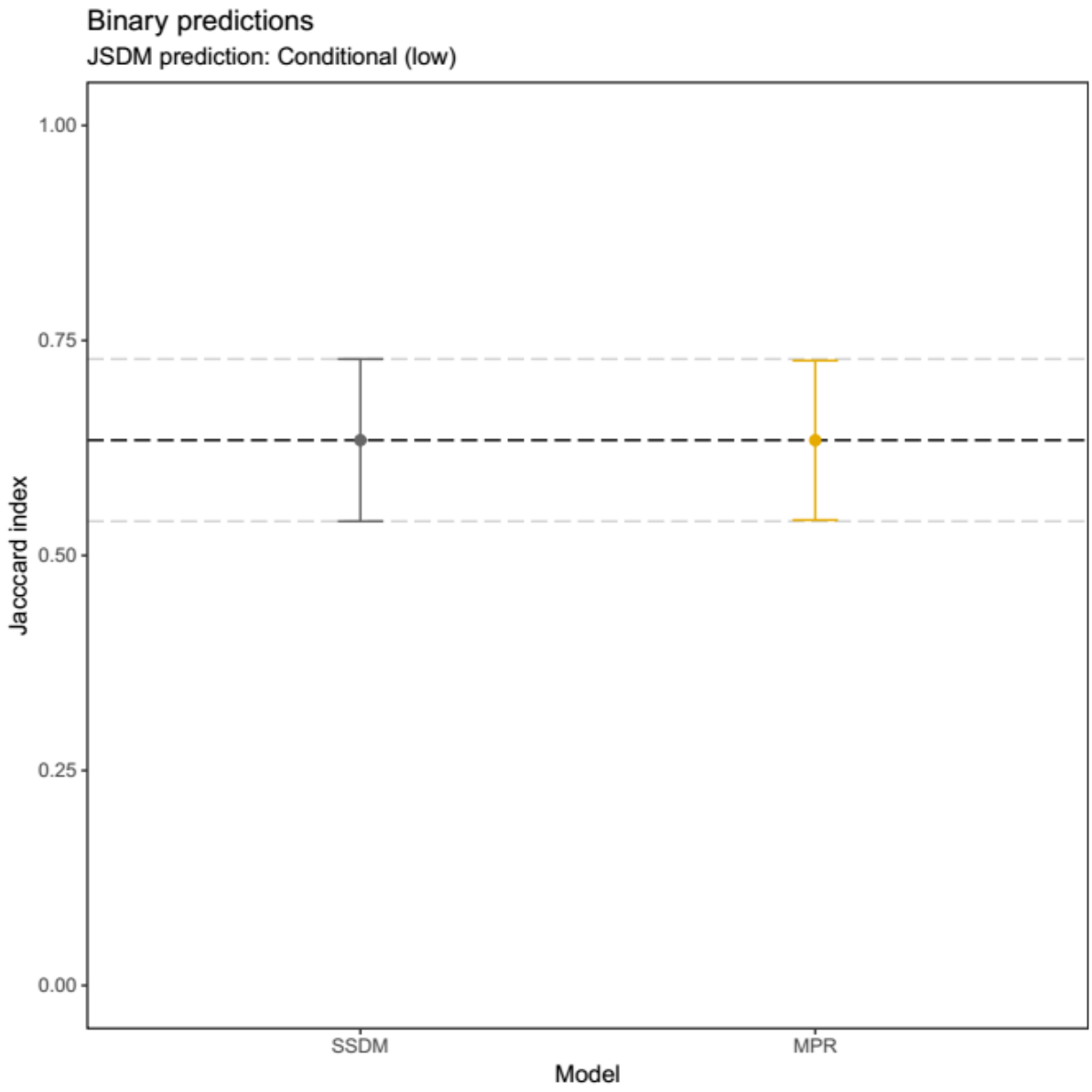


Figure 89: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

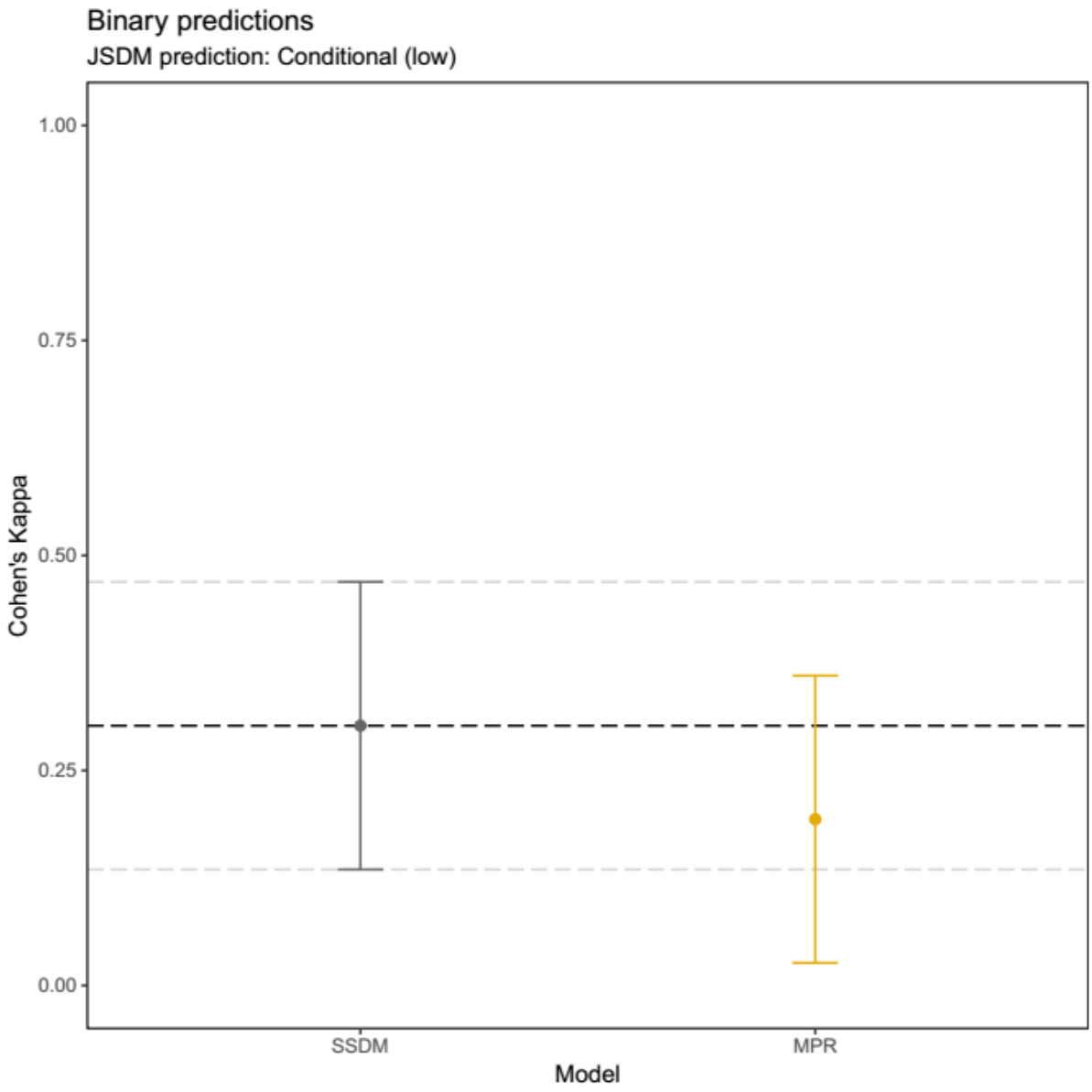


Figure 90: Performance of the SSDM and JSDM for the Cohen's Kappa evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

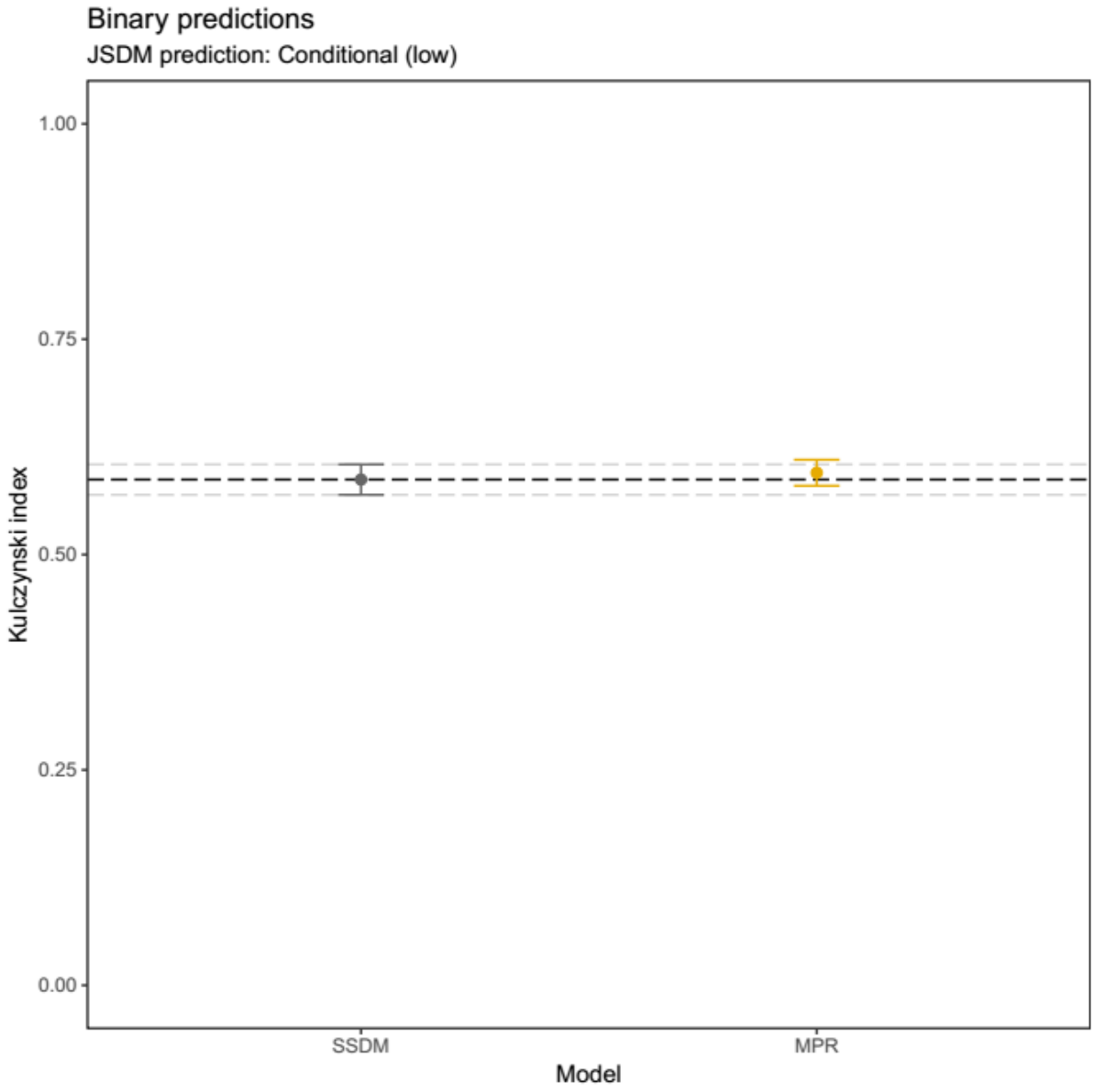


Figure 91: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

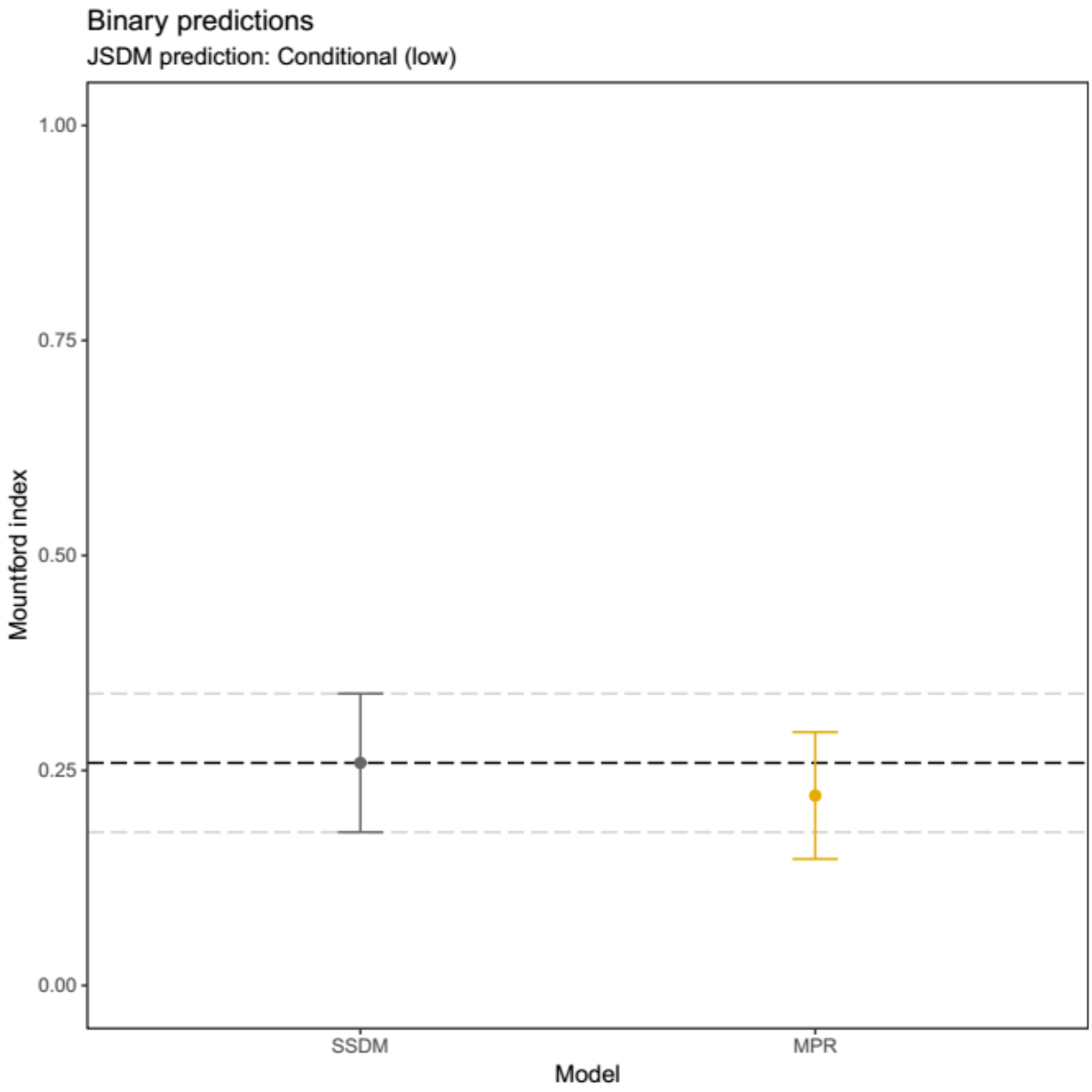


Figure 92: Performance of the SSDM and JSDM for the Mountford index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

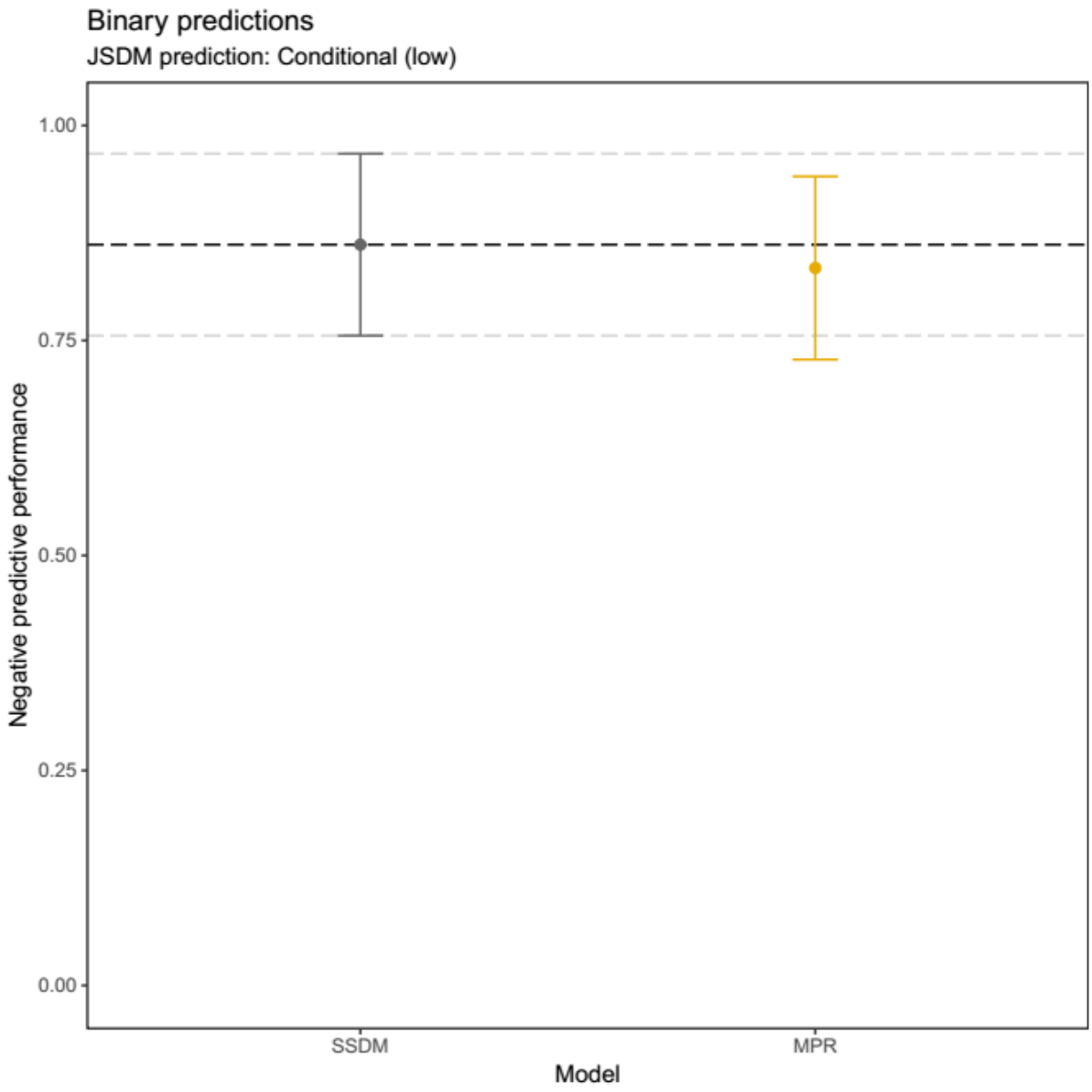


Figure 93: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

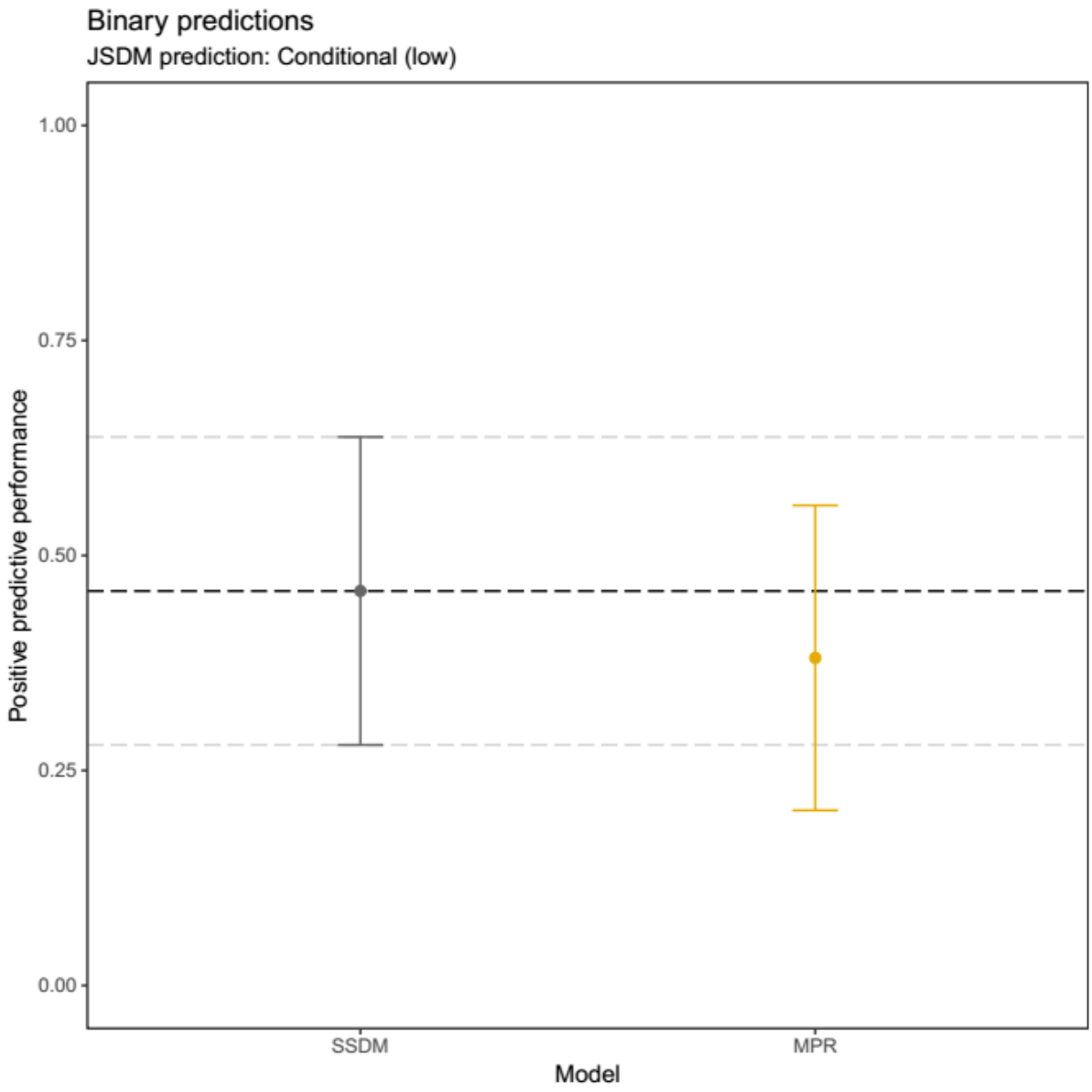


Figure 94: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

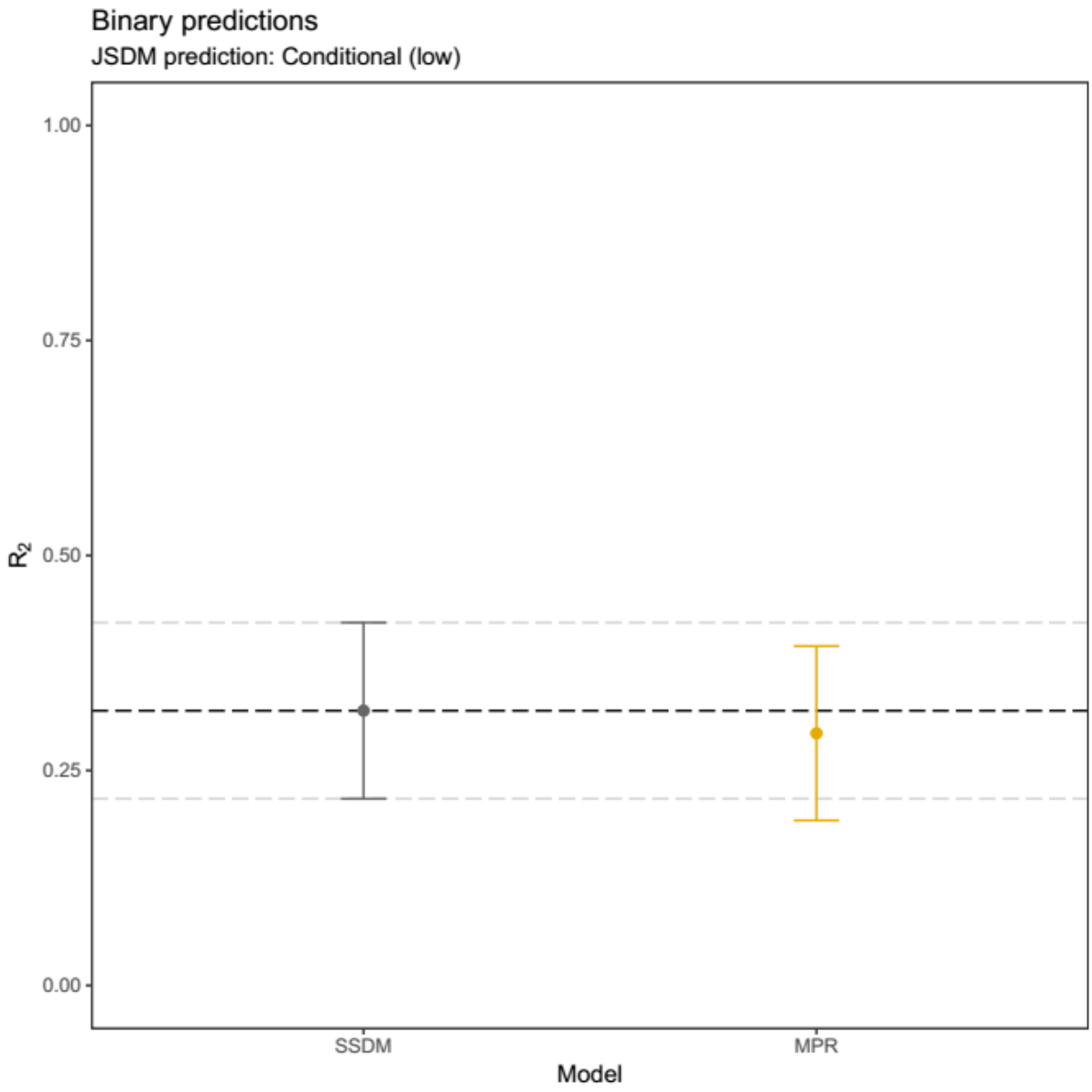


Figure 95: Performance of the SSDM and JSDM for the R^2 evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

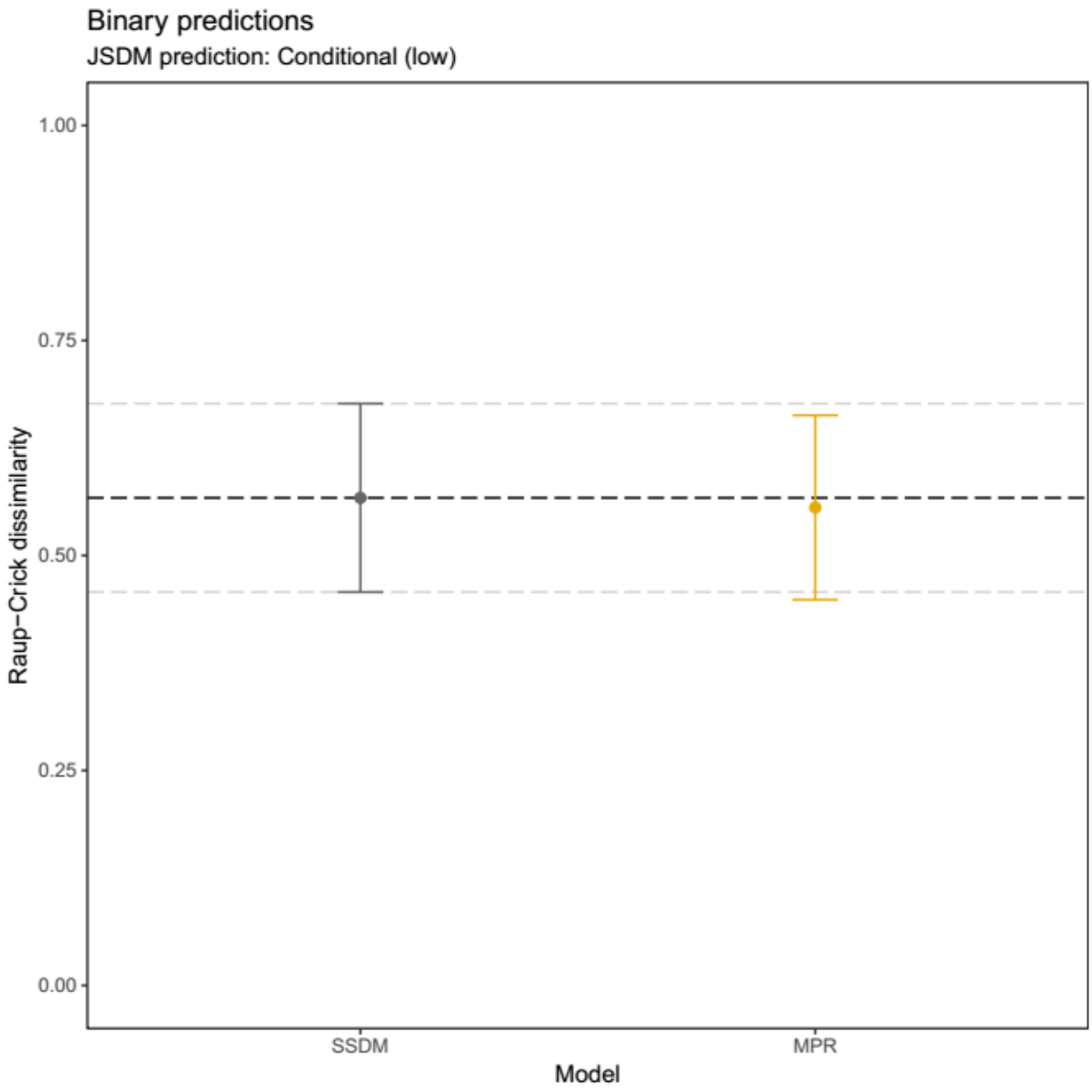


Figure 96: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

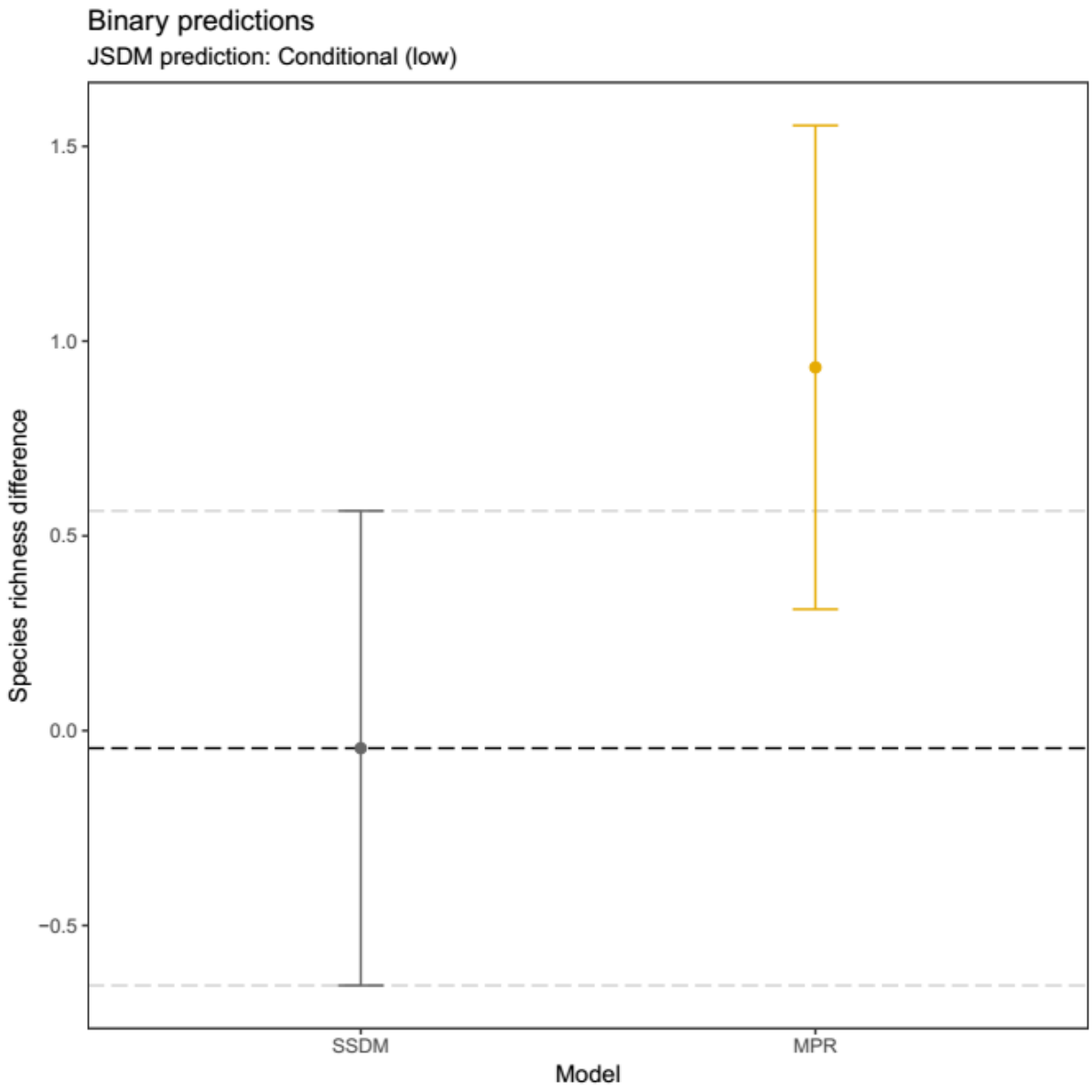


Figure 97: Performance of the SSDM and JSDM for the species richness difference evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

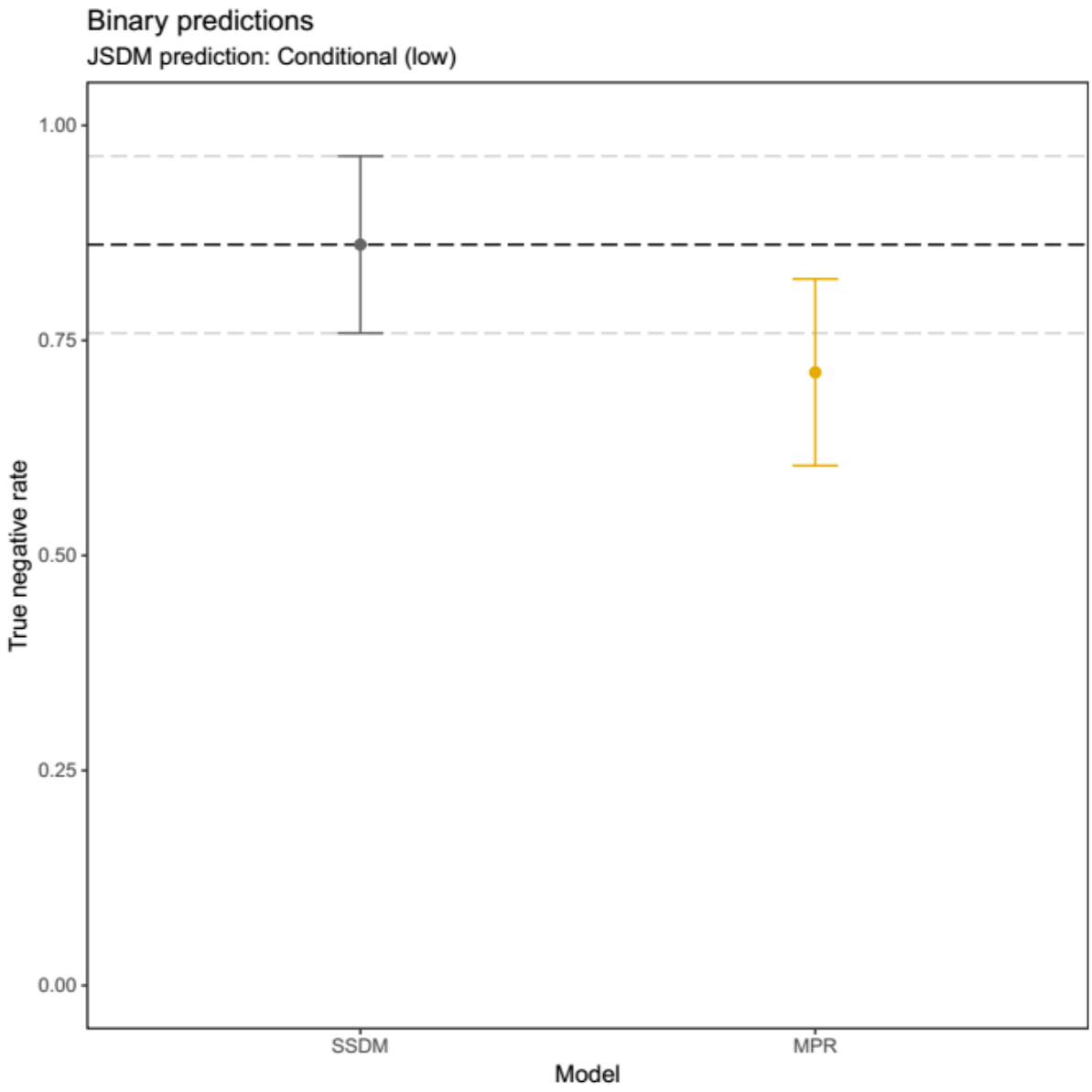


Figure 98: Performance of the SSDM and JSDM for the true negative rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

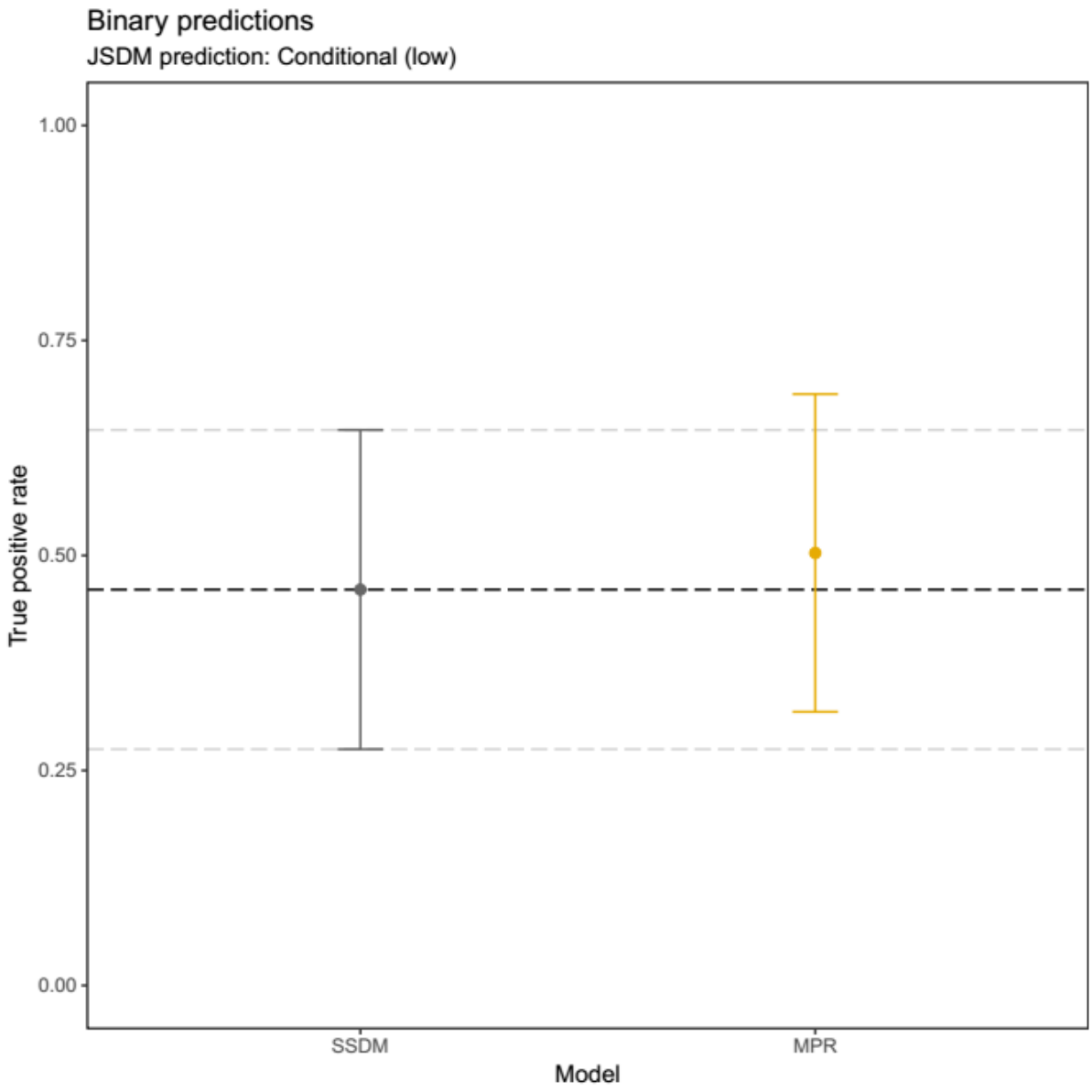


Figure 99: Performance of the SSDM and JSDM for the true positive rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

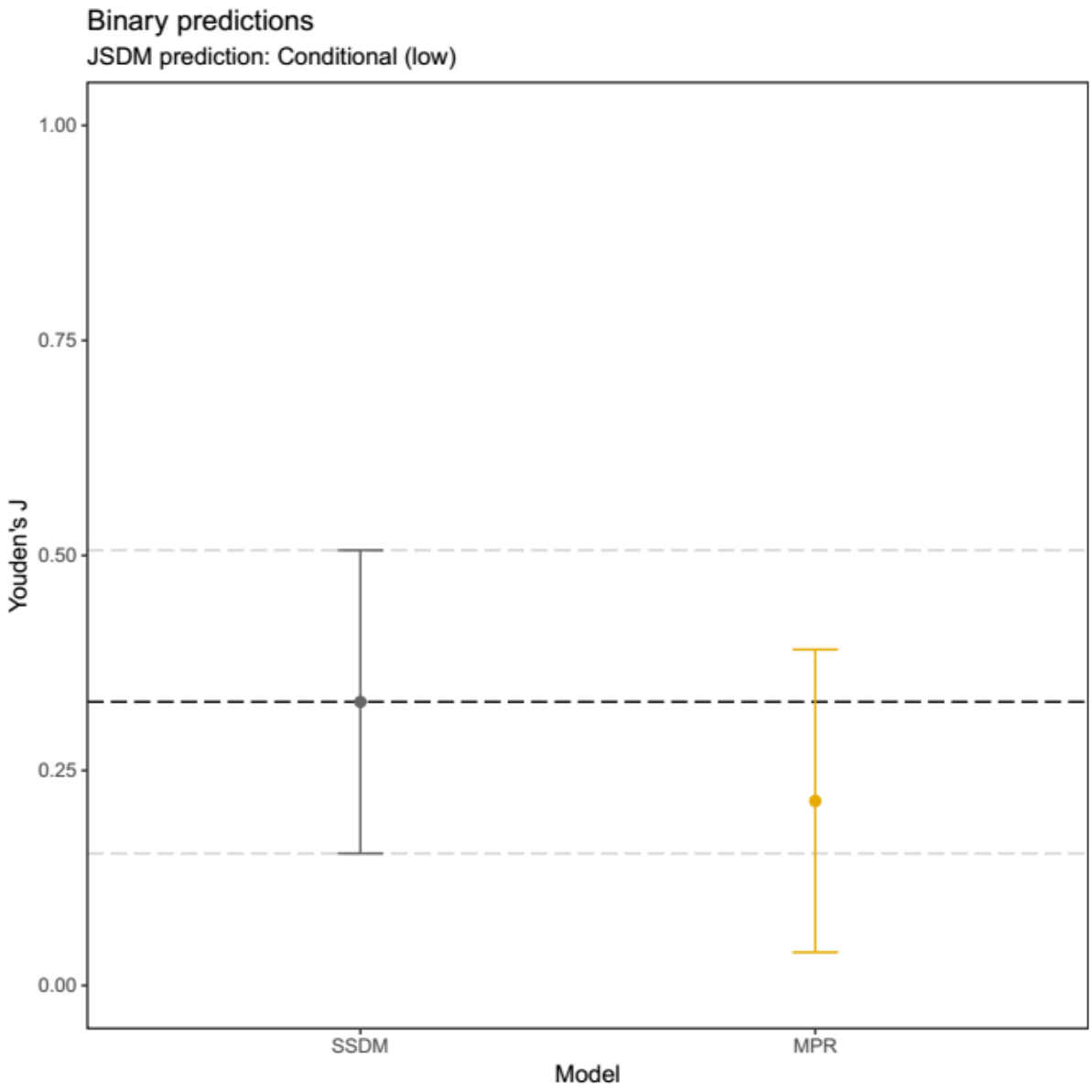


Figure 100: Performance of the SSDM and JSDM for the Youden's J evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

5. Conditional (medium-prevalence) predictions

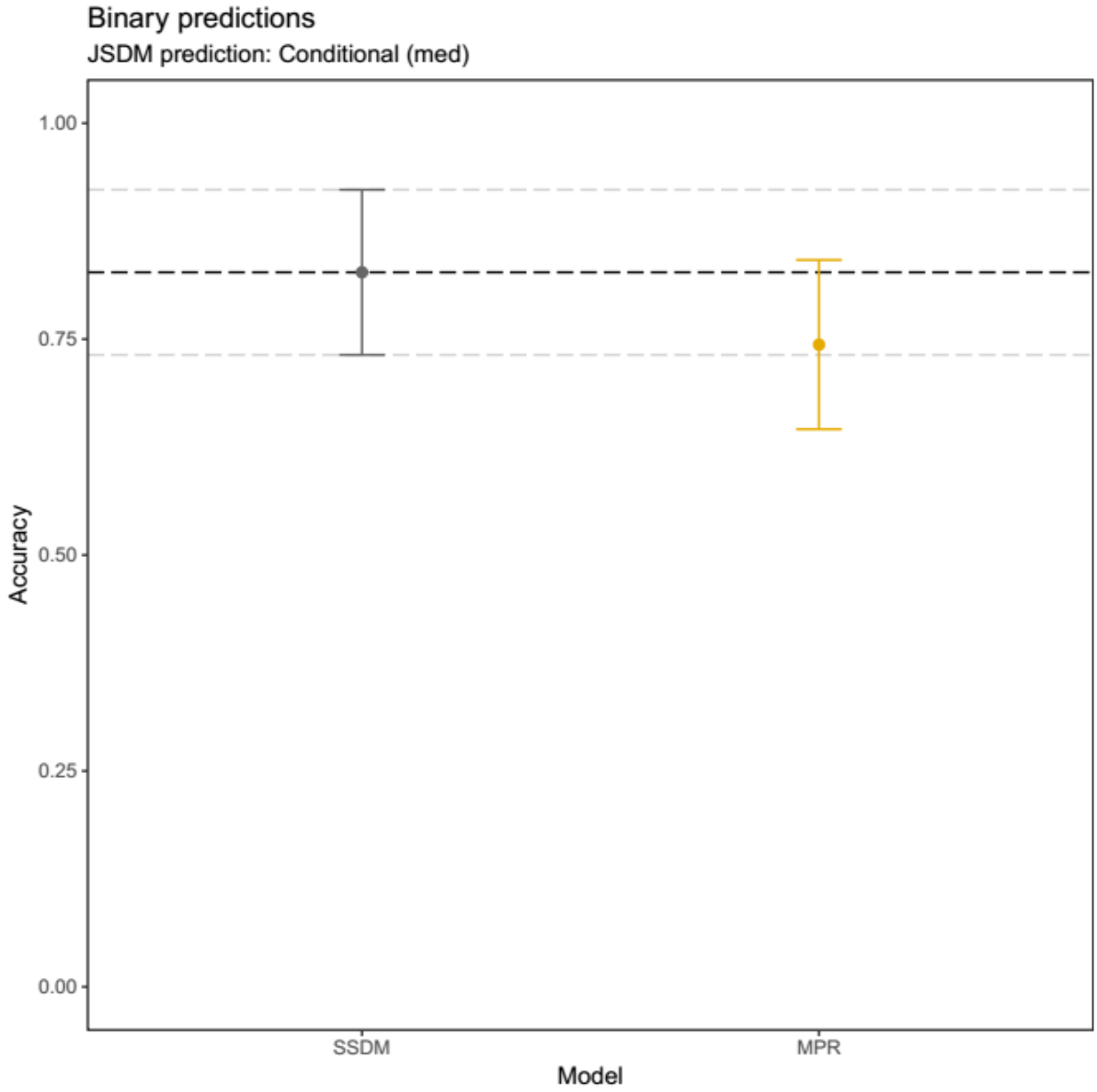


Figure 101: Performance of the SSDM and JSDM for the accuracy evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

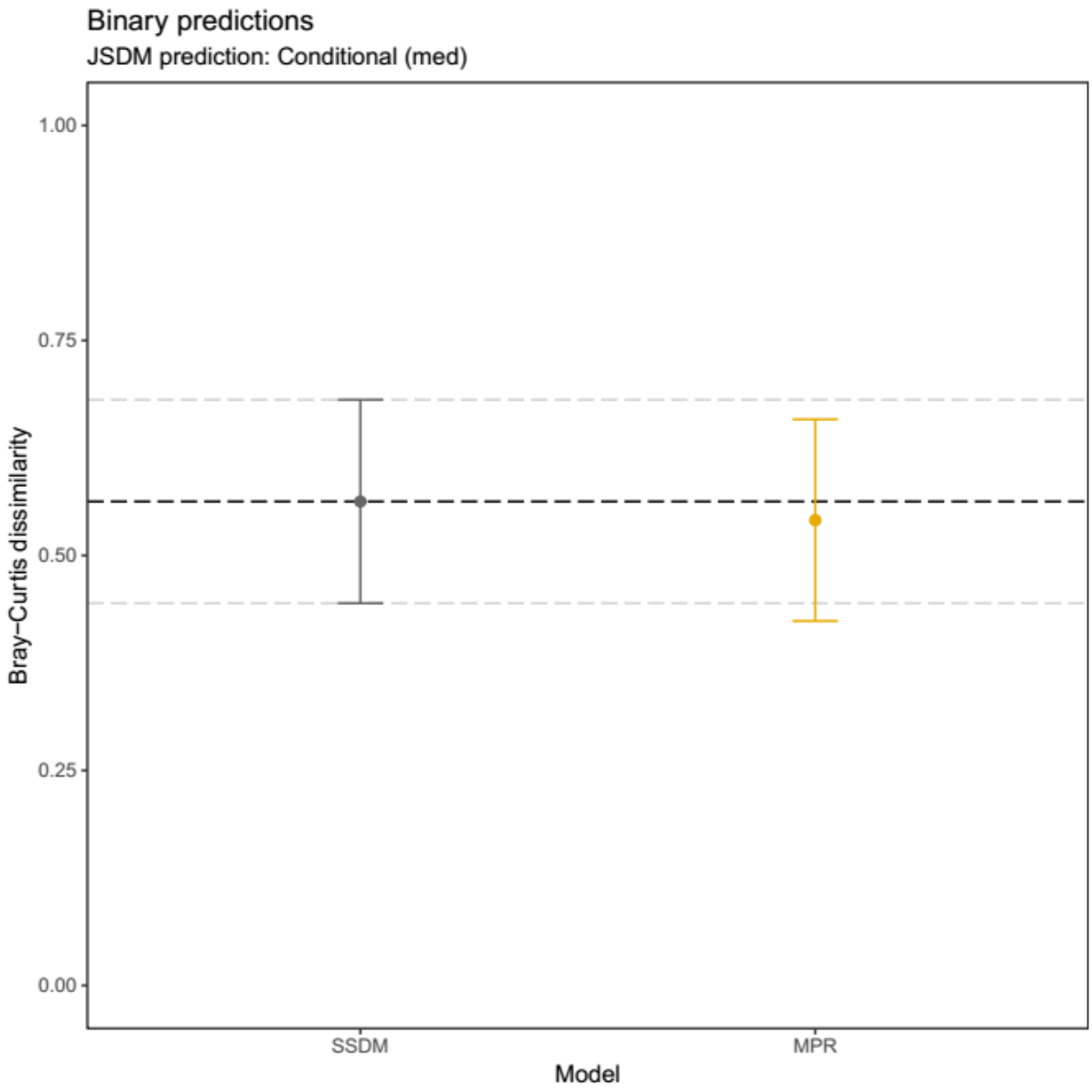


Figure 102: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

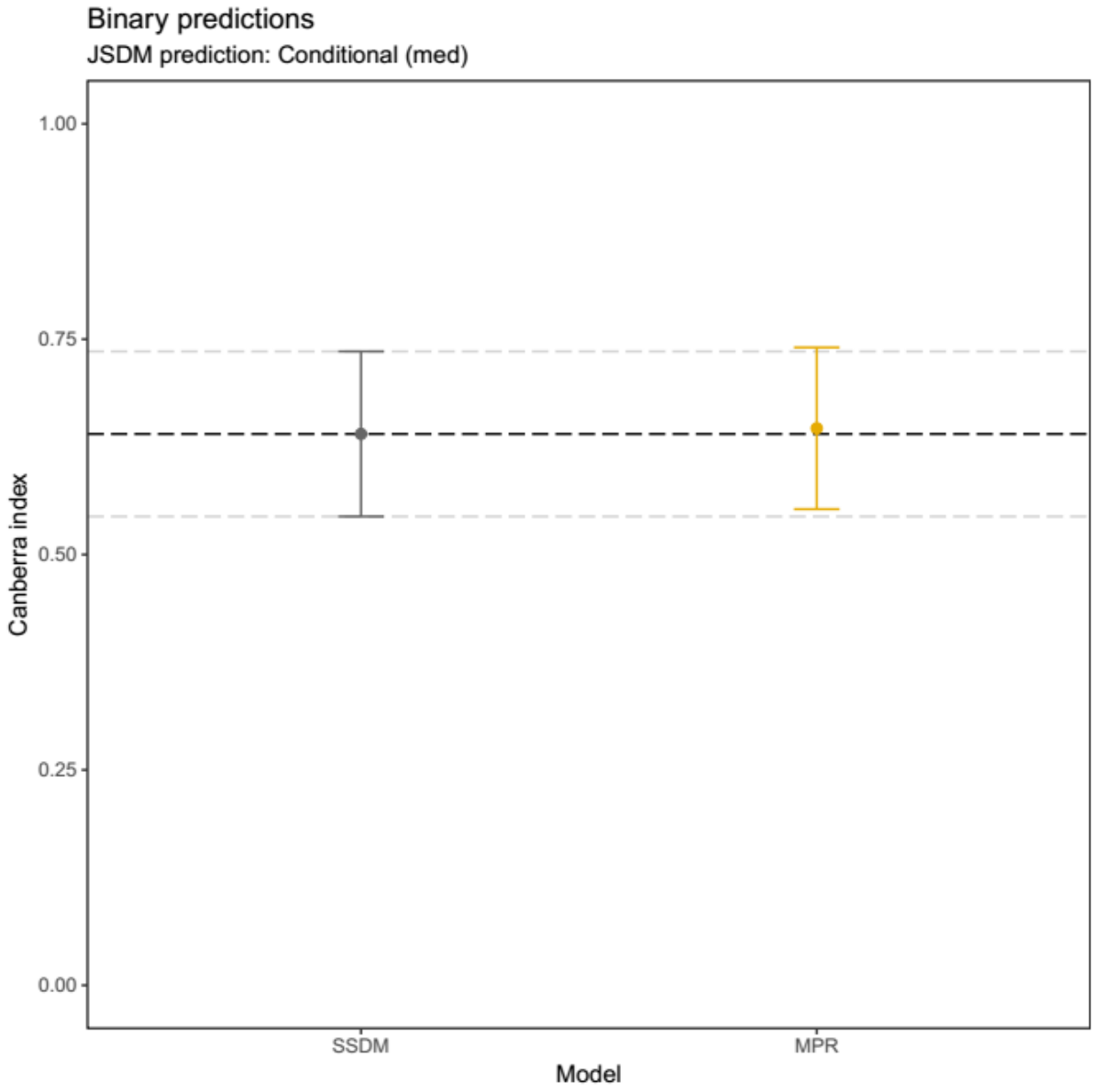


Figure 103: Performance of the SSDM and JSDM for the Canberra index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

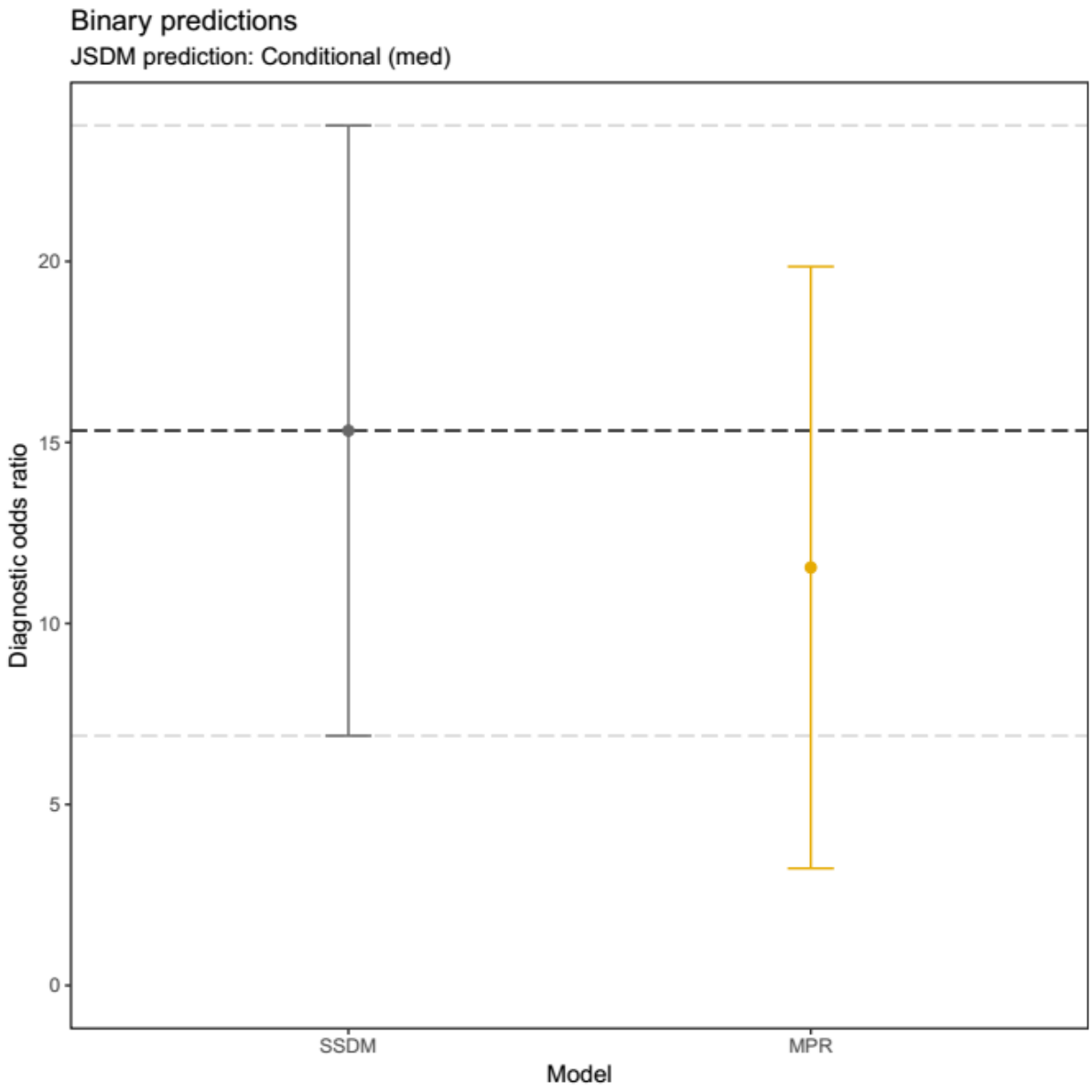


Figure 104: Performance of the SSDM and JSDM for the diagnostic odds ratio evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

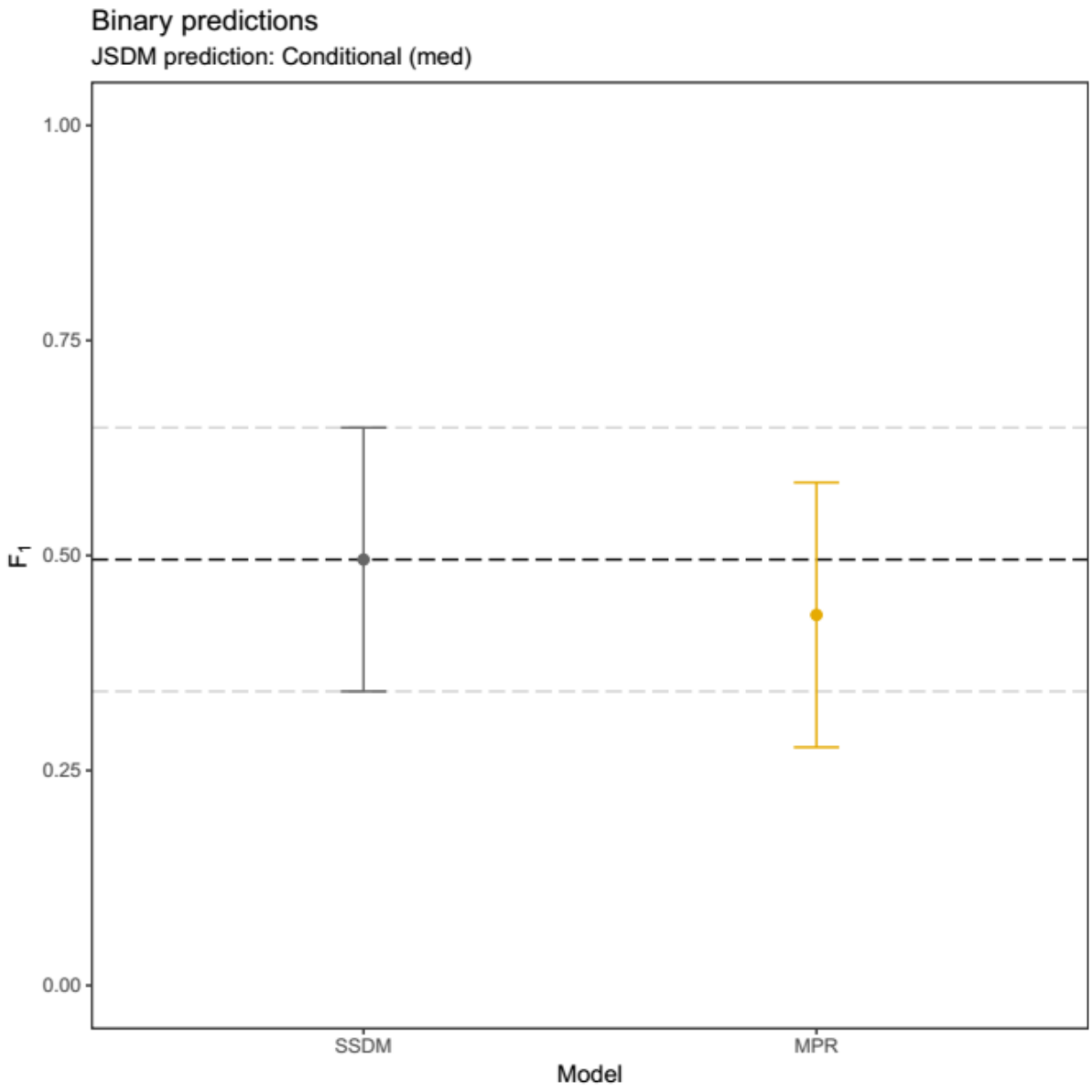


Figure 105: Performance of the SSDM and JSDM for the F_1 evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

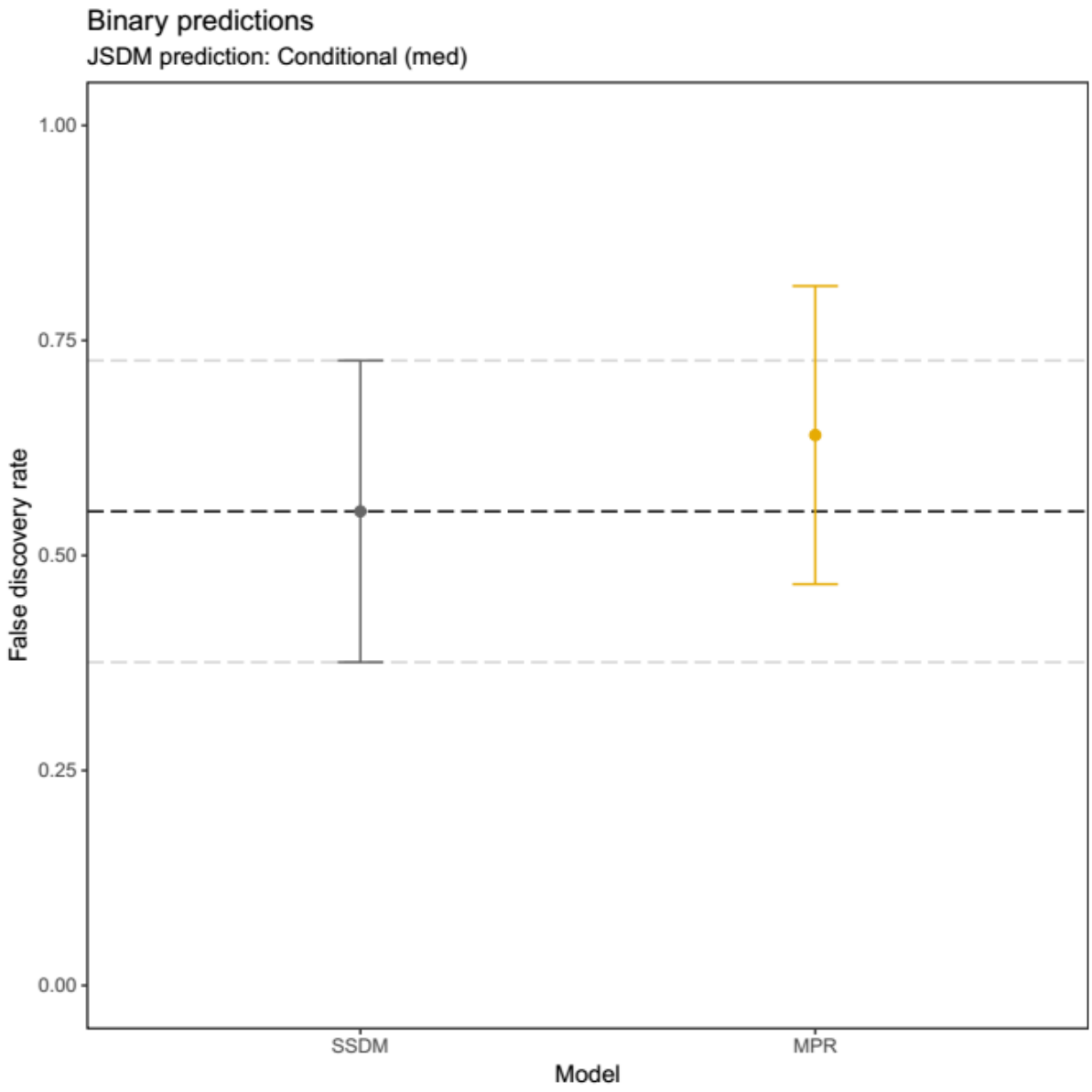


Figure 106: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

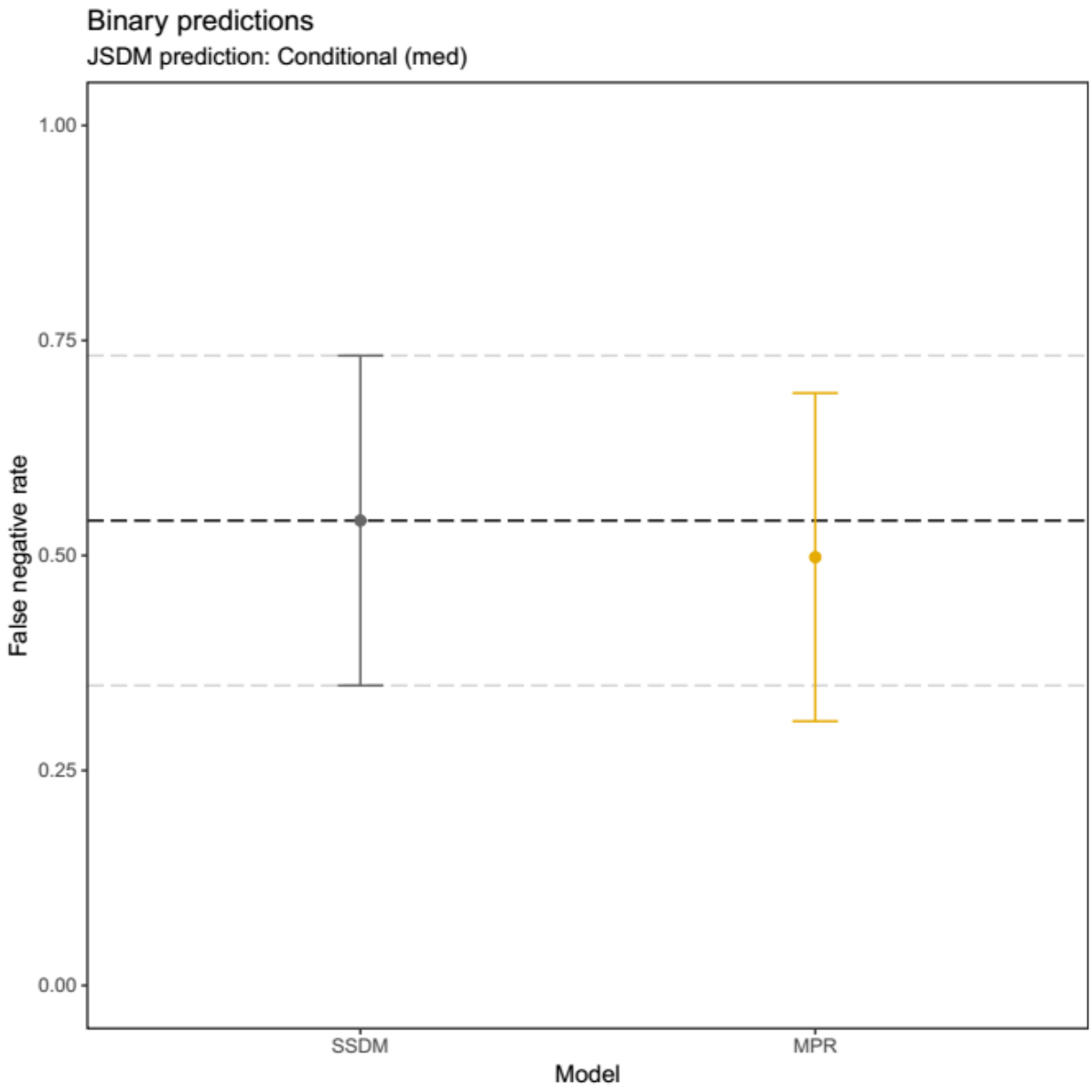


Figure 107: Performance of the SSDM and JSDM for the false negative rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

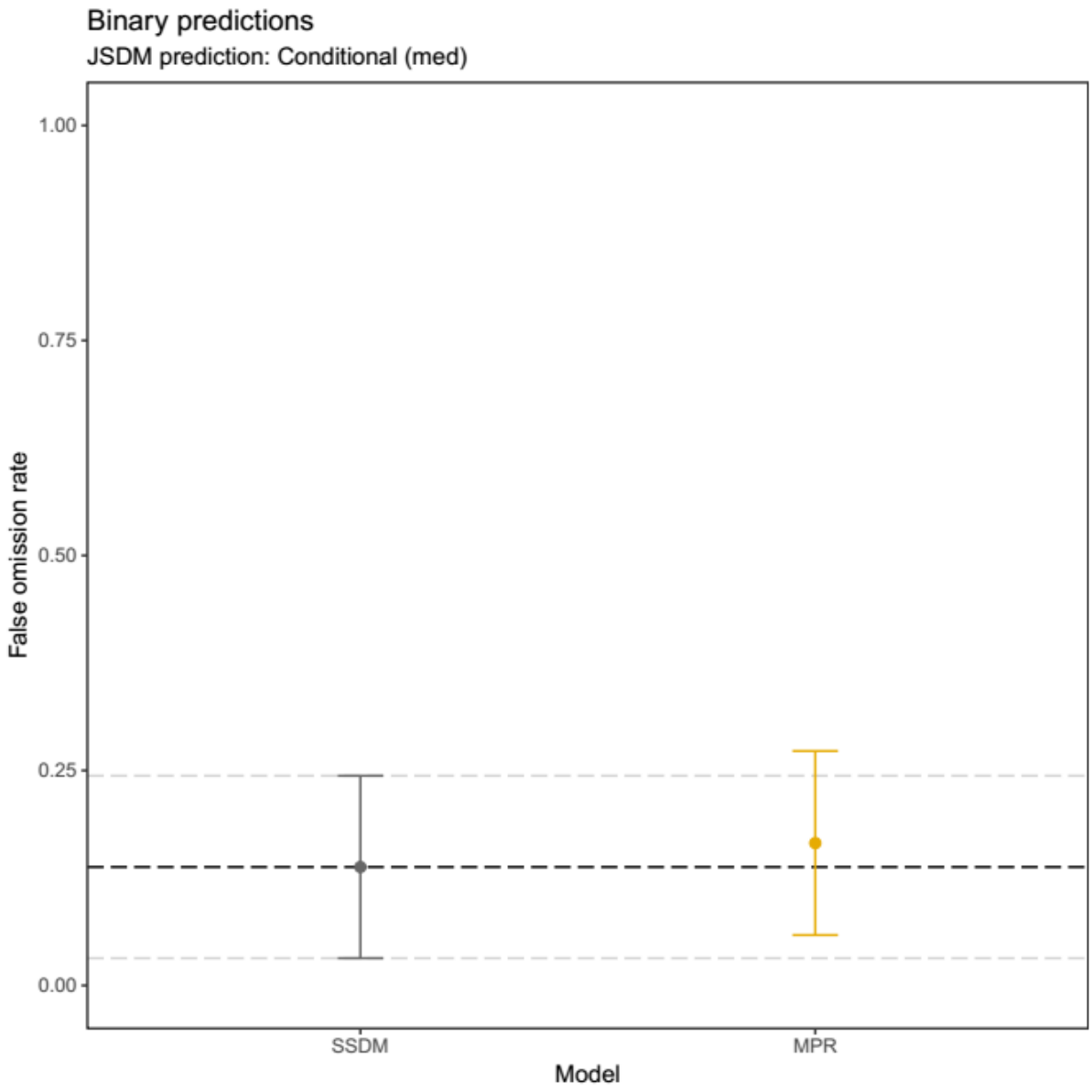


Figure 108: Performance of the SSDM and JSDM for the false omission rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

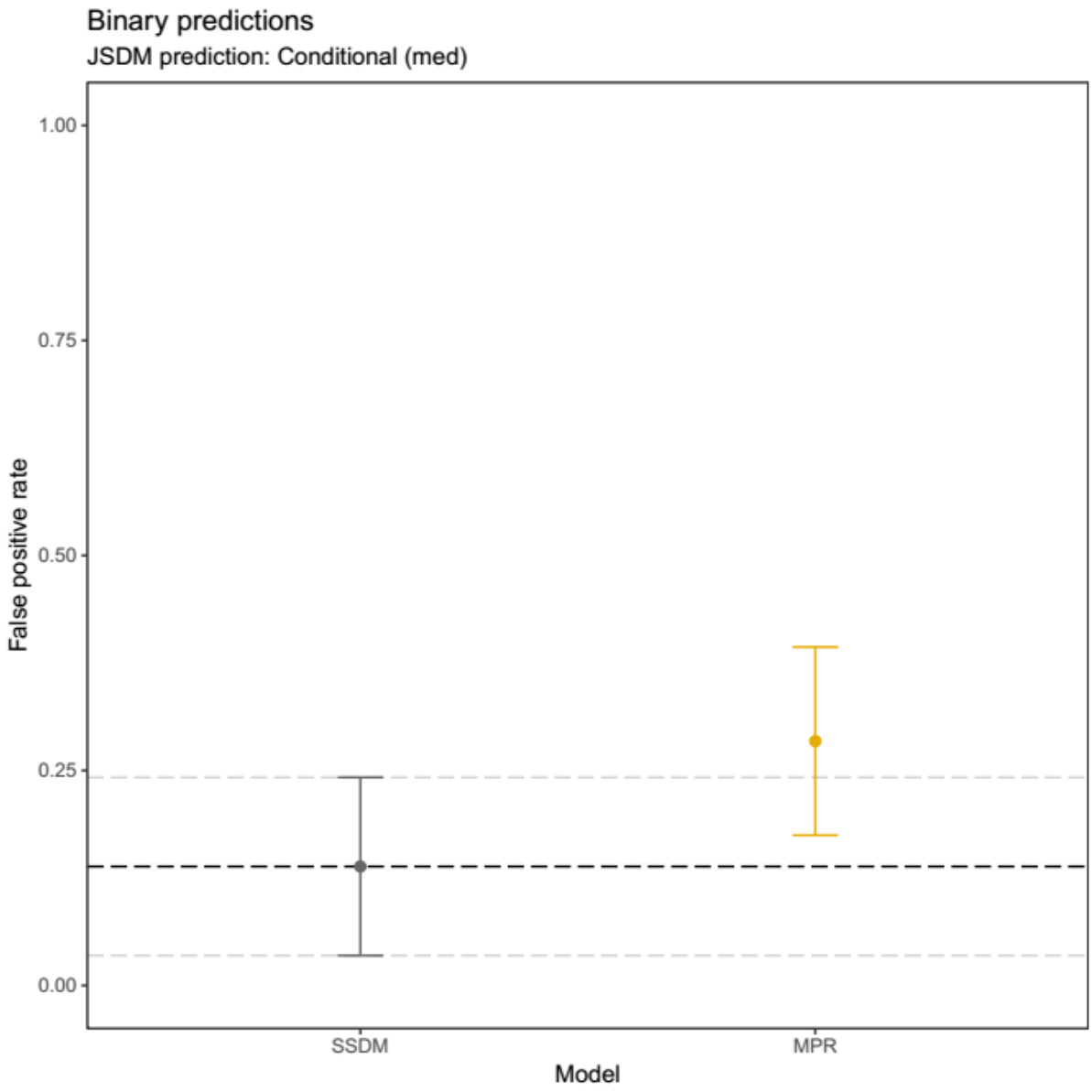


Figure 109: Performance of the SSDM and JSDM for the false positive rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

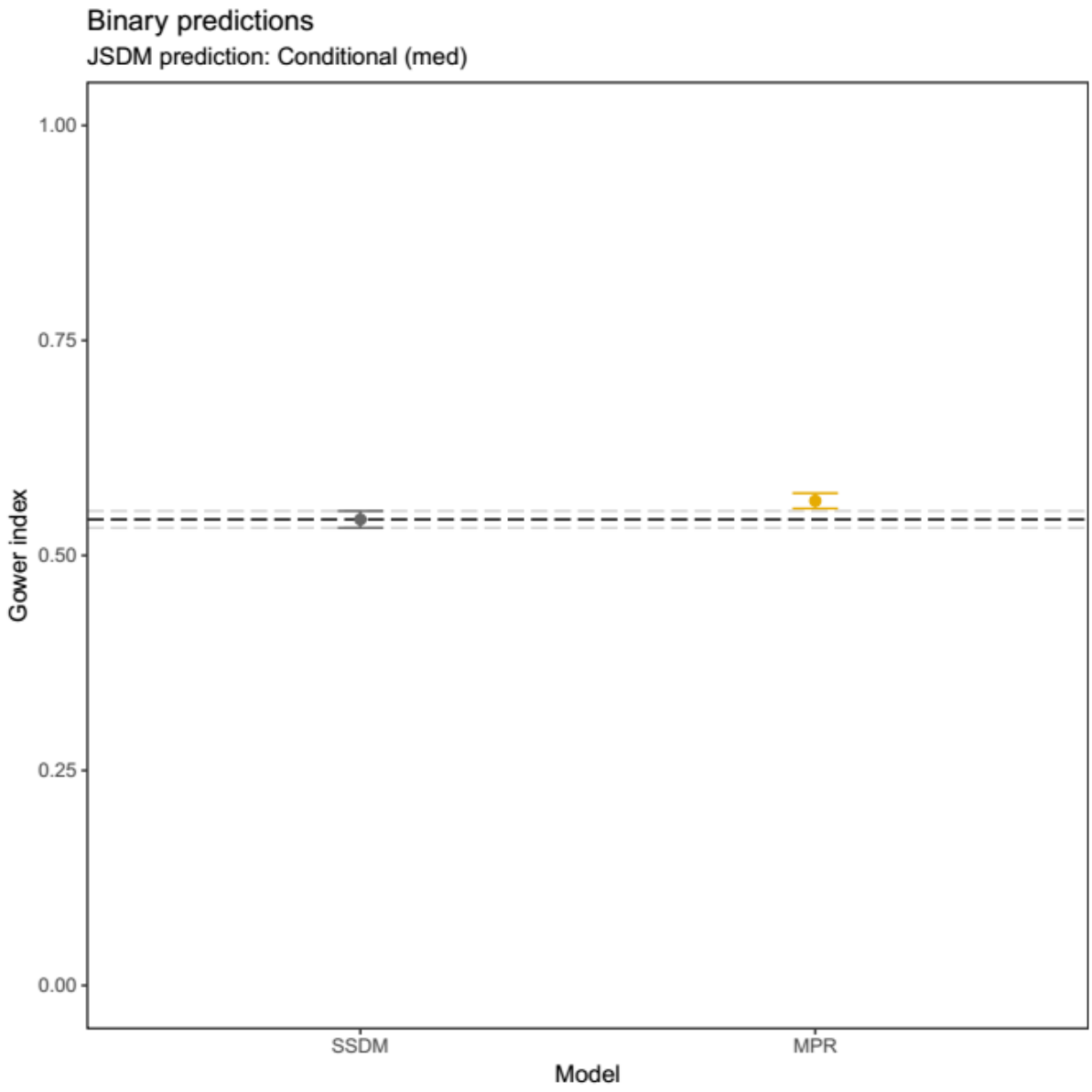


Figure 110: Performance of the SSDM and JSDM for the Gower index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

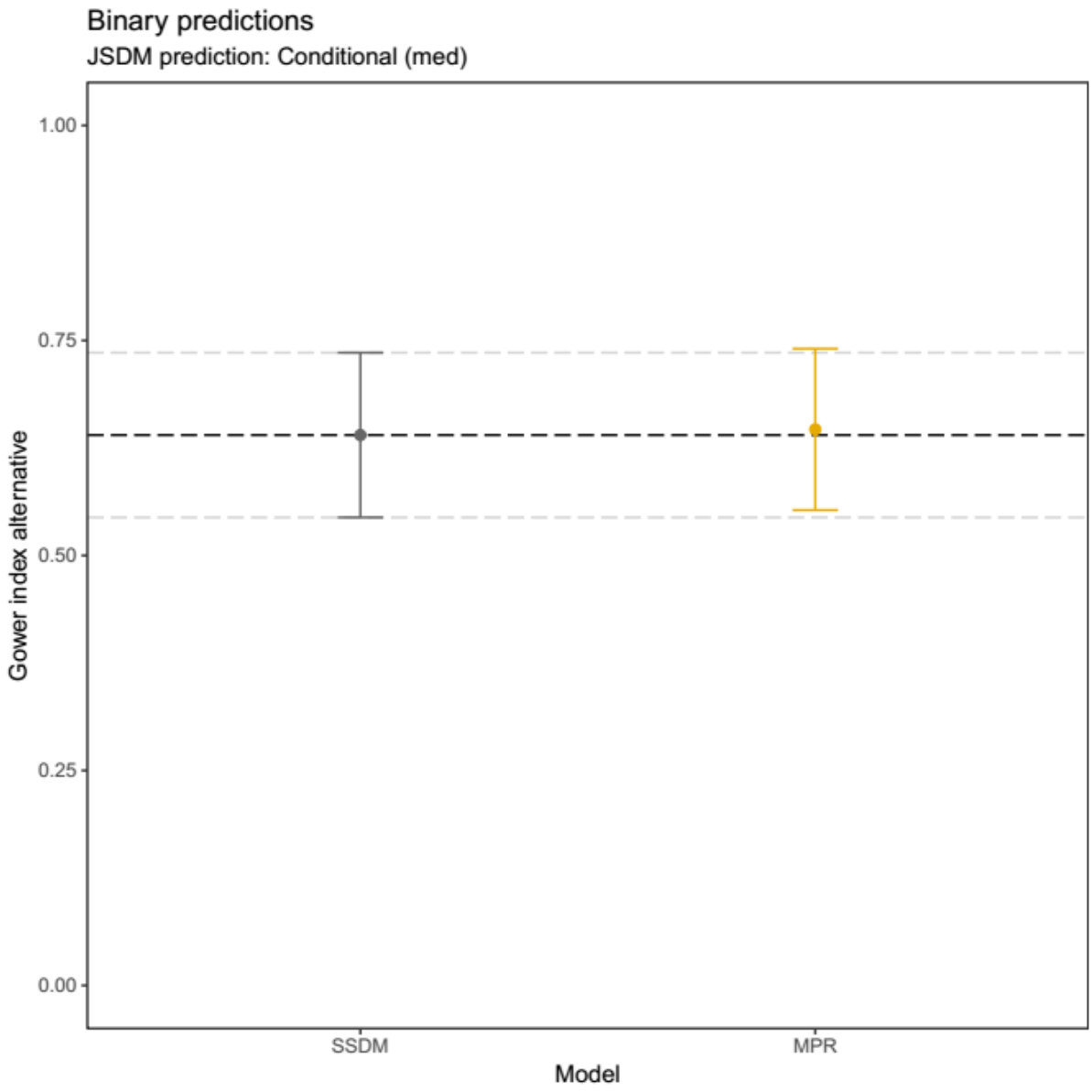


Figure 111: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

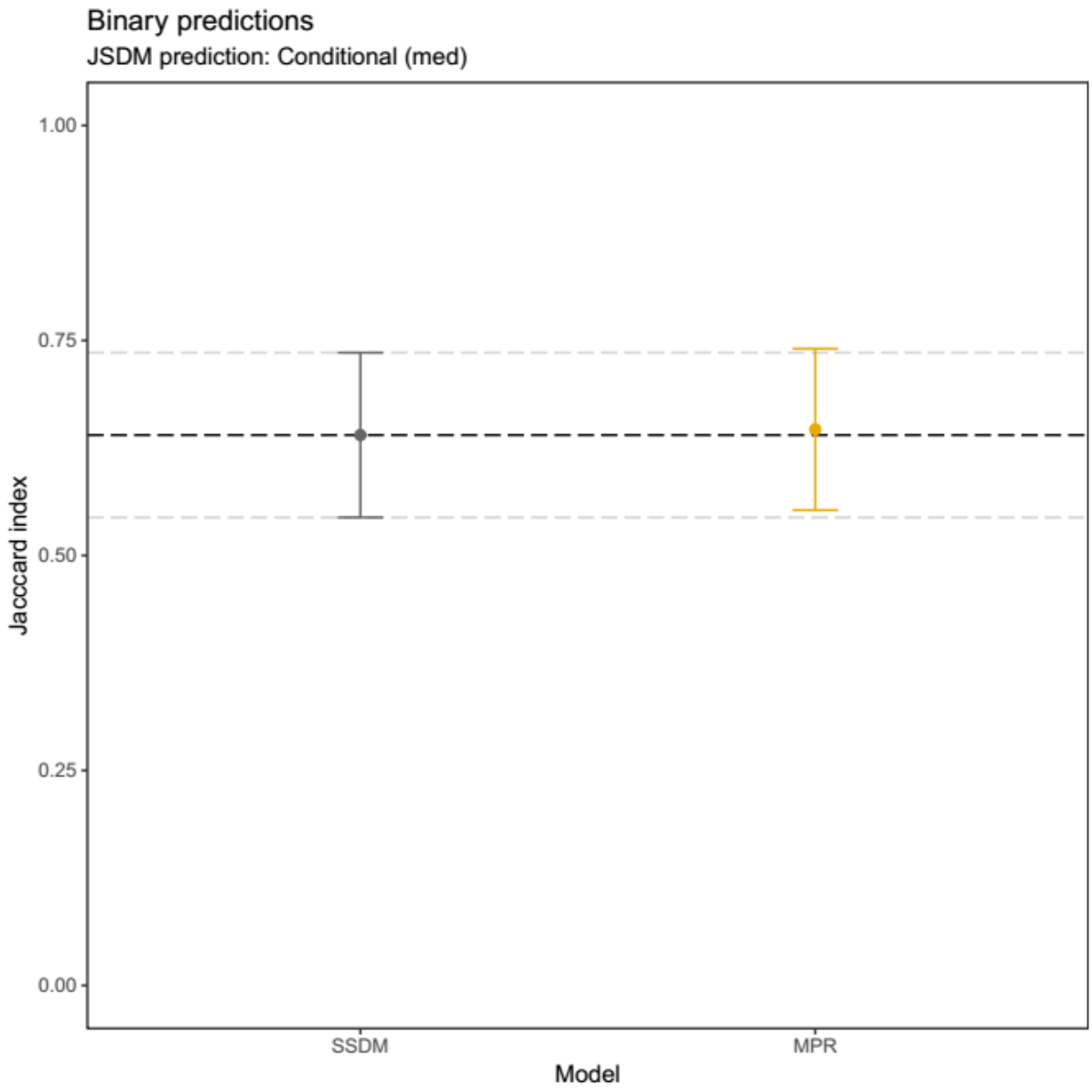


Figure 112: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

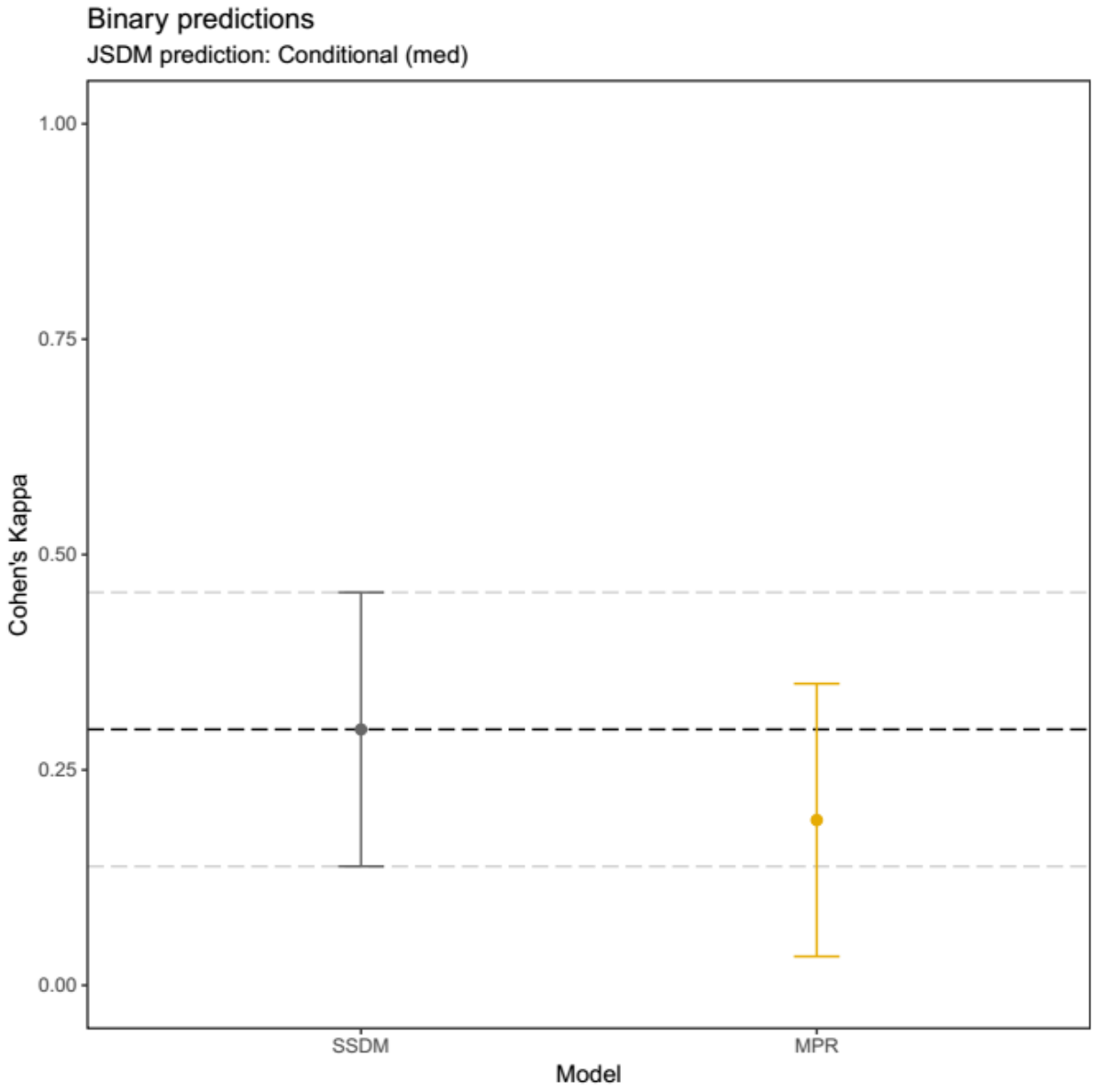


Figure 113: Performance of the SSDM and JSDM for the Cohen's Kappa evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

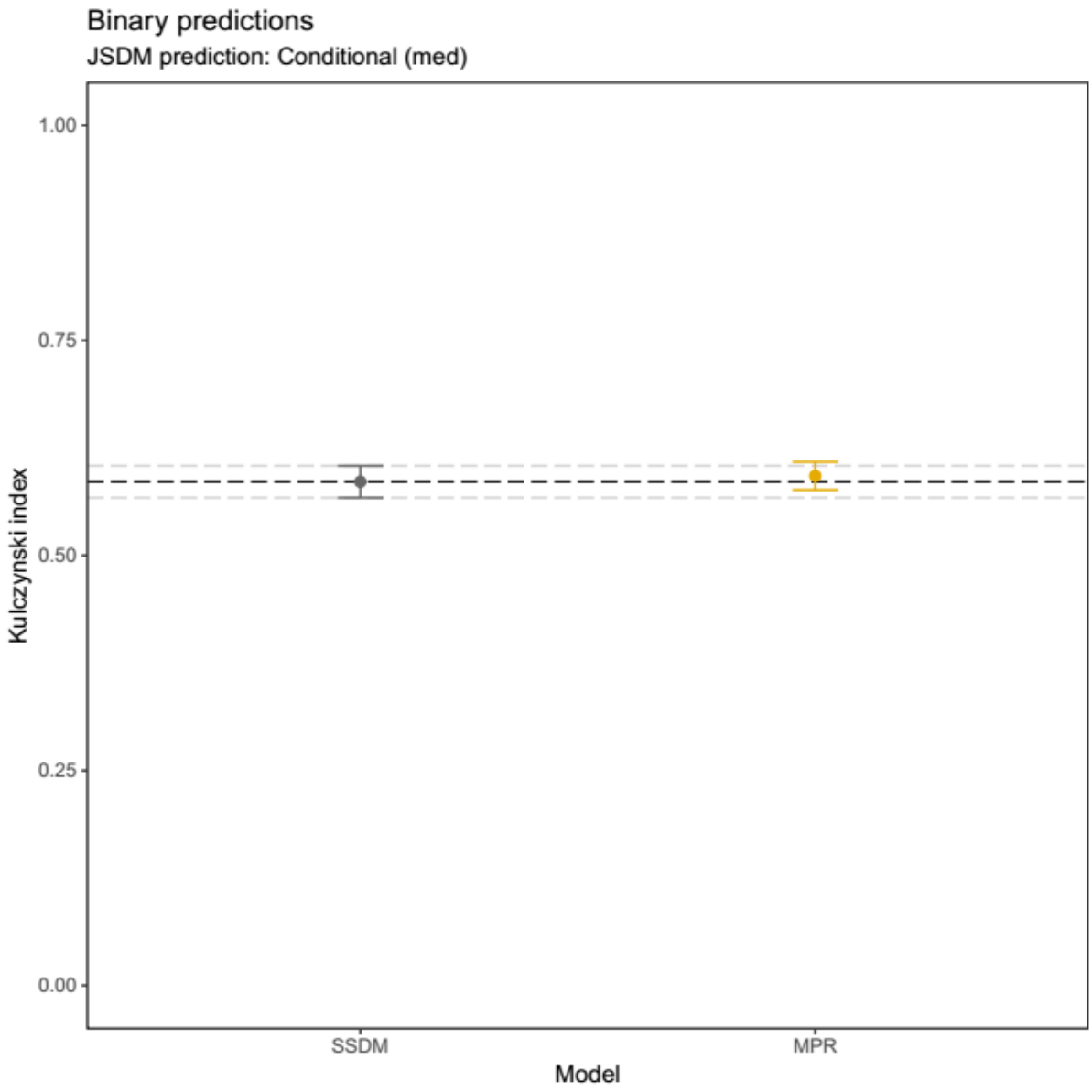


Figure 114: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

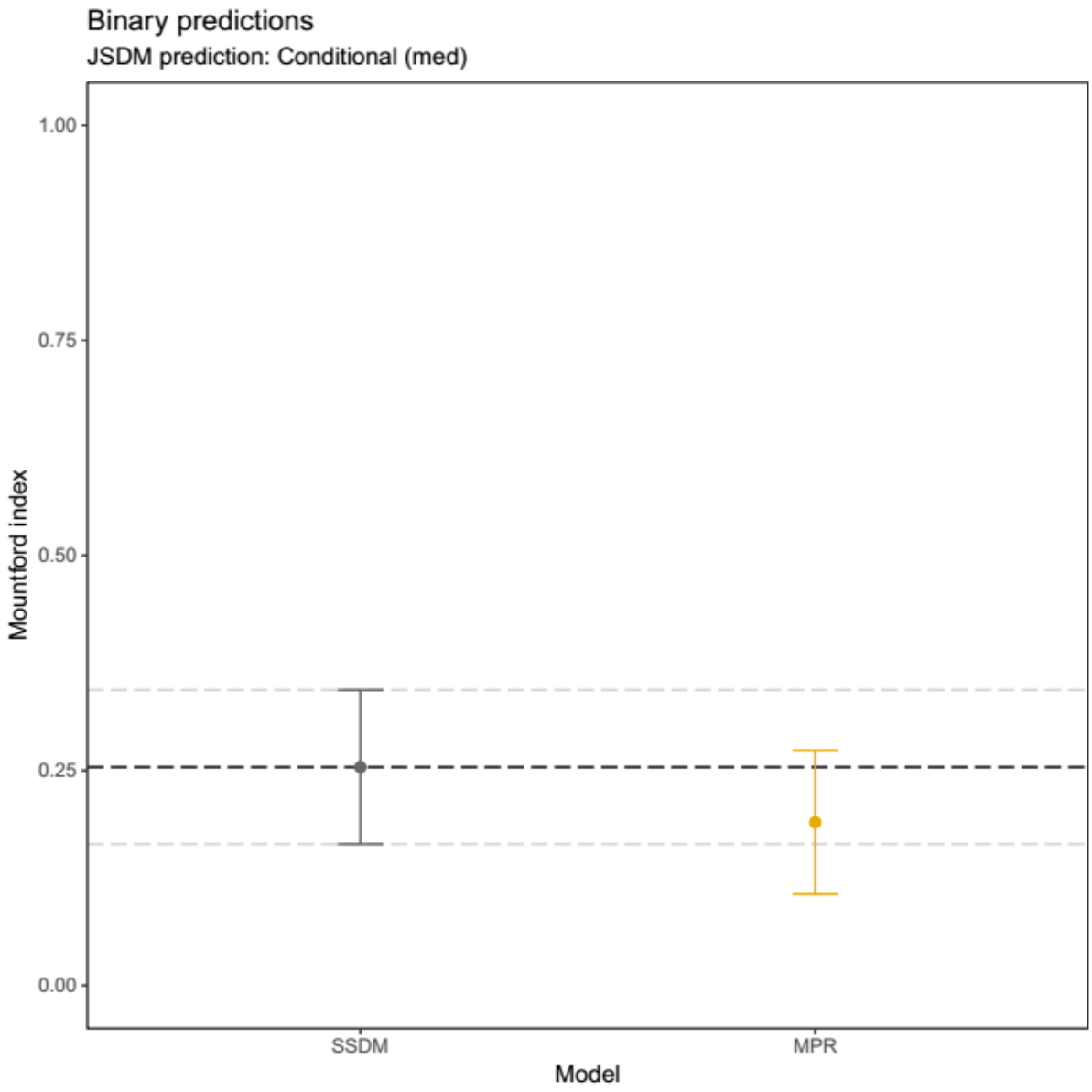


Figure 115: Performance of the SSDM and JSDM for the Mountford index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

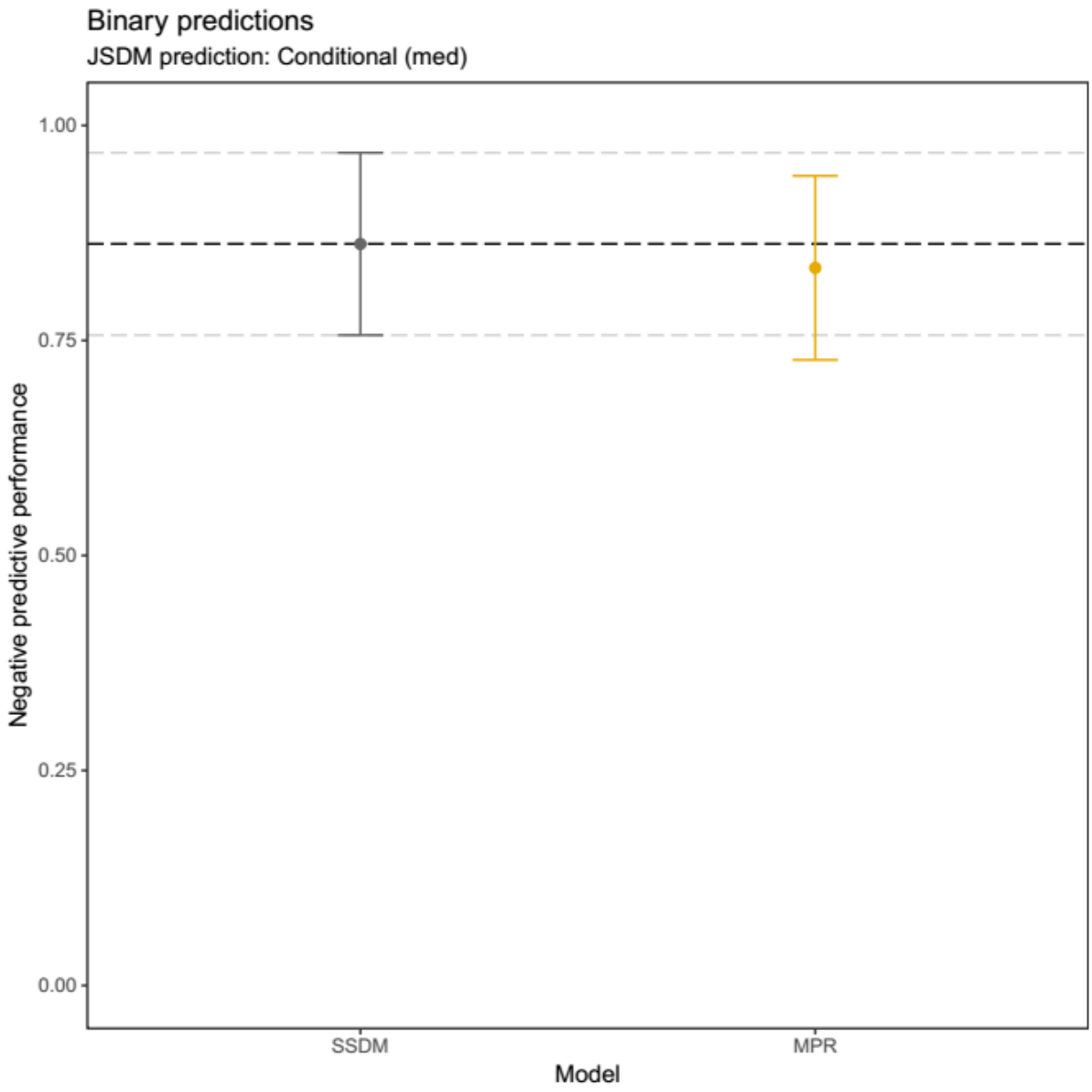


Figure 116: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

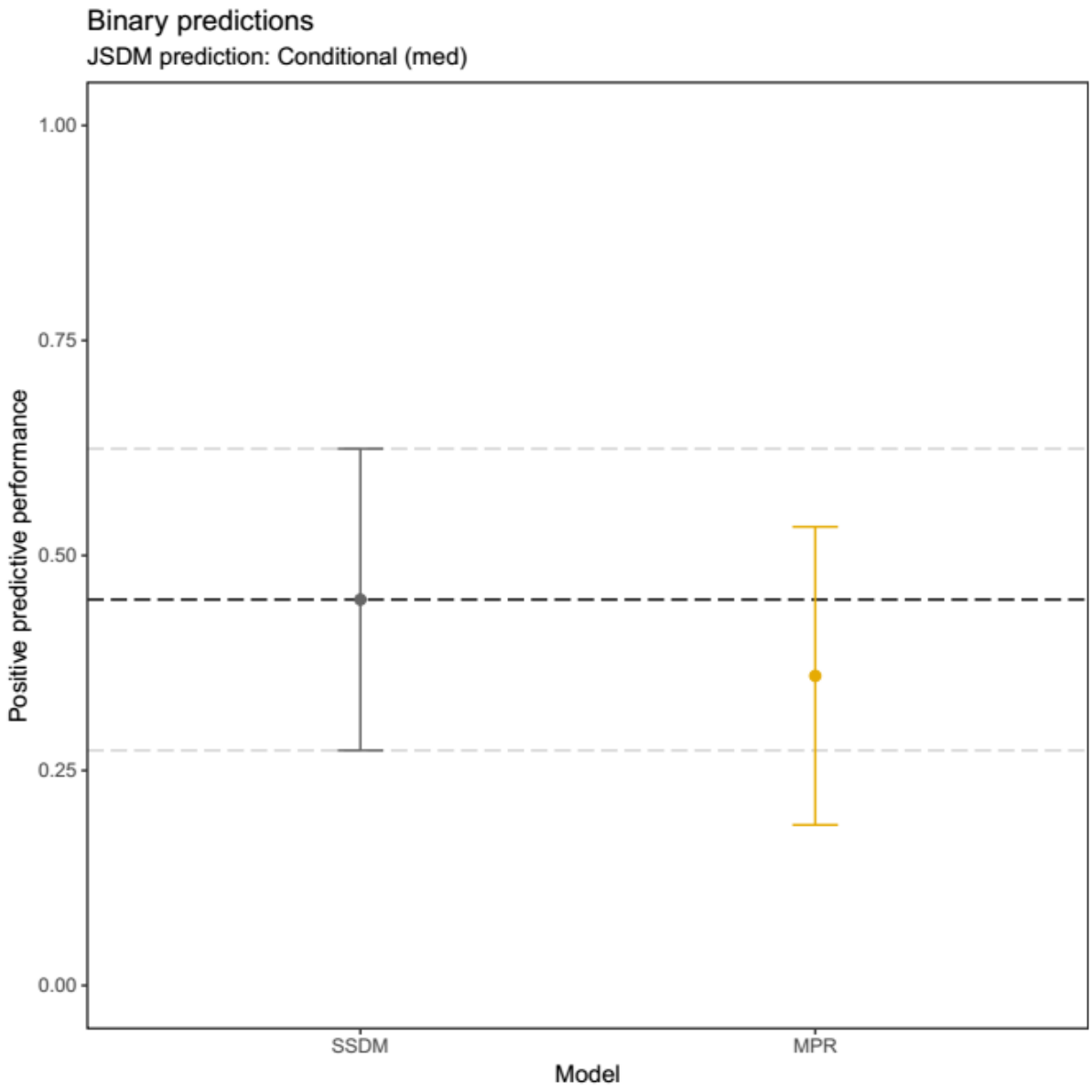


Figure 117: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

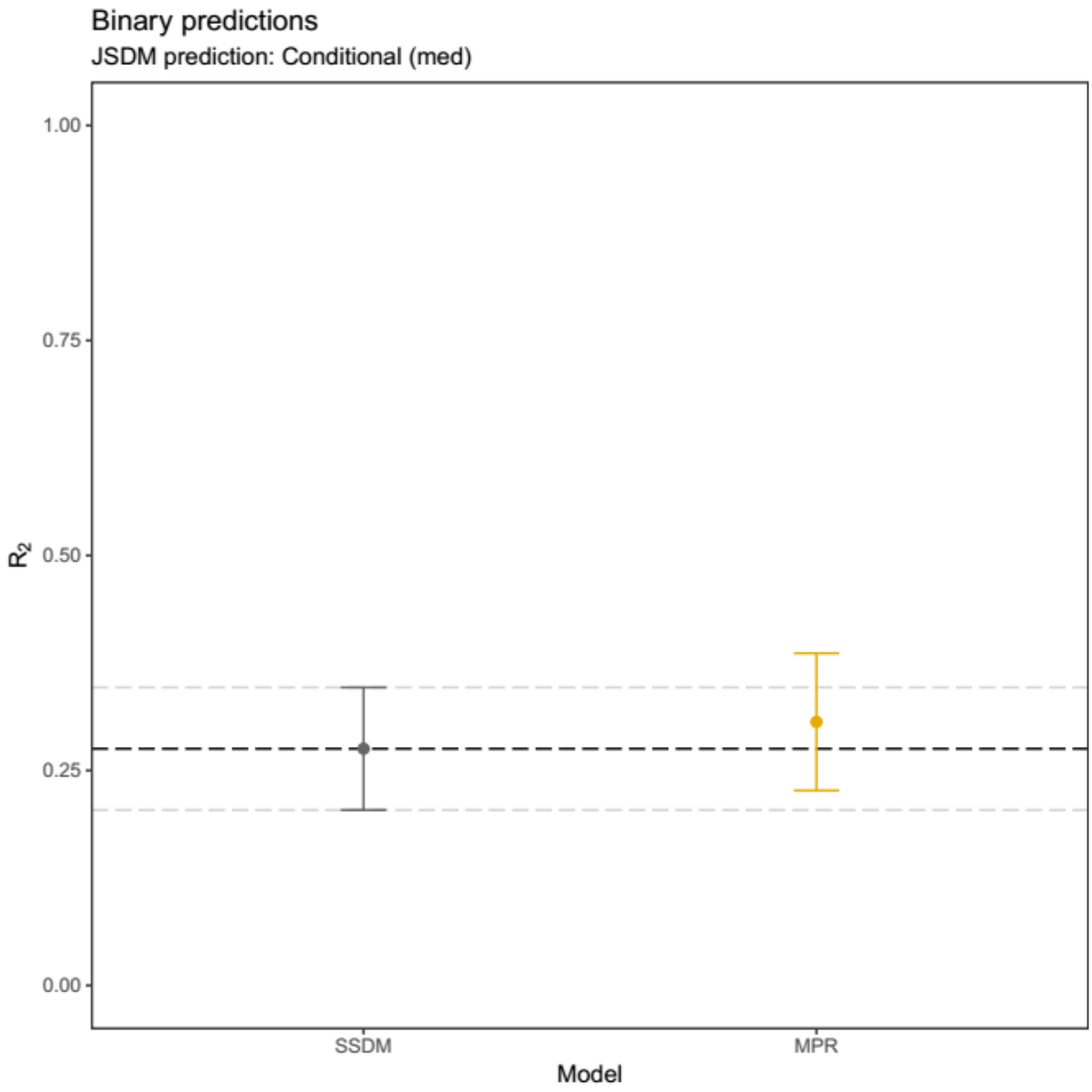


Figure 118: Performance of the SSDM and JSDM for the R^2 evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

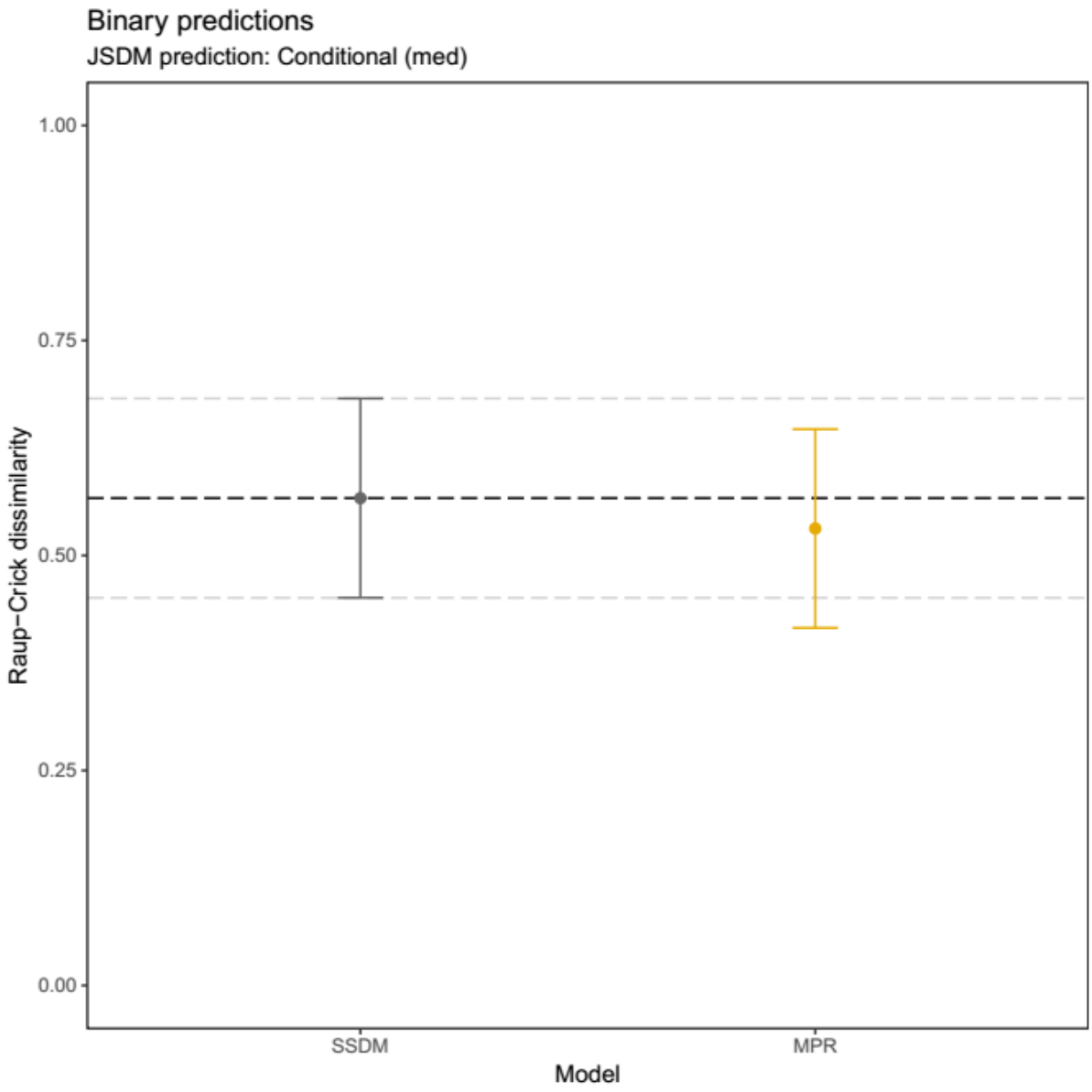


Figure 119: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

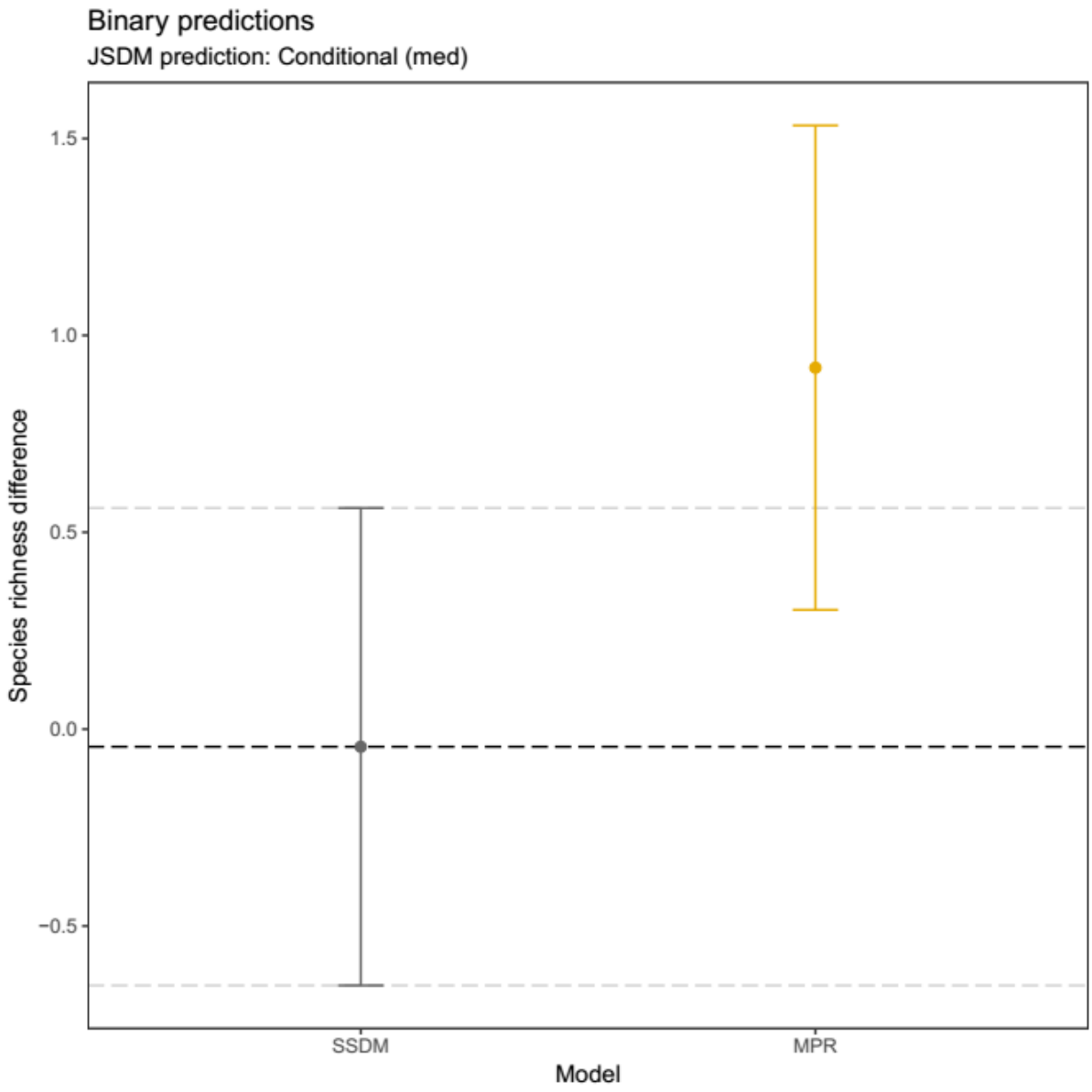


Figure 120: Performance of the SSDM and JSDM for the species richness difference evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

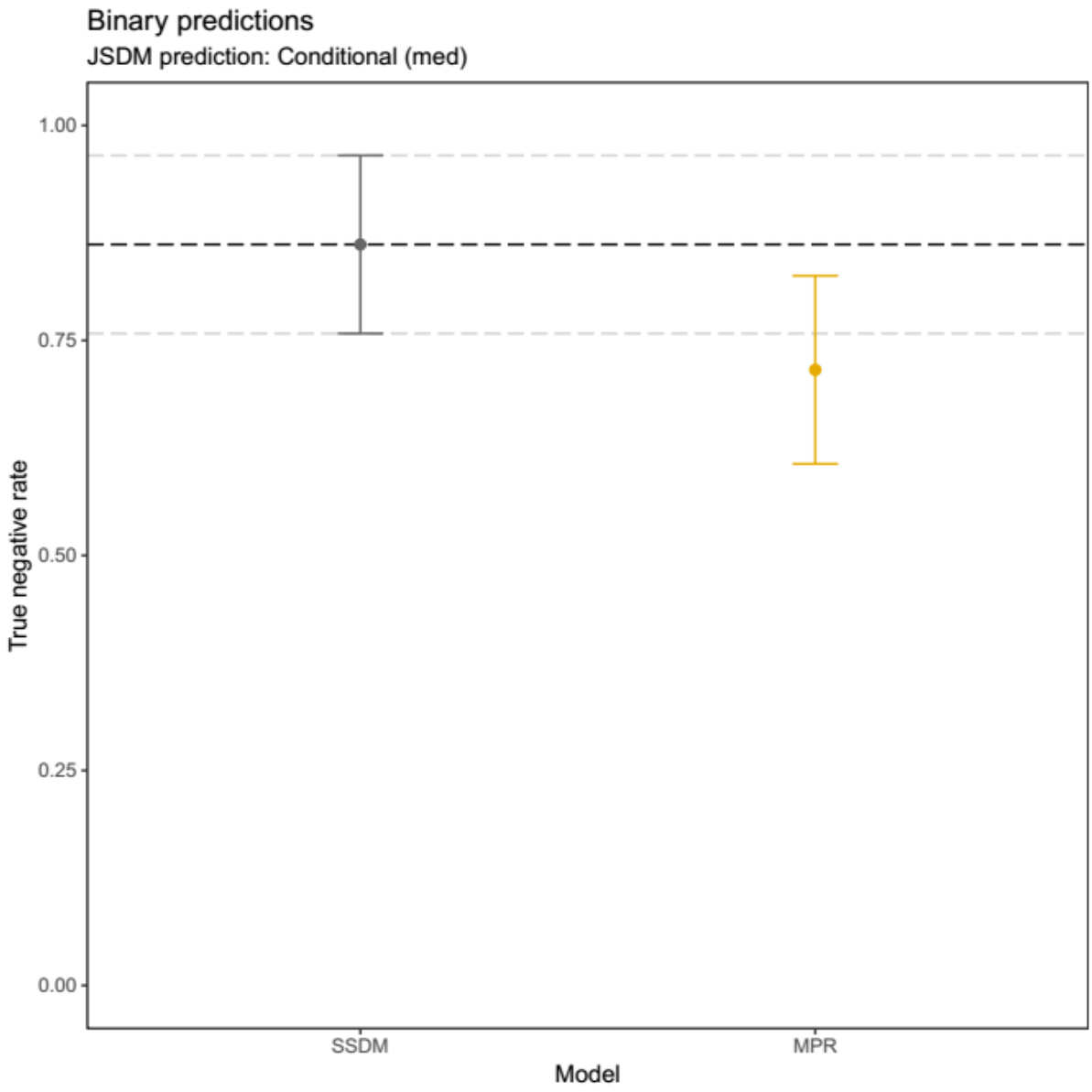


Figure 121: Performance of the SSDM and JSDM for the true negative rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

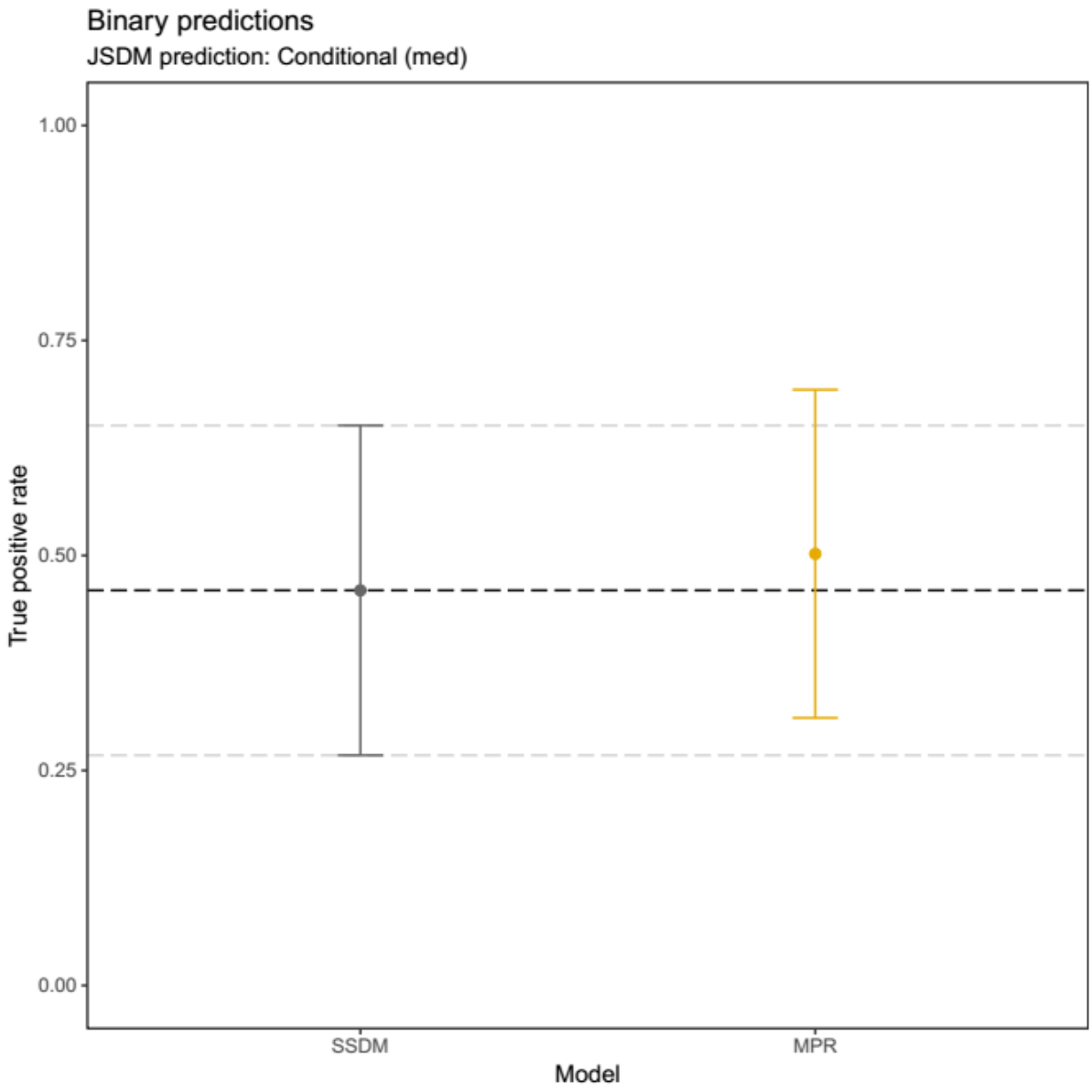


Figure 122: Performance of the SSDM and JSDM for the true positive rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

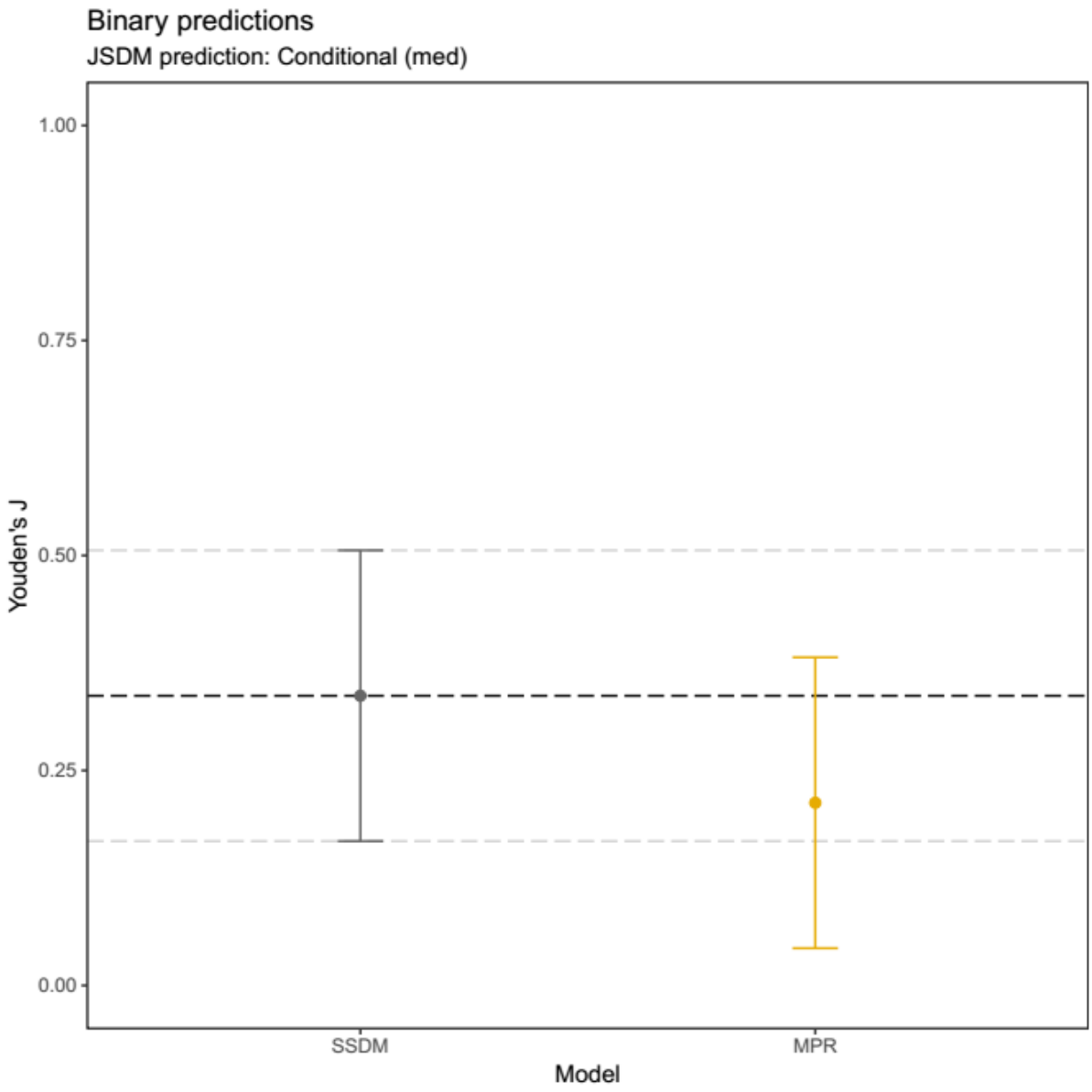


Figure 123: Performance of the SSDM and JSDM for the Youden's J evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

6. Conditional (high-prevalence) predictions

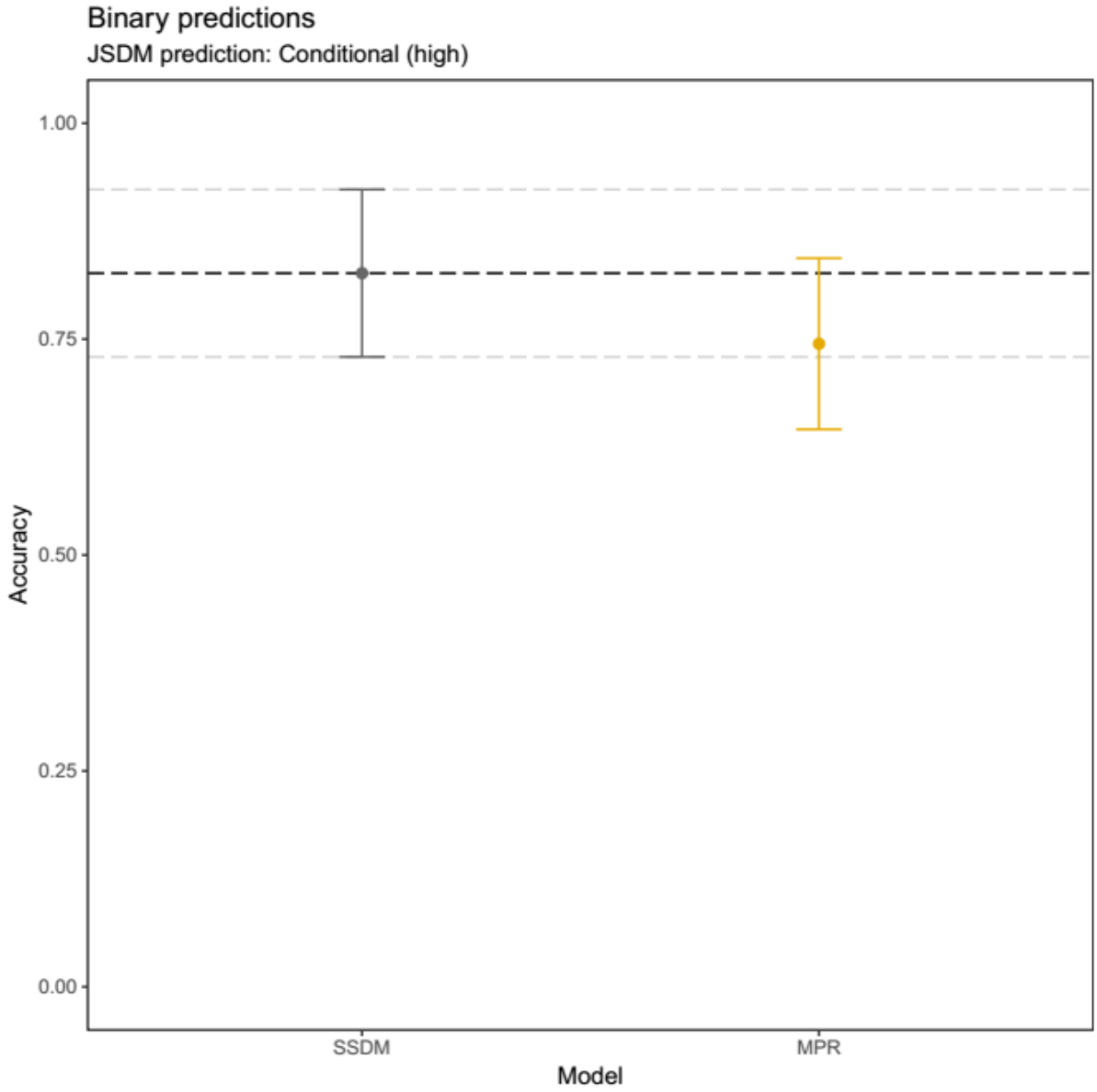


Figure 124: Performance of the SSDM and JSDM for the accuracy evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

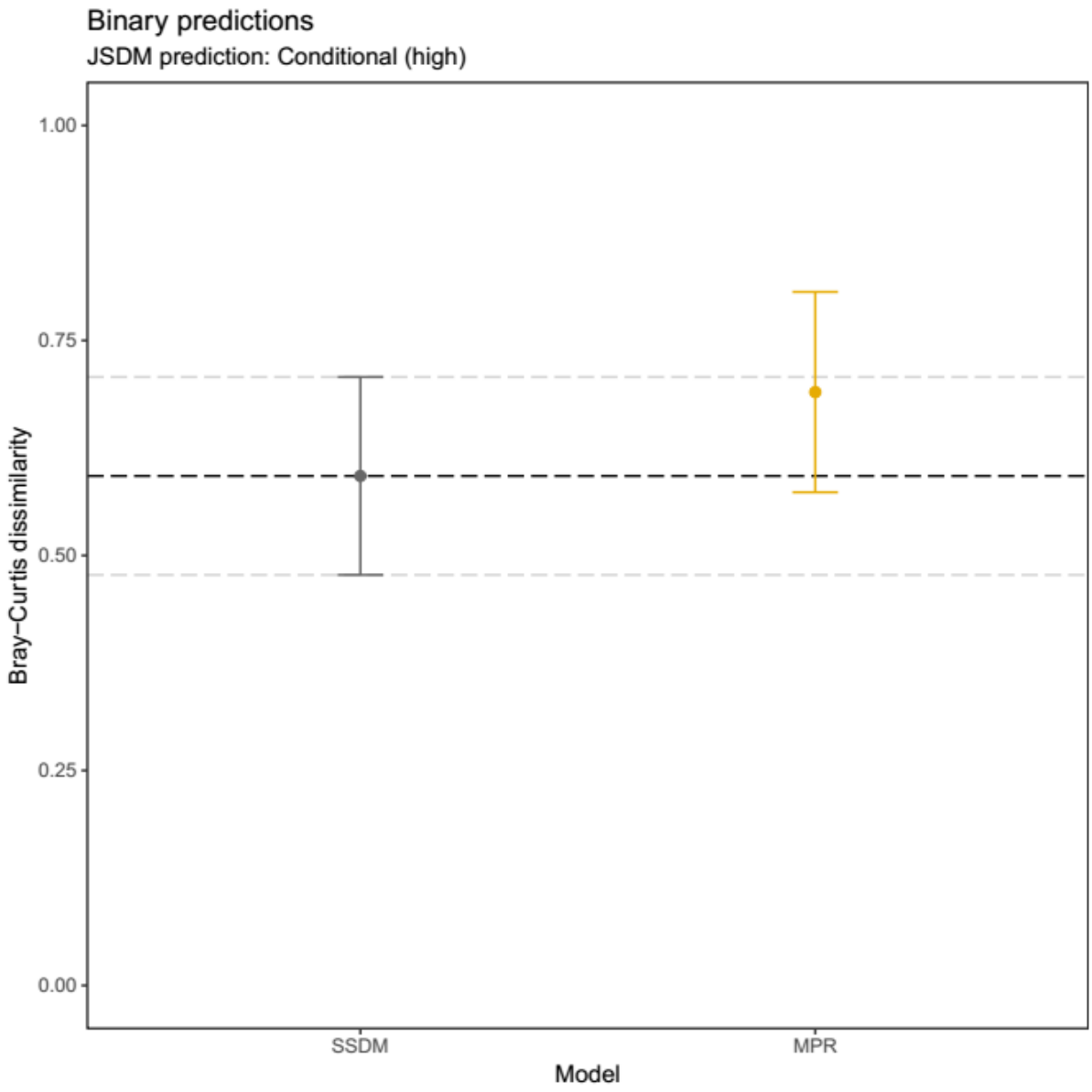


Figure 125: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

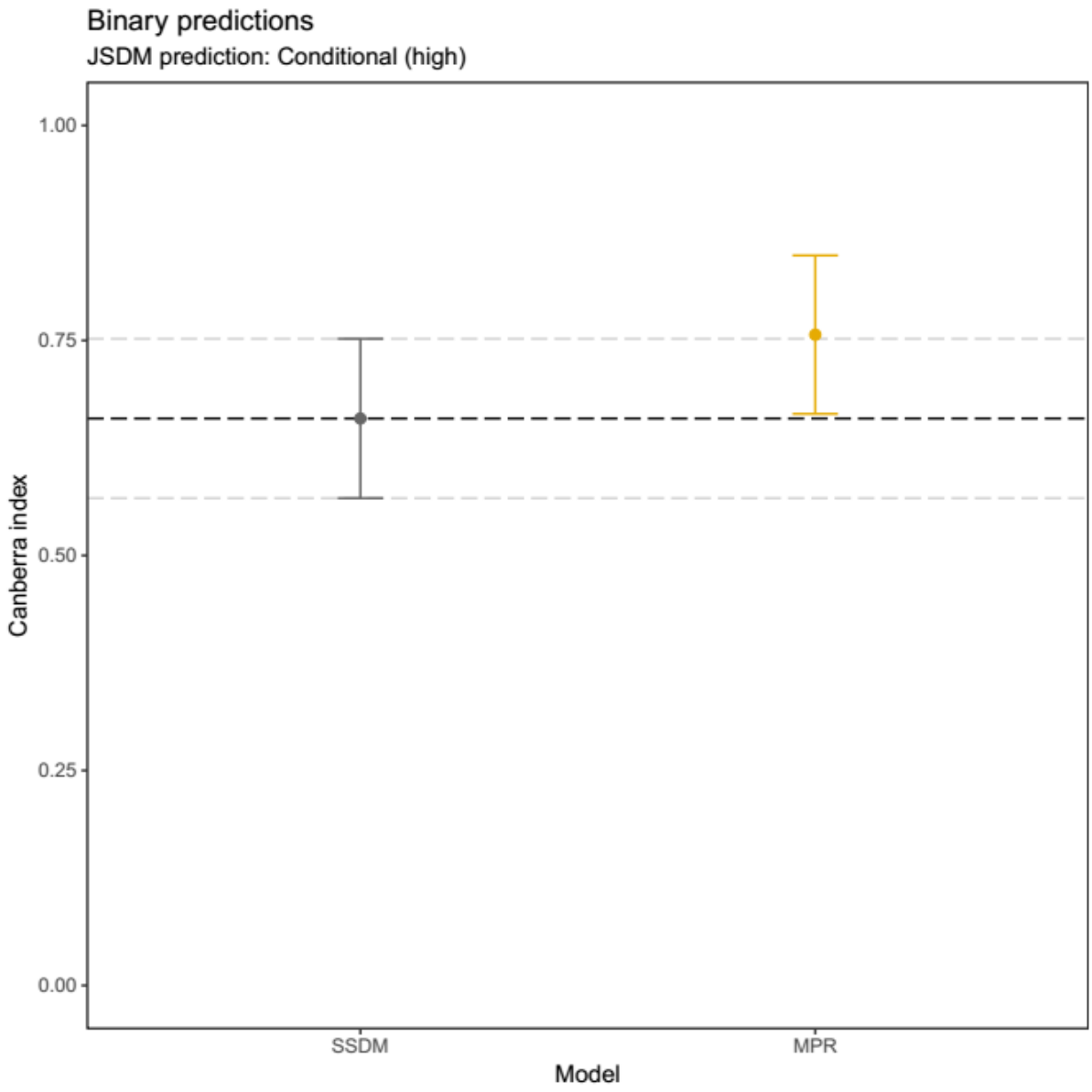


Figure 126: Performance of the SSDM and JSDM for the Canberra index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

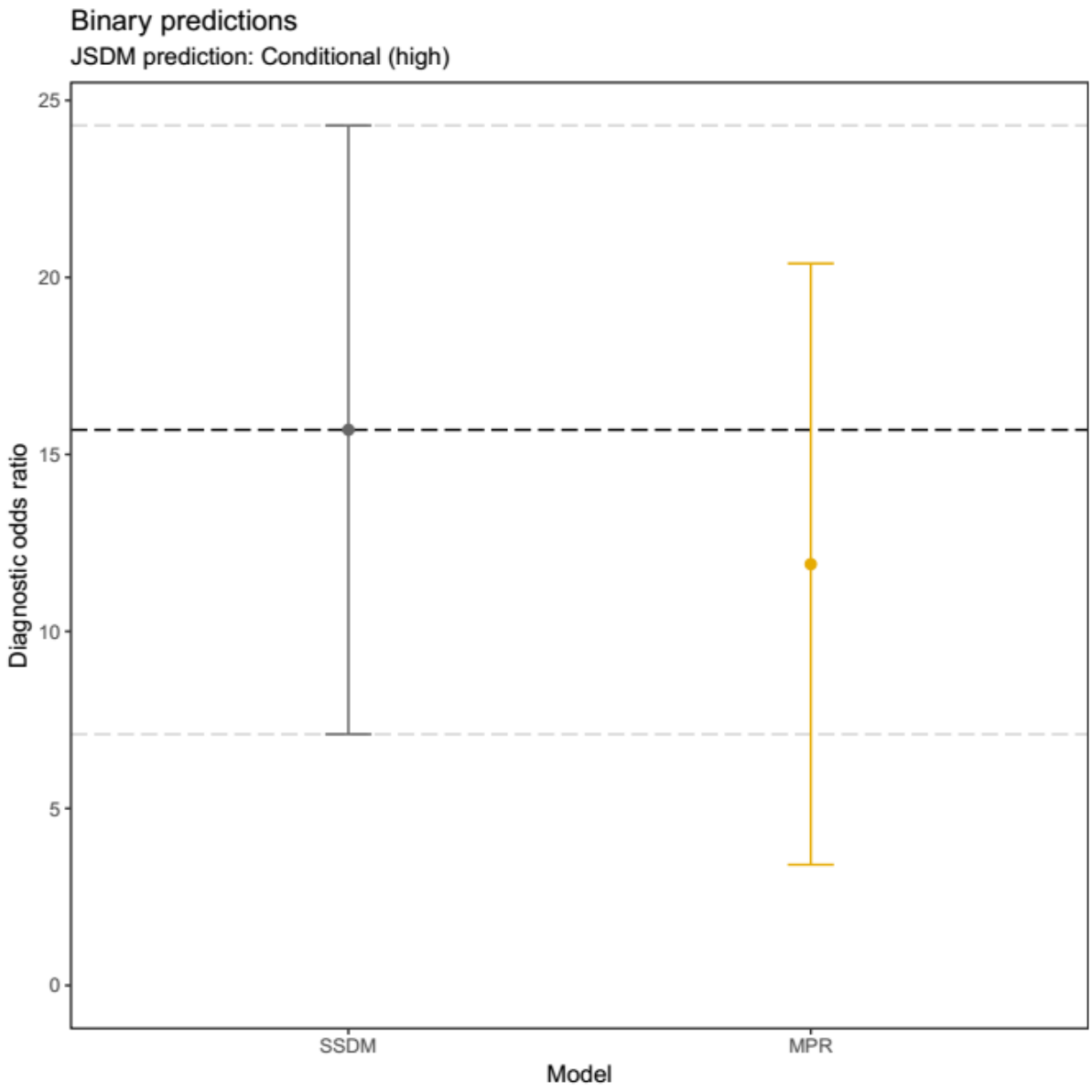


Figure 127: Performance of the SSDM and JSDM for the diagnostic odds ratio evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

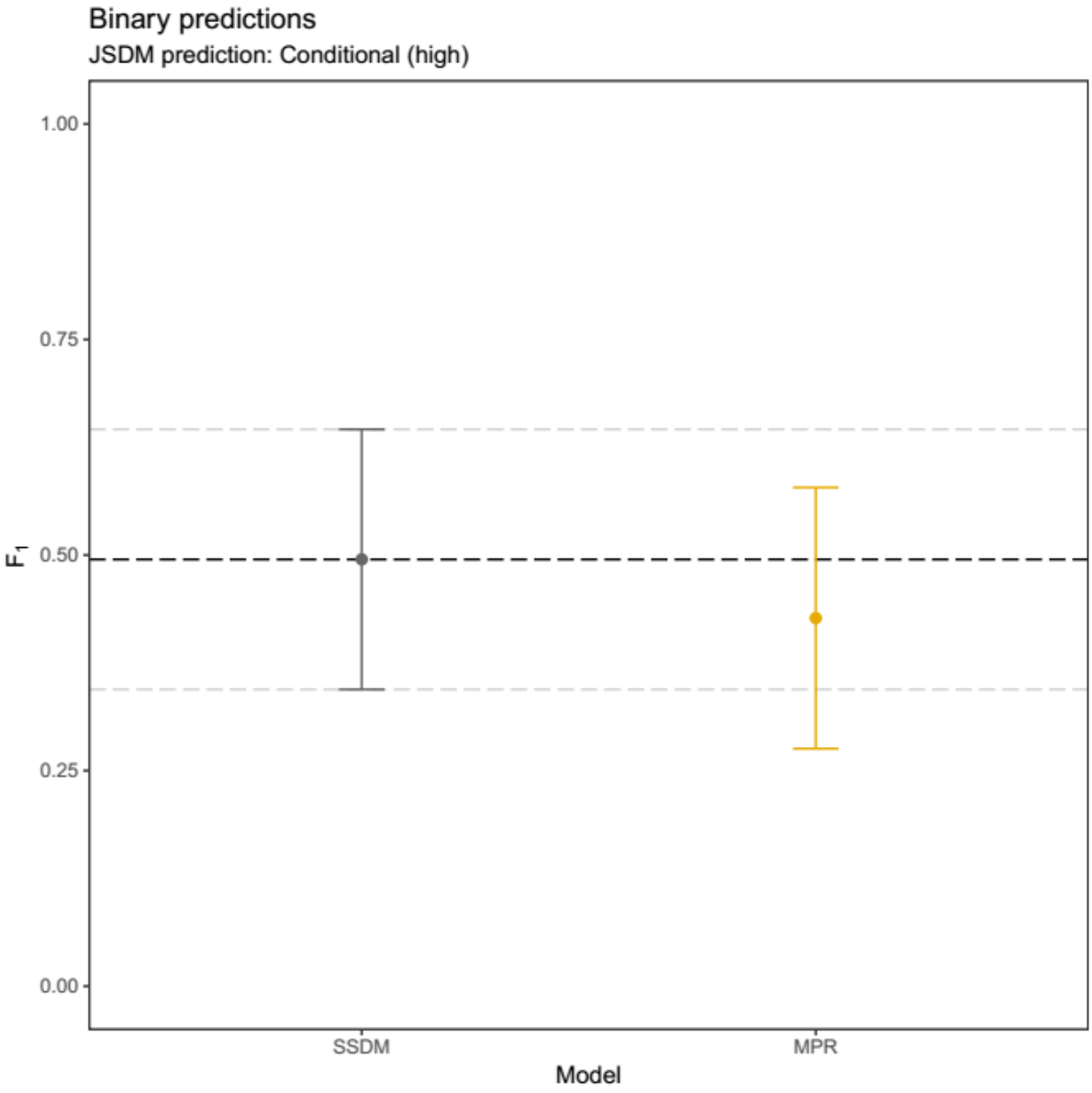


Figure 128: Performance of the SSDM and JSDM for the F_1 evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

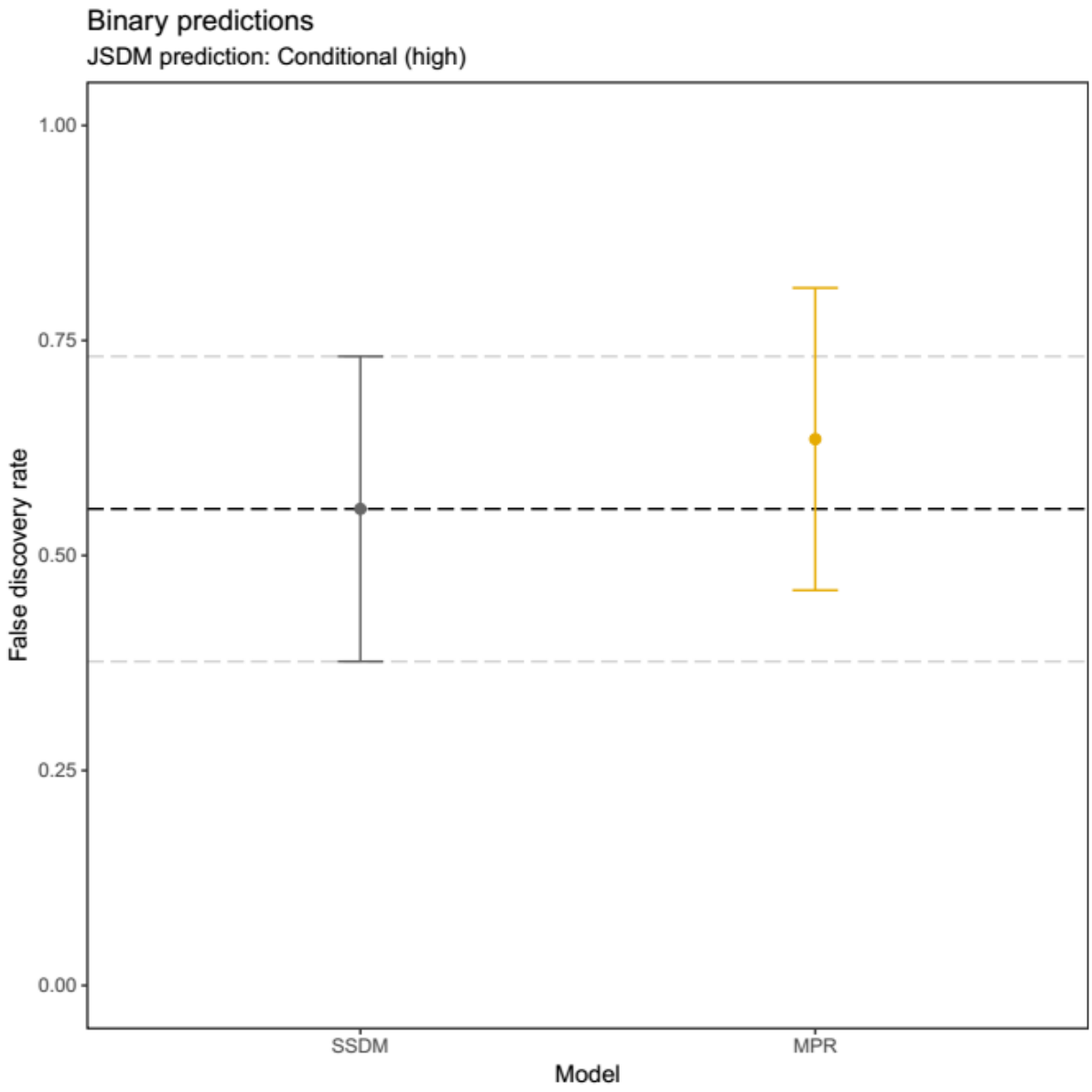


Figure 129: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

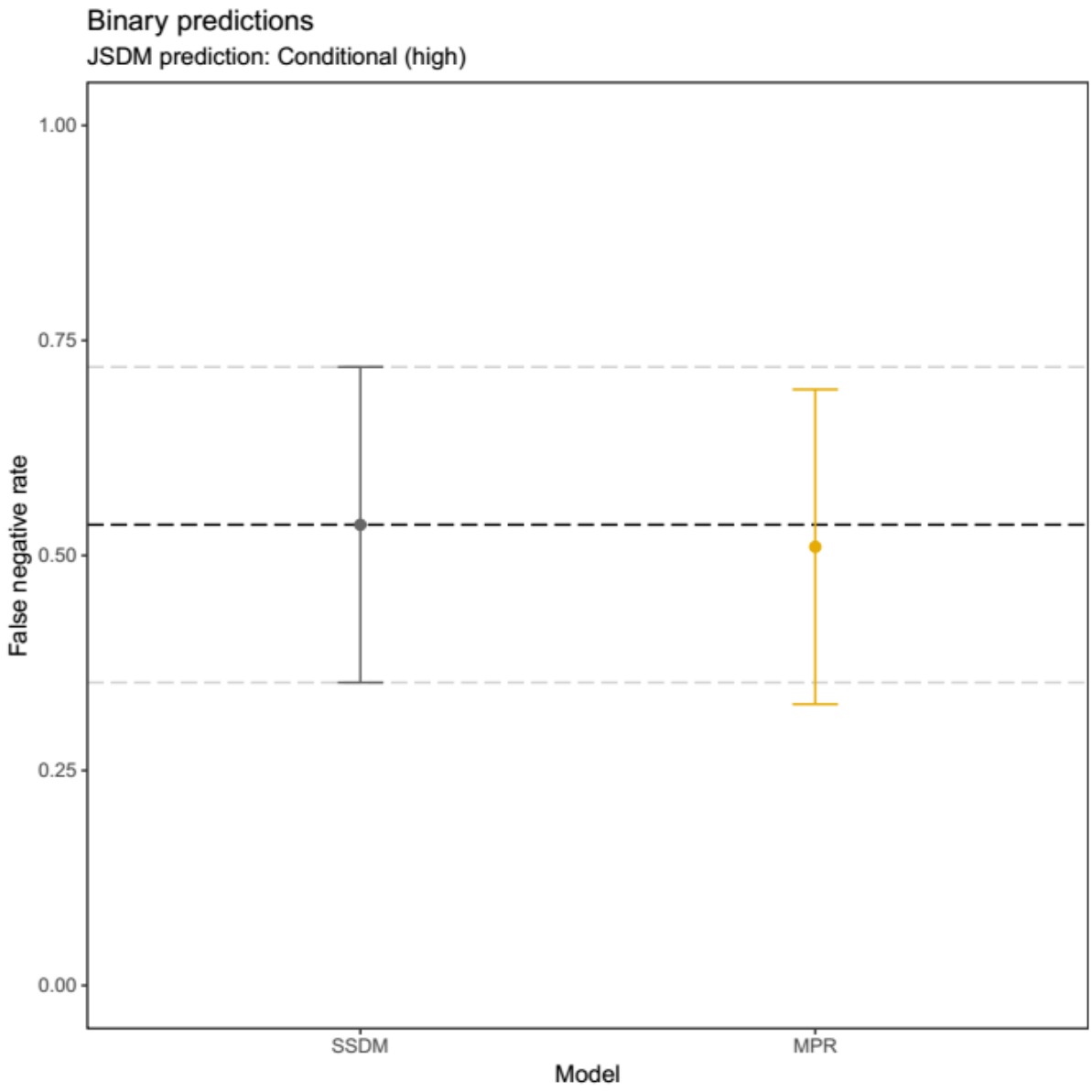


Figure 130: Performance of the SSDM and JSDM for the false negative rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

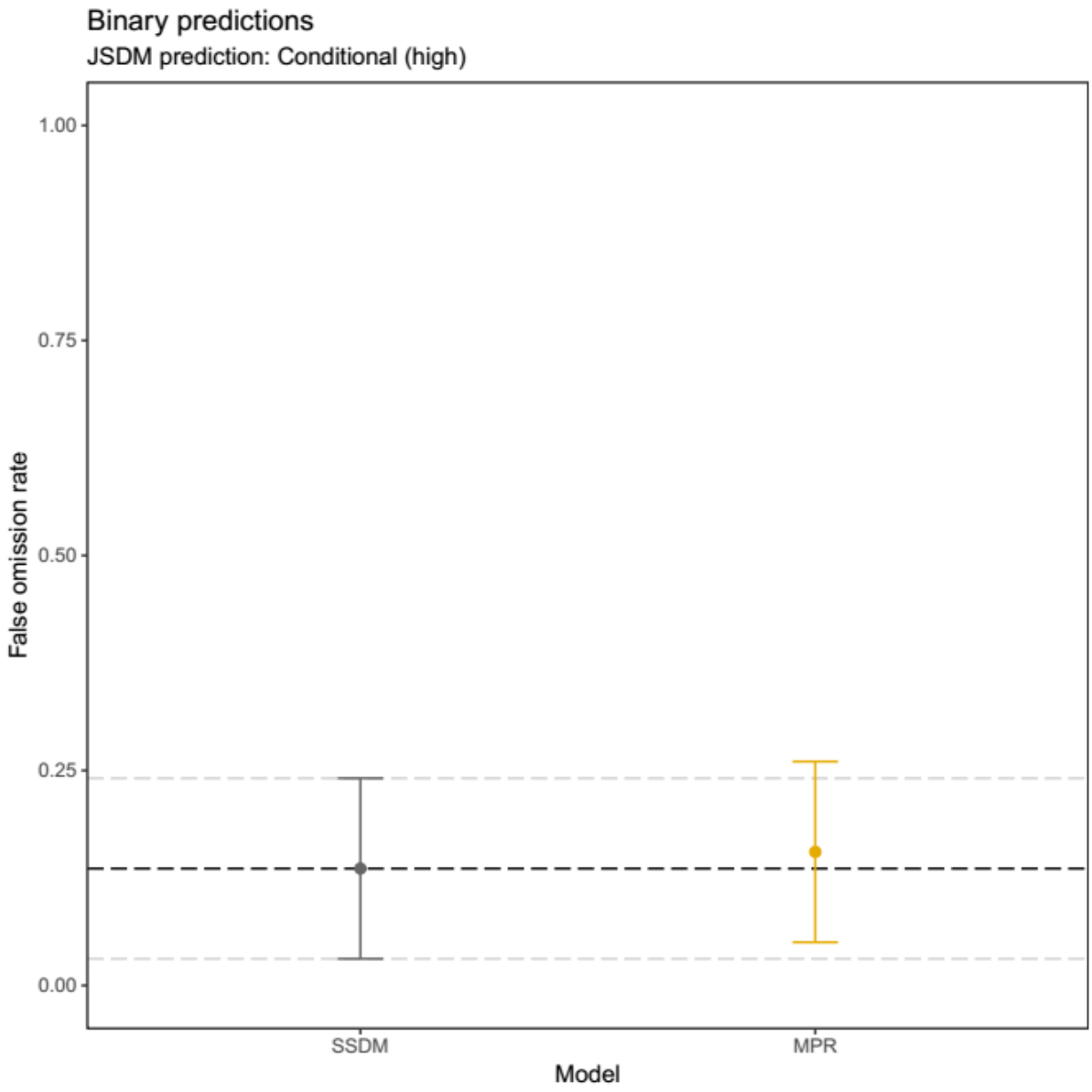


Figure 131: Performance of the SSDM and JSDM for the false omission rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

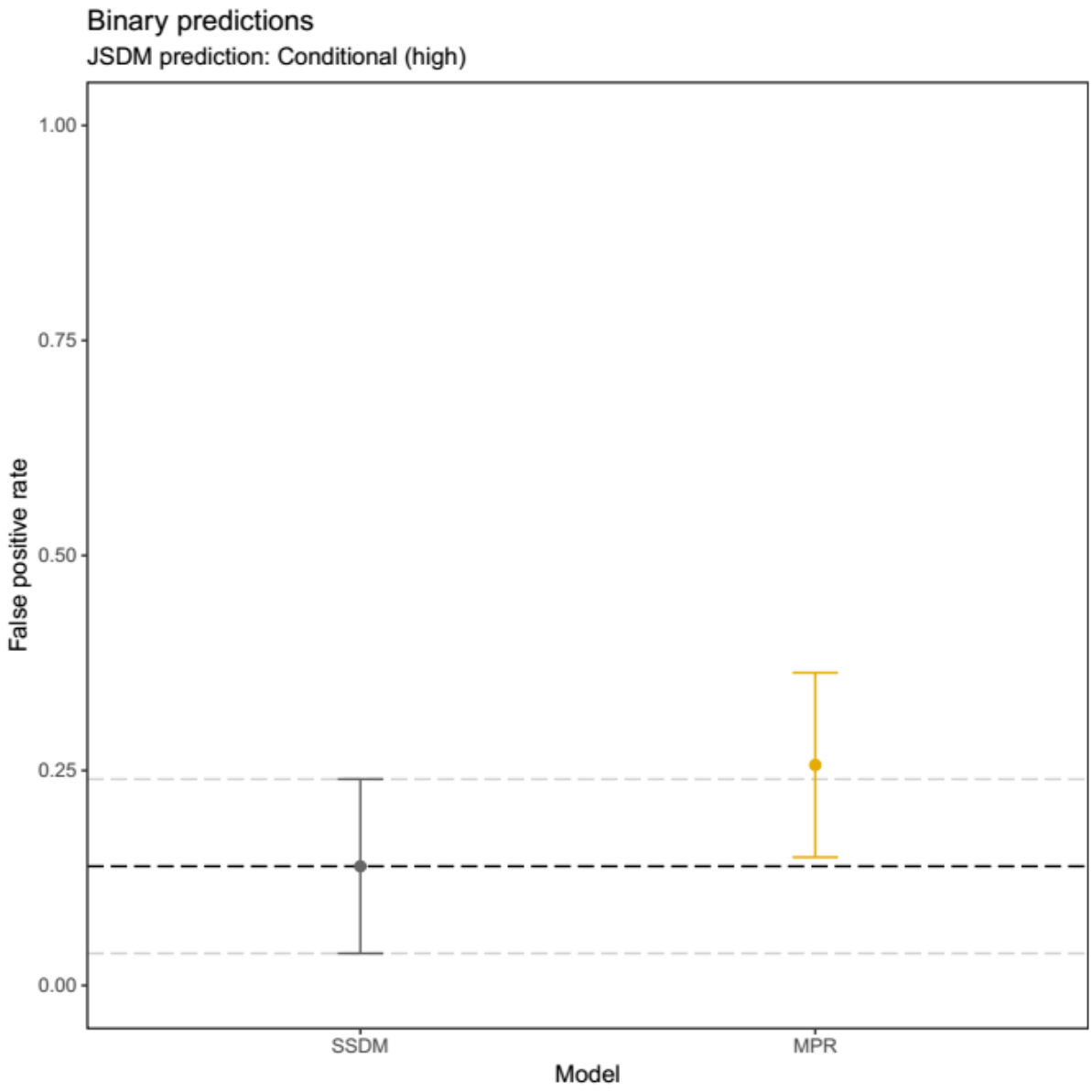


Figure 132: Performance of the SSDM and JSDM for the false positive rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

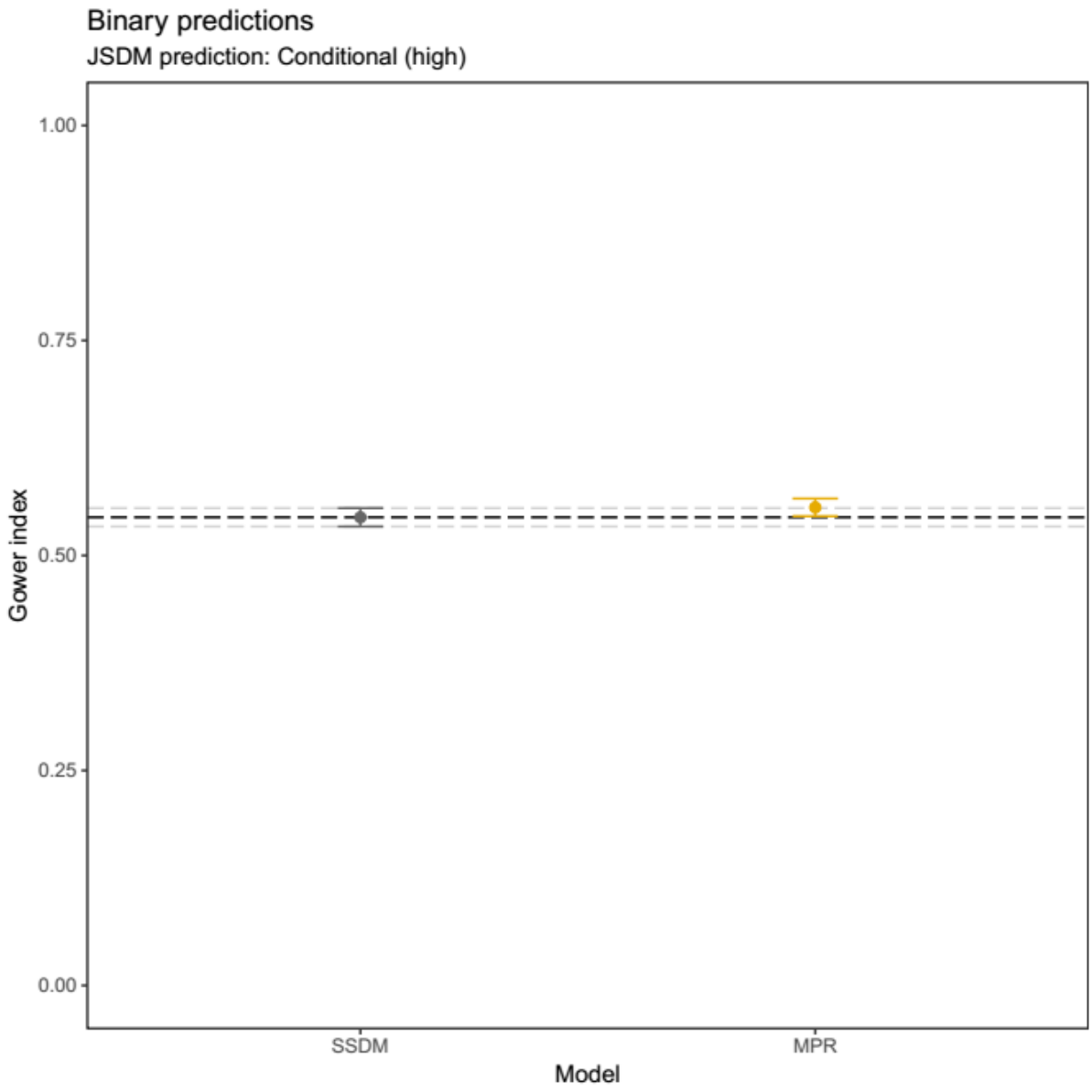


Figure 133: Performance of the SSDM and JSDM for the Gower index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

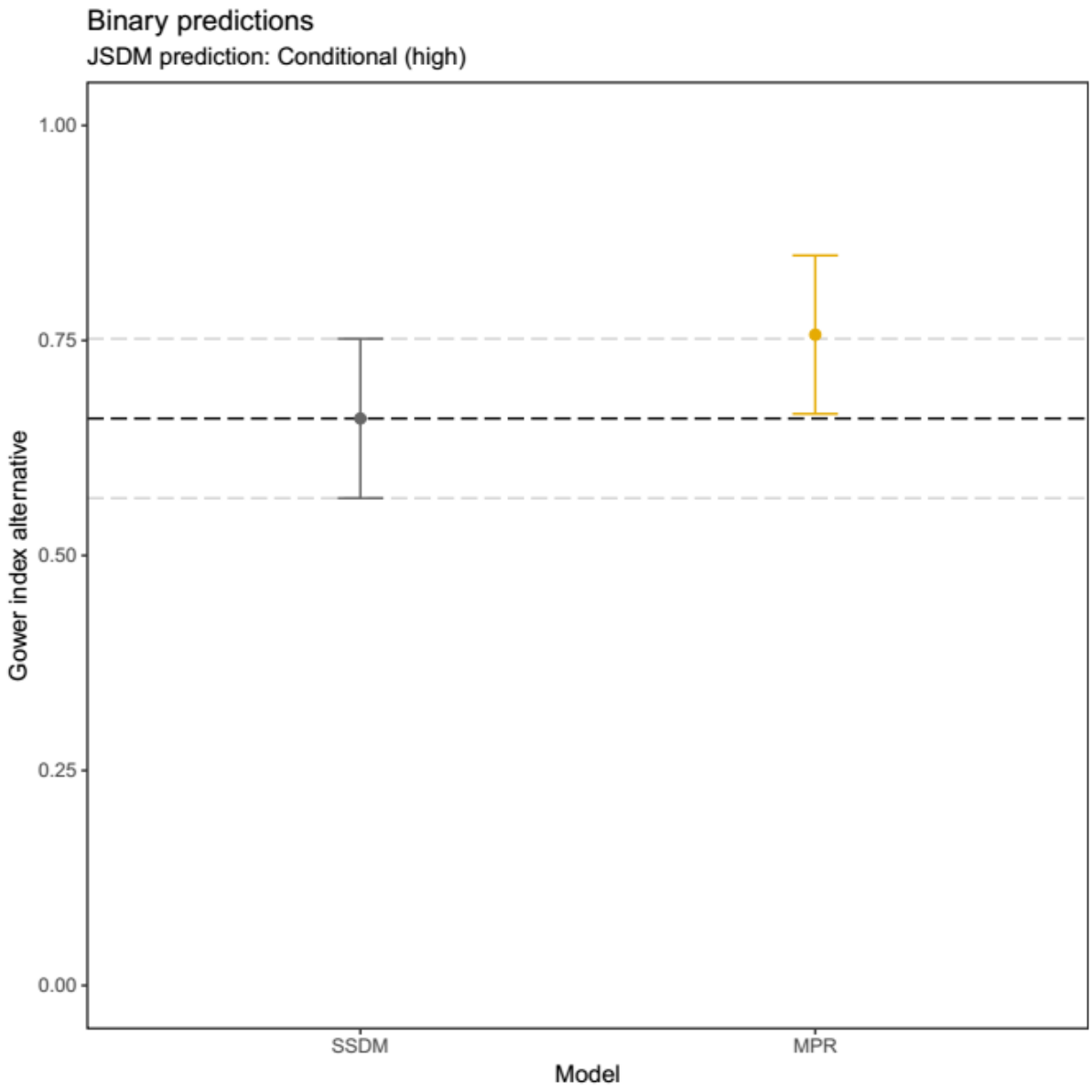


Figure 134: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

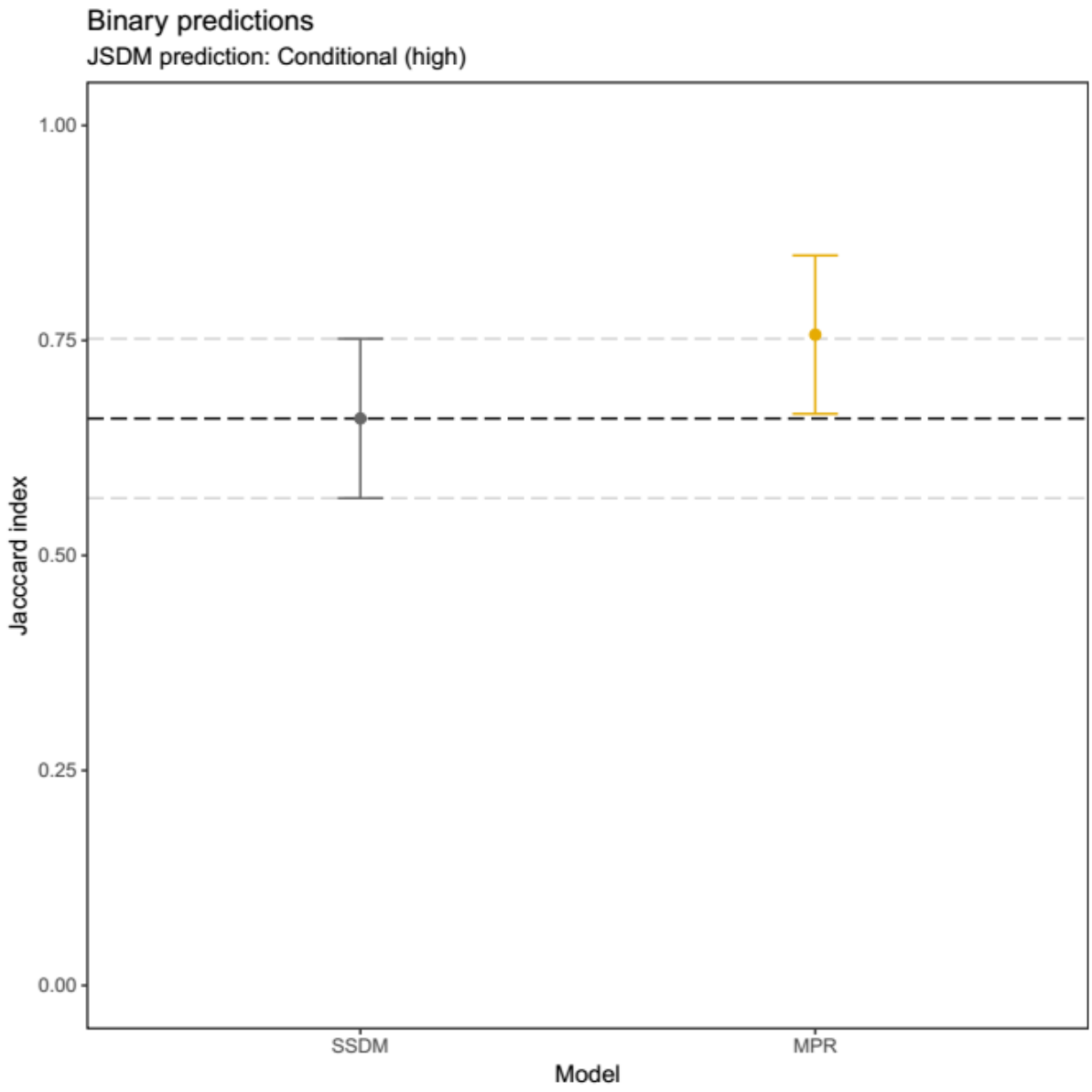


Figure 135: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

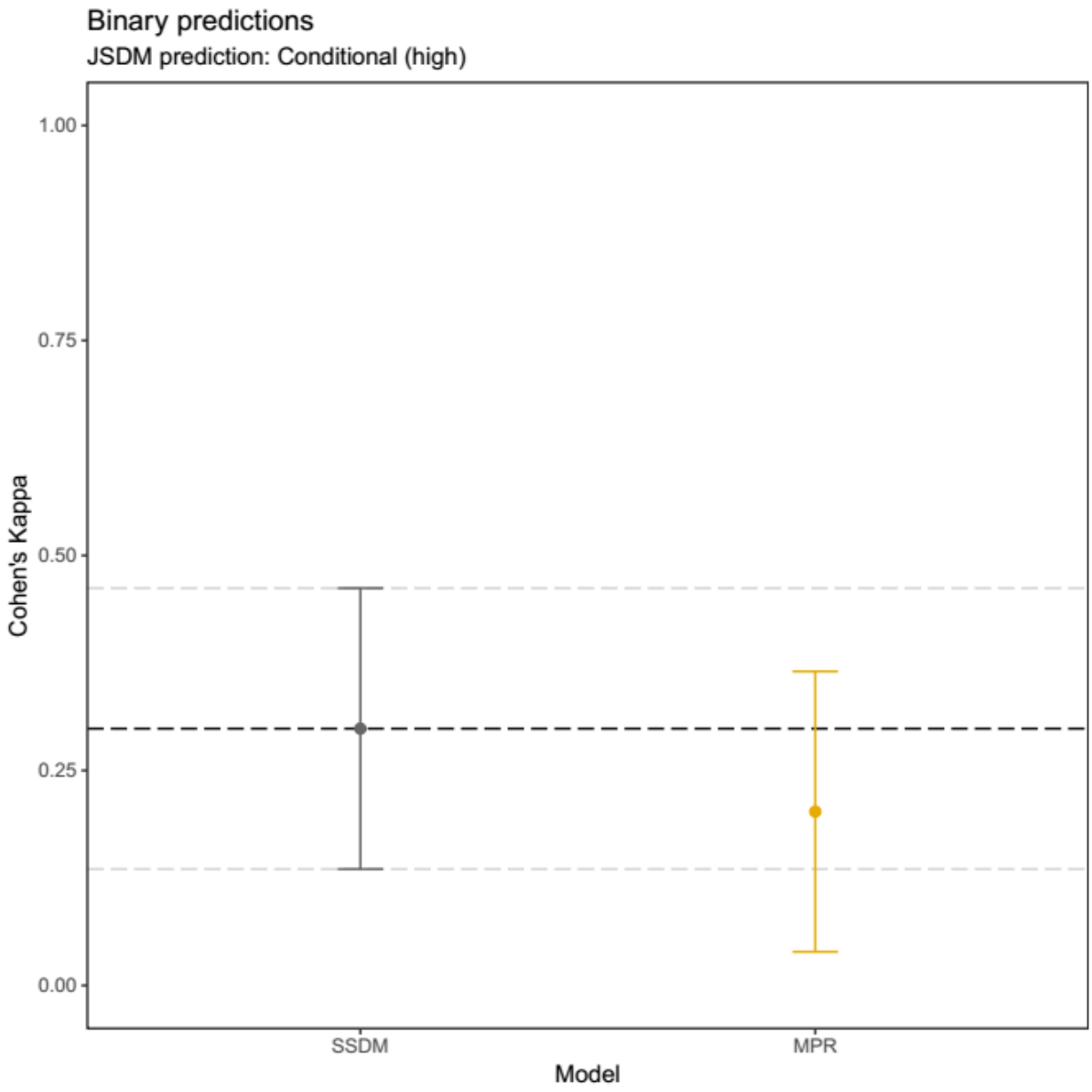


Figure 136: Performance of the SSDM and JSDM for the Cohen's Kappa evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

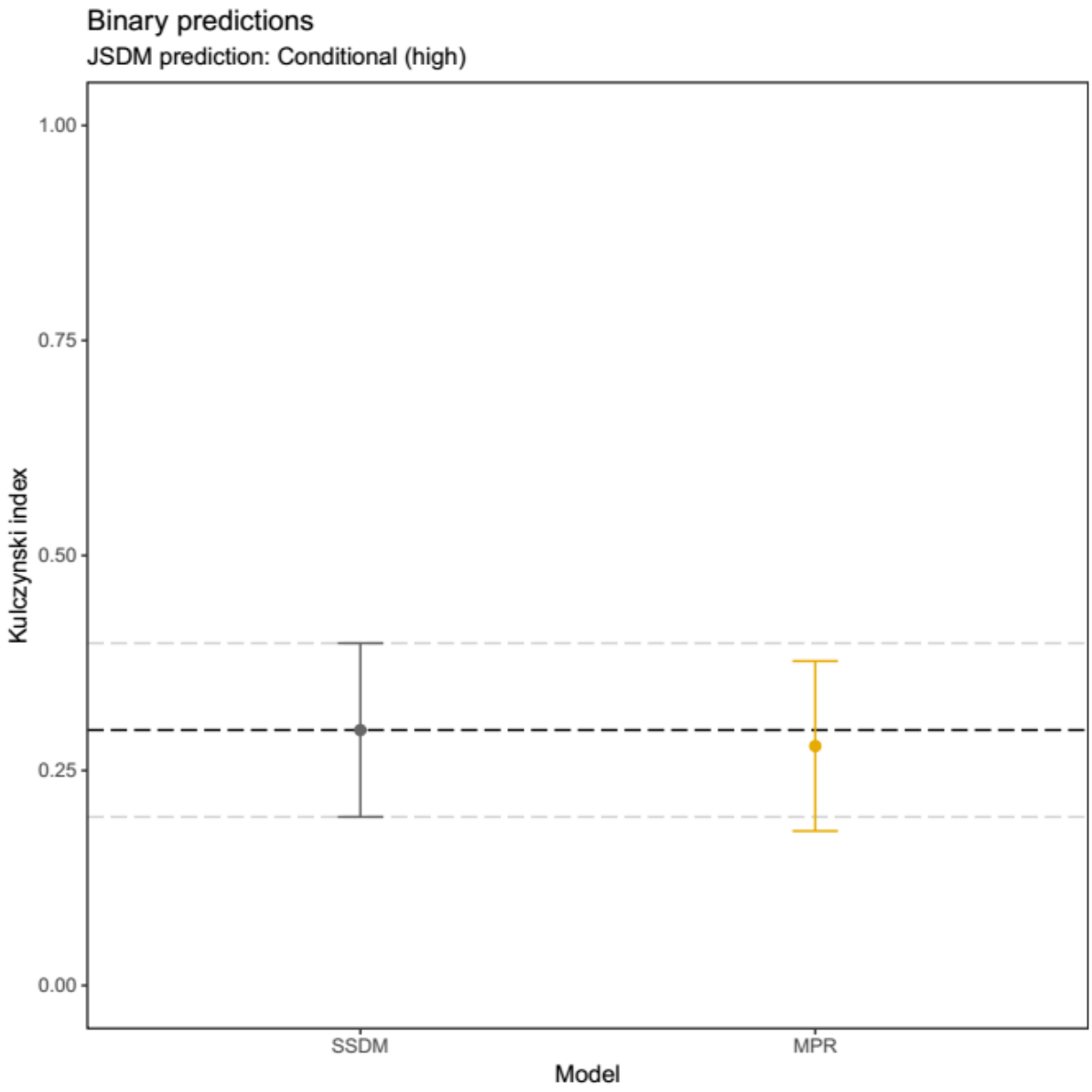


Figure 137: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

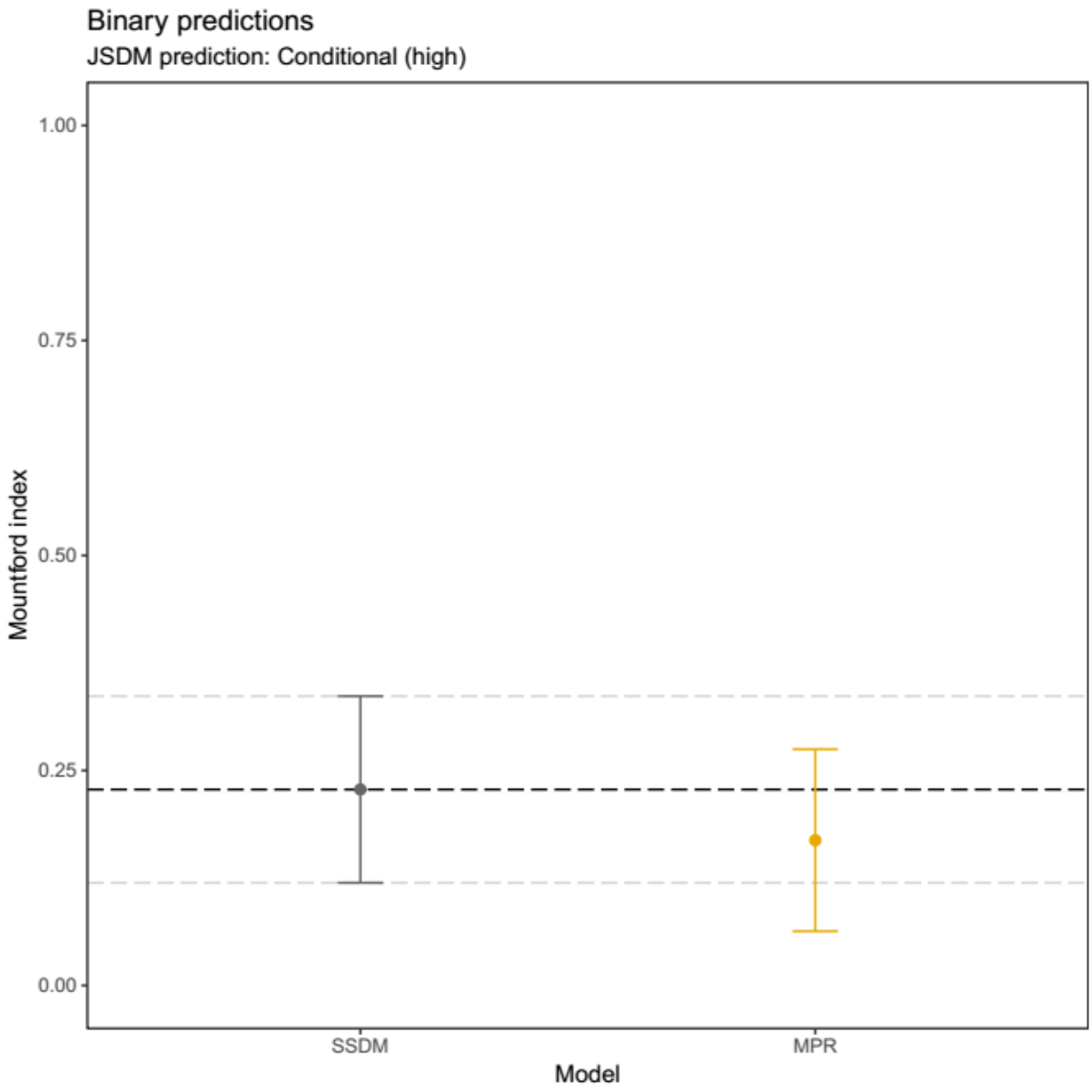


Figure 138: Performance of the SSDM and JSDM for the Mountford index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

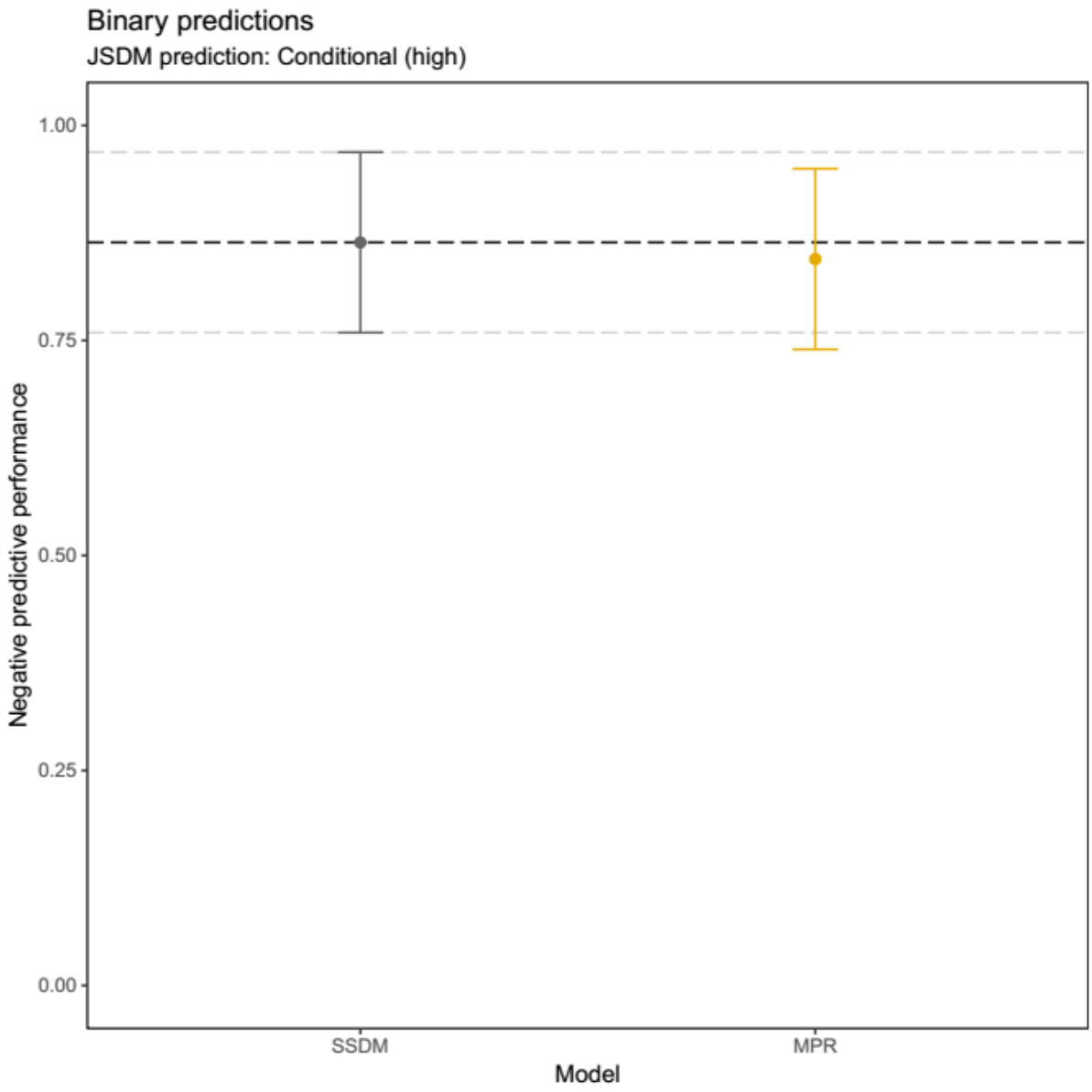


Figure 139: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

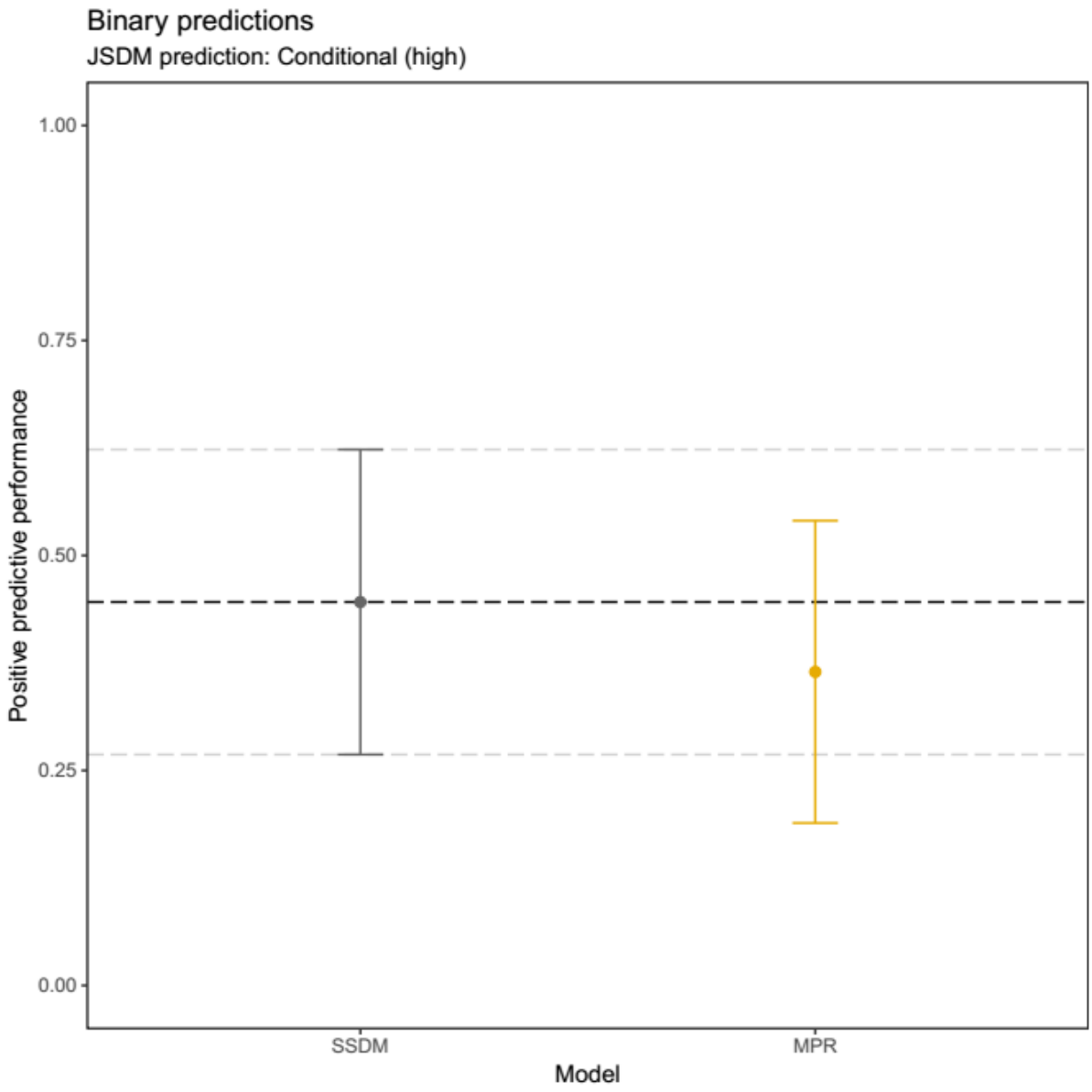


Figure 140: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

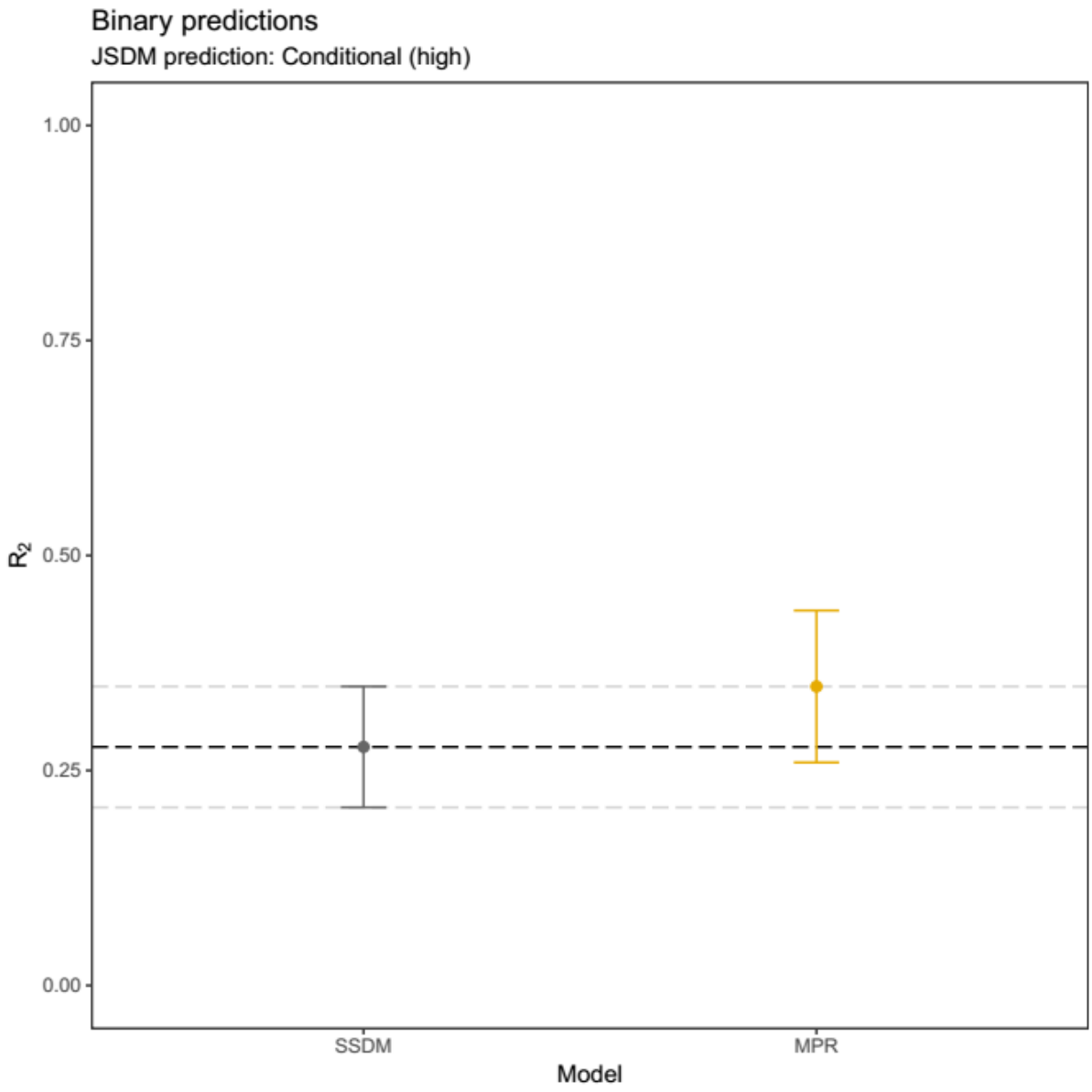


Figure 141: Performance of the SSDM and JSDM for the R^2 evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

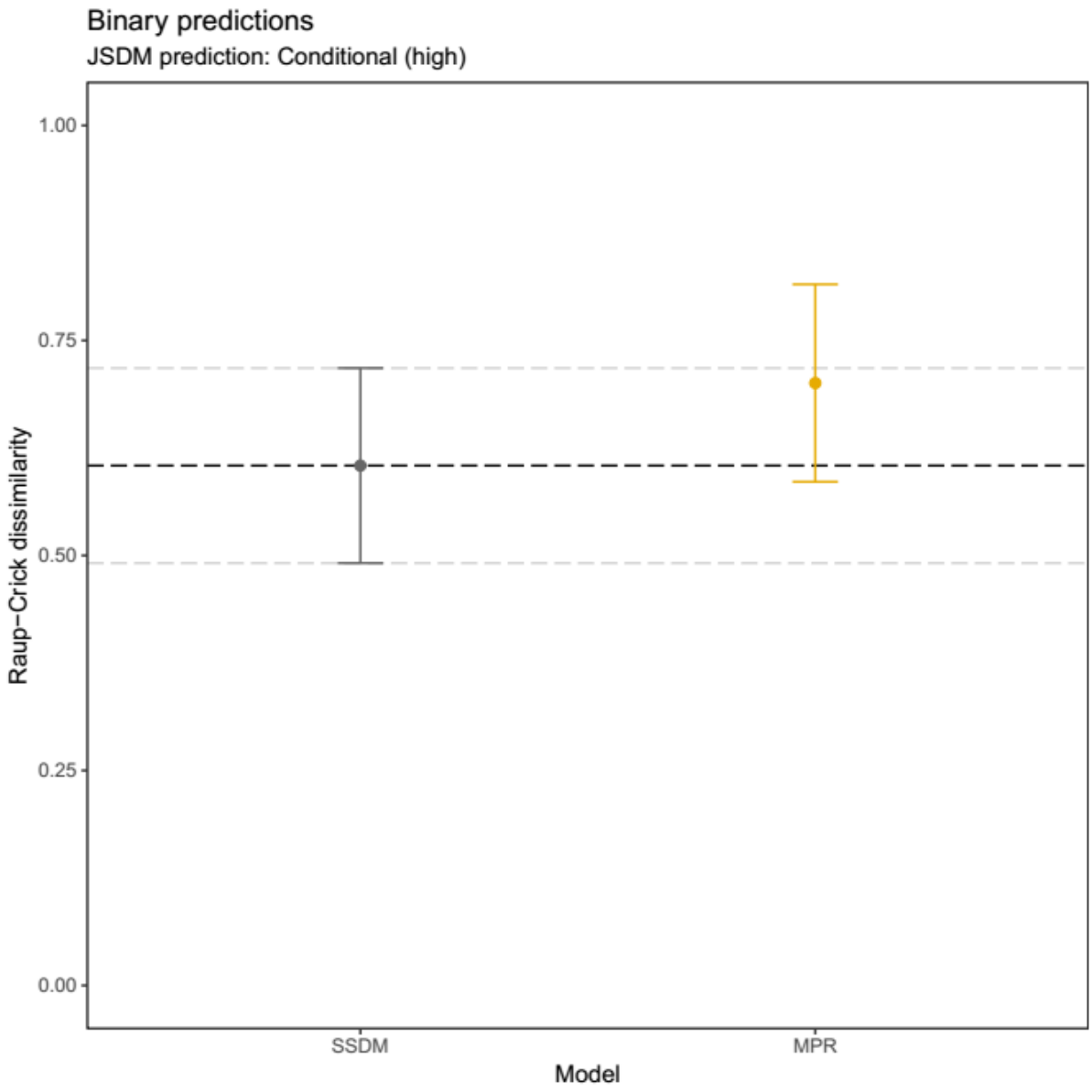


Figure 142: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

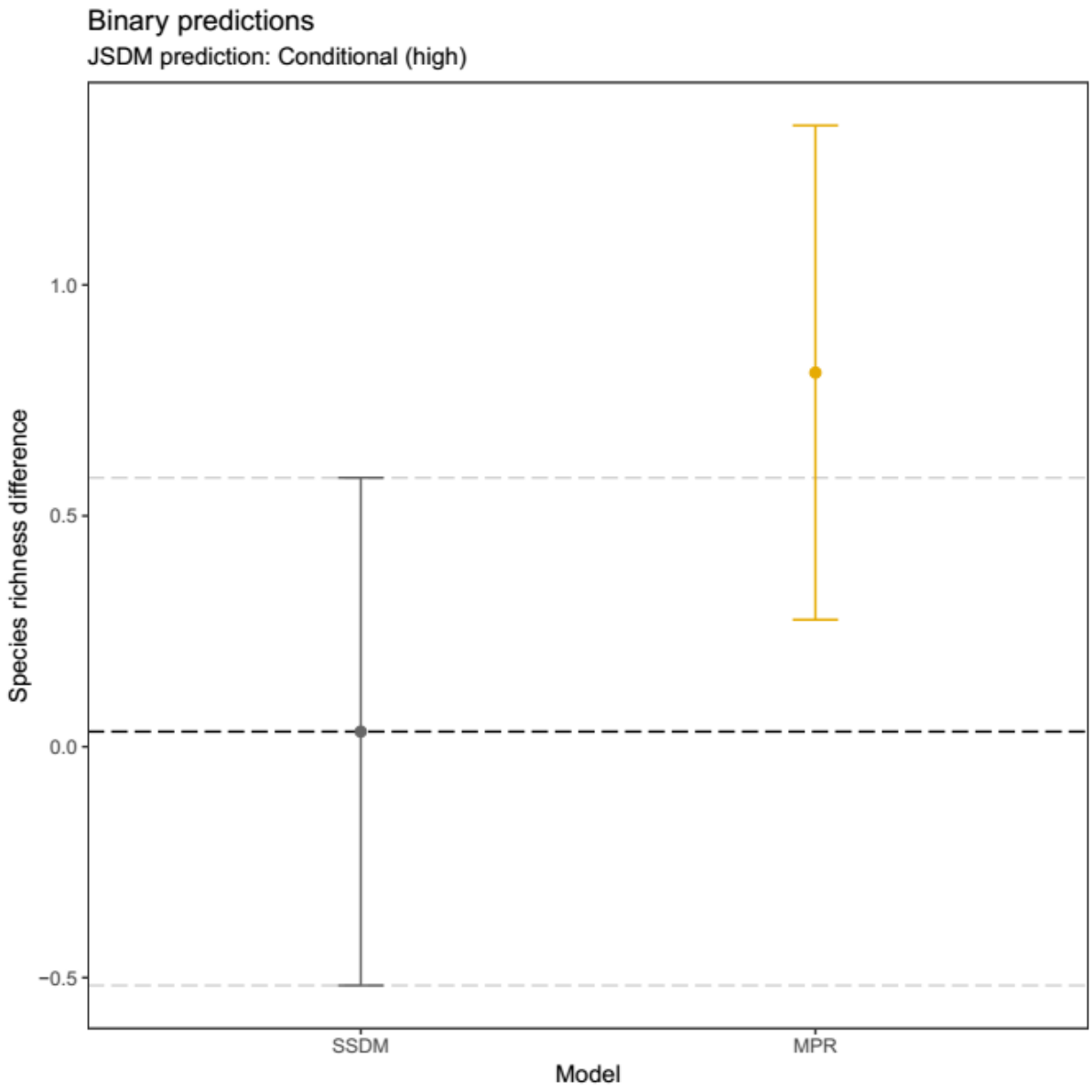


Figure 143: Performance of the SSDM and JSDM for the species richness difference evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

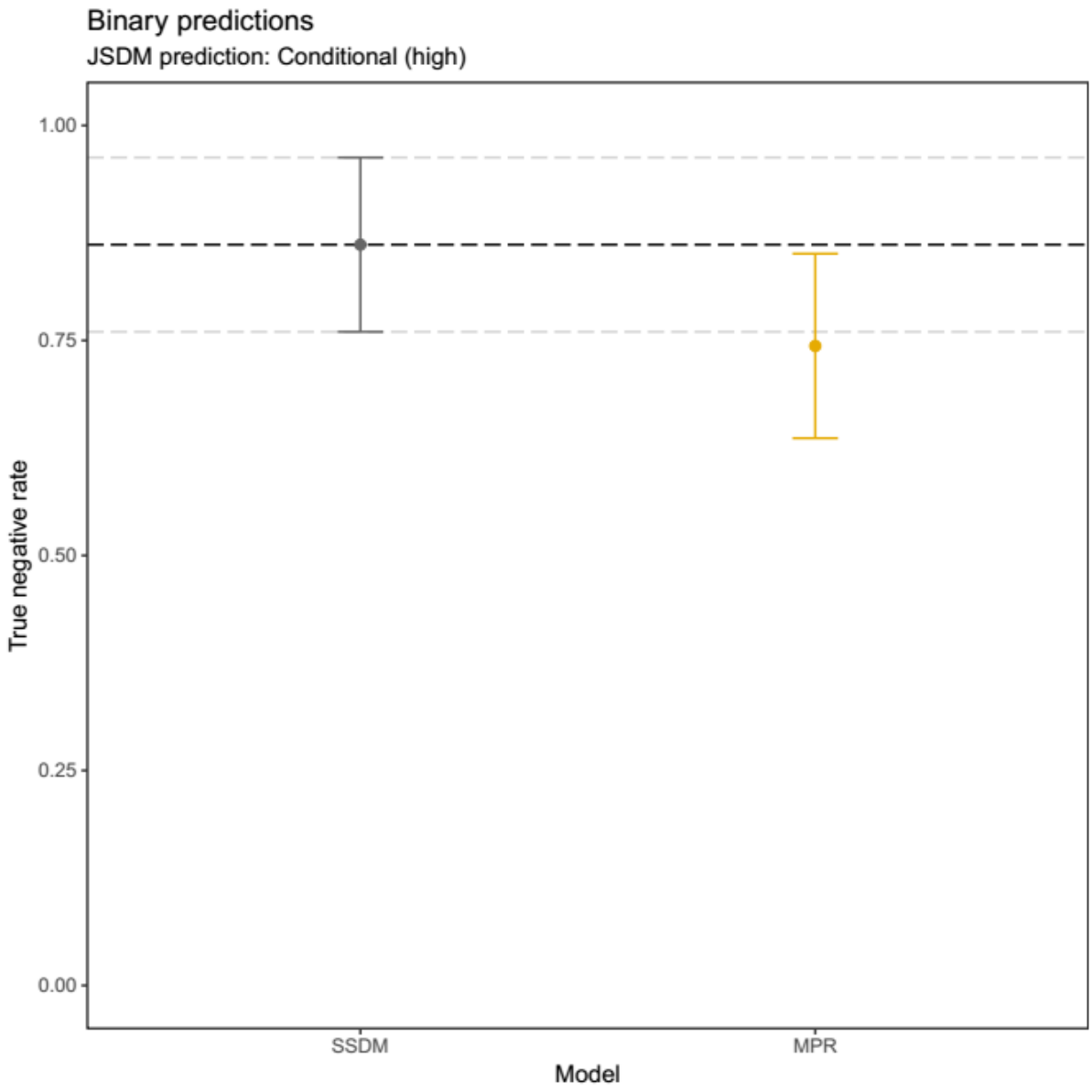


Figure 144: Performance of the SSDM and JSDM for the true negative rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

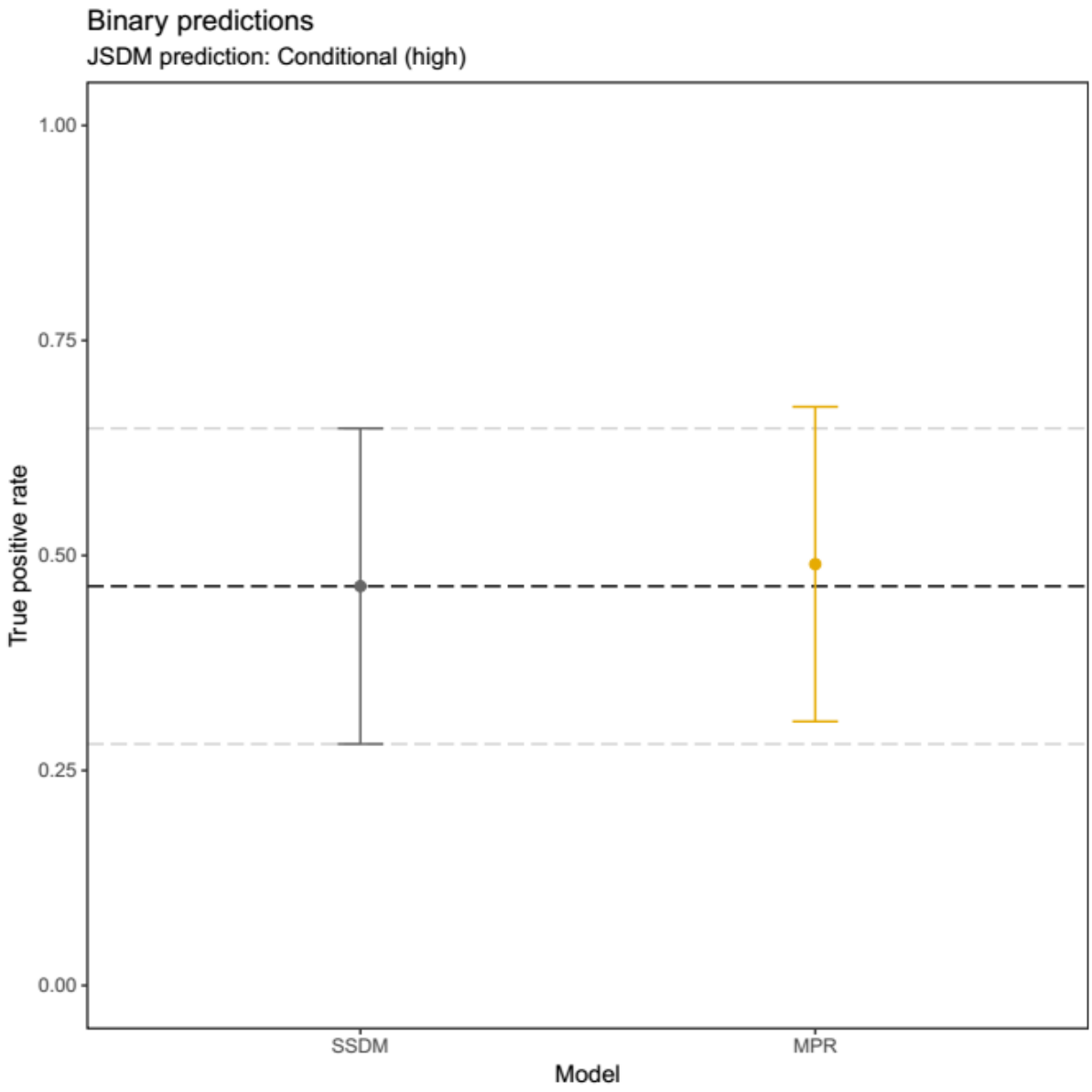


Figure 145: Performance of the SSDM and JSDM for the true positive rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

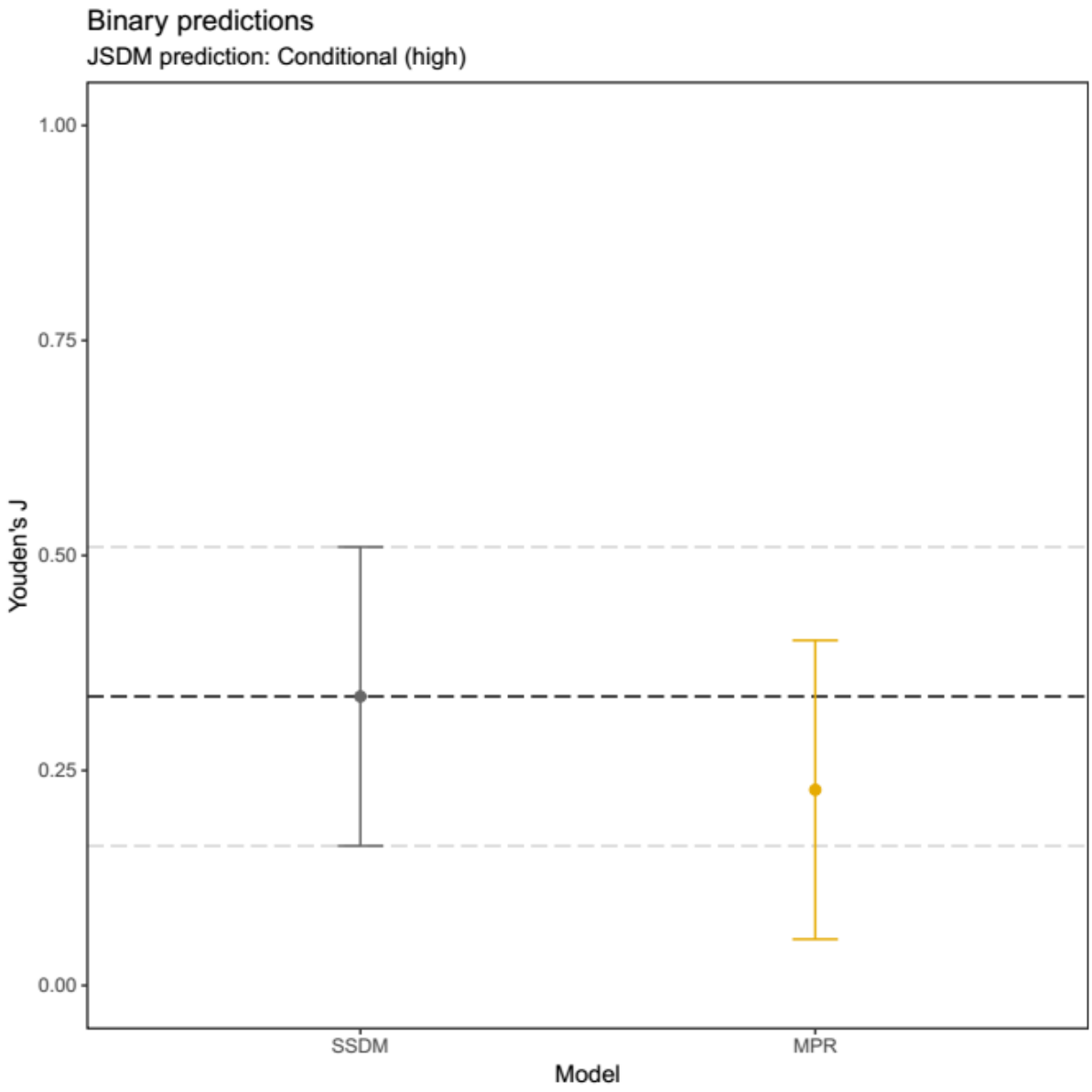


Figure 146: Performance of the SSDM and JSDM for the Youden's J evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

7. Conditional marginal (low-prevalence) predictions

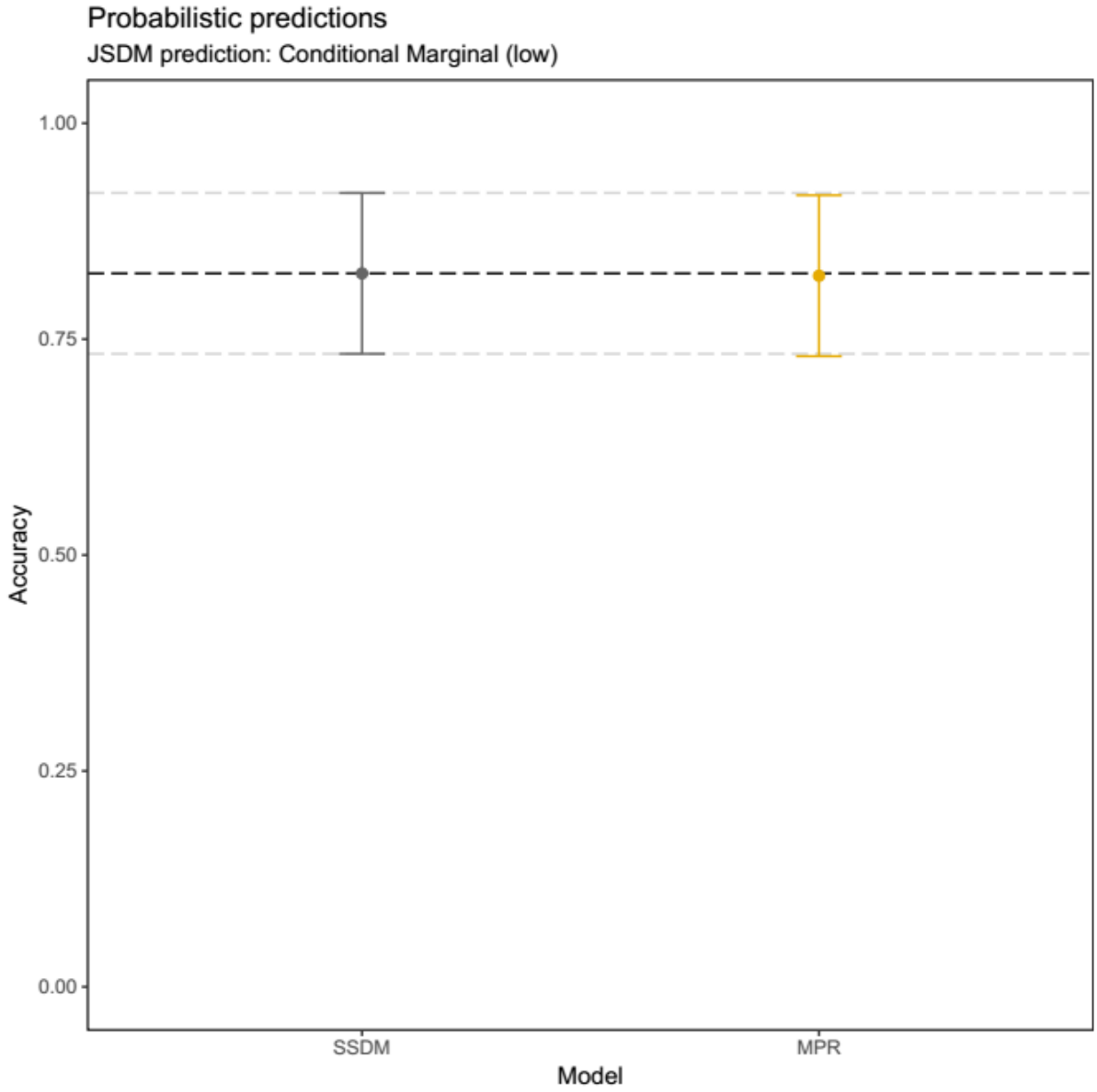


Figure 147: Performance of the SSDM and JSDM for the accuracy evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

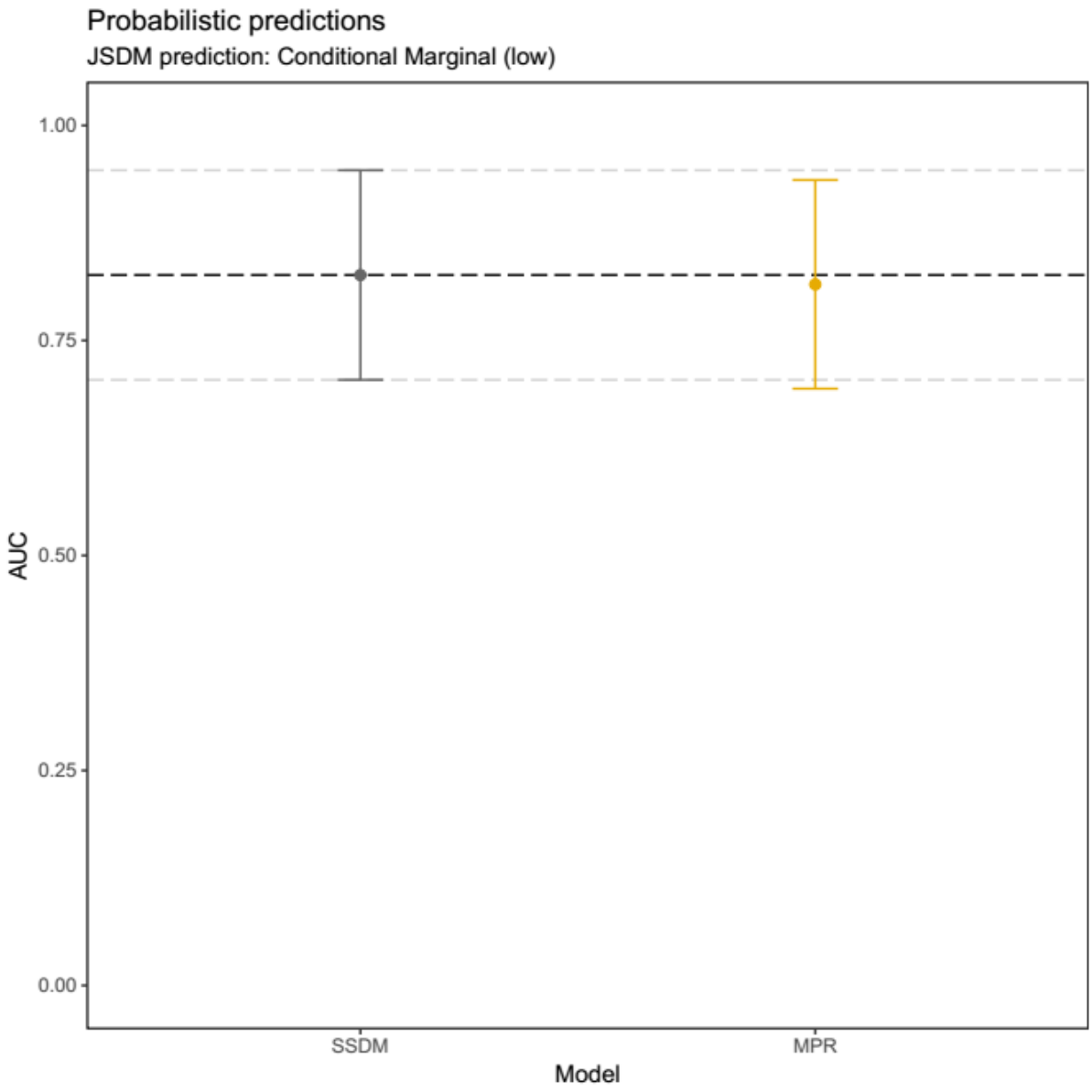


Figure 148: Performance of the SSDM and JSDM for the AUC evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

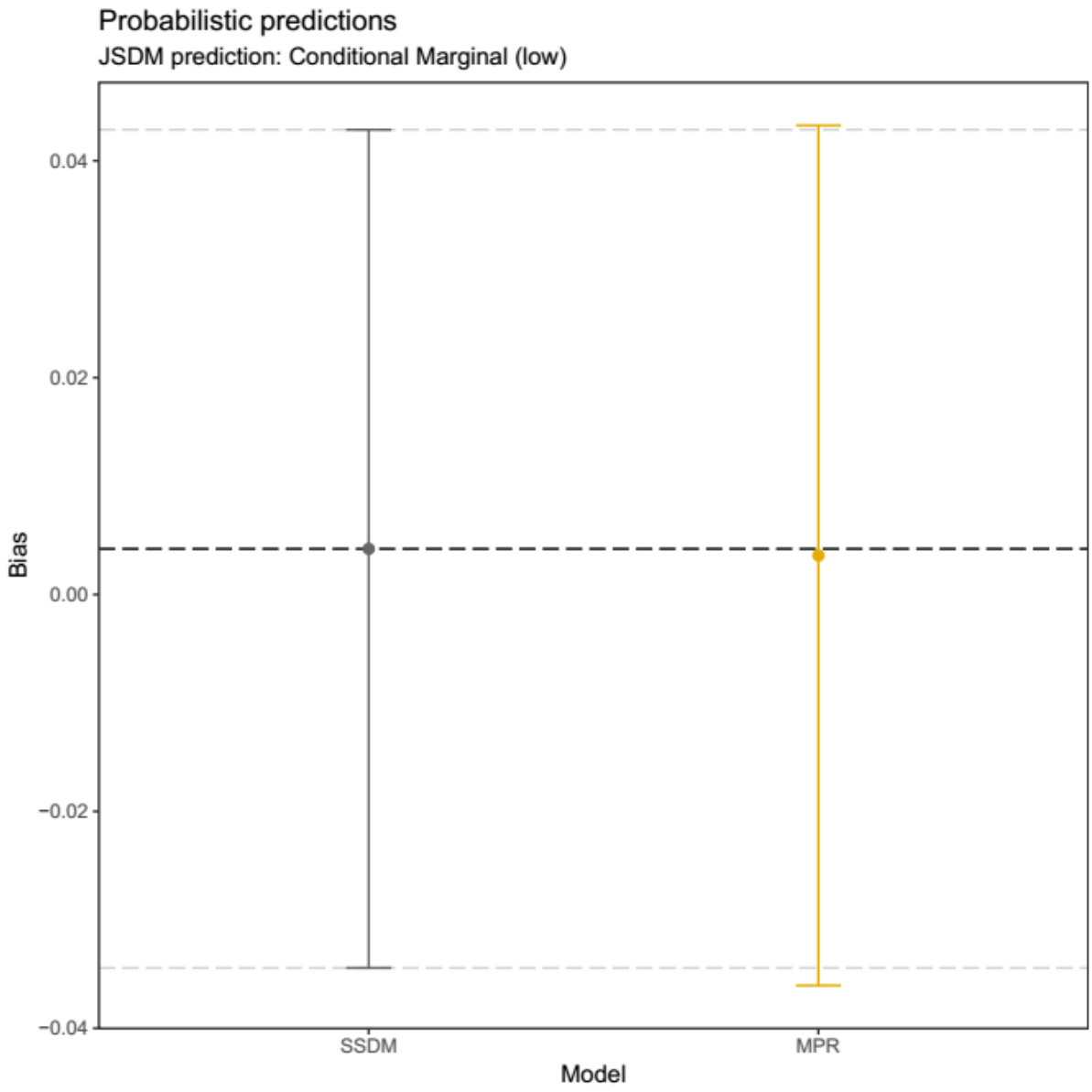


Figure 149: Performance of the SSDM and JSDM for the mean error/bias evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

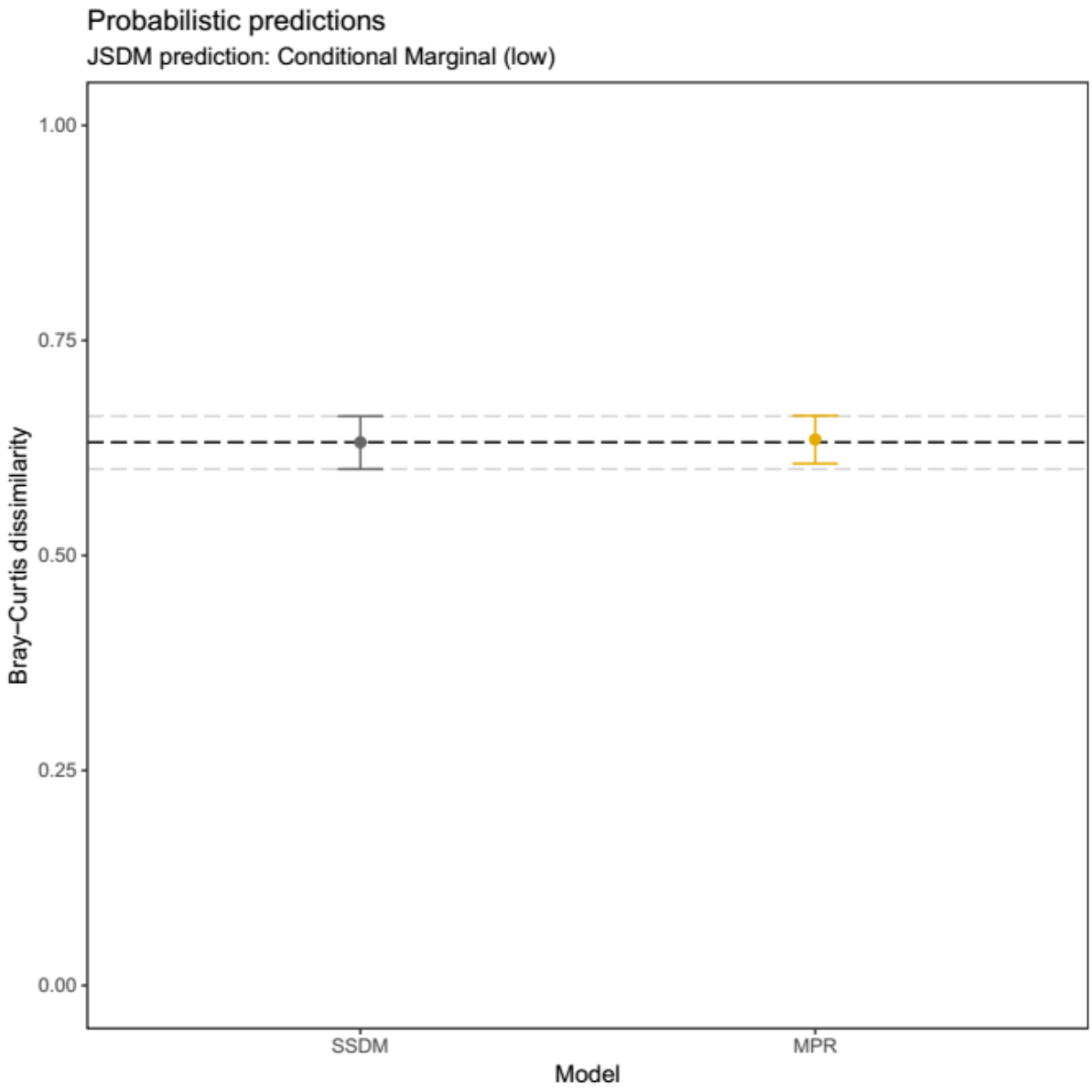


Figure 150: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

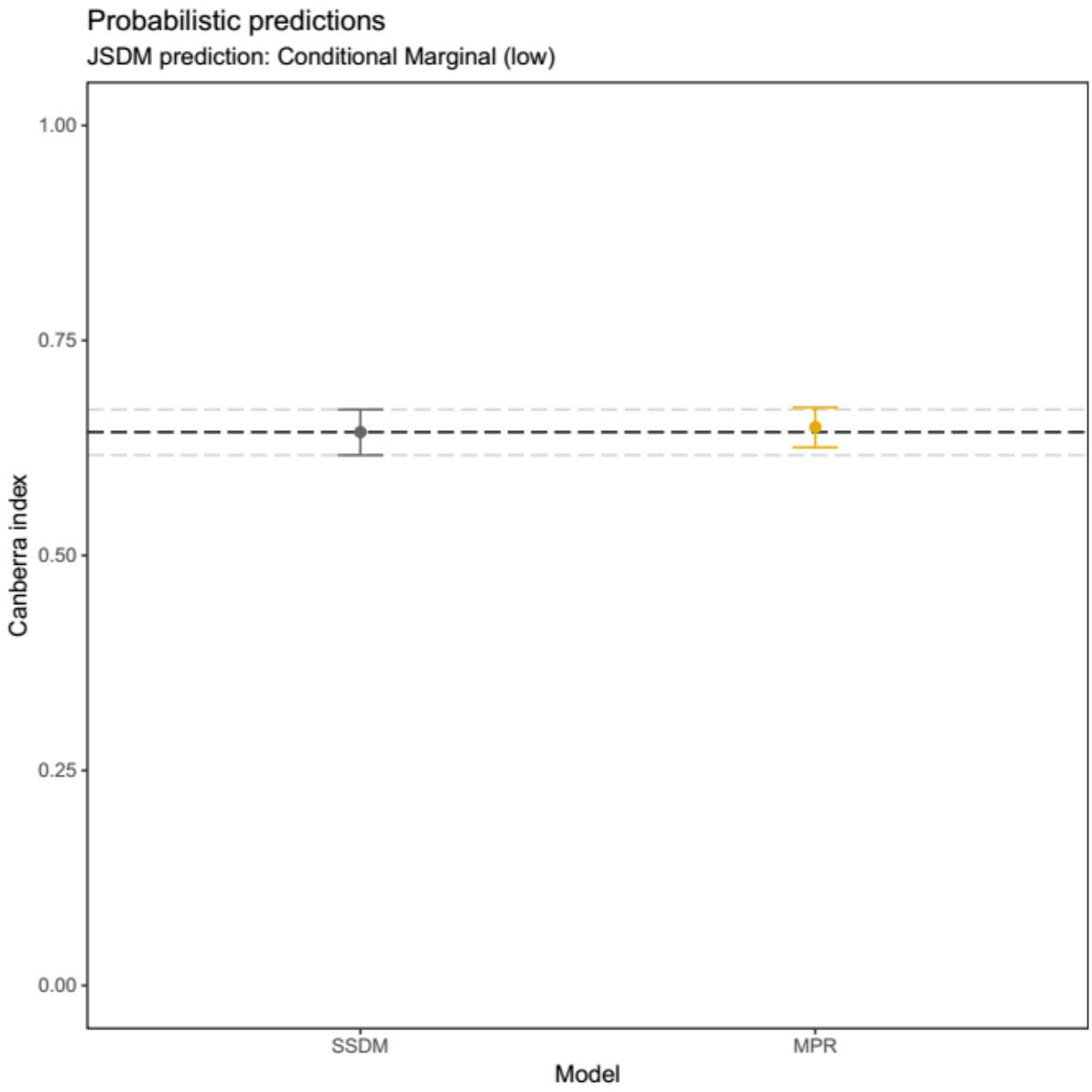


Figure 151: Performance of the SSDM and JSDM for the Canberra index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

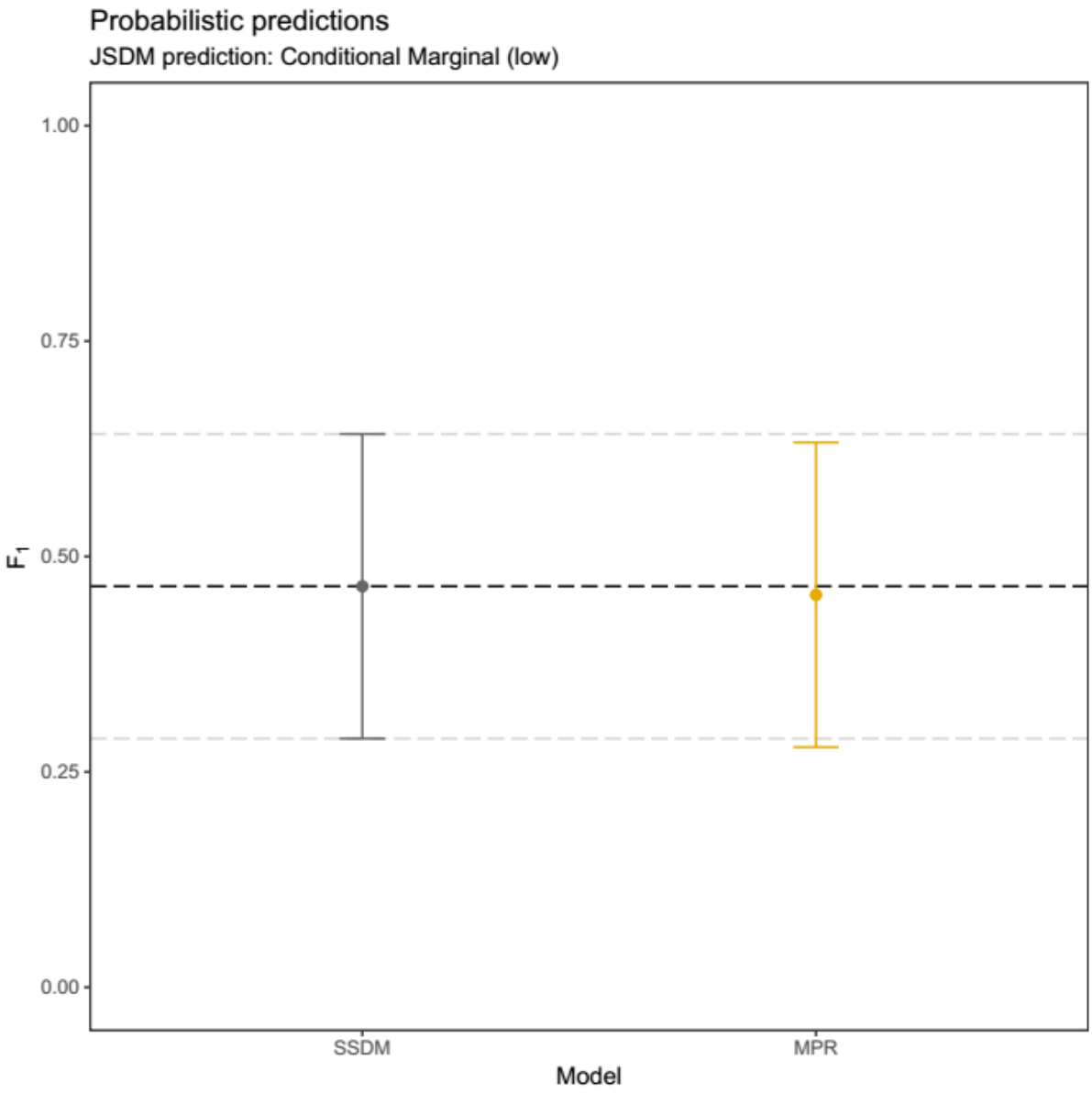


Figure 152: Performance of the SSDM and JSDM for the F_1 evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

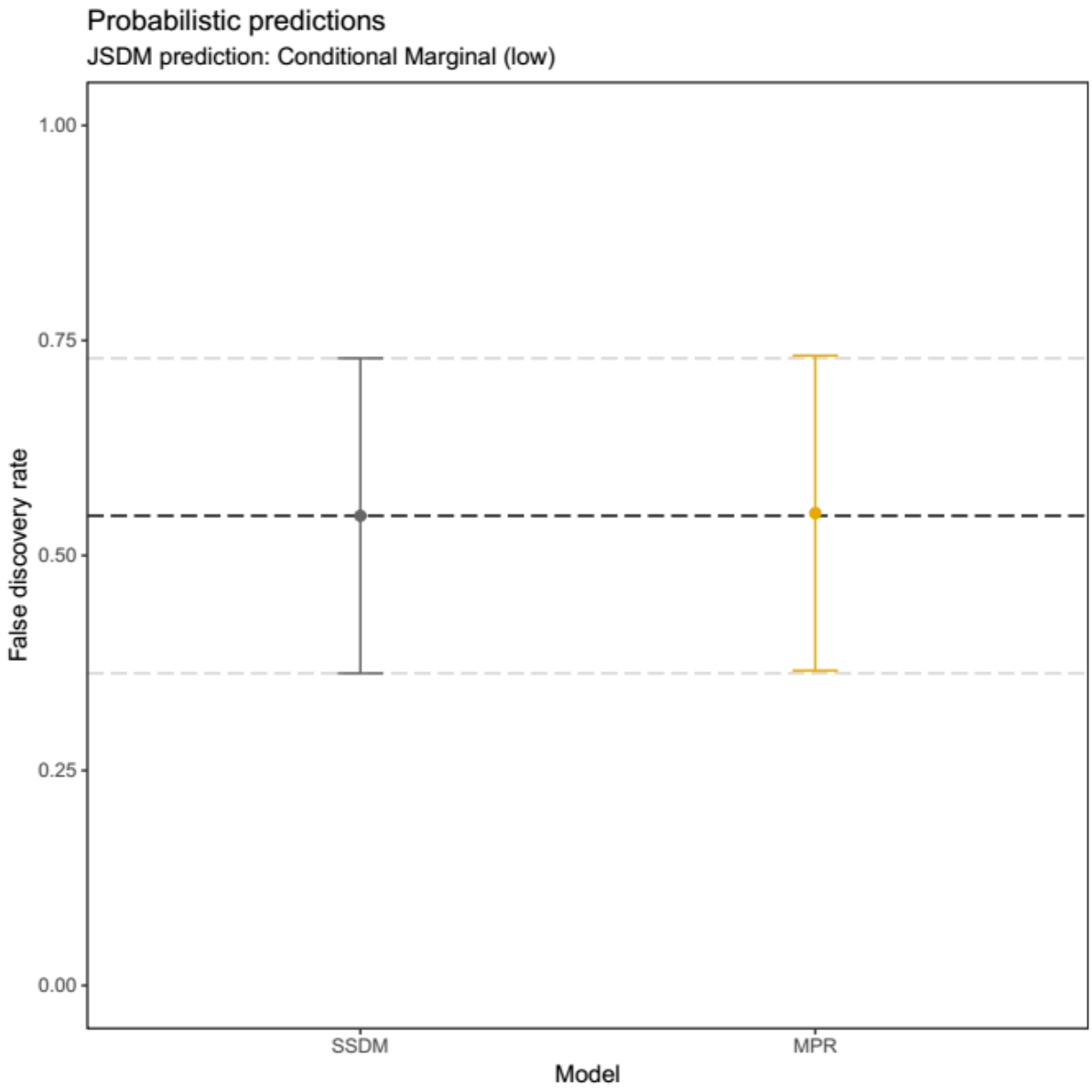


Figure 153: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

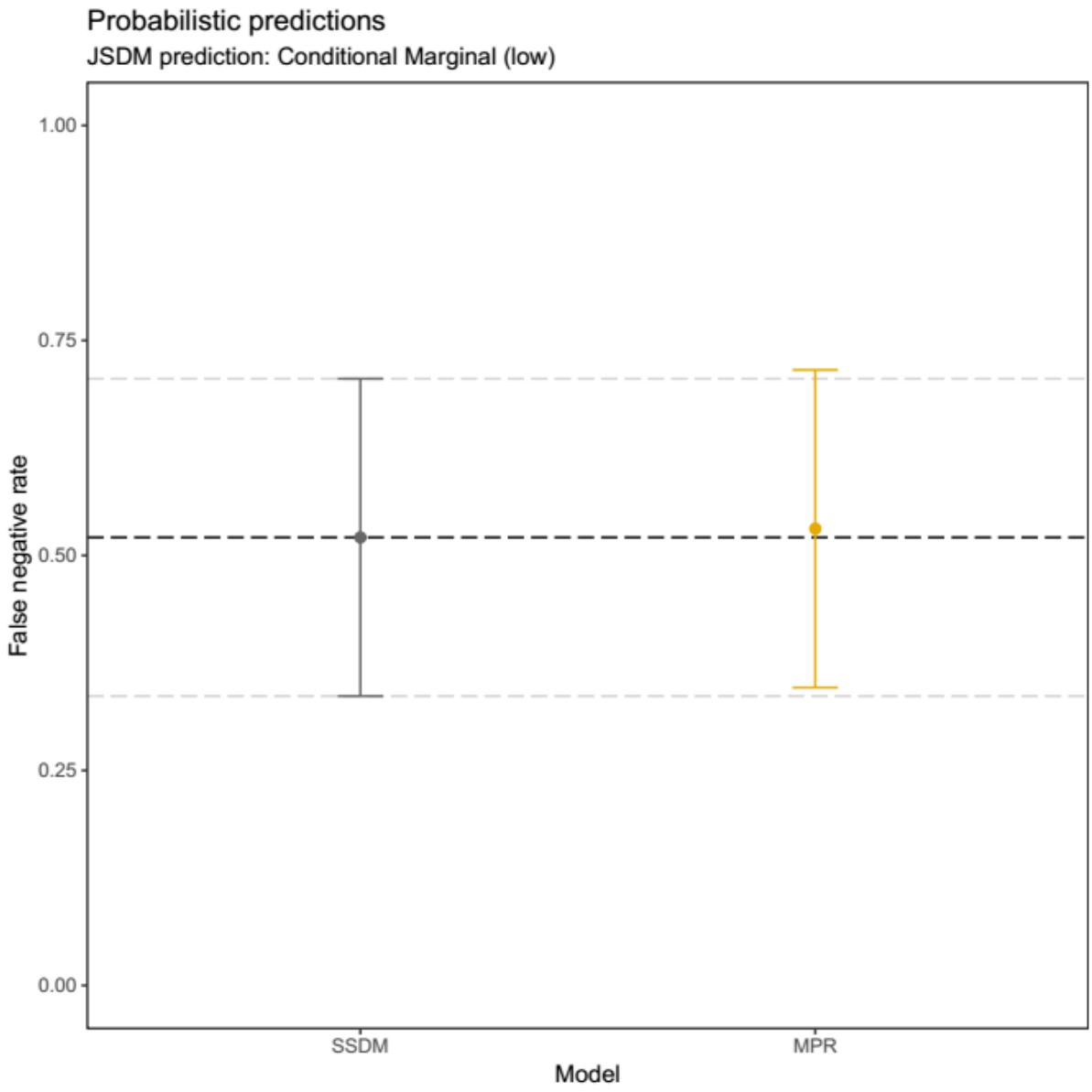


Figure 154: Performance of the SSDM and JSDM for the false negative rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

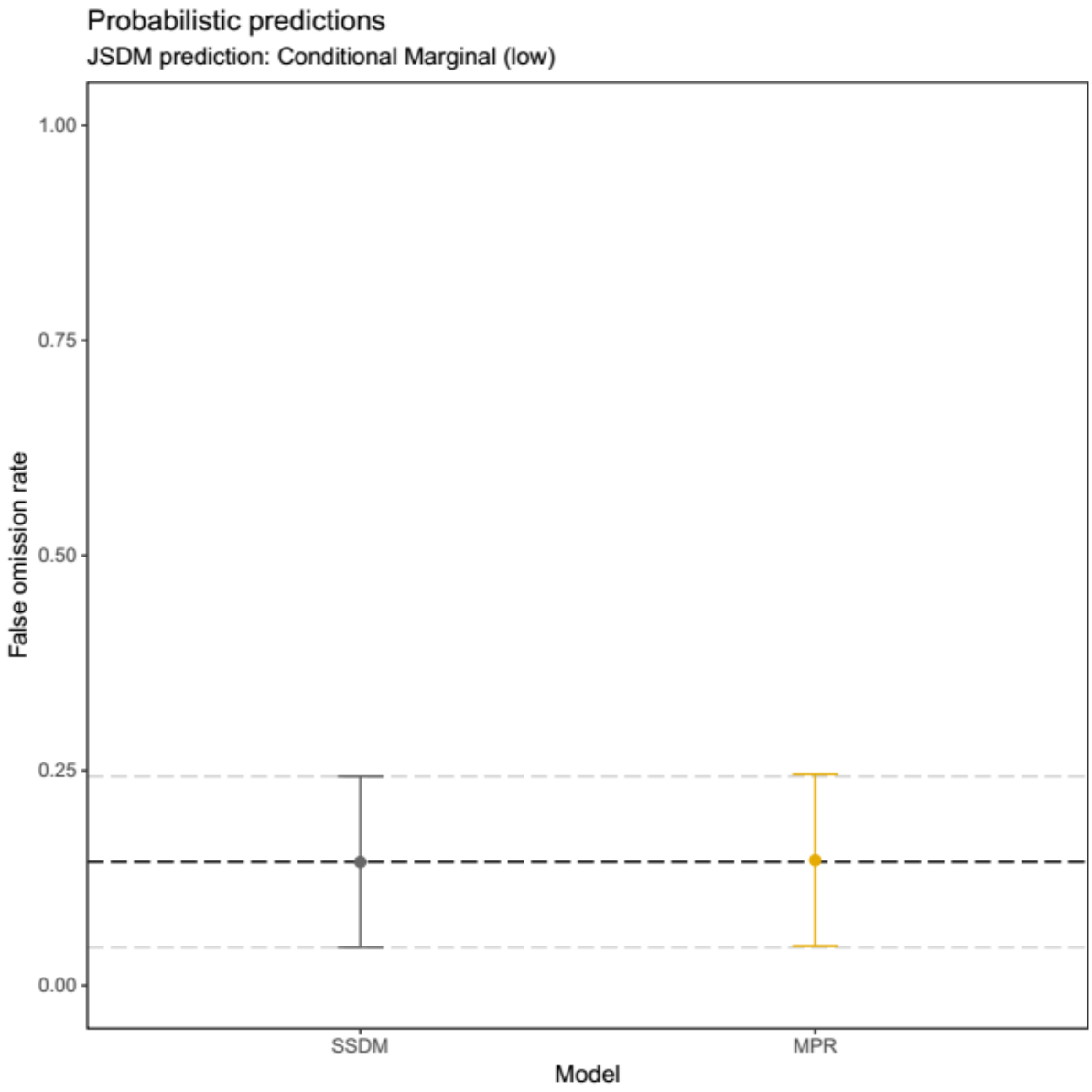


Figure 155: Performance of the SSDM and JSDM for the false omission rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

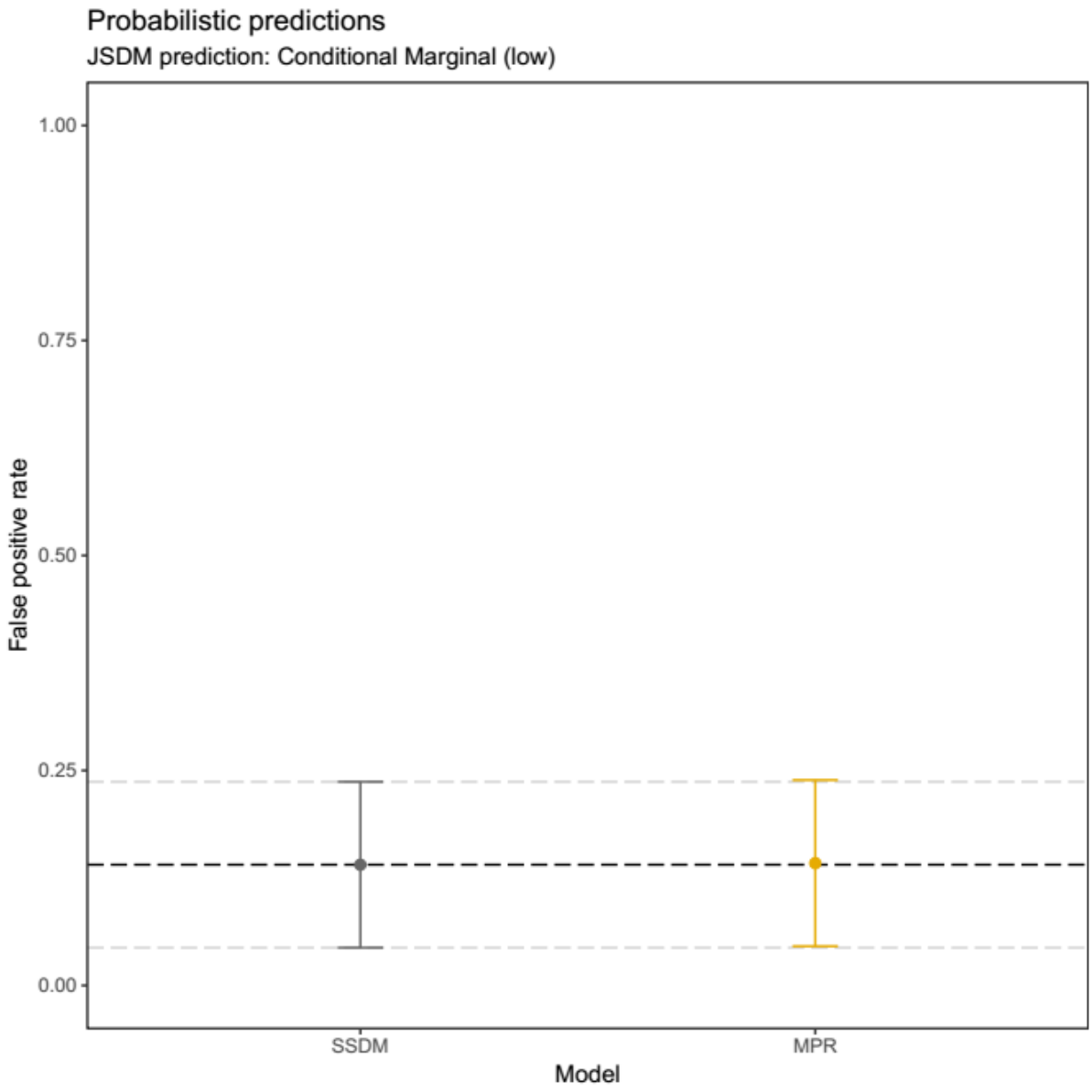


Figure 156: Performance of the SSDM and JSDM for the false positive rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

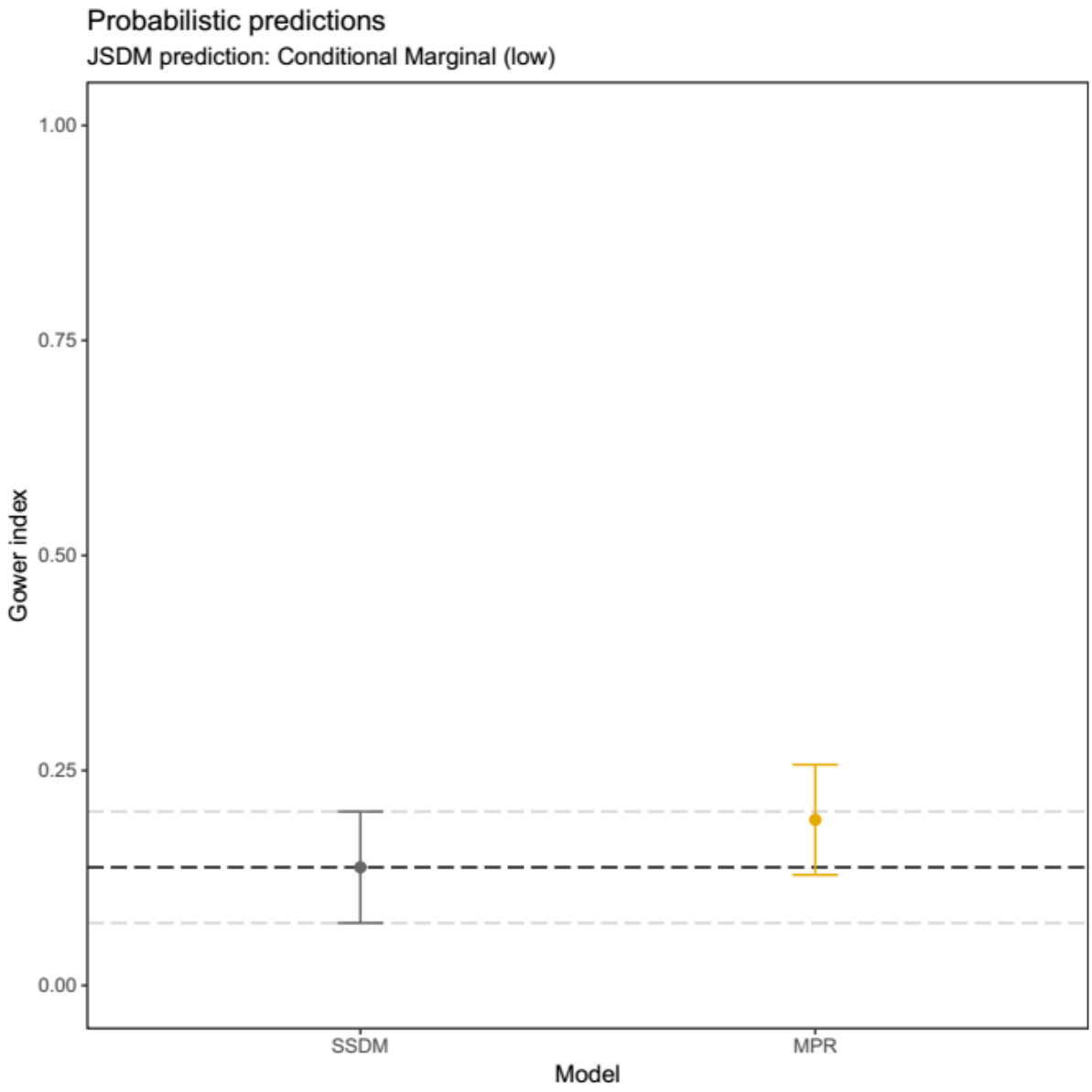


Figure 157: Performance of the SSDM and JSDM for the Gower index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

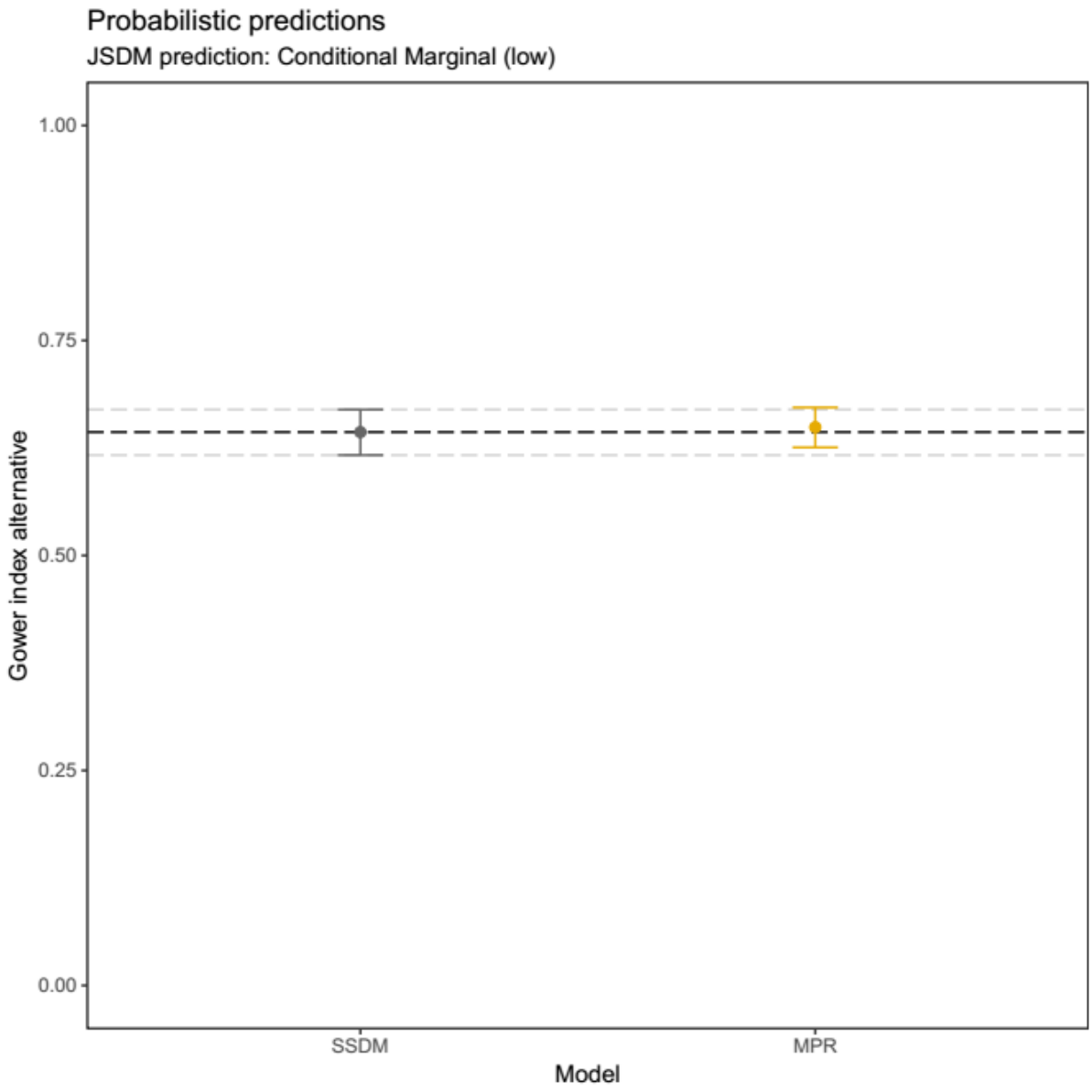


Figure 158: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

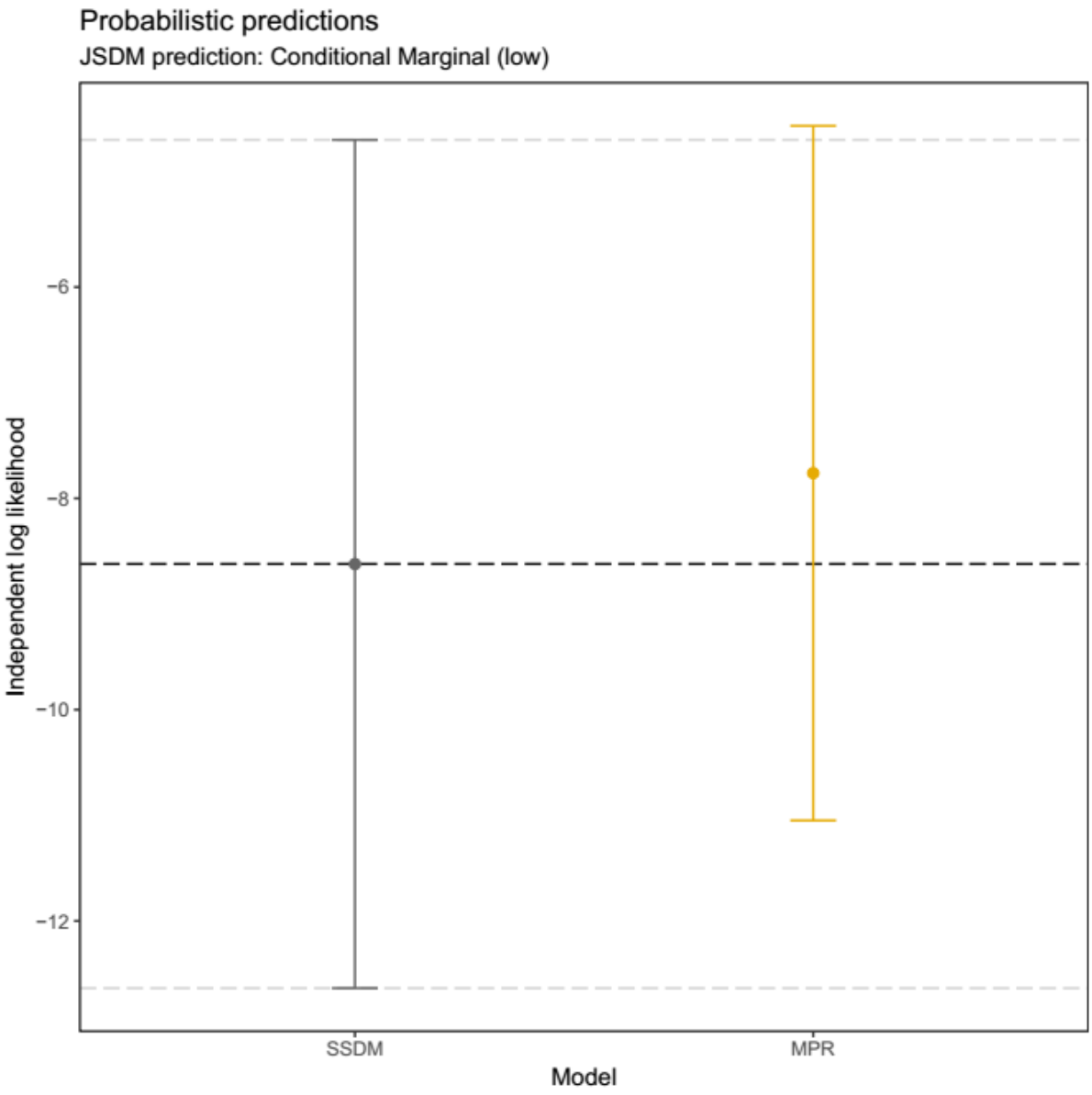


Figure 159: Performance of the SSDM and JSDM for the independent log likelihood evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

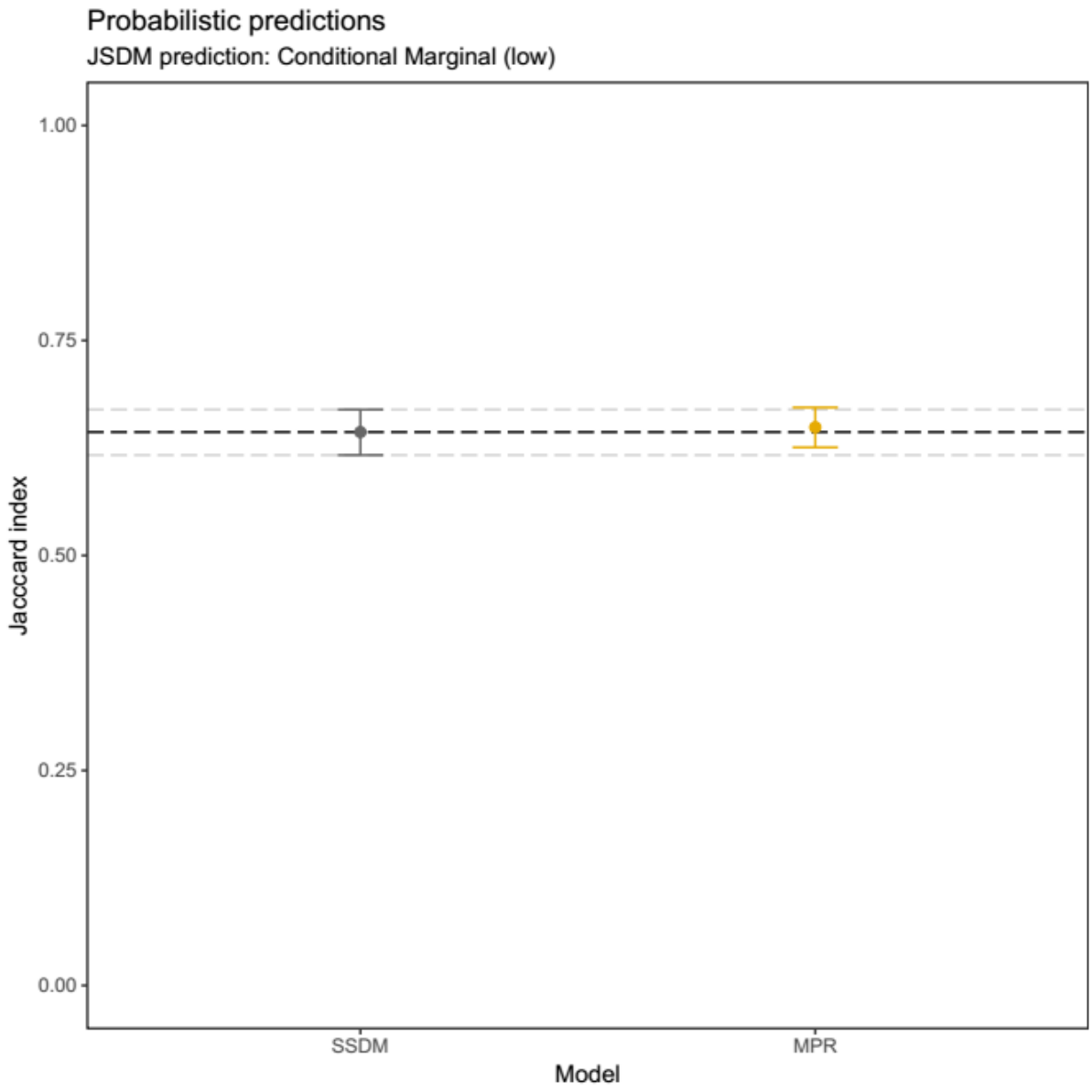


Figure 160: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

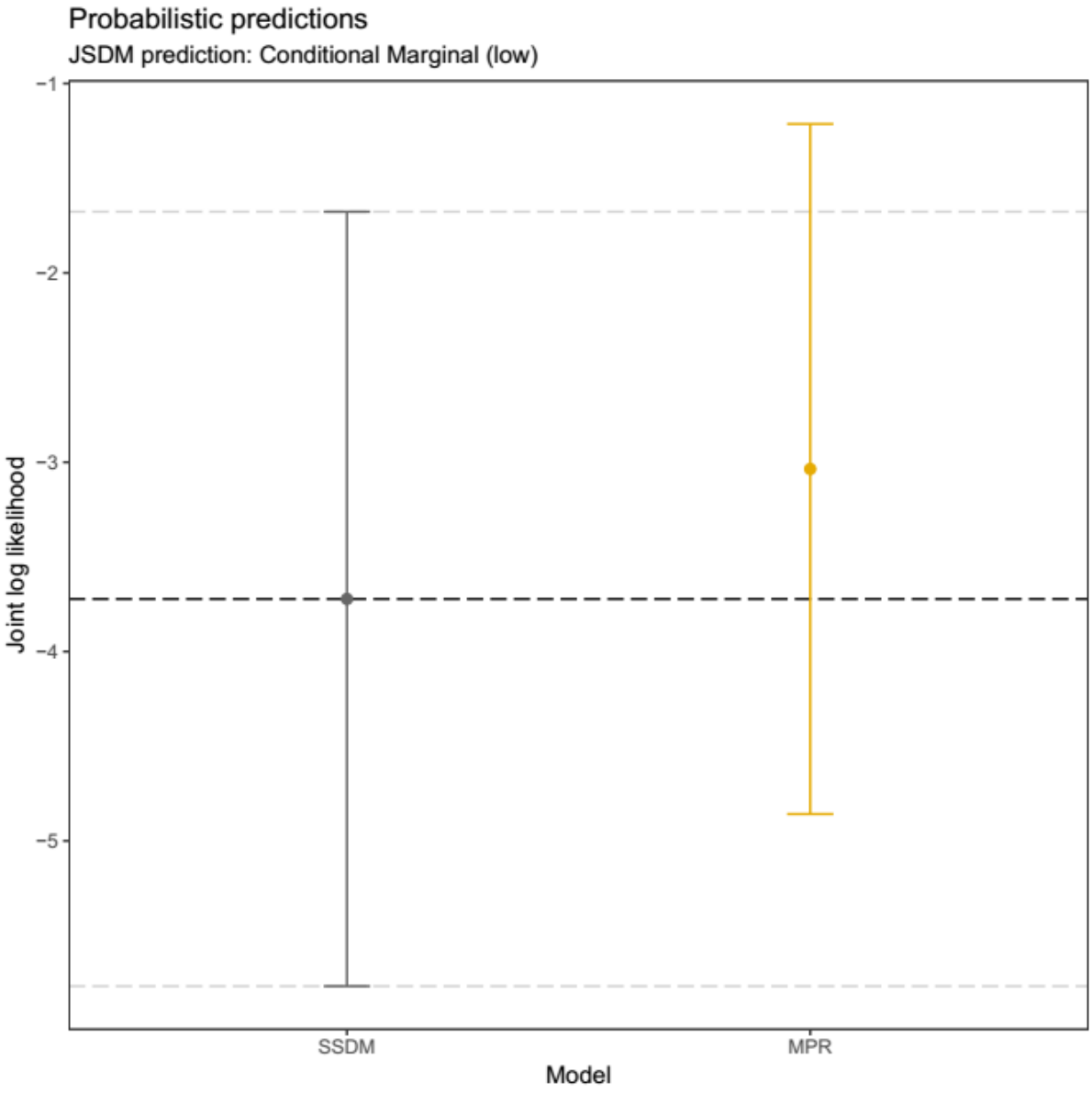


Figure 161: Performance of the SSDM and JSDM for the joint log likelihood evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

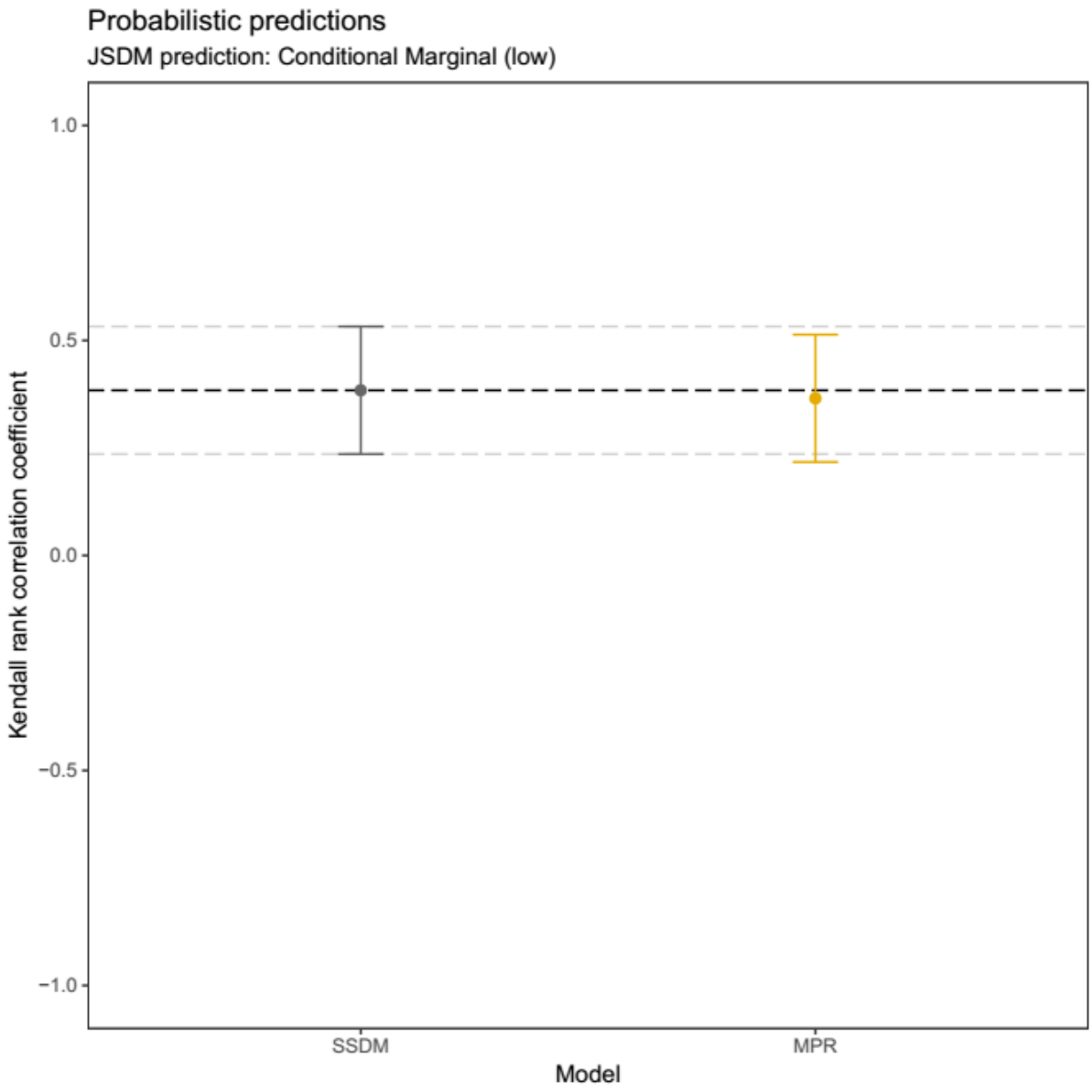


Figure 162: Performance of the SSDM and JSDM for the Kendall rank correlation coefficient evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

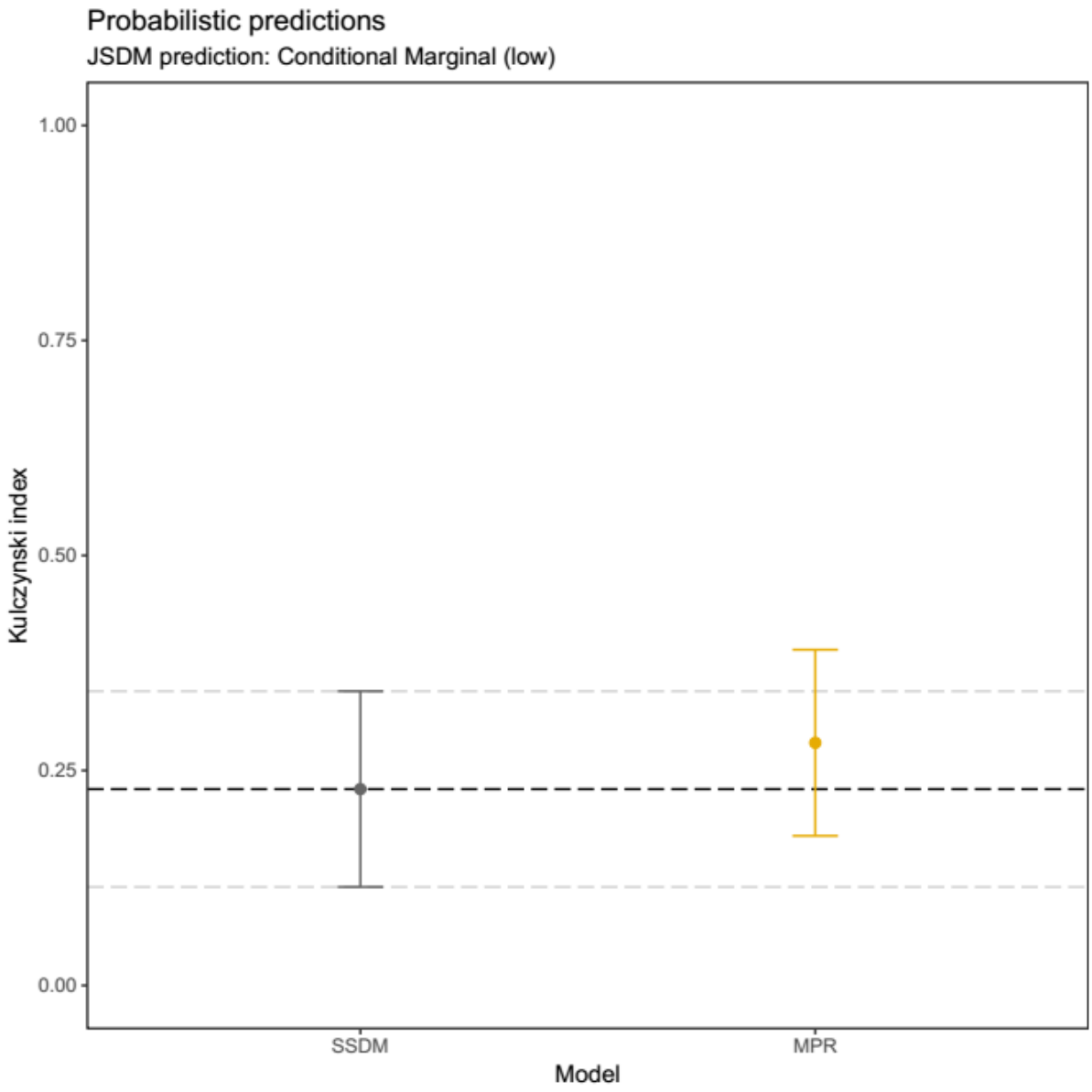


Figure 163: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

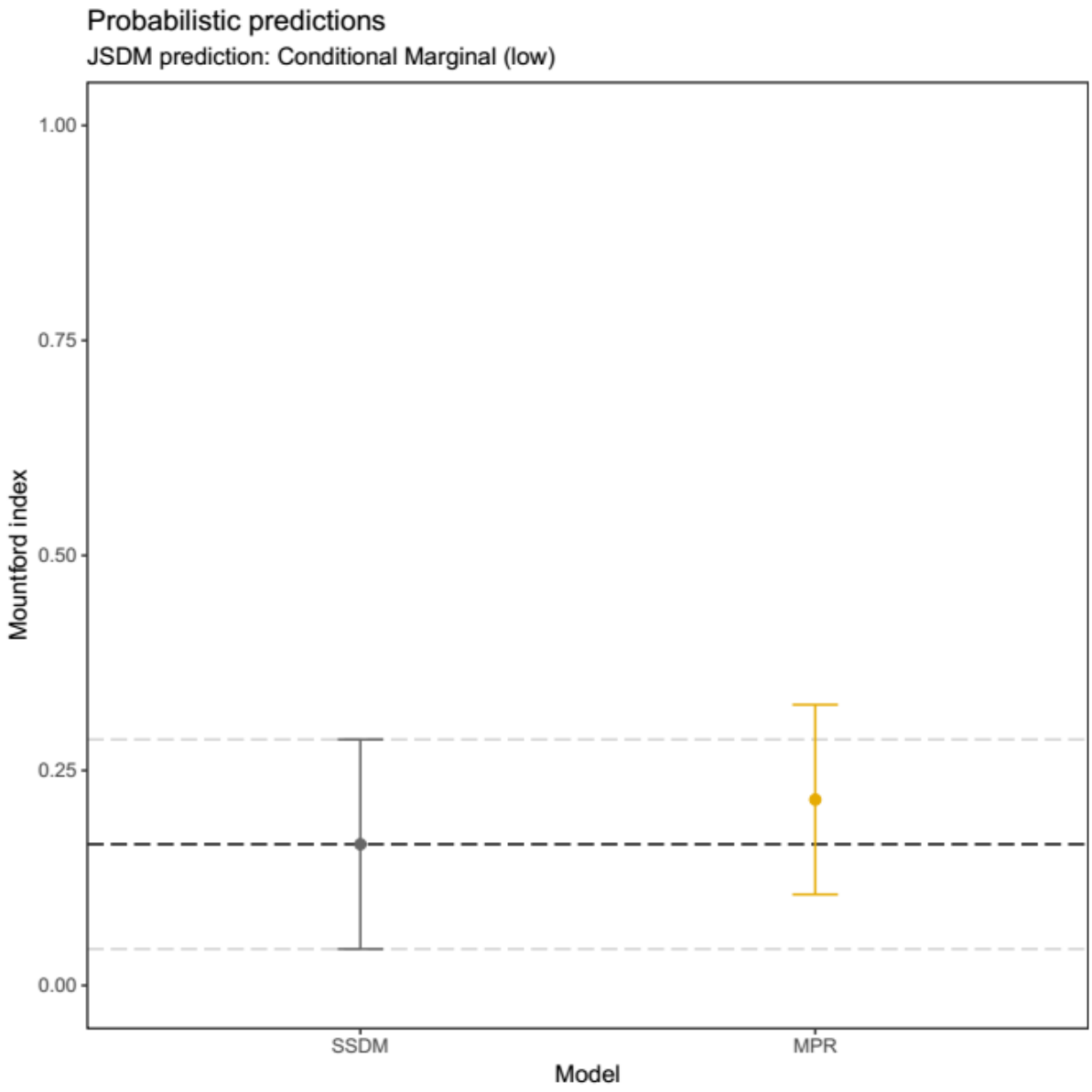


Figure 164: Performance of the SSDM and JSDM for the Mountford index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

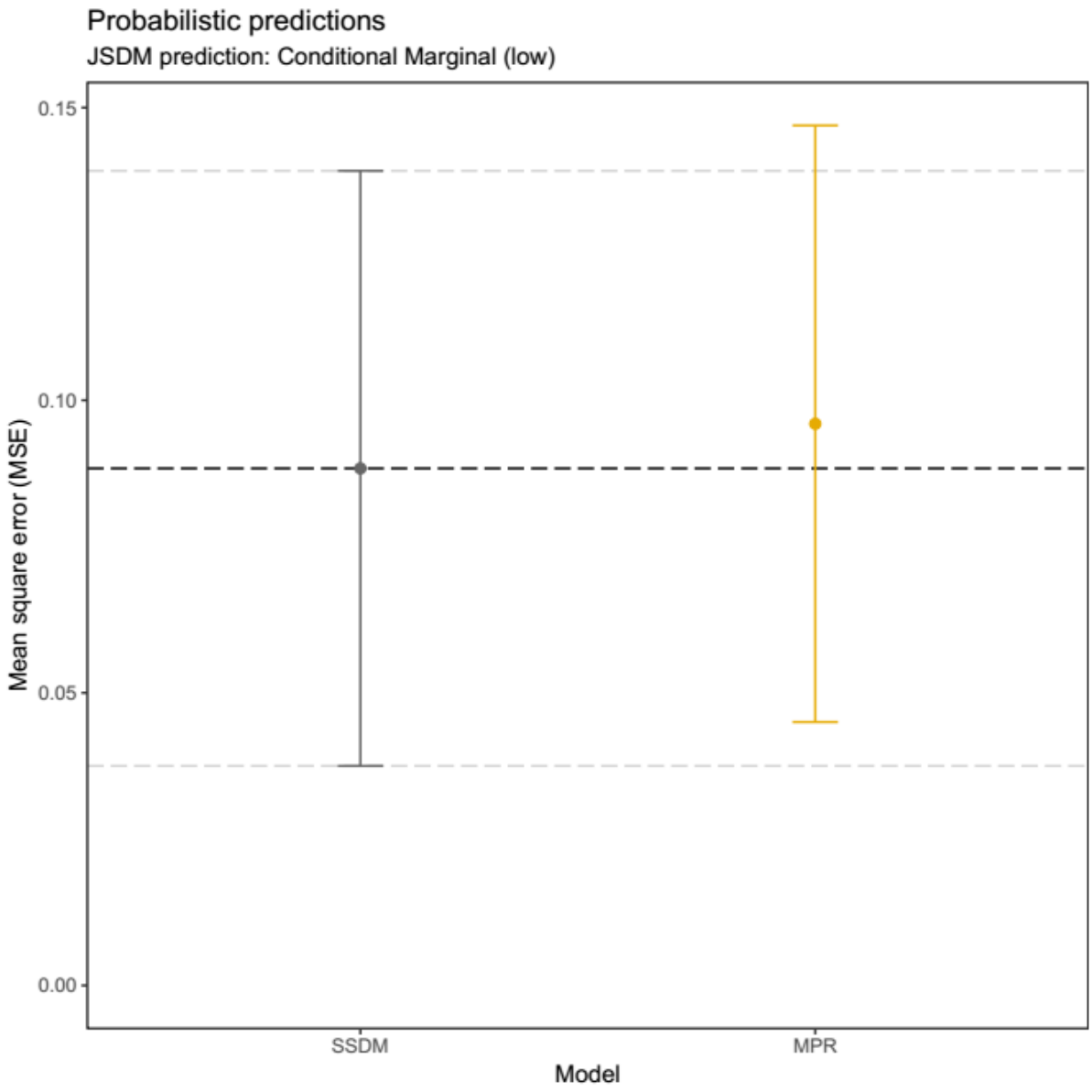


Figure 165: Performance of the SSDM and JSDM for the mean square error evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

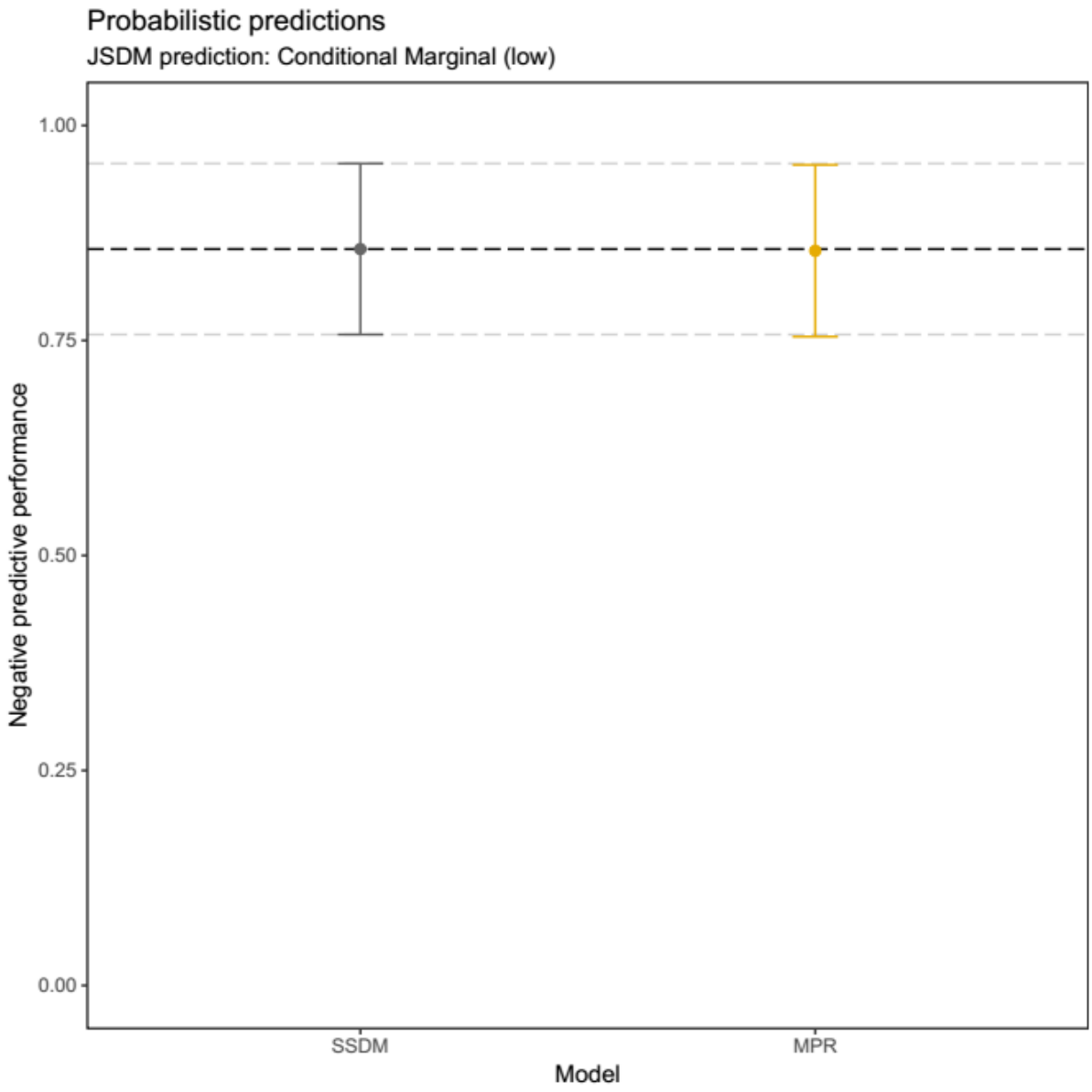


Figure 166: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

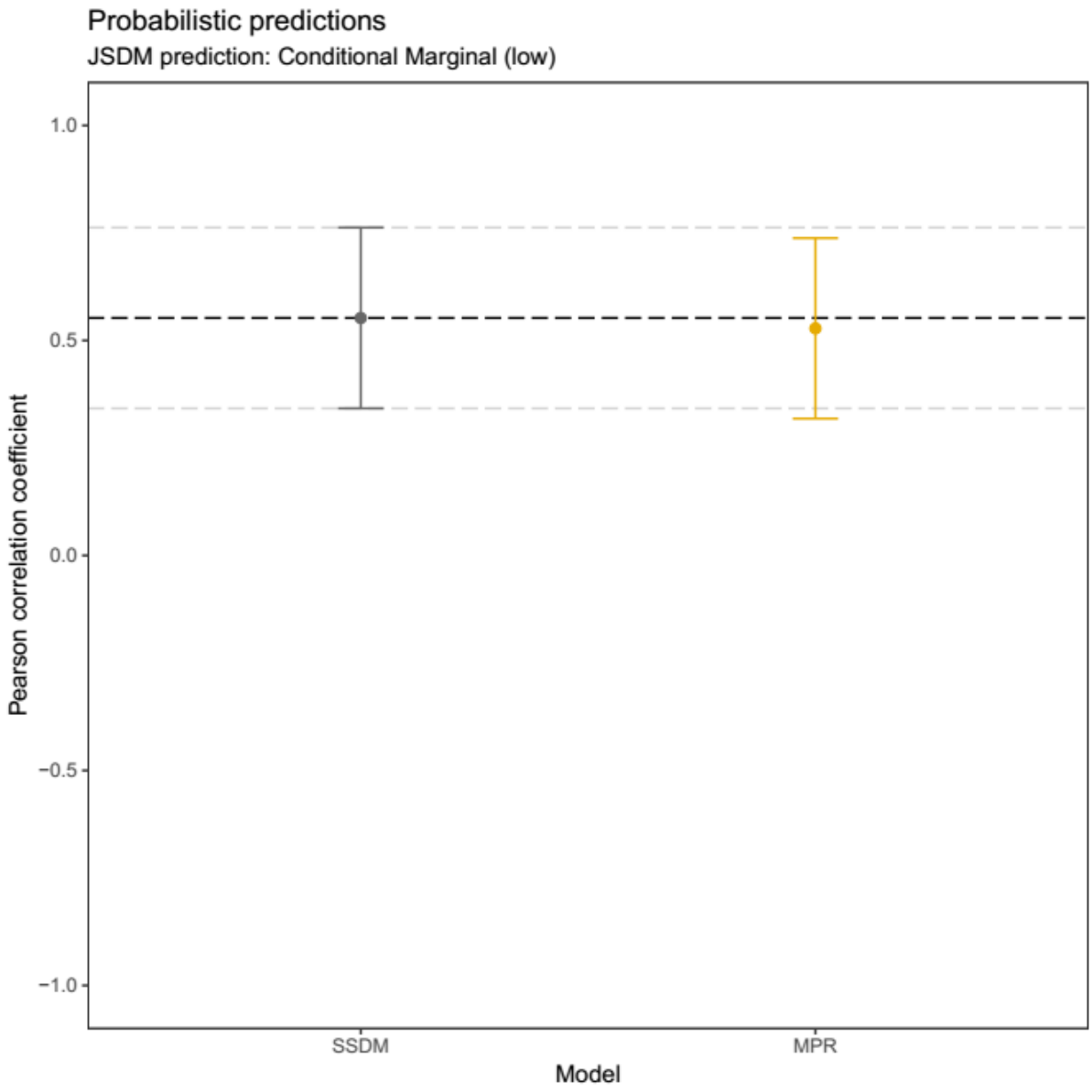


Figure 167: Performance of the SSDM and JSDM for the Pearson correlation coefficient evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

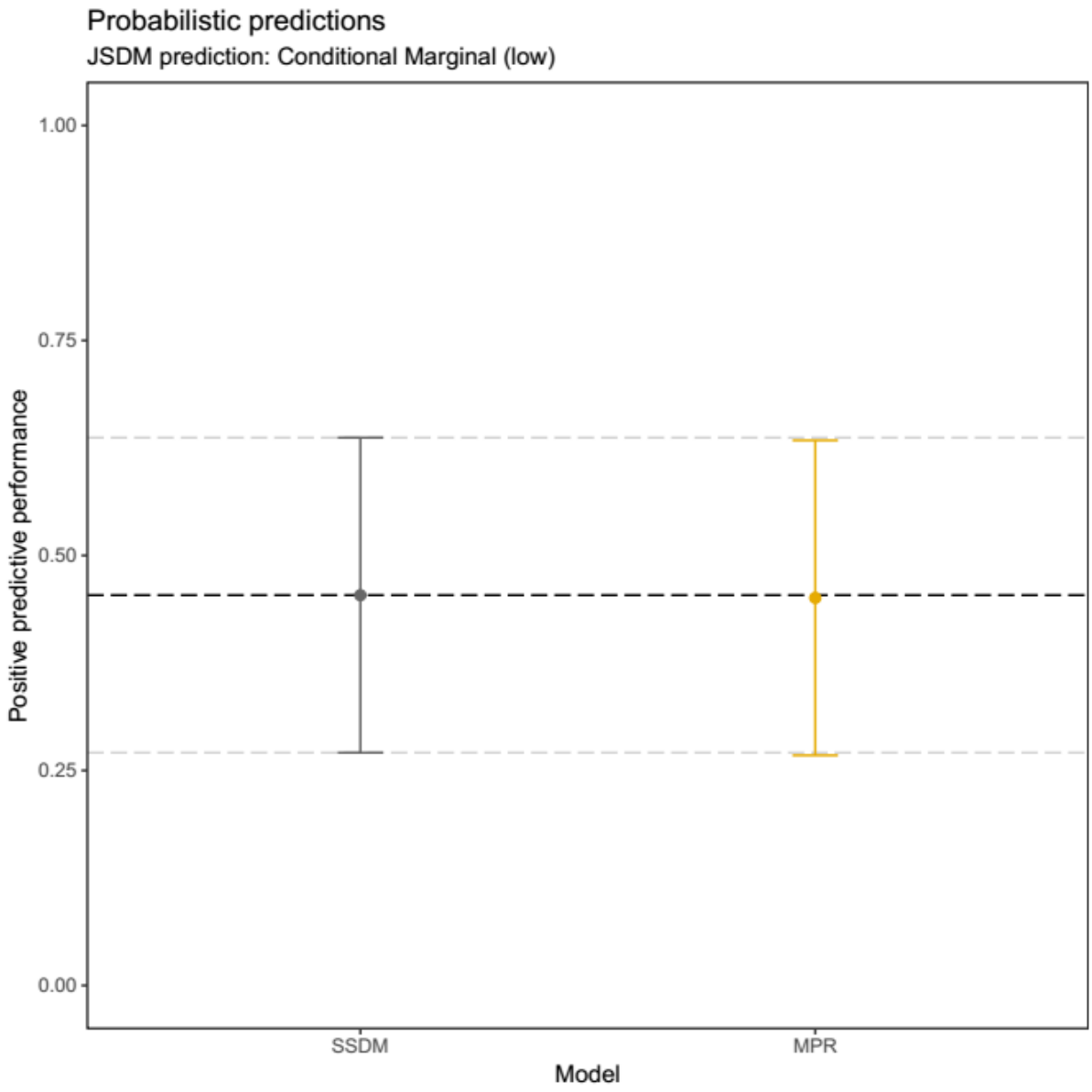


Figure 168: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

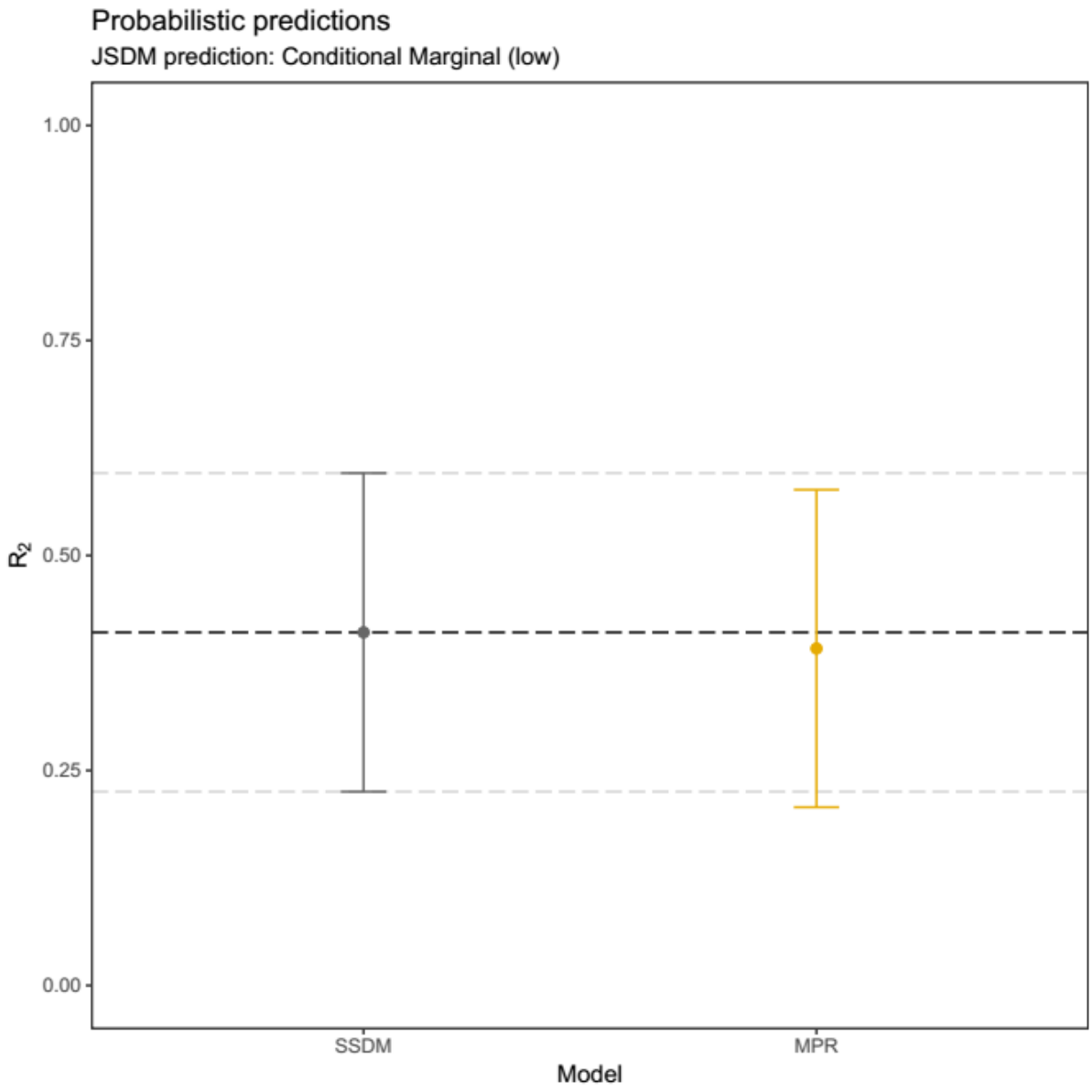


Figure 169: Performance of the SSDM and JSDM for the R^2 evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

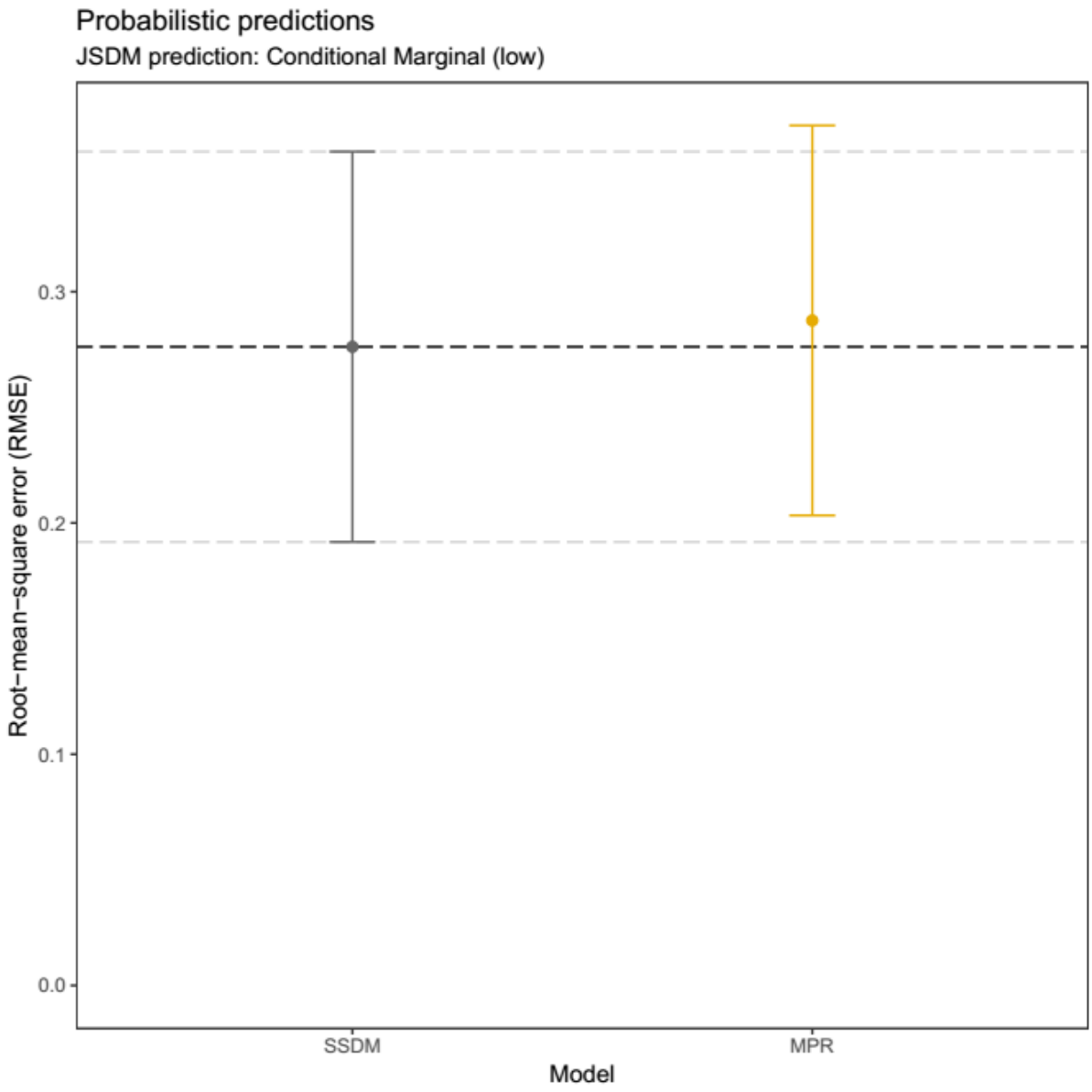


Figure 170: Performance of the SSDM and JSDM for the root-mean-square error evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

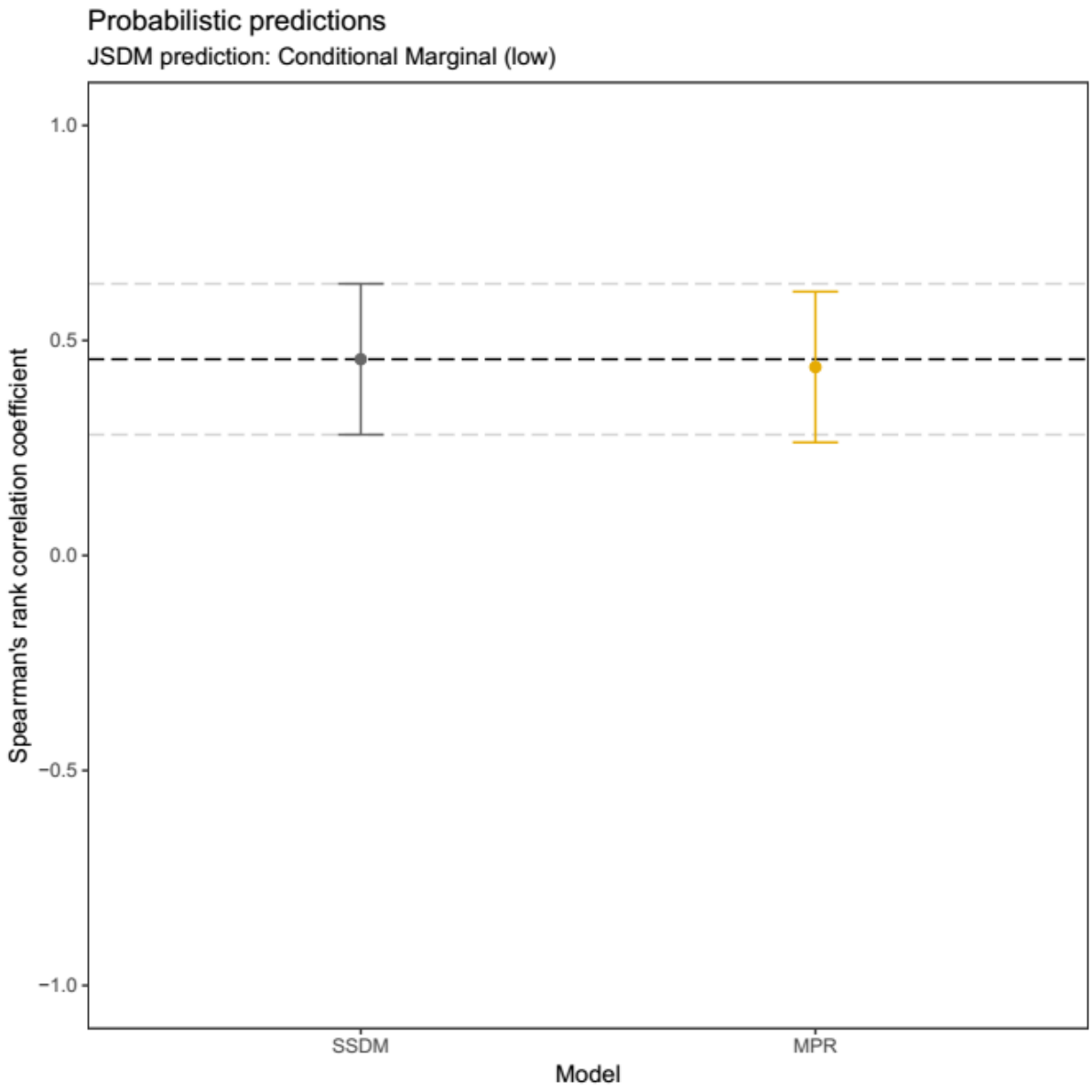


Figure 171: Performance of the SSDM and JSDM for the Spearman's rank correlation coefficient evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

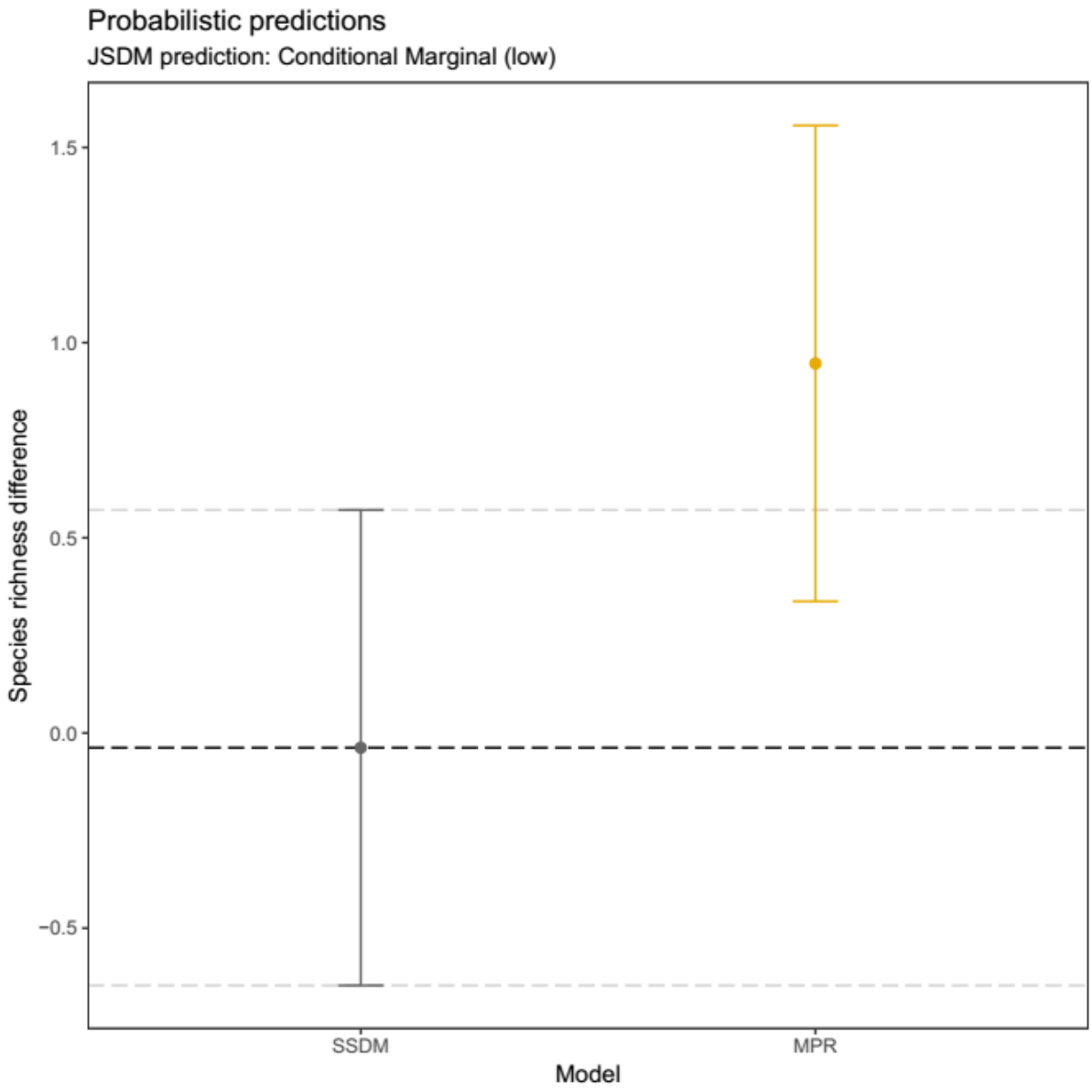


Figure 172: Performance of the SSDM and JSDM for the species richness difference evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

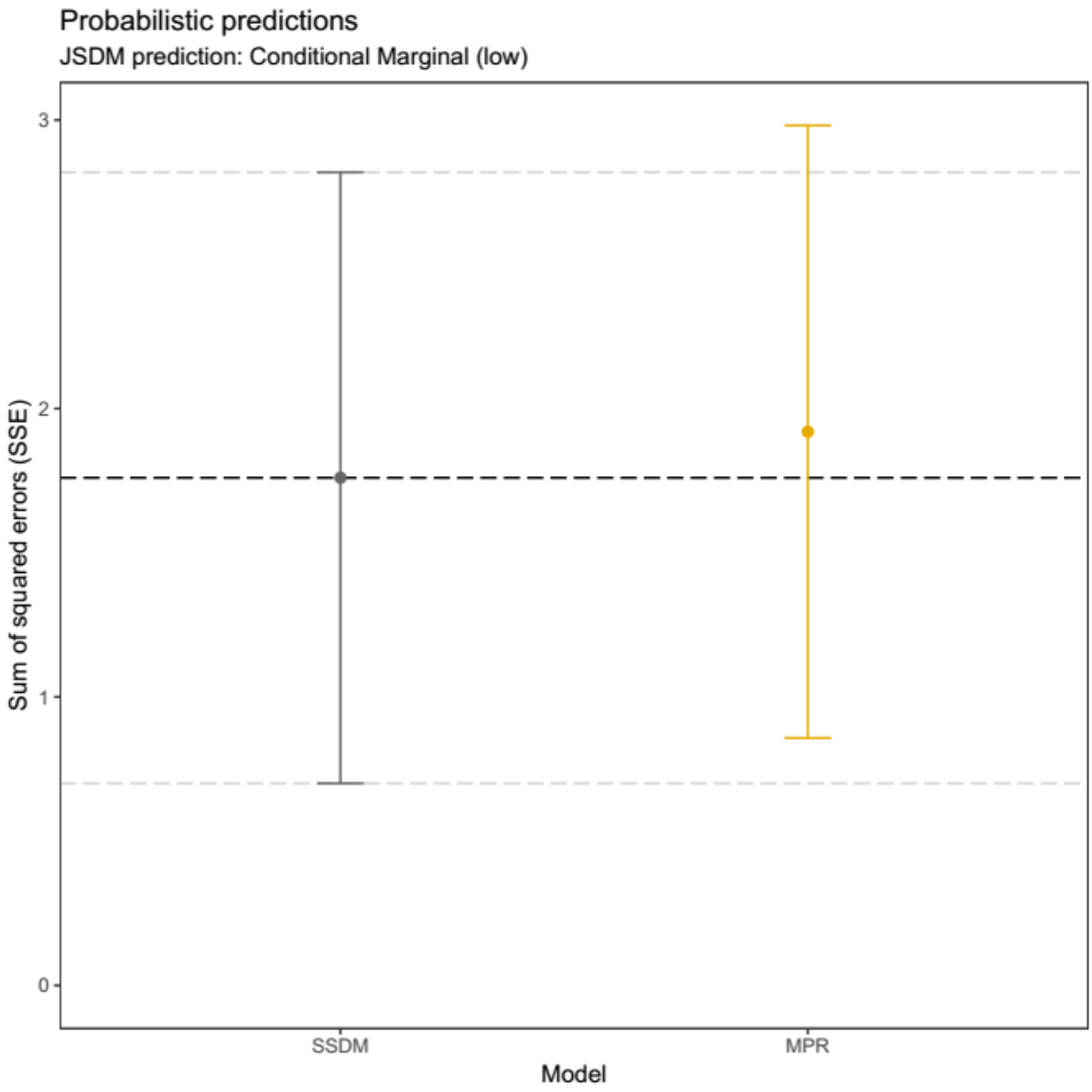


Figure 173: Performance of the SSDM and JSDM for the sum of squared errors evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

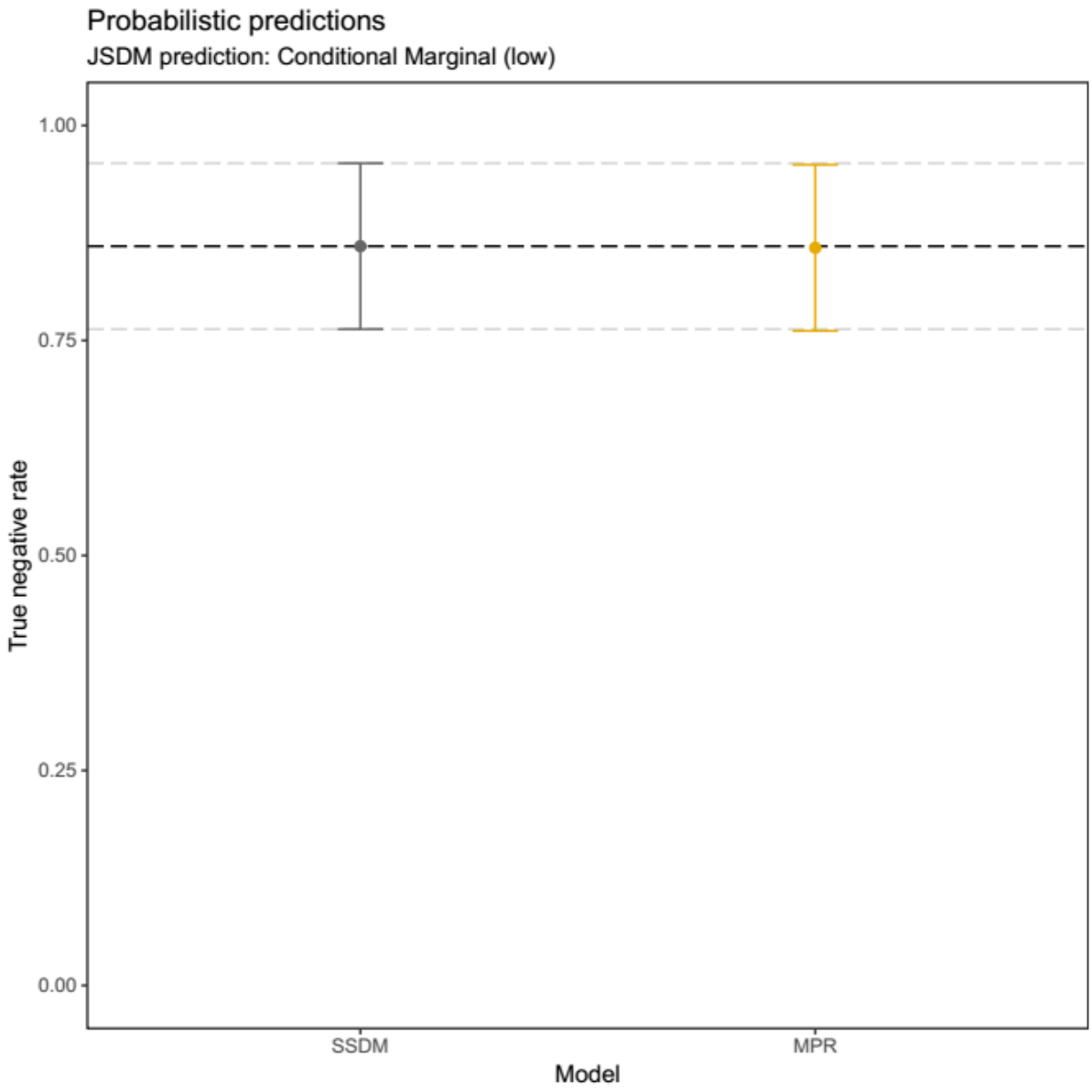


Figure 174: Performance of the SSDM and JSDM for the true negative rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

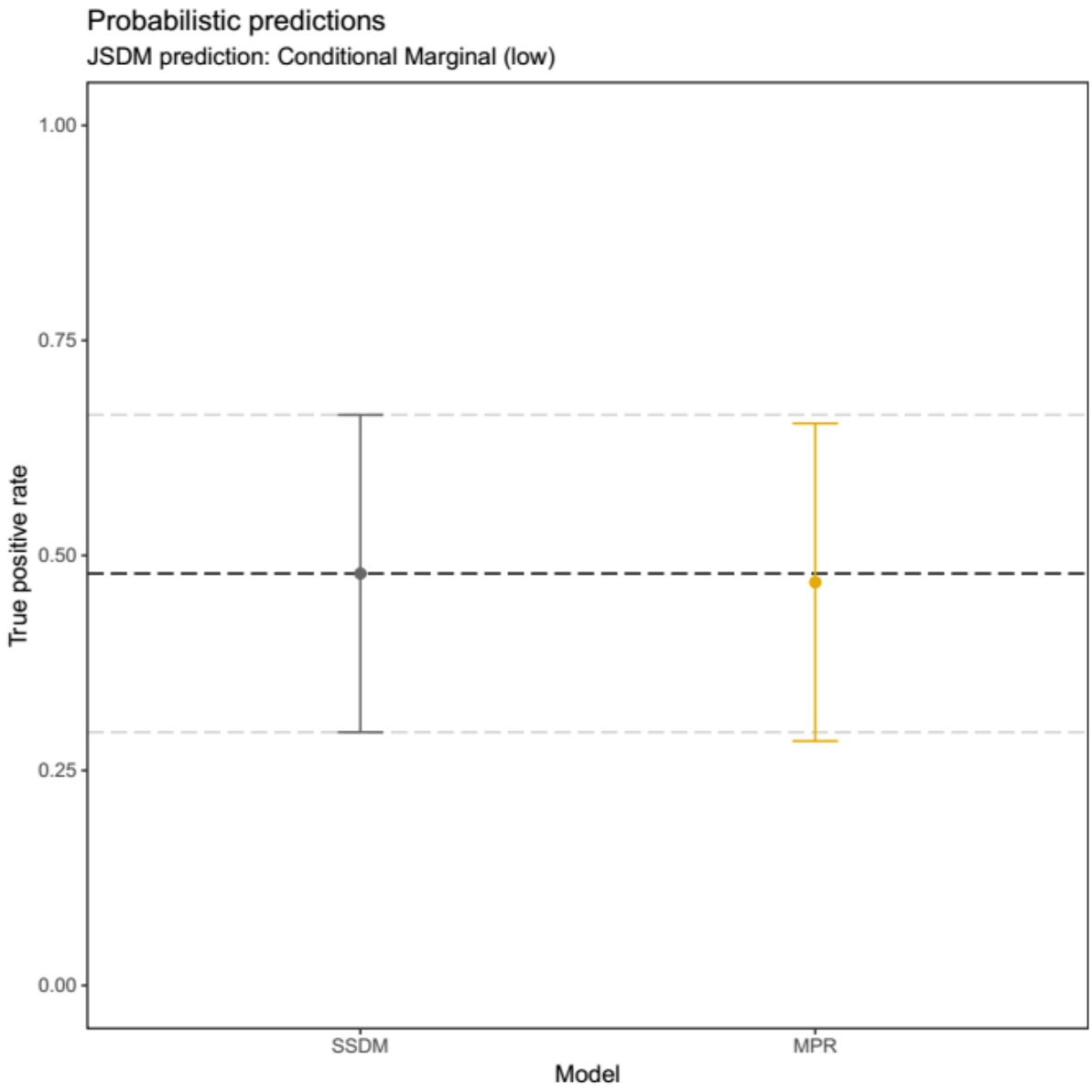


Figure 175: Performance of the SSDM and JSDM for the true positive rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

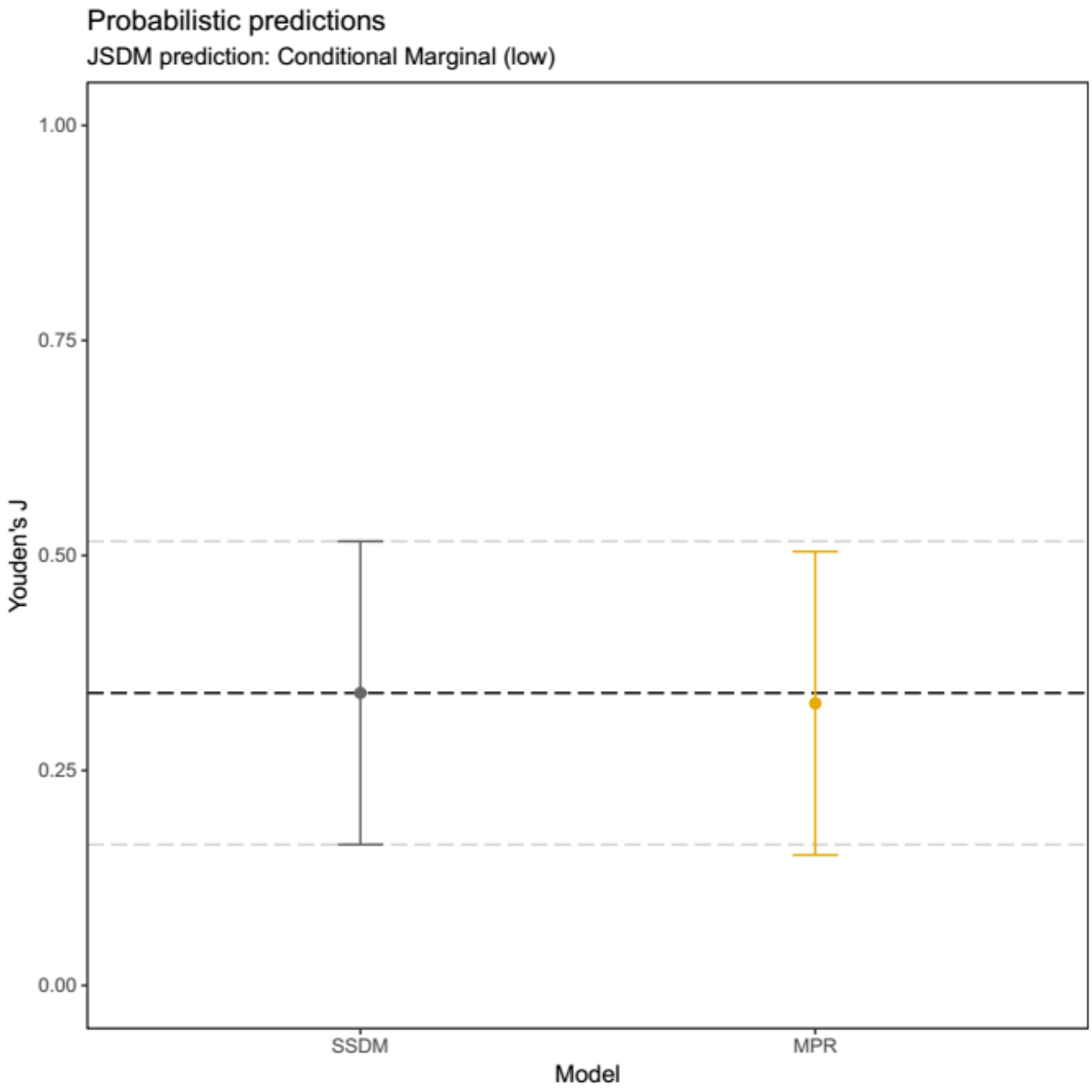


Figure 176: Performance of the SSDM and JSDM for the Youden's J evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

8. Conditional marginal (medium-prevalence) predictions

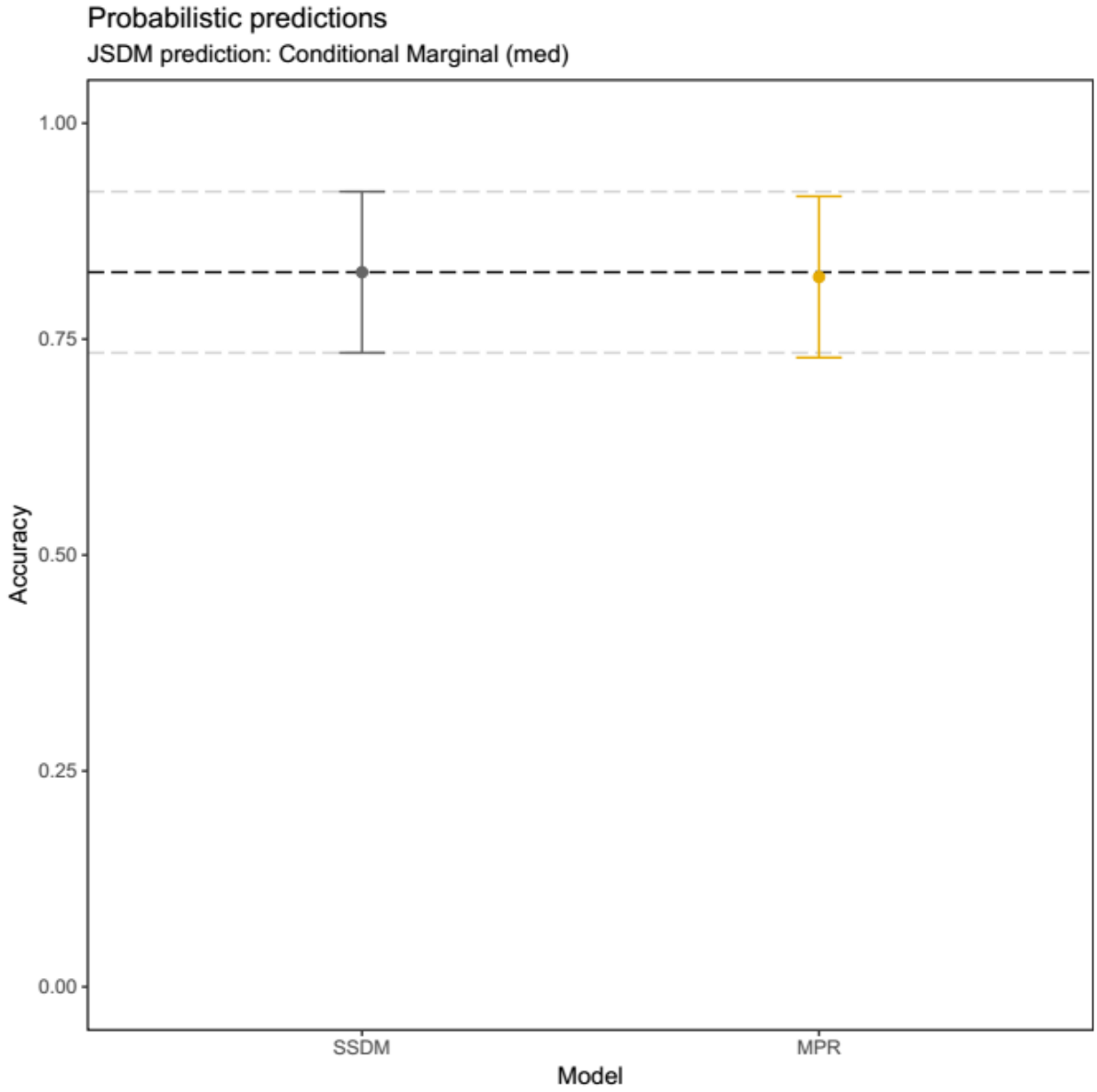


Figure 177: Performance of the SSDM and JSDM for the accuracy evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

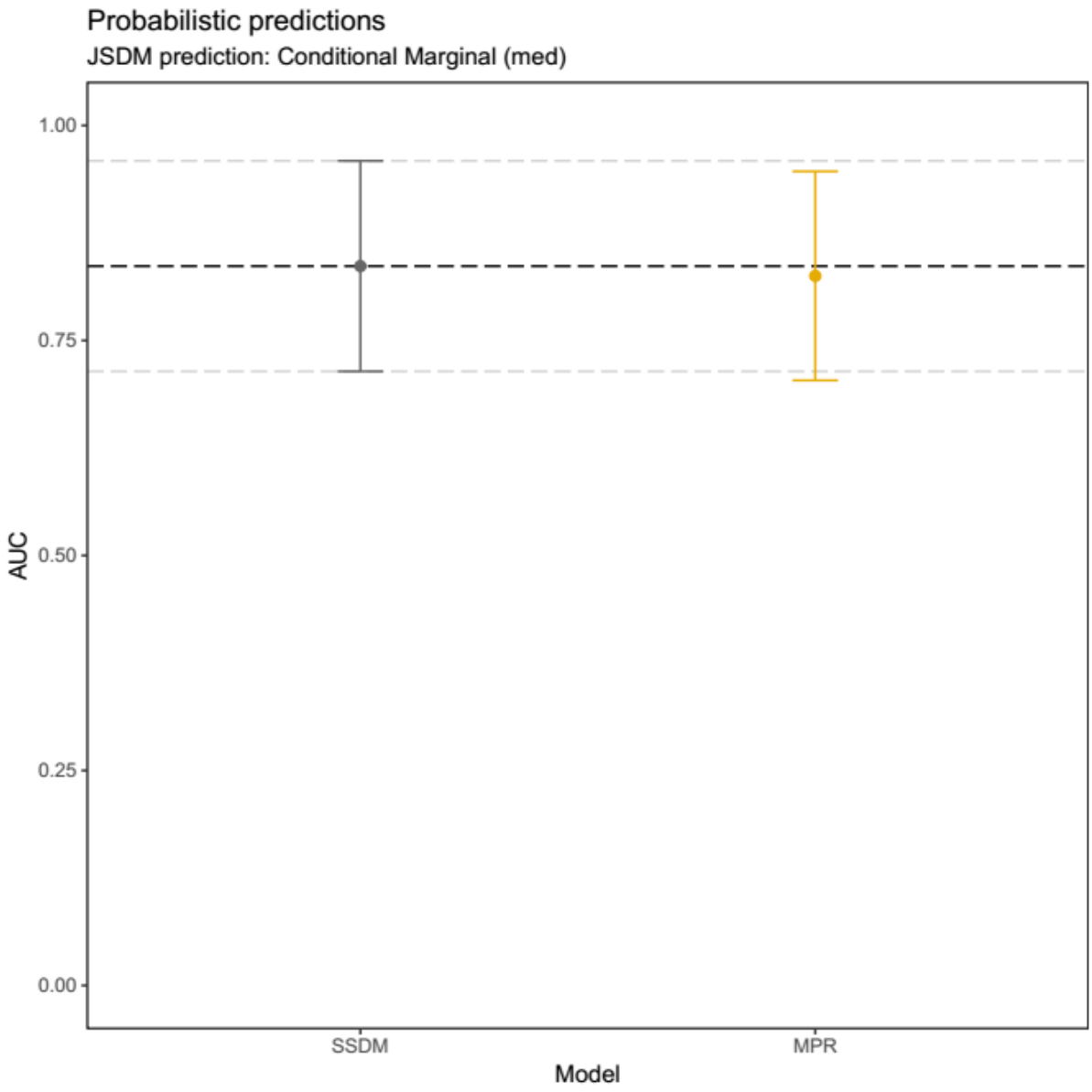


Figure 178: Performance of the SSDM and JSDM for the AUC evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

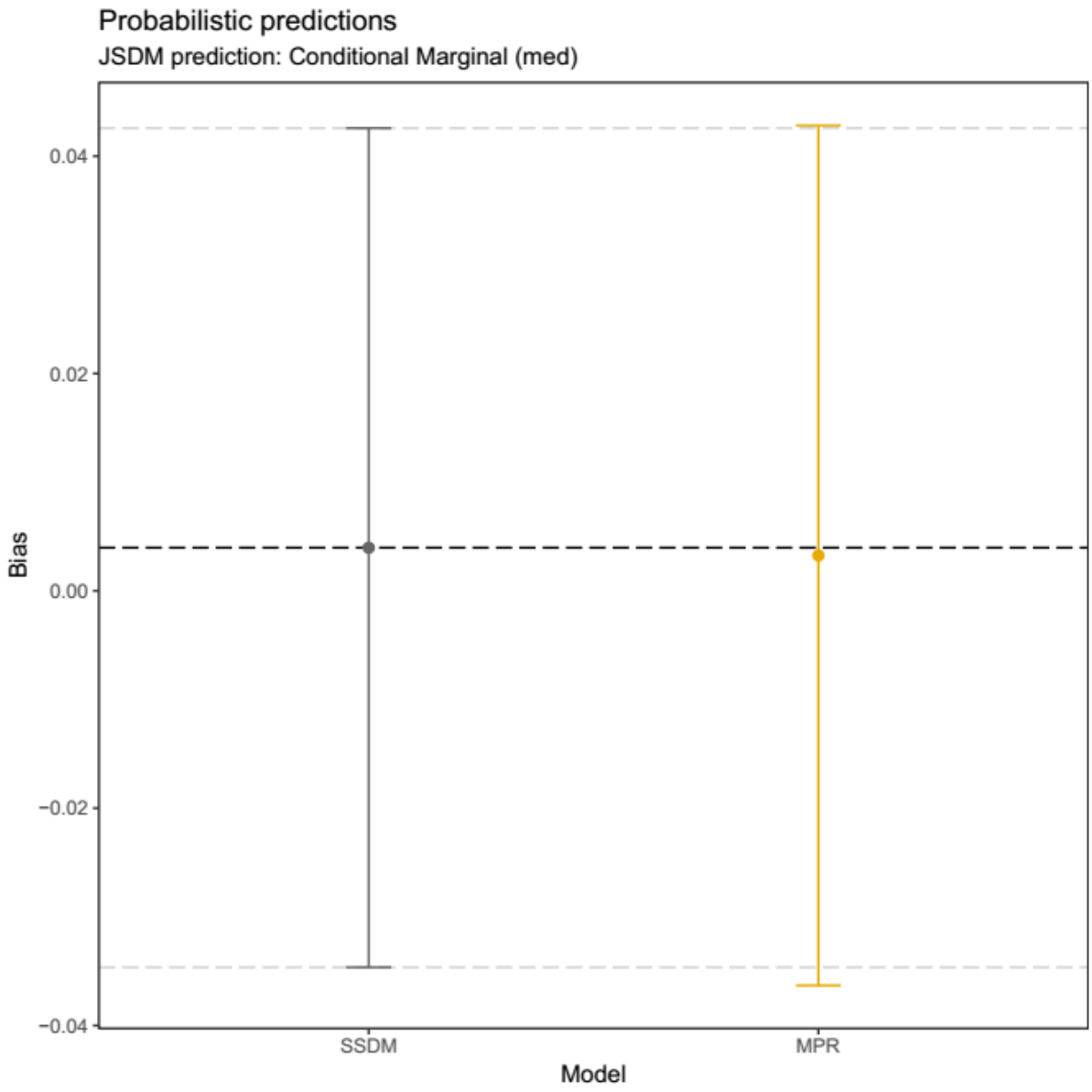


Figure 179: Performance of the SSDM and JSDM for the mean error/bias evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

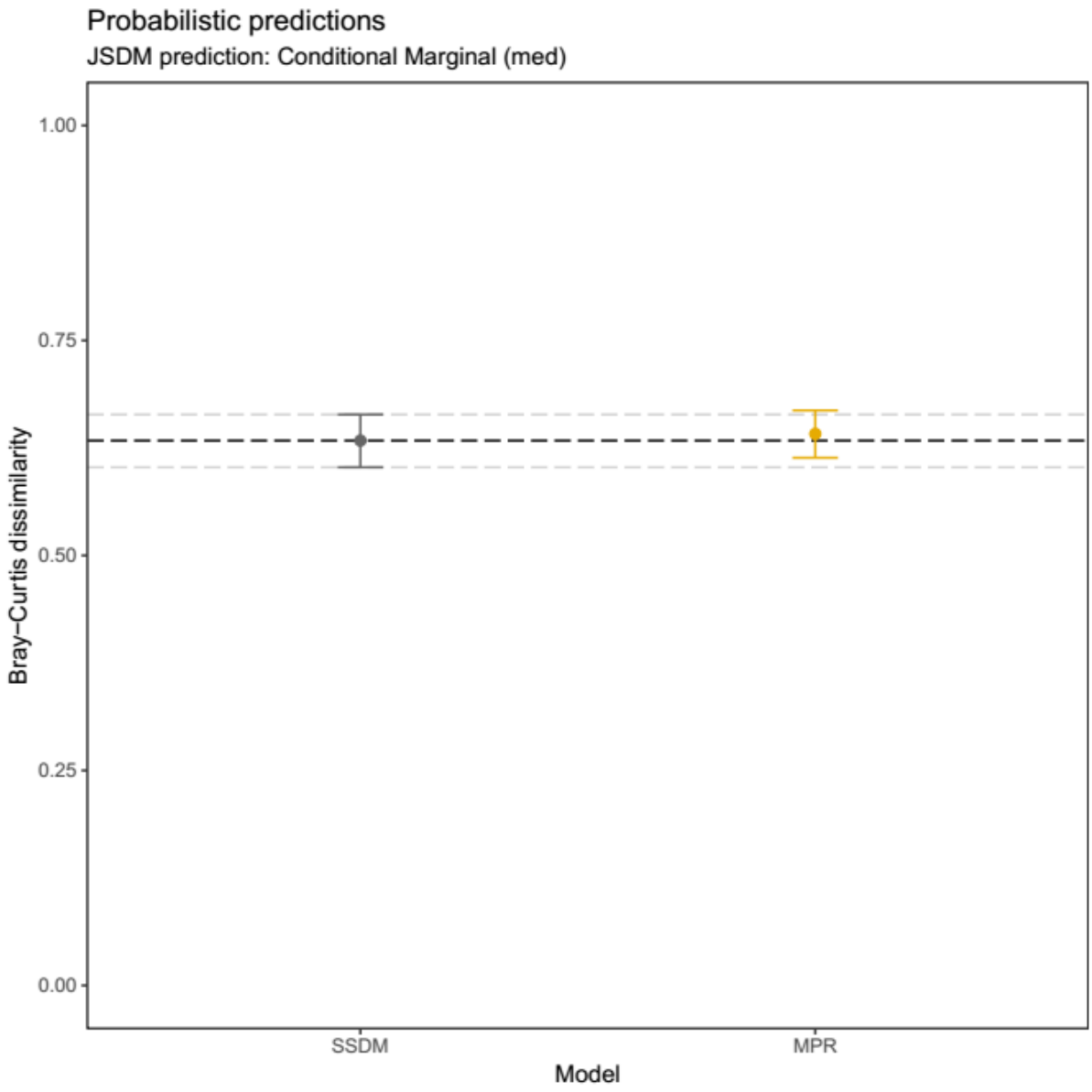


Figure 180: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

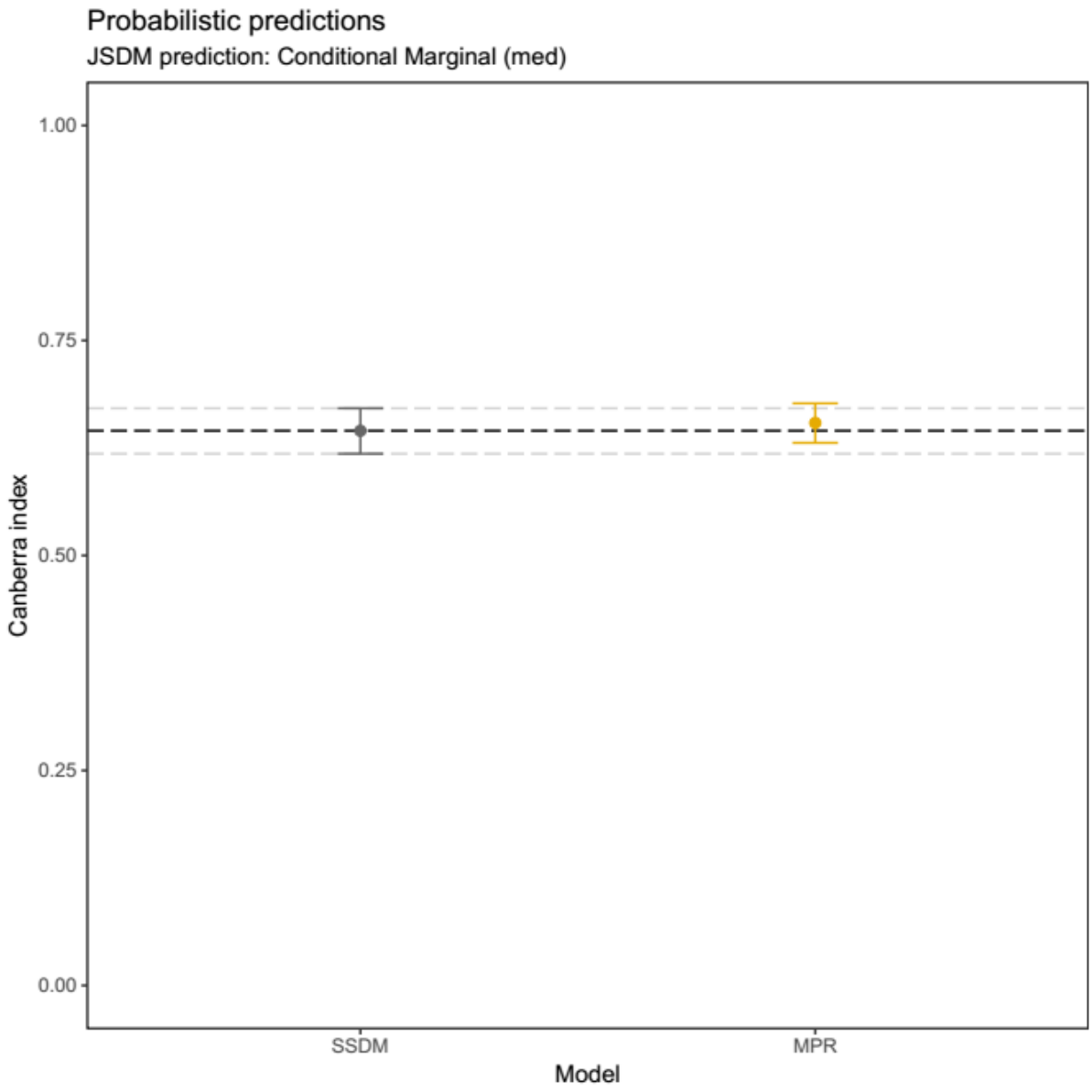


Figure 181: Performance of the SSDM and JSDM for the Canberra index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

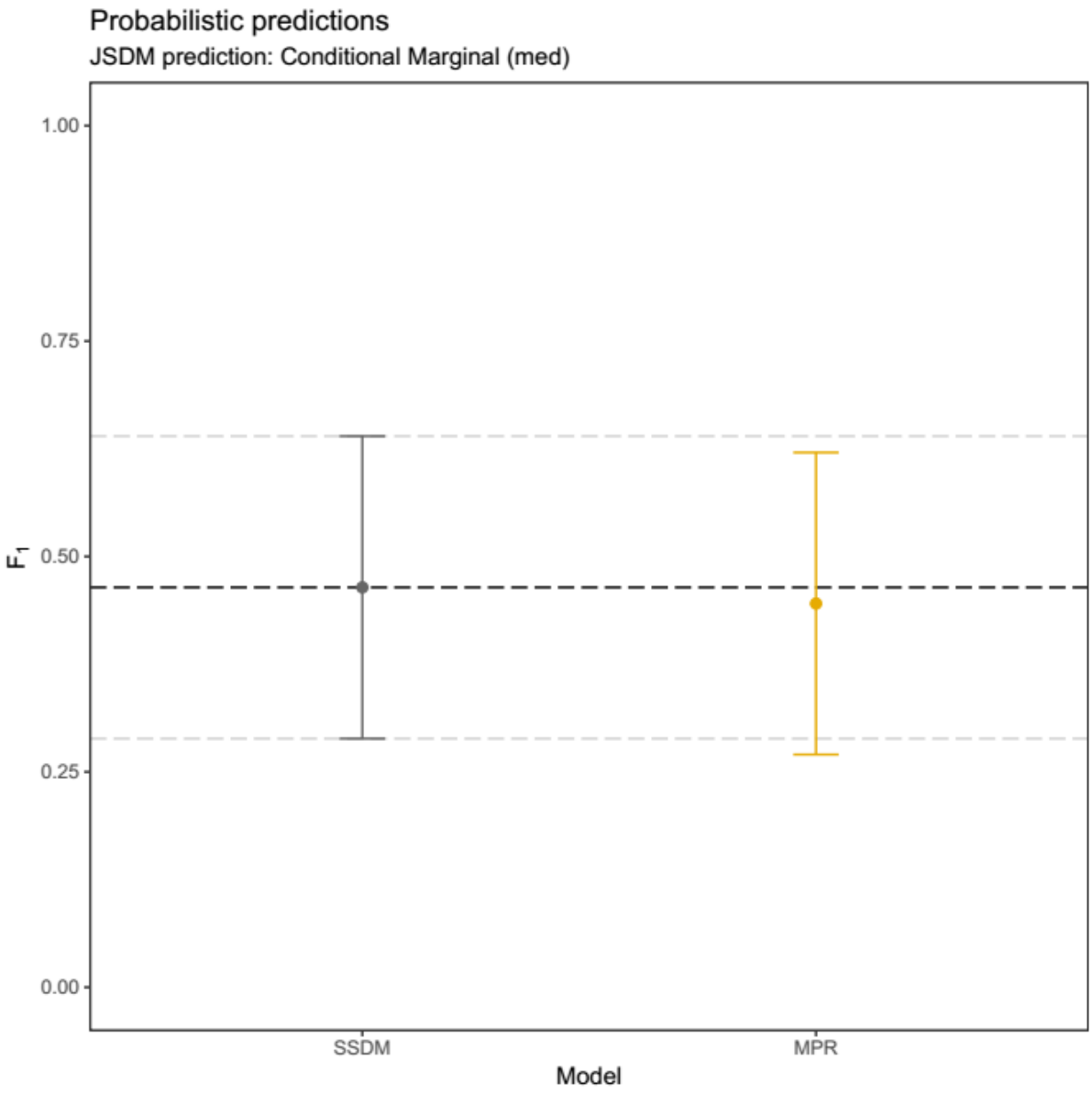


Figure 182: Performance of the SSDM and JSDM for the F_1 evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

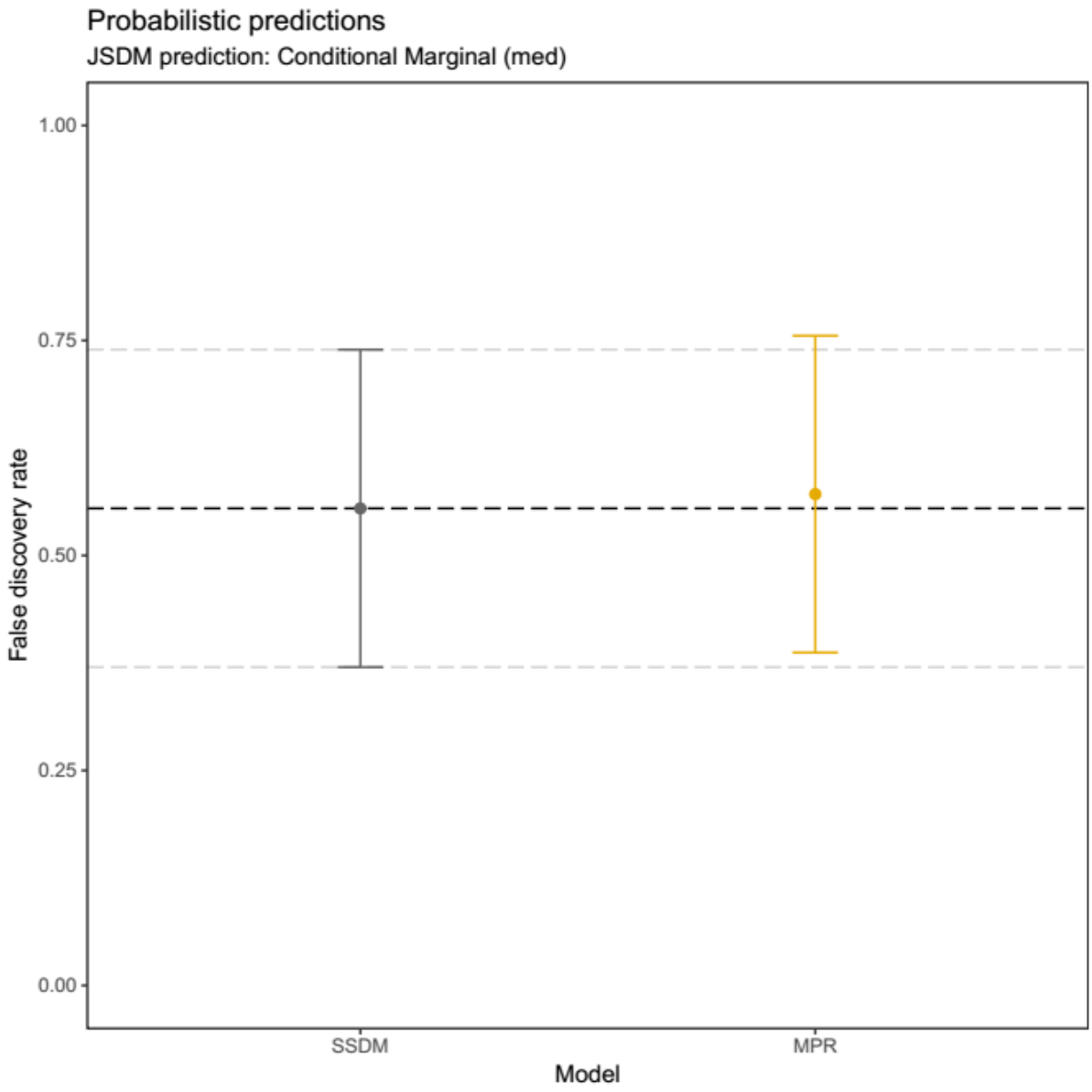


Figure 183: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

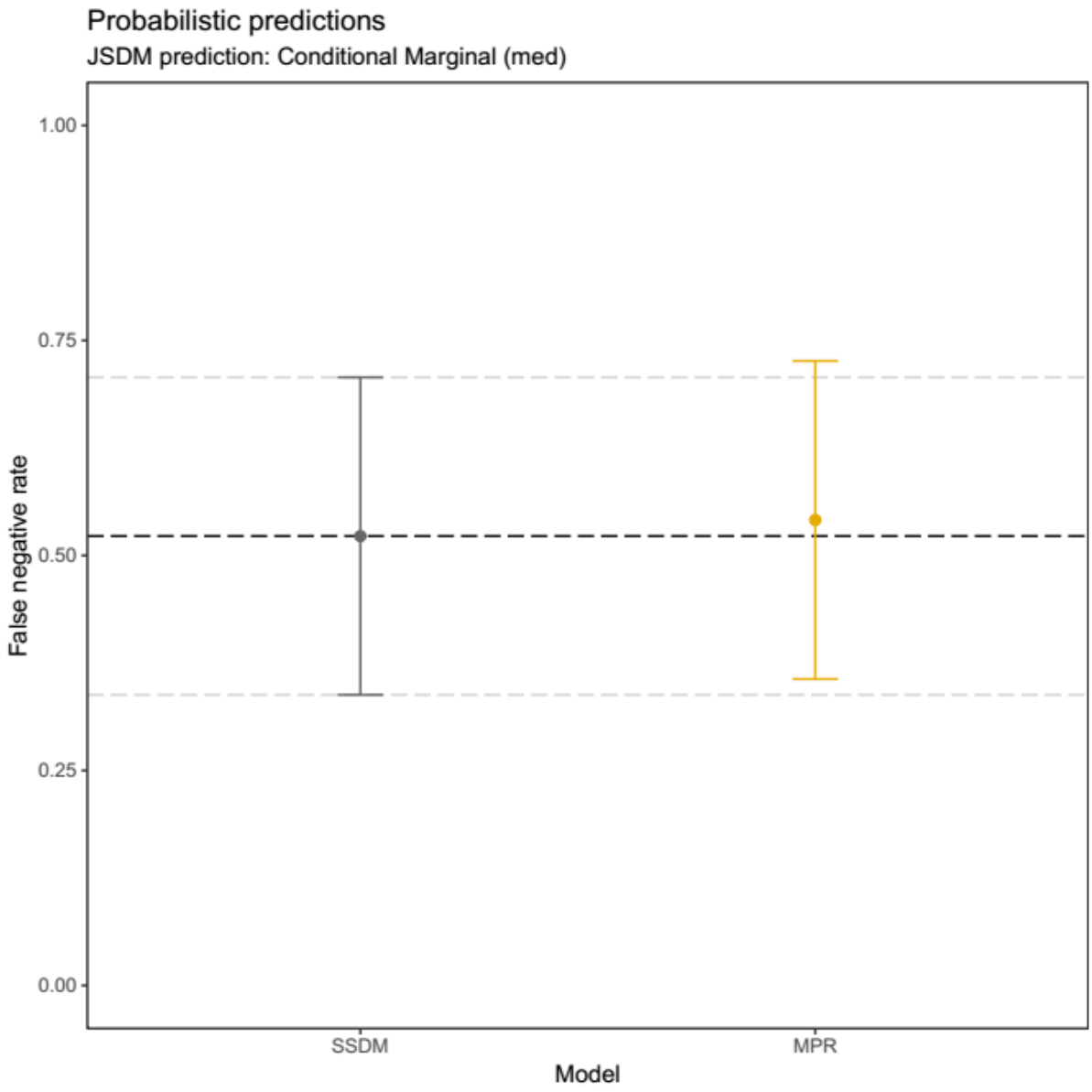


Figure 184: Performance of the SSDM and JSDM for the false negative rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

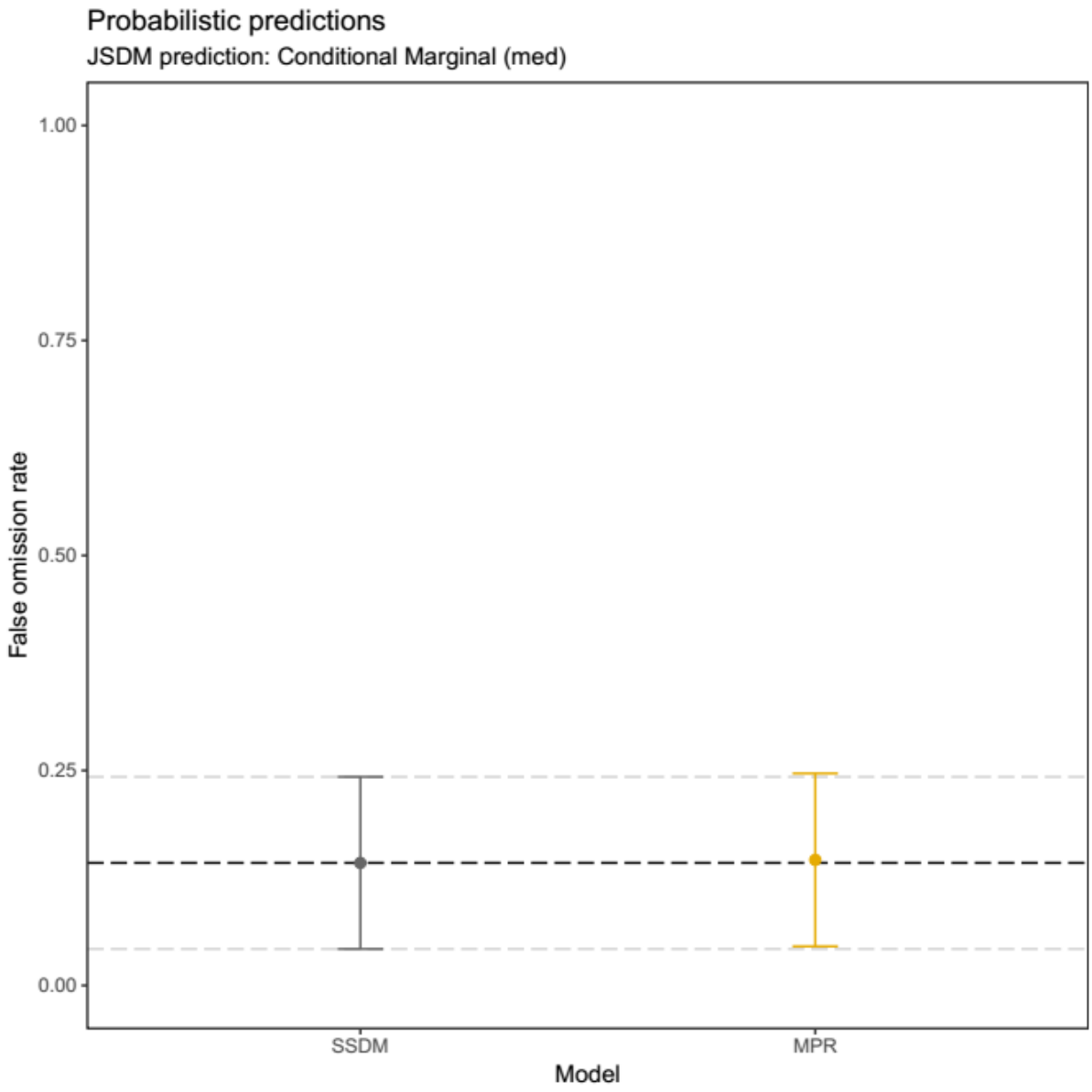


Figure 185: Performance of the SSDM and JSDM for the false omission rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

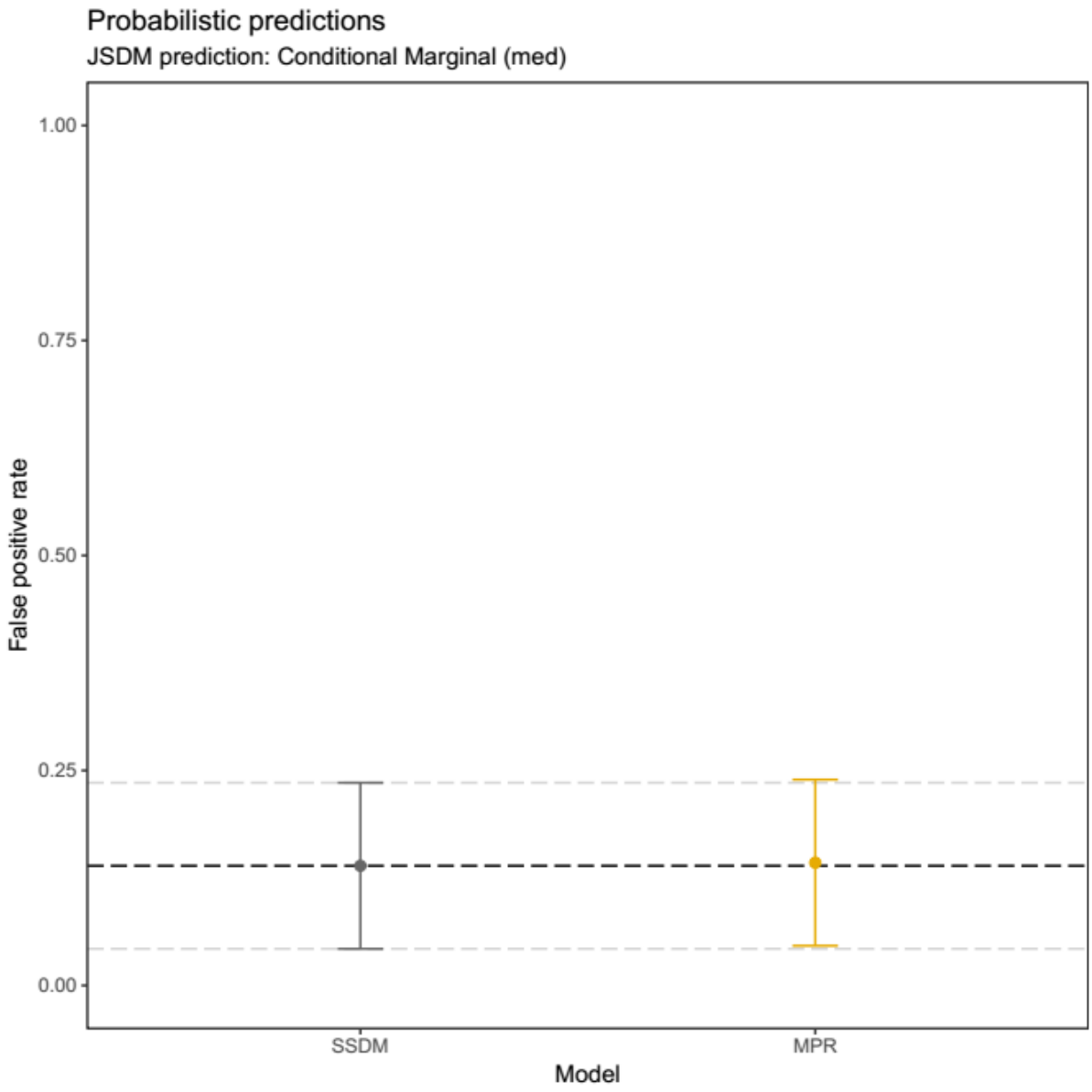


Figure 186: Performance of the SSDM and JSDM for the false positive rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

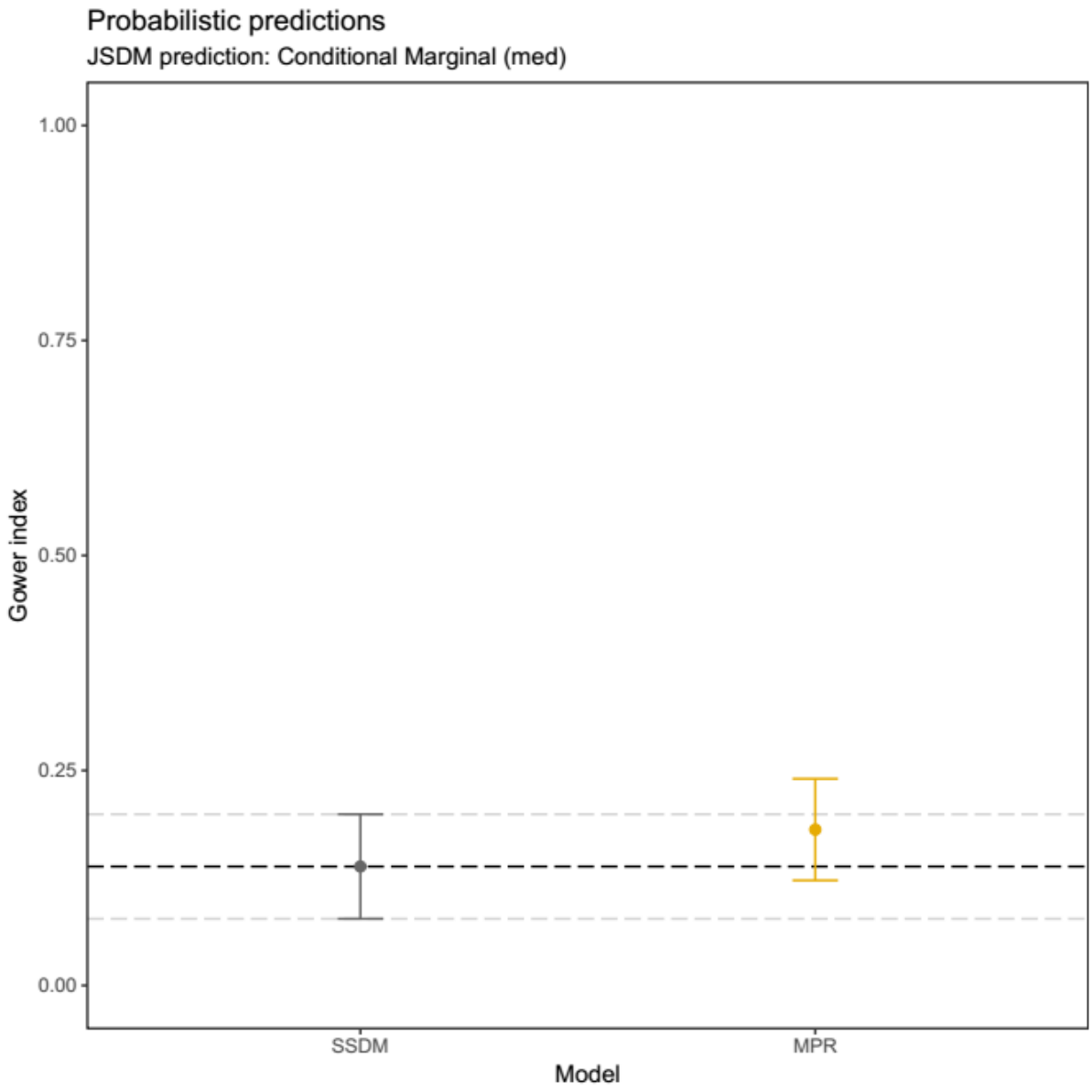


Figure 187: Performance of the SSDM and JSDM for the Gower index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

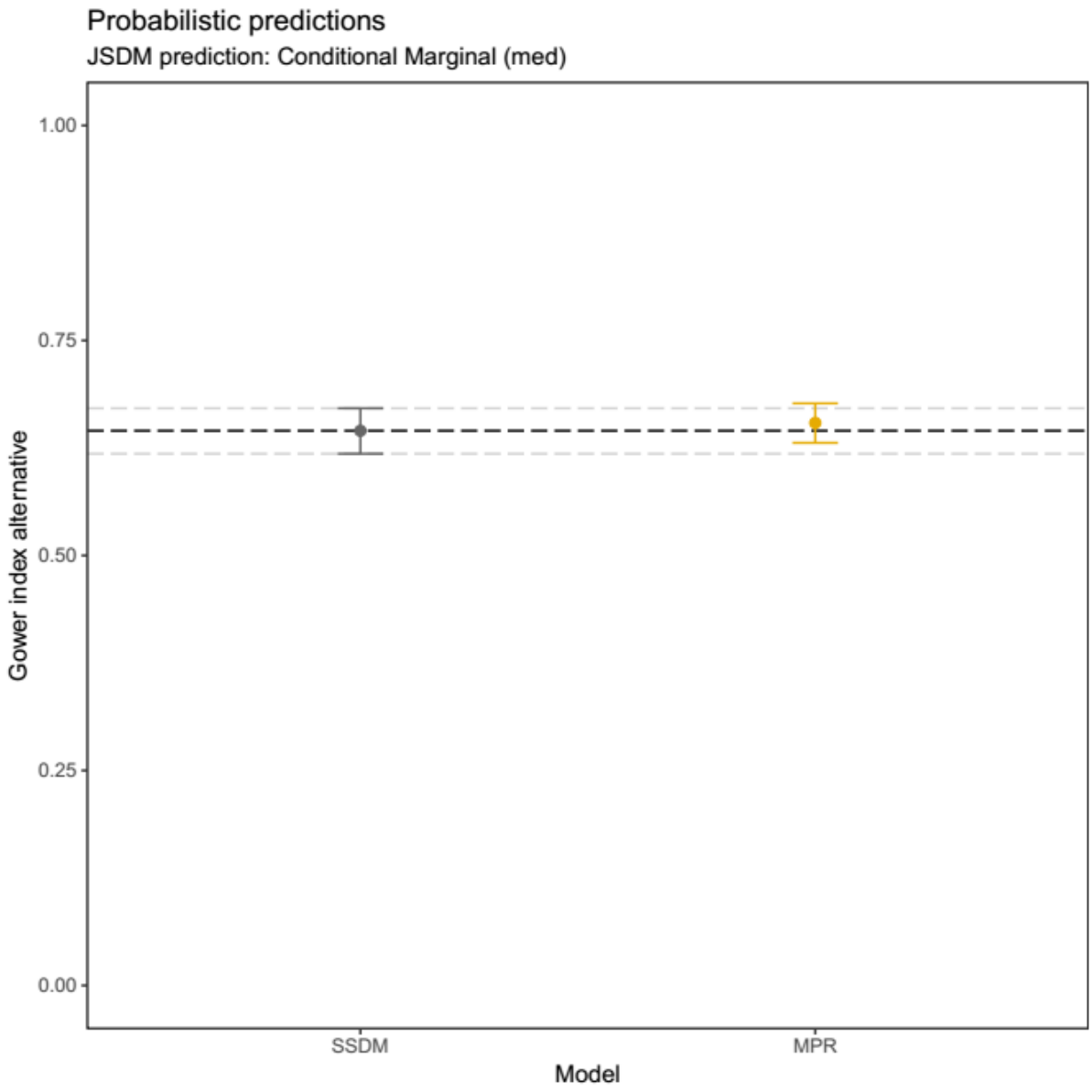


Figure 188: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

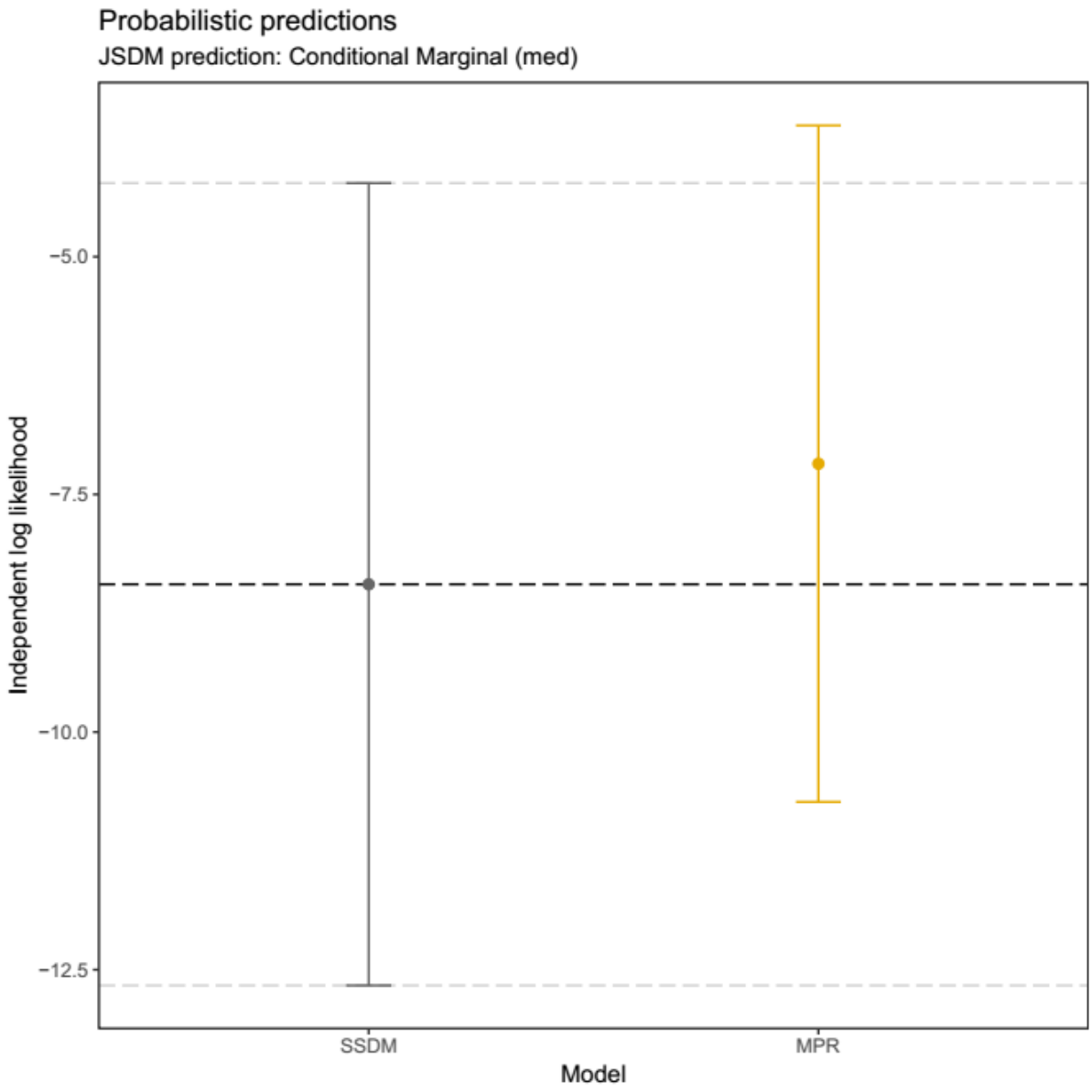


Figure 189: Performance of the SSDM and JSDM for the independent log likelihood evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

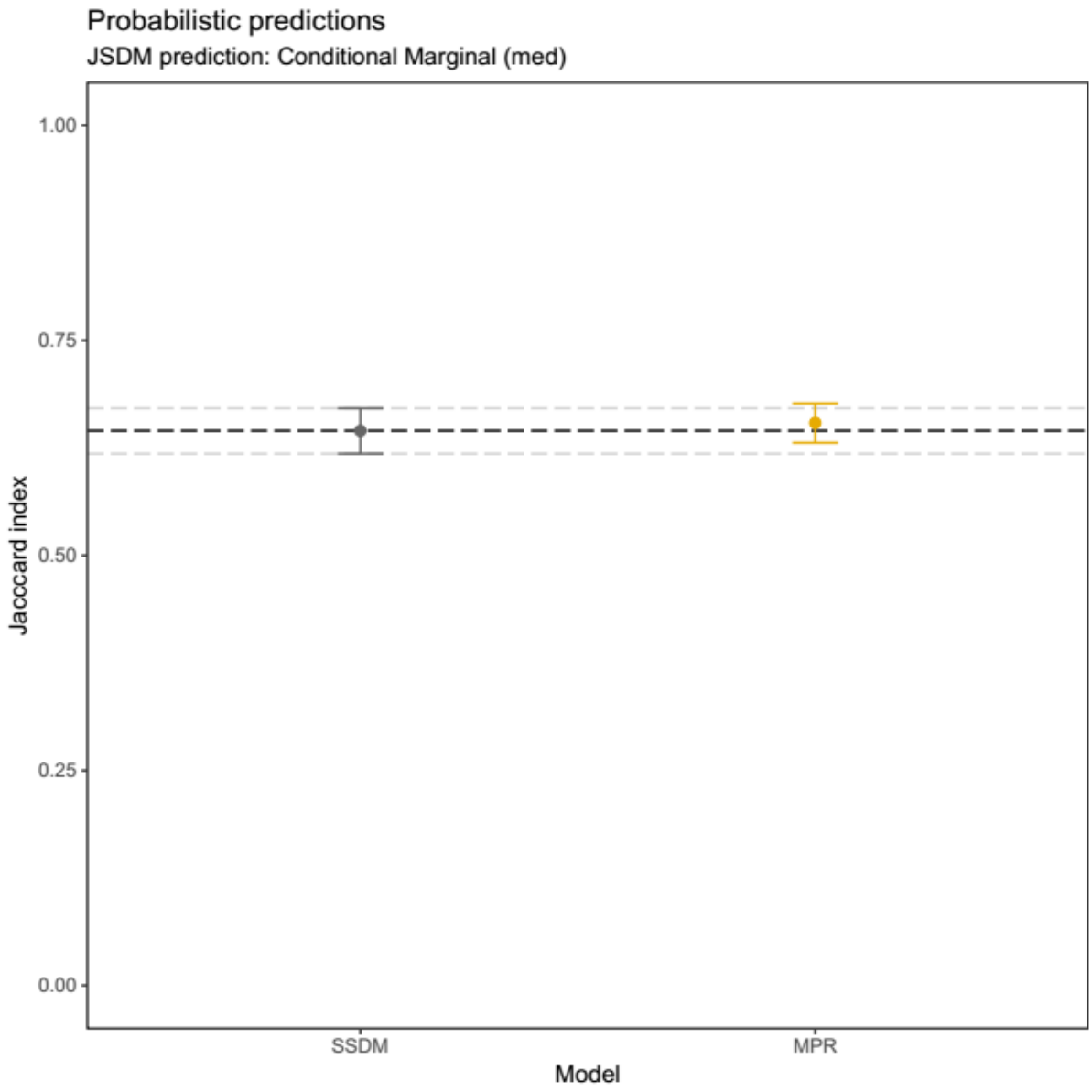


Figure 190: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

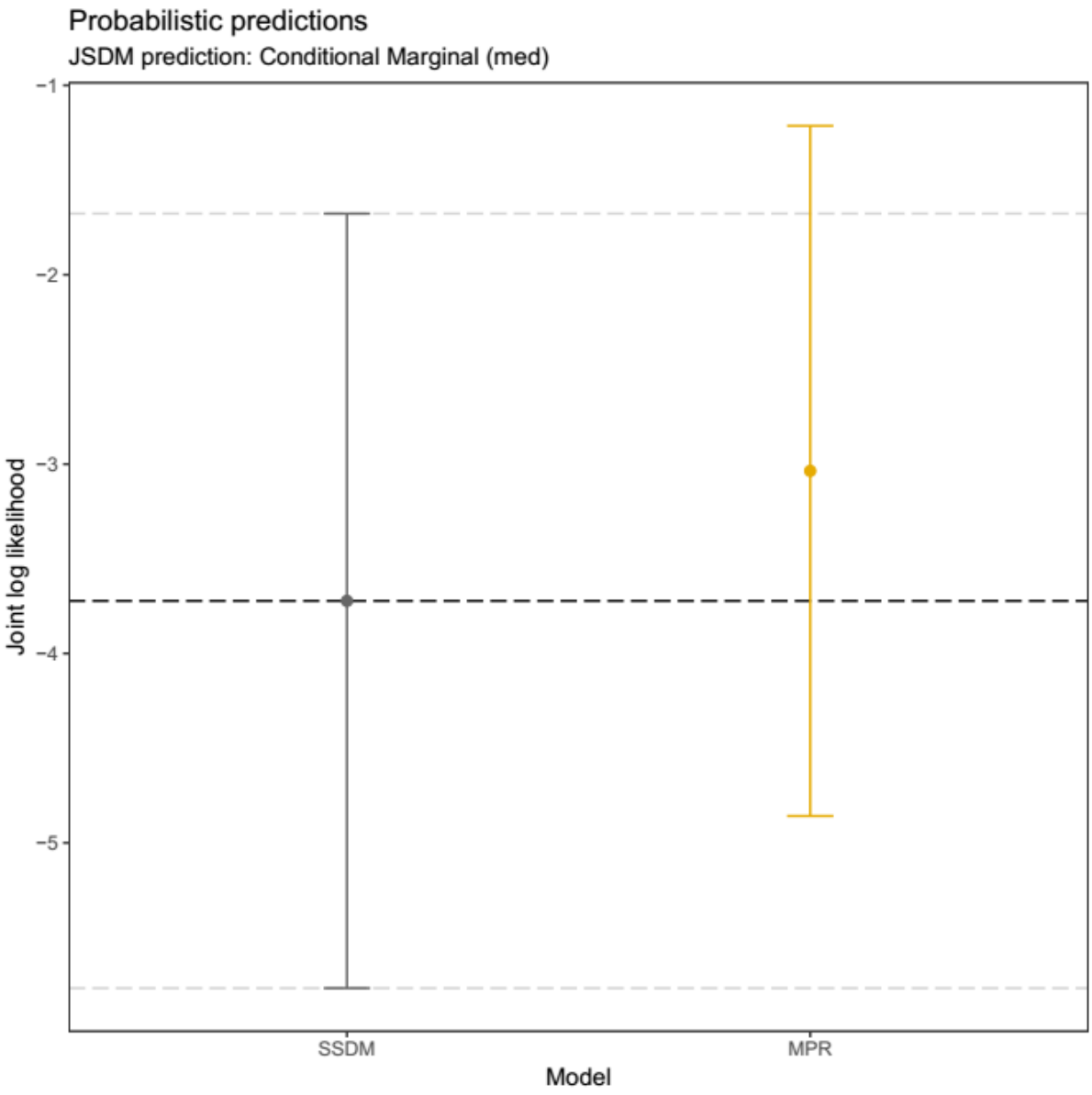


Figure 191: Performance of the SSDM and JSDM for the joint log likelihood evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

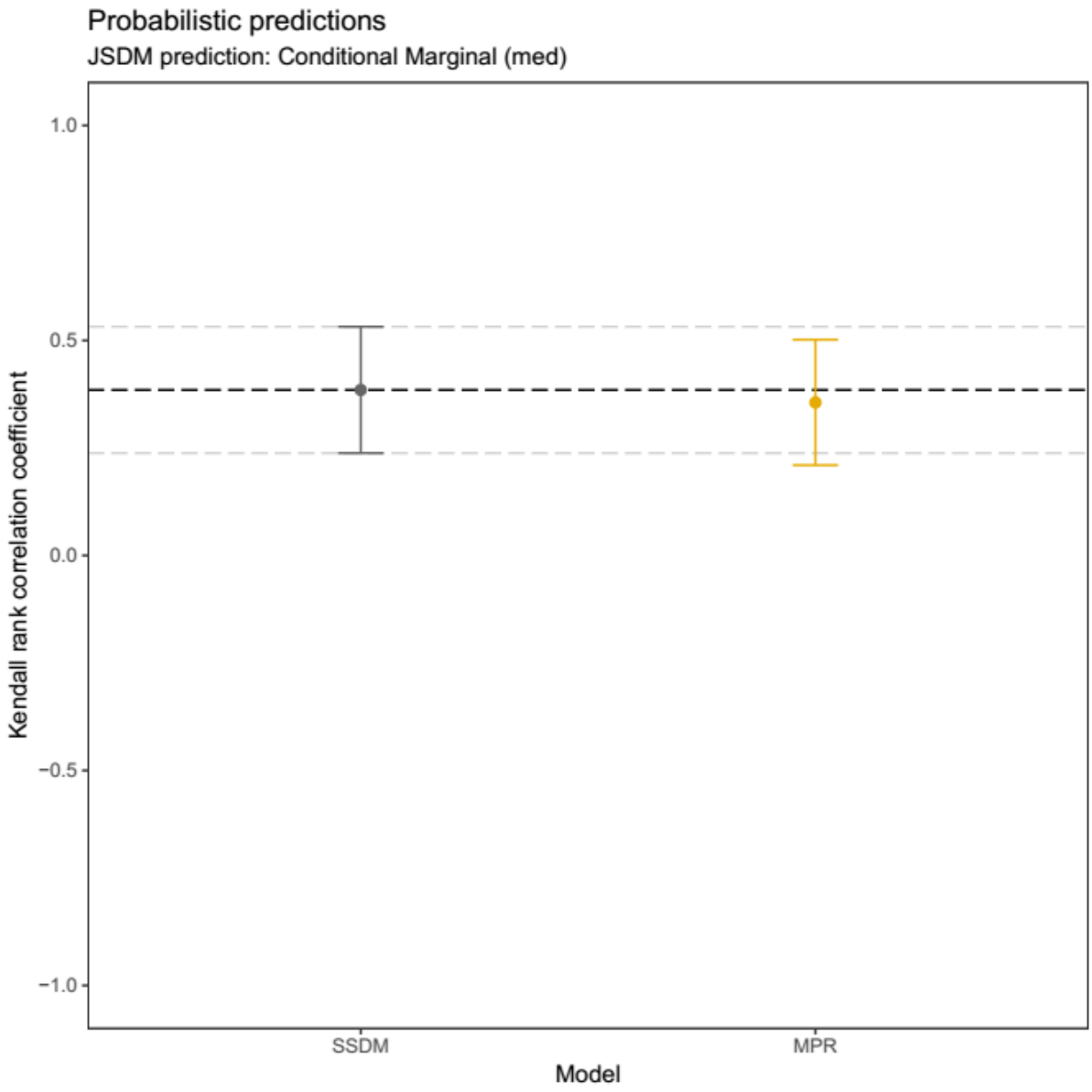


Figure 192: Performance of the SSDM and JSDM for the Kendall rank correlation coefficient evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

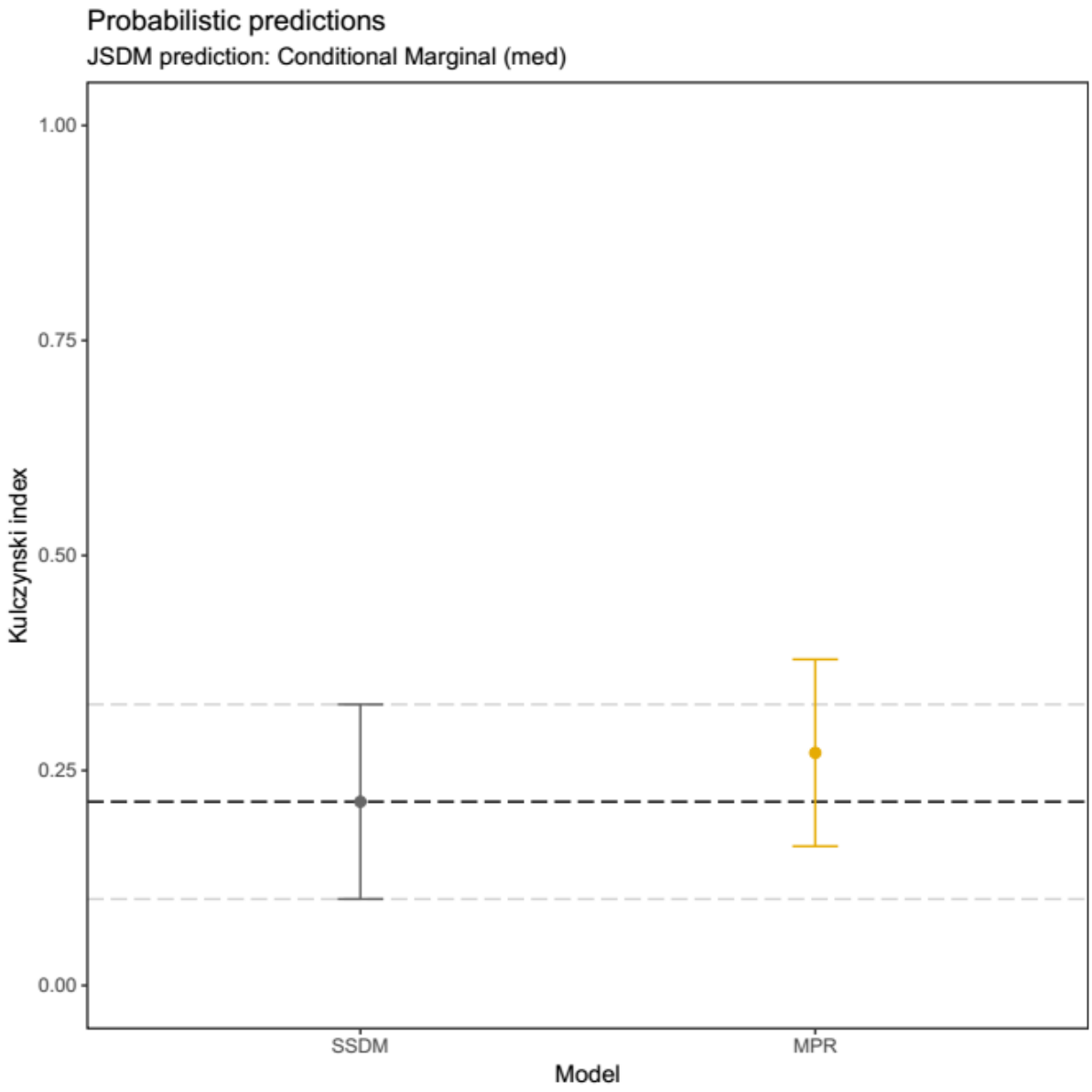


Figure 193: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

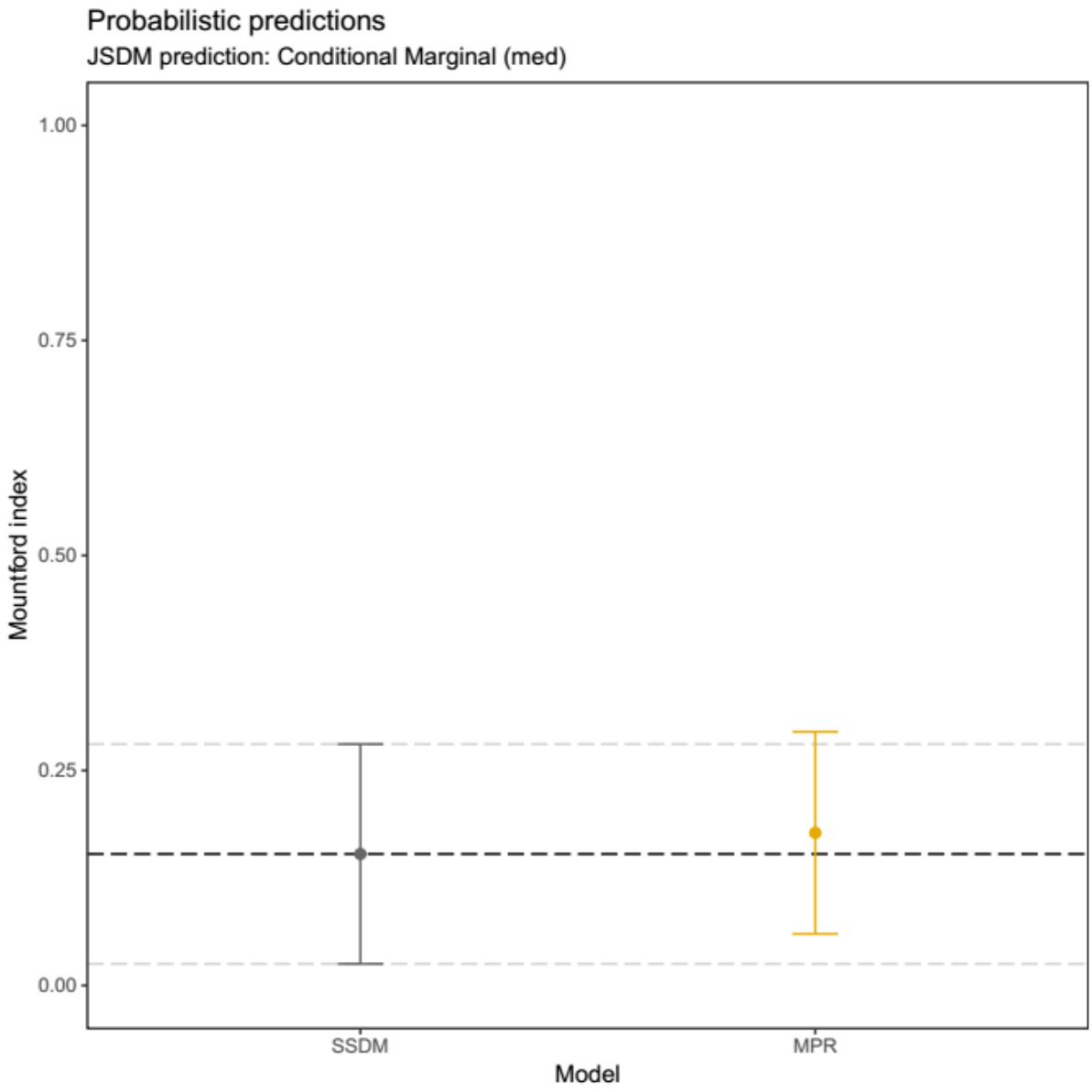


Figure 194: Performance of the SSDM and JSDM for the Mountford index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

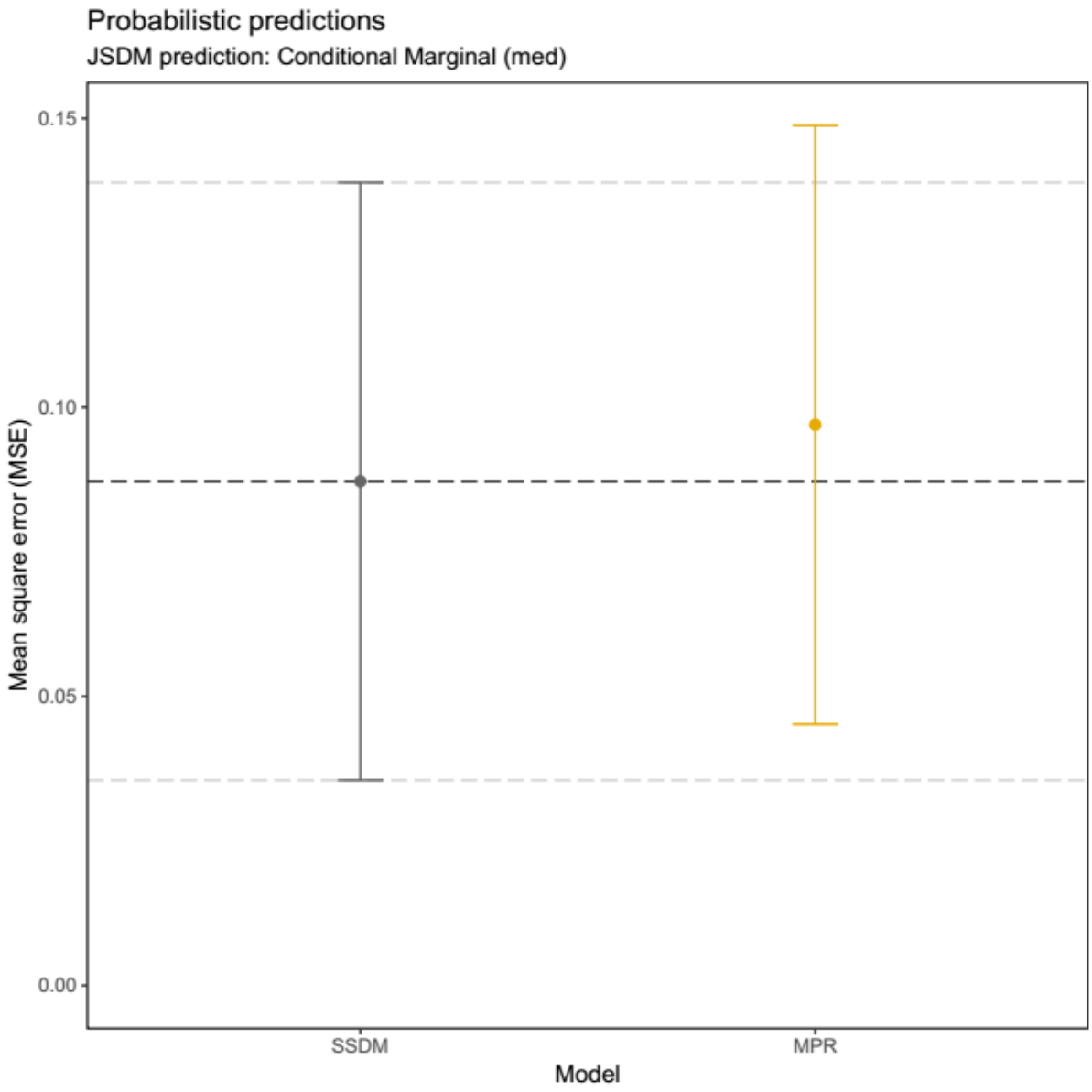


Figure 195: Performance of the SSDM and JSDM for the mean square error evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

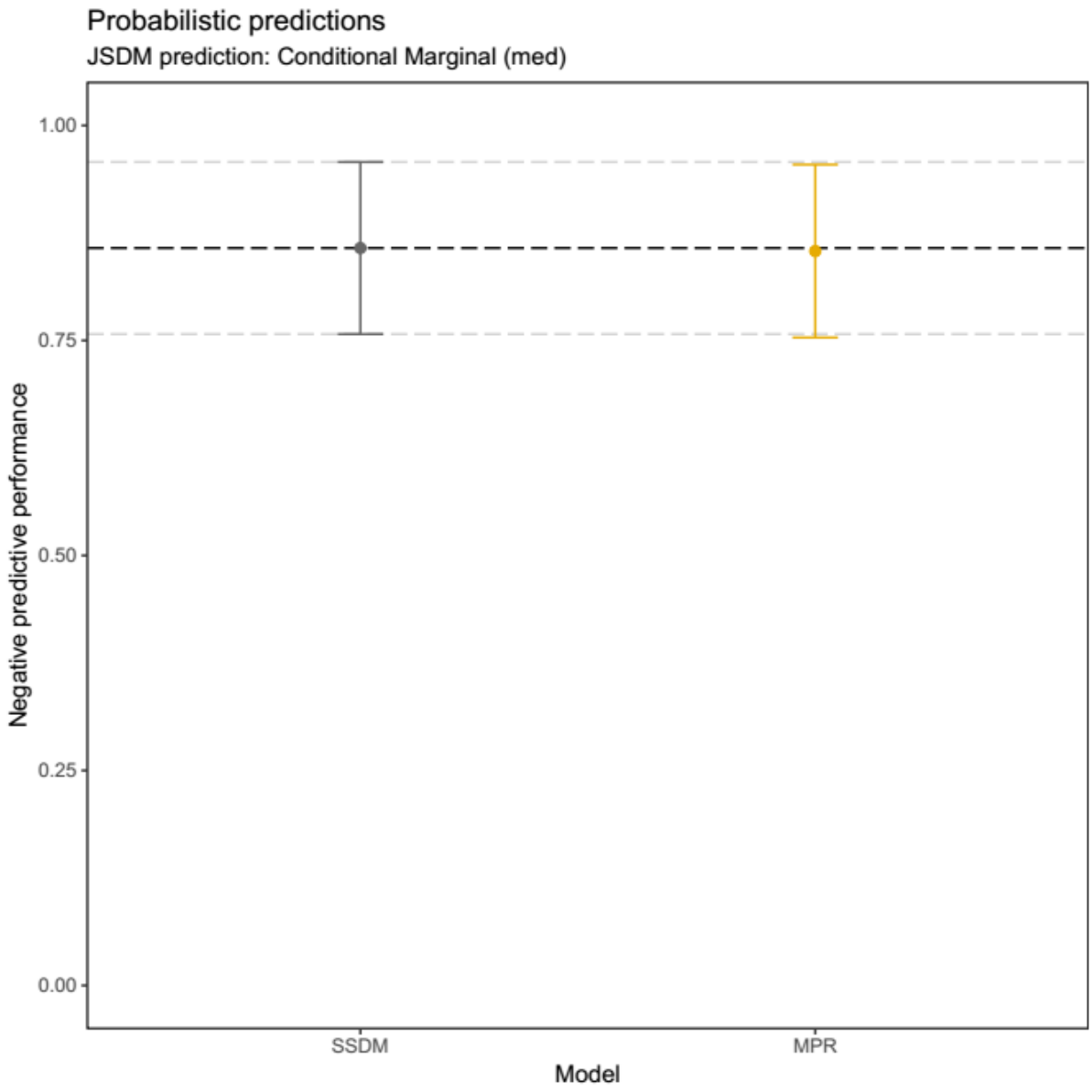


Figure 195: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

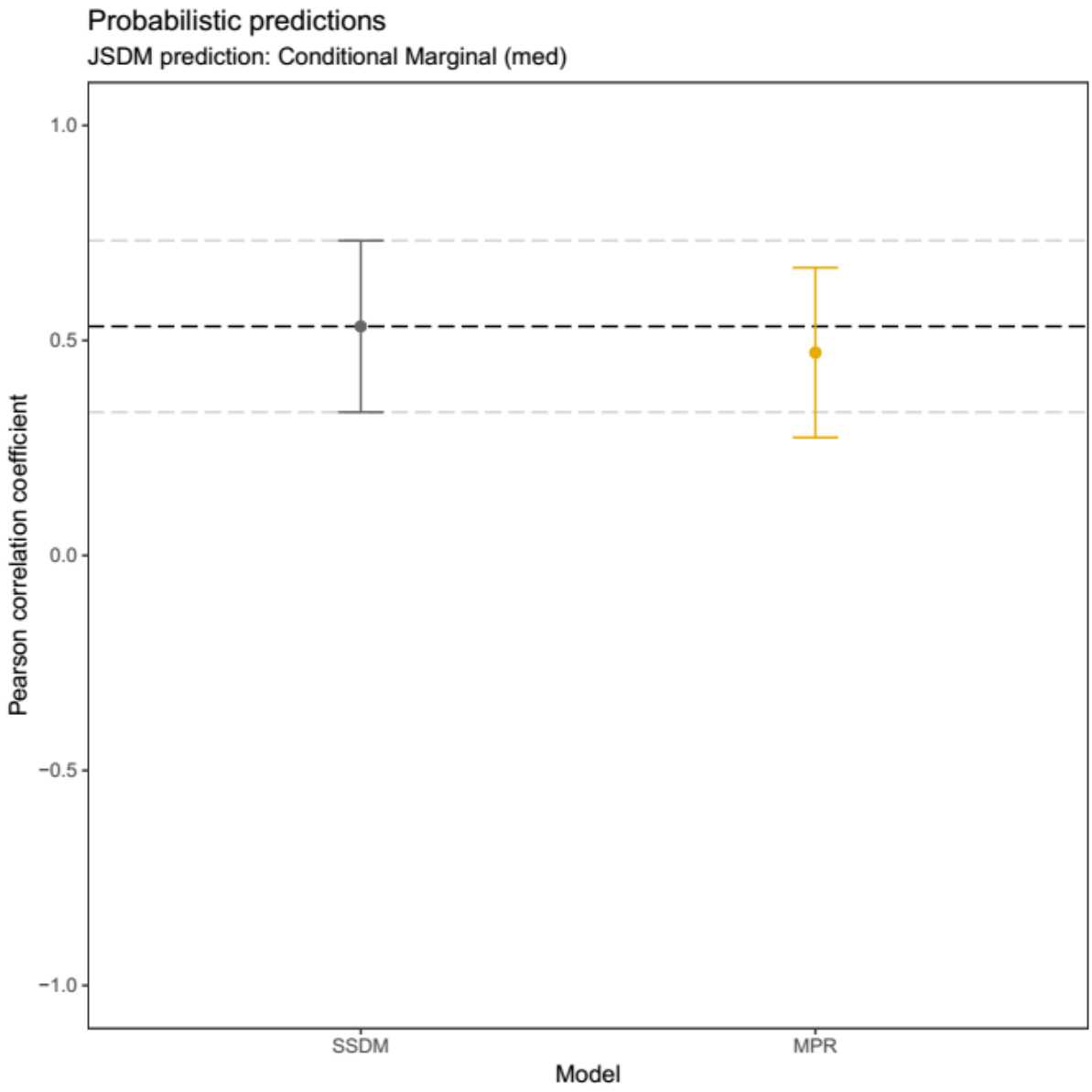


Figure 197: Performance of the SSDM and JSDM for the Pearson correlation coefficient evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

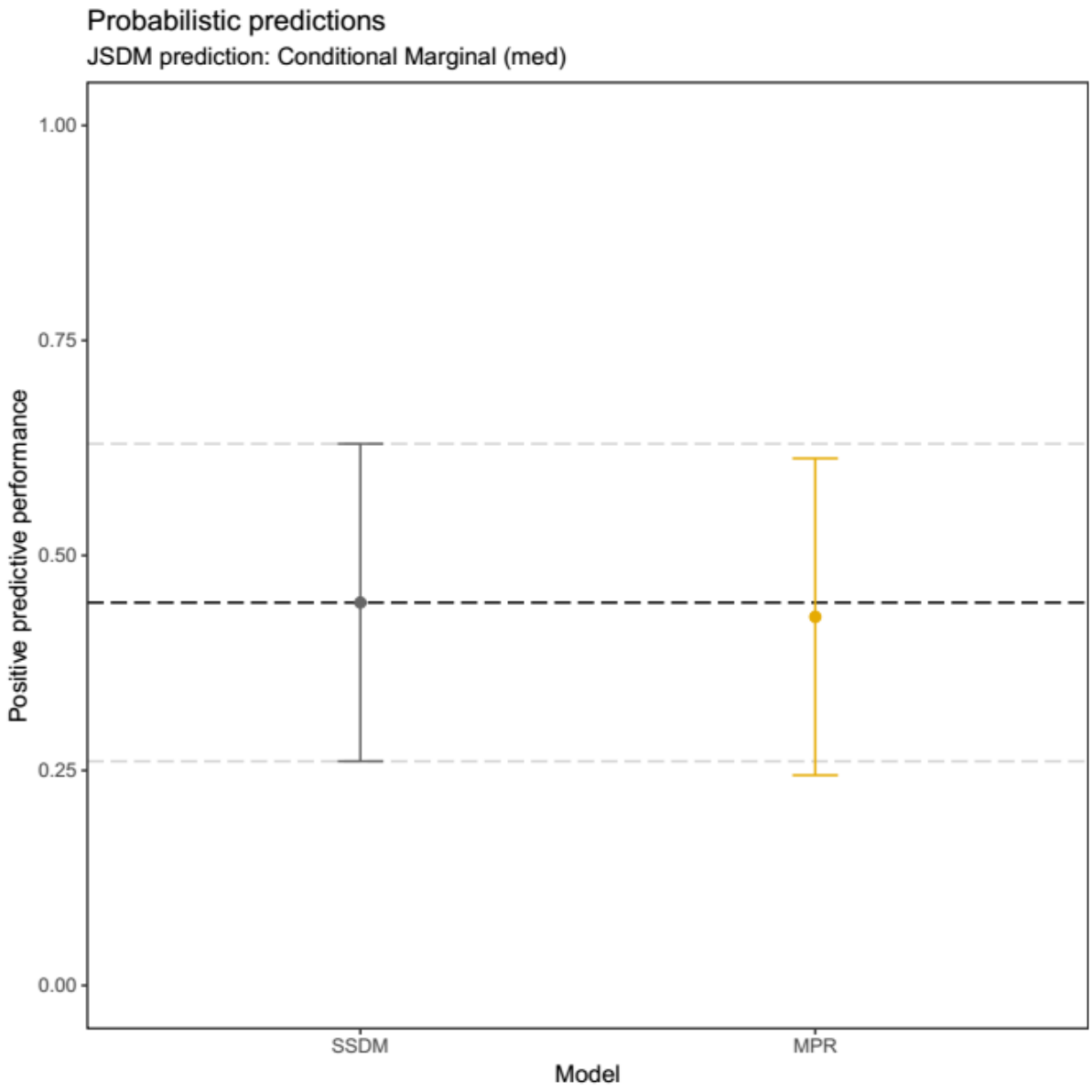


Figure 198: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

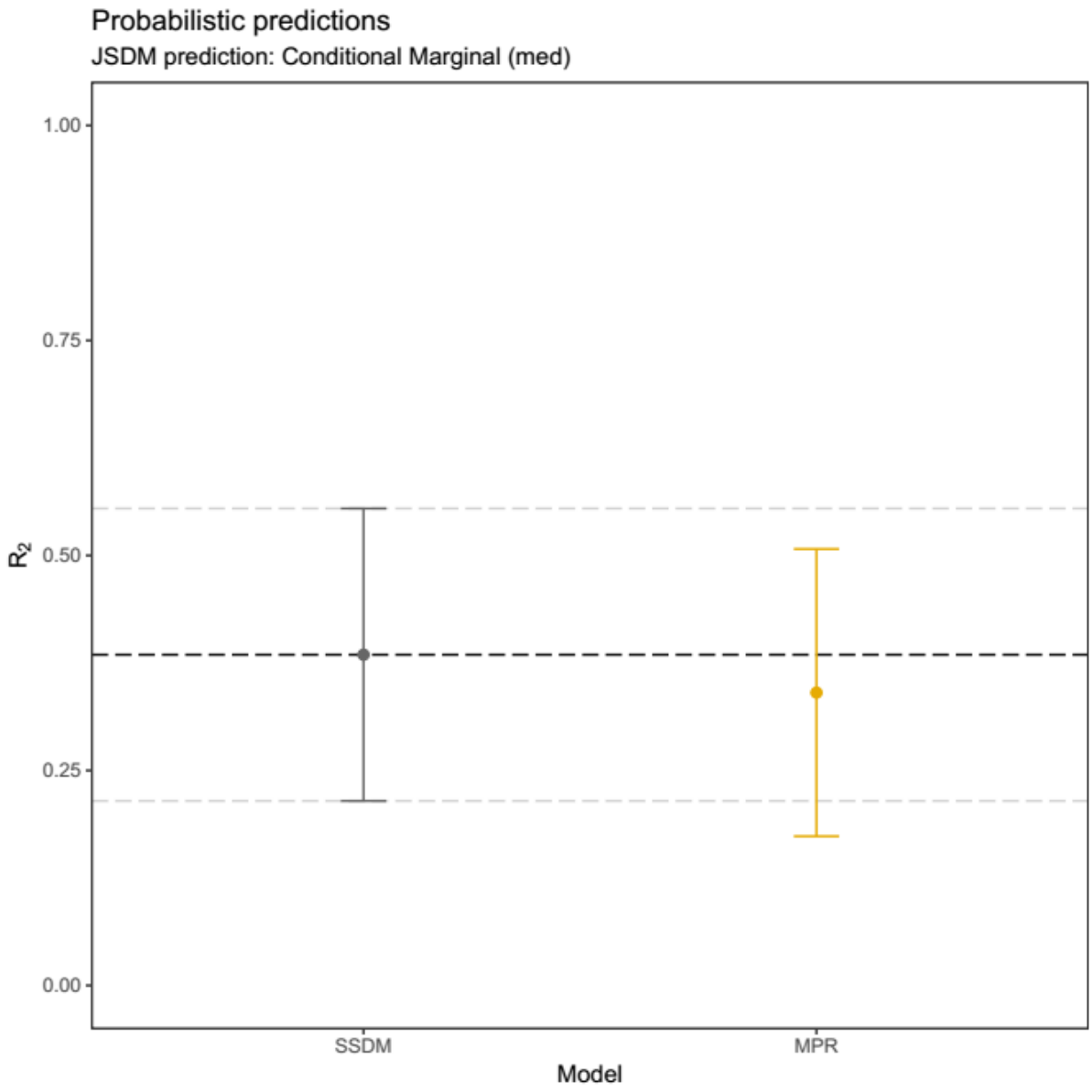


Figure 199: Performance of the SSDM and JSDM for the R^2 evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

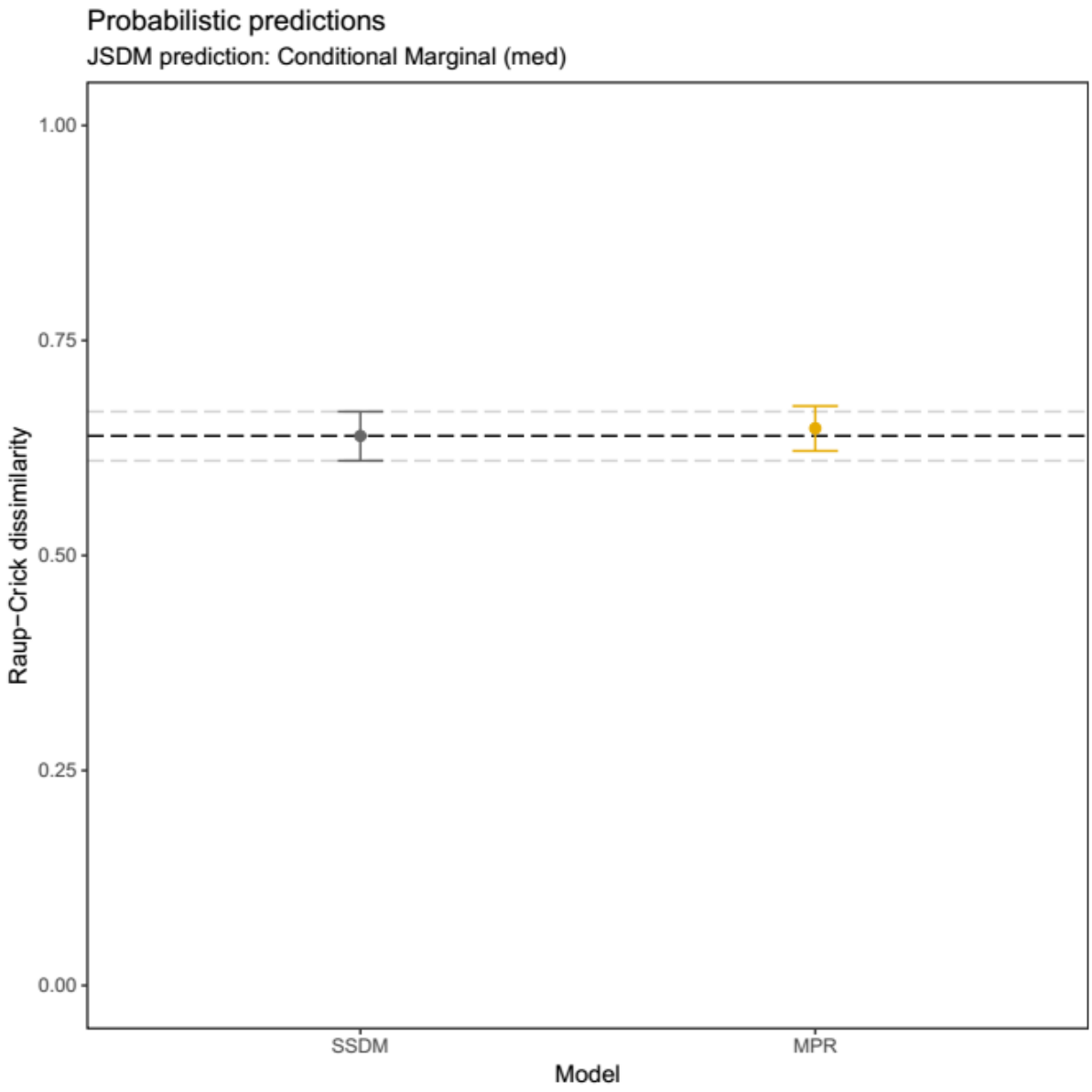


Figure 200: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

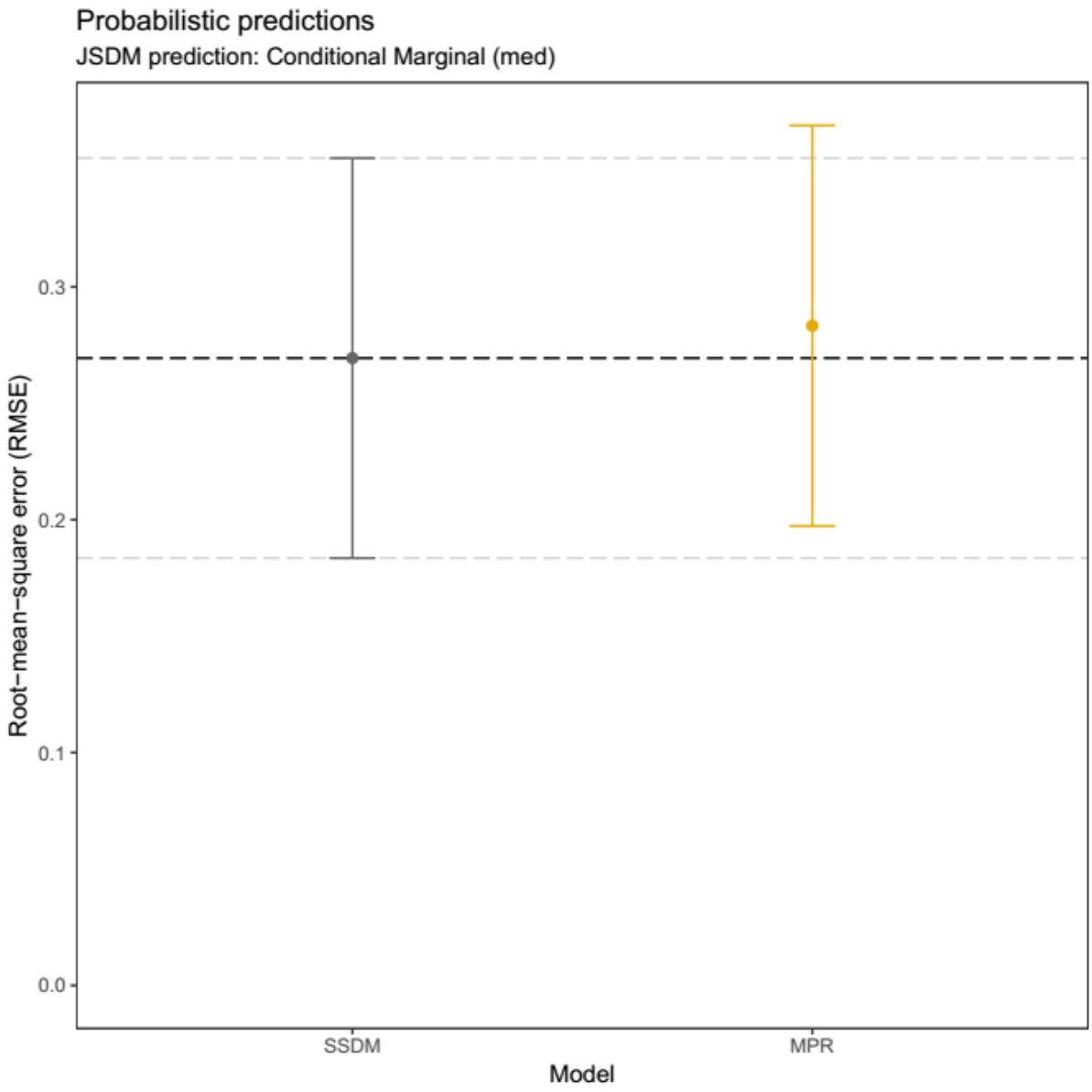


Figure 201: Performance of the SSDM and JSDM for the root-mean-square error evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

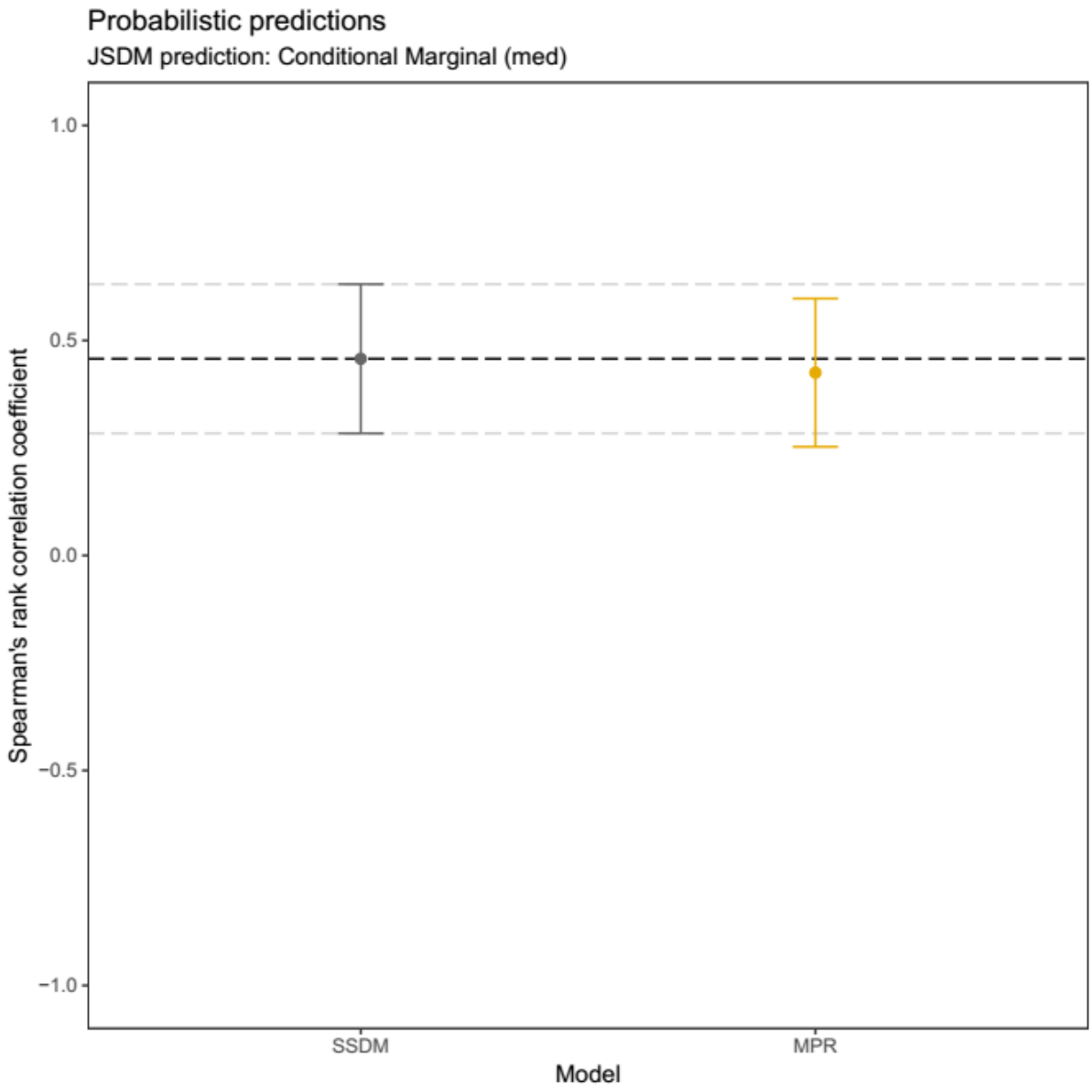


Figure 202: Performance of the SSDM and JSDM for the Spearman's rank correlation coefficient evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

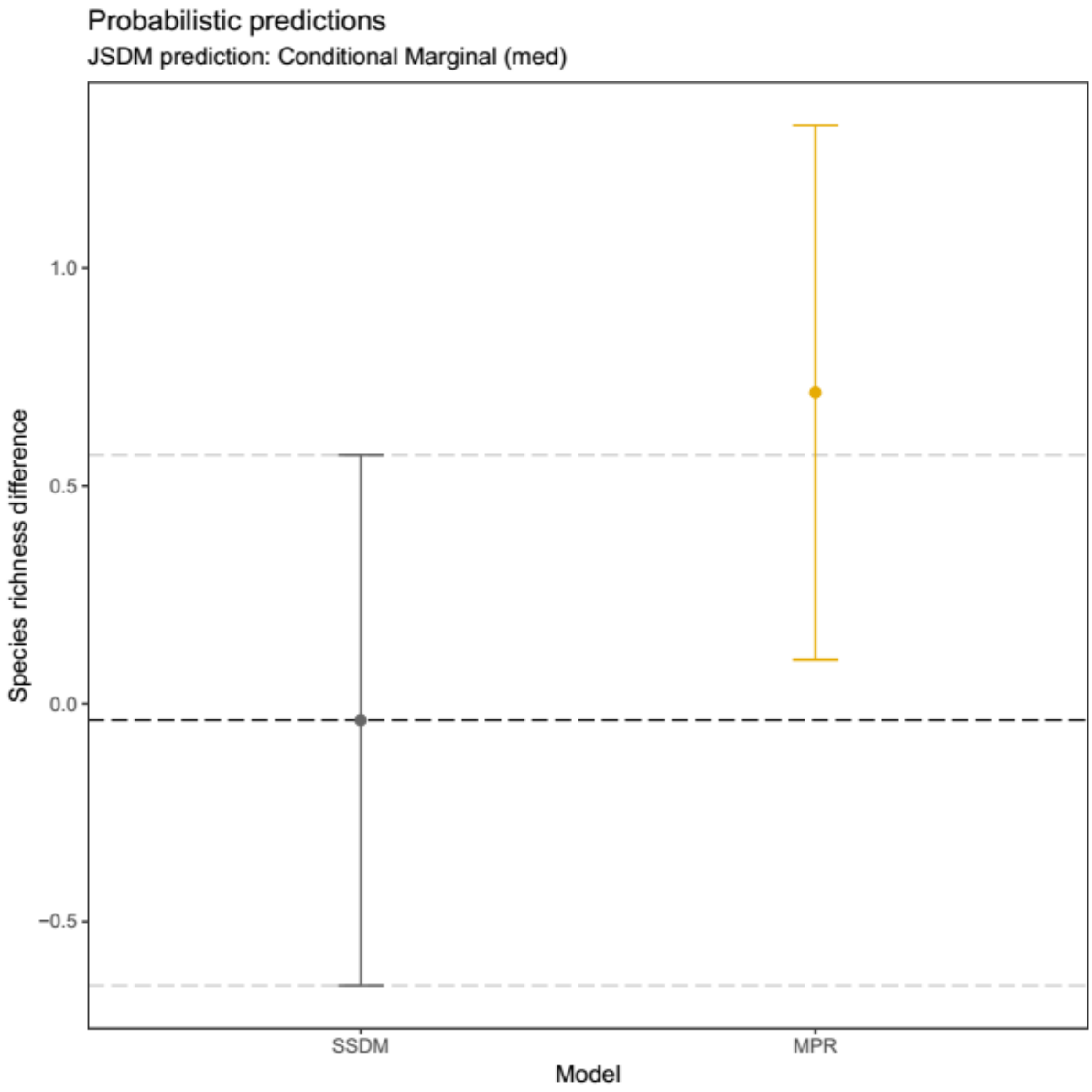


Figure 203: Performance of the SSDM and JSDM for the species richness difference evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

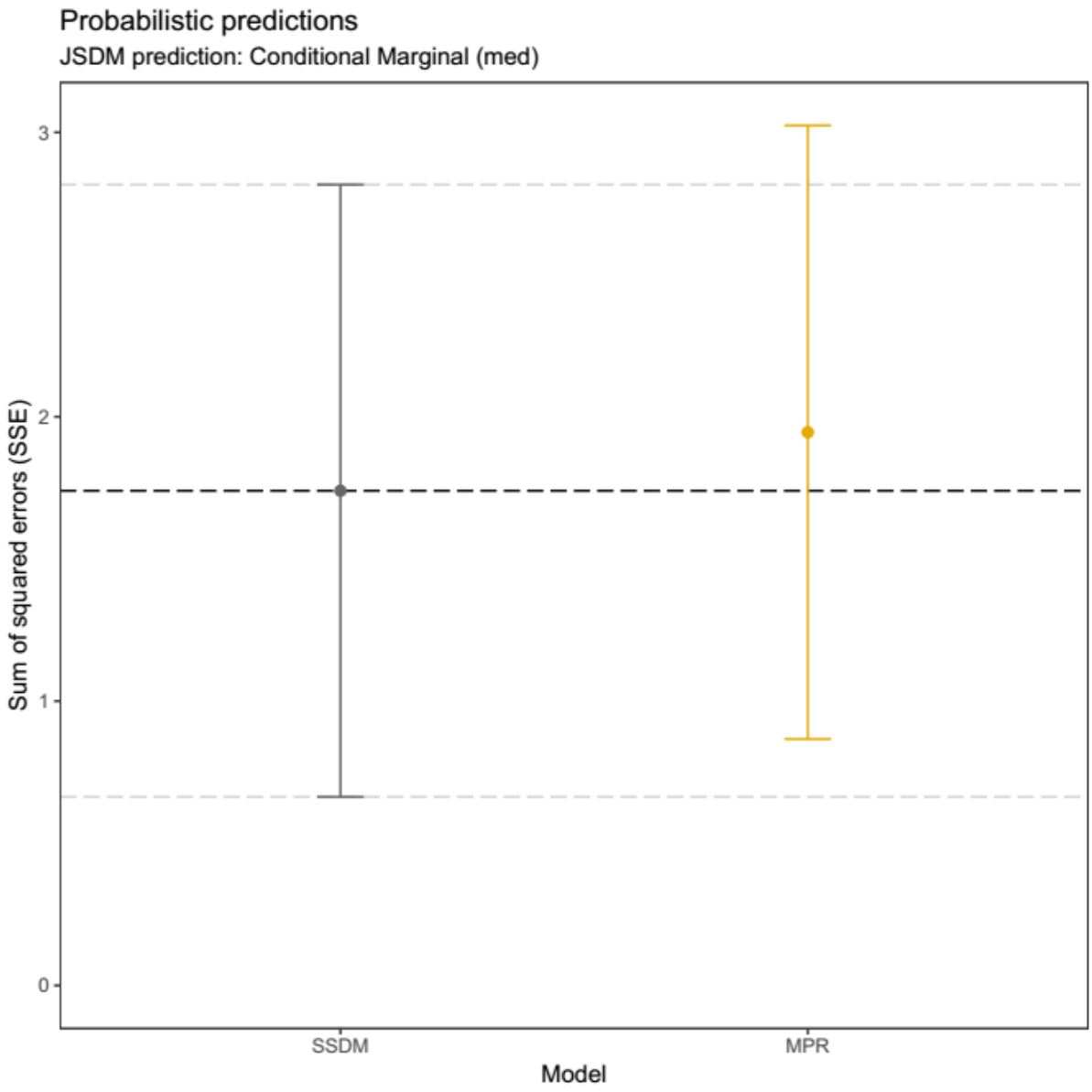


Figure 204: Performance of the SSDM and JSDM for the sum of squared errors evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

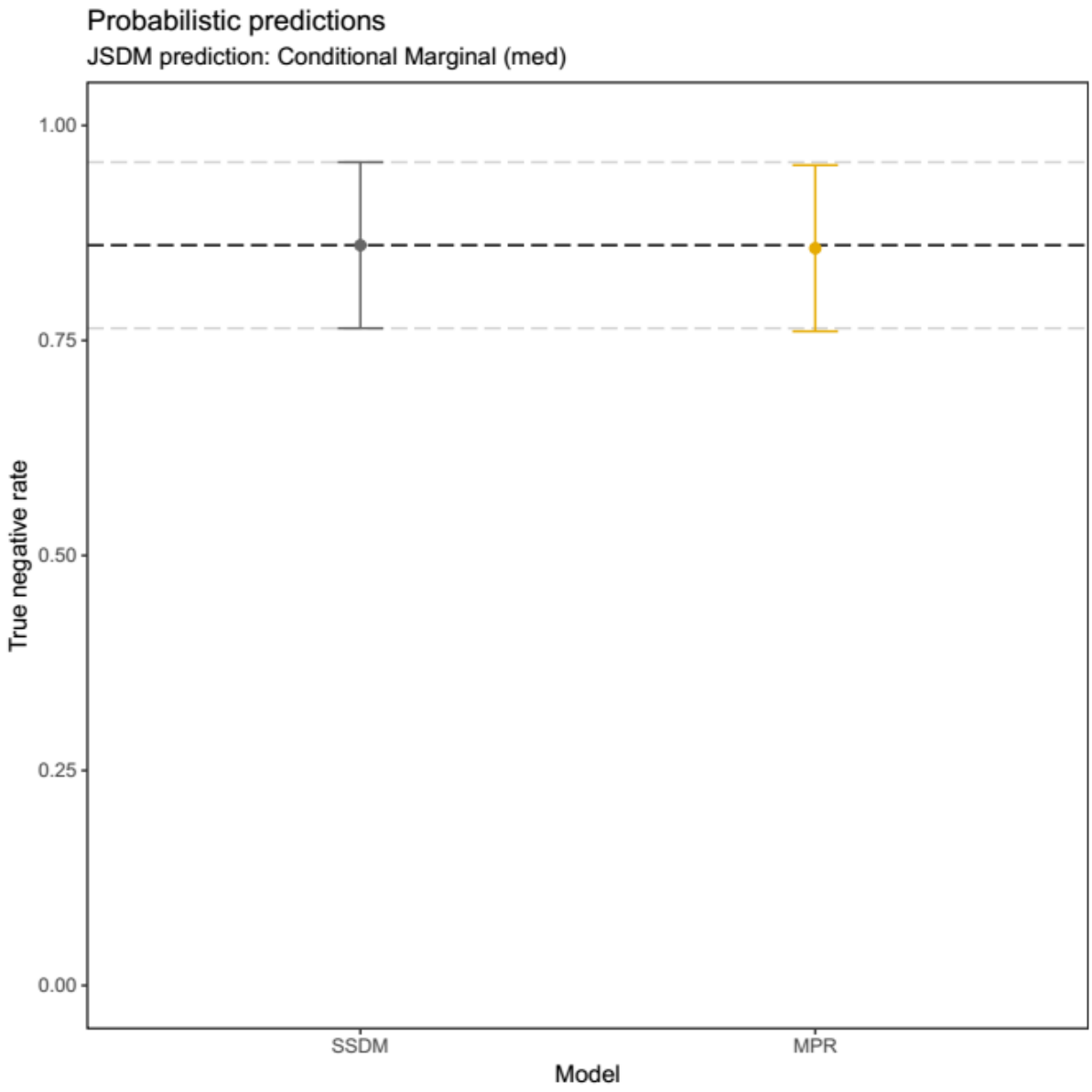


Figure 205: Performance of the SSDM and JSDM for the true negative rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

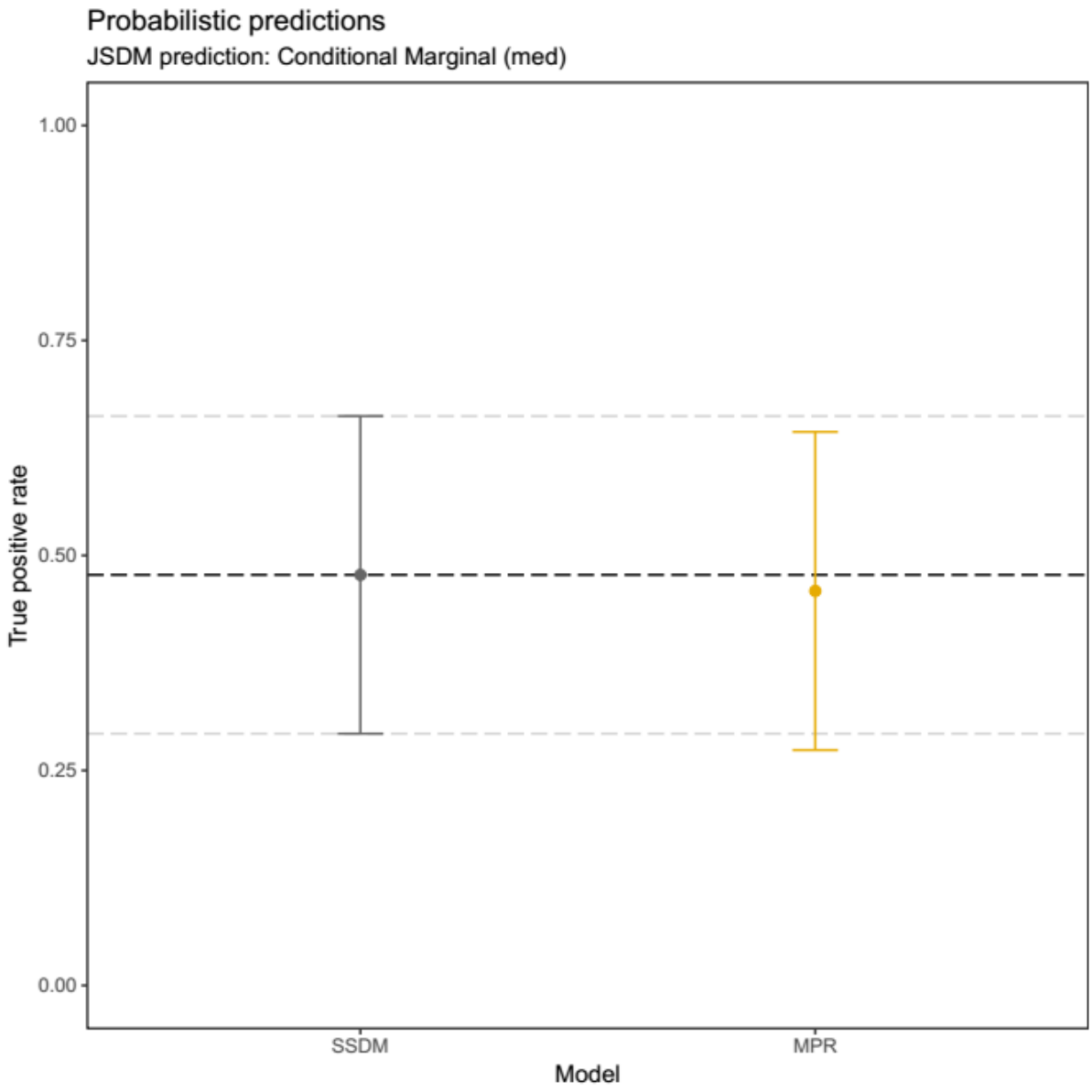


Figure 206: Performance of the SSDM and JSDM for the true positive rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

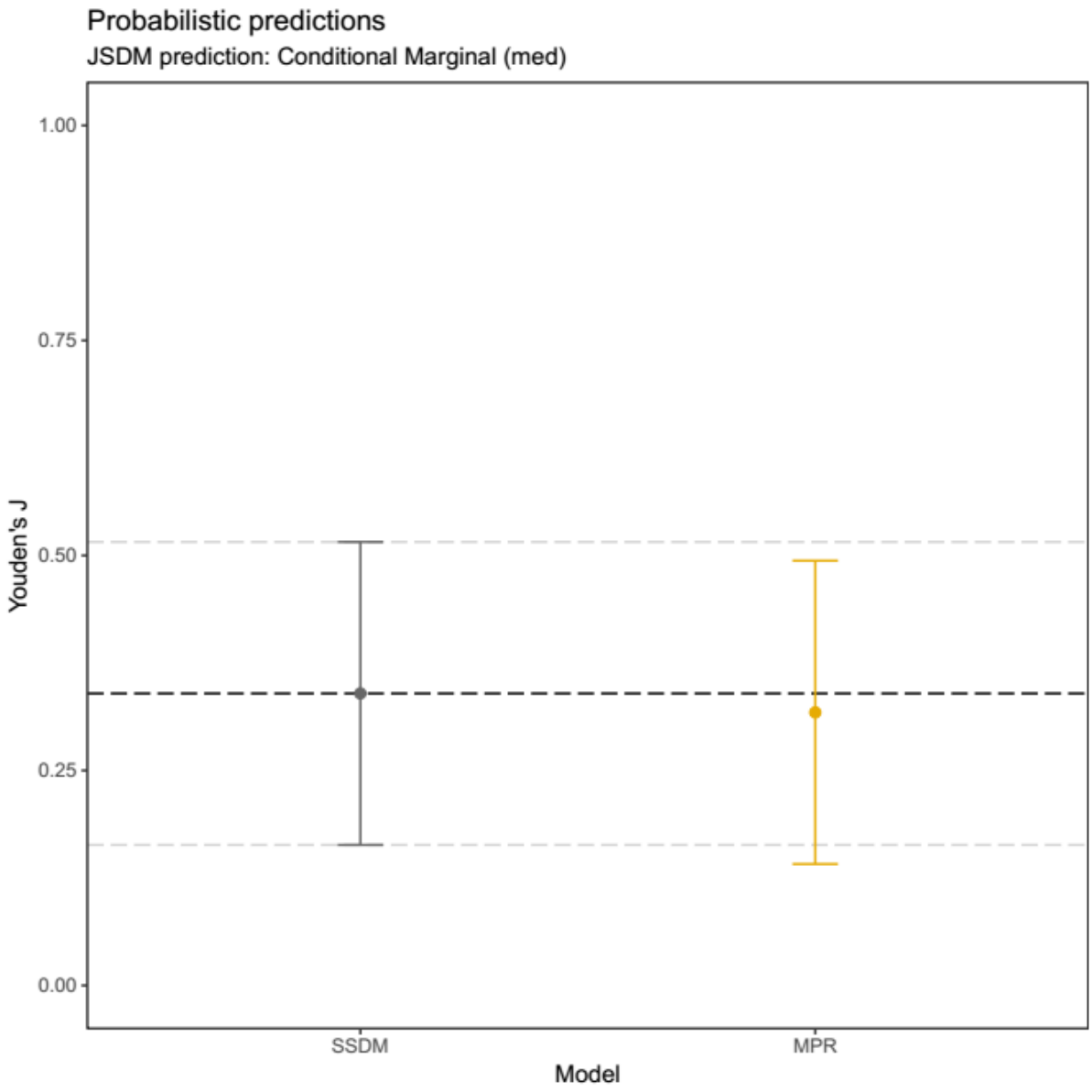


Figure 207: Performance of the SSDM and JSDM for the Youden's J evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

9. Conditional marginal (high-prevalence) predictions

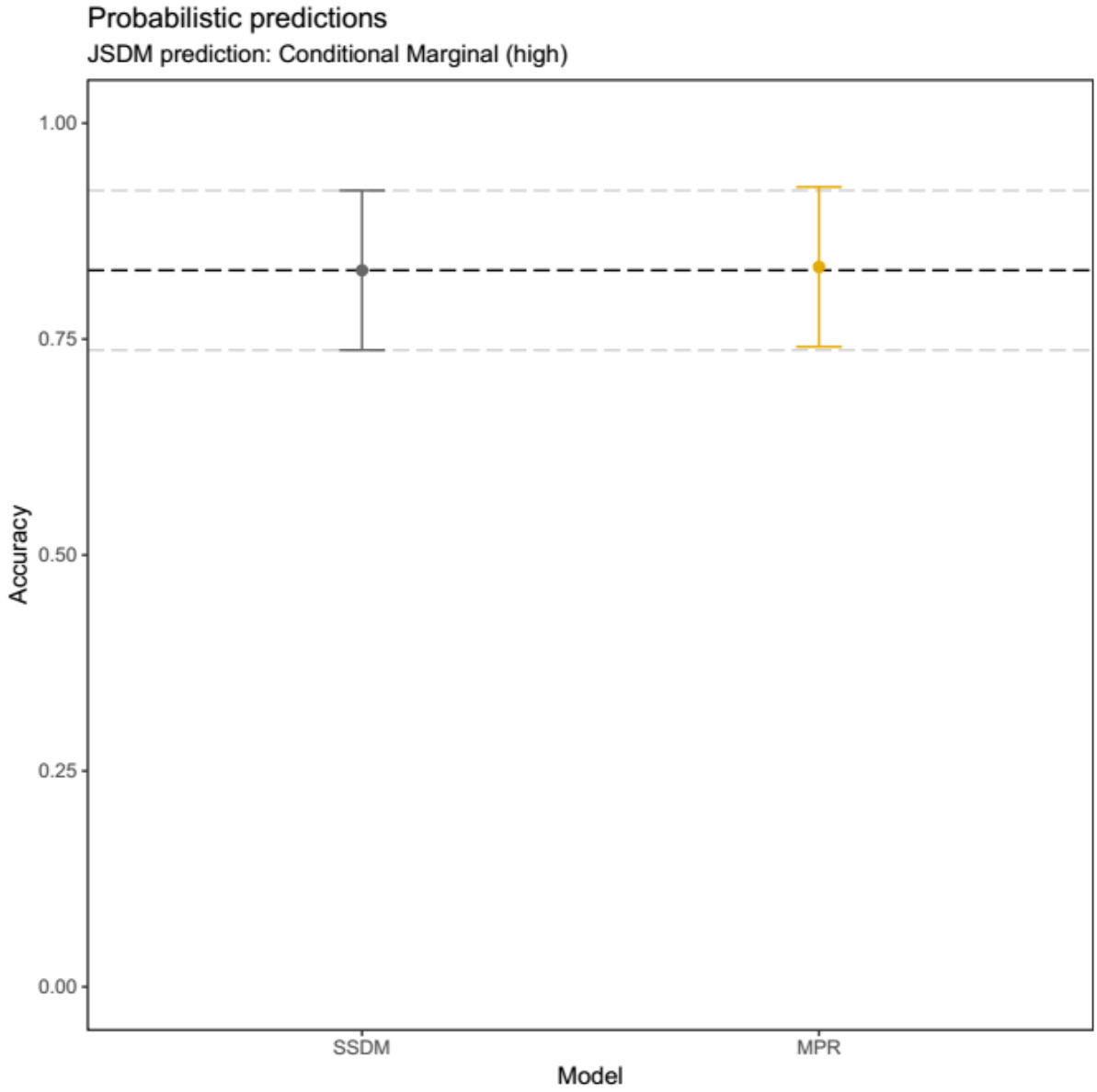


Figure 208: Performance of the SSDM and JSDM for the accuracy evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

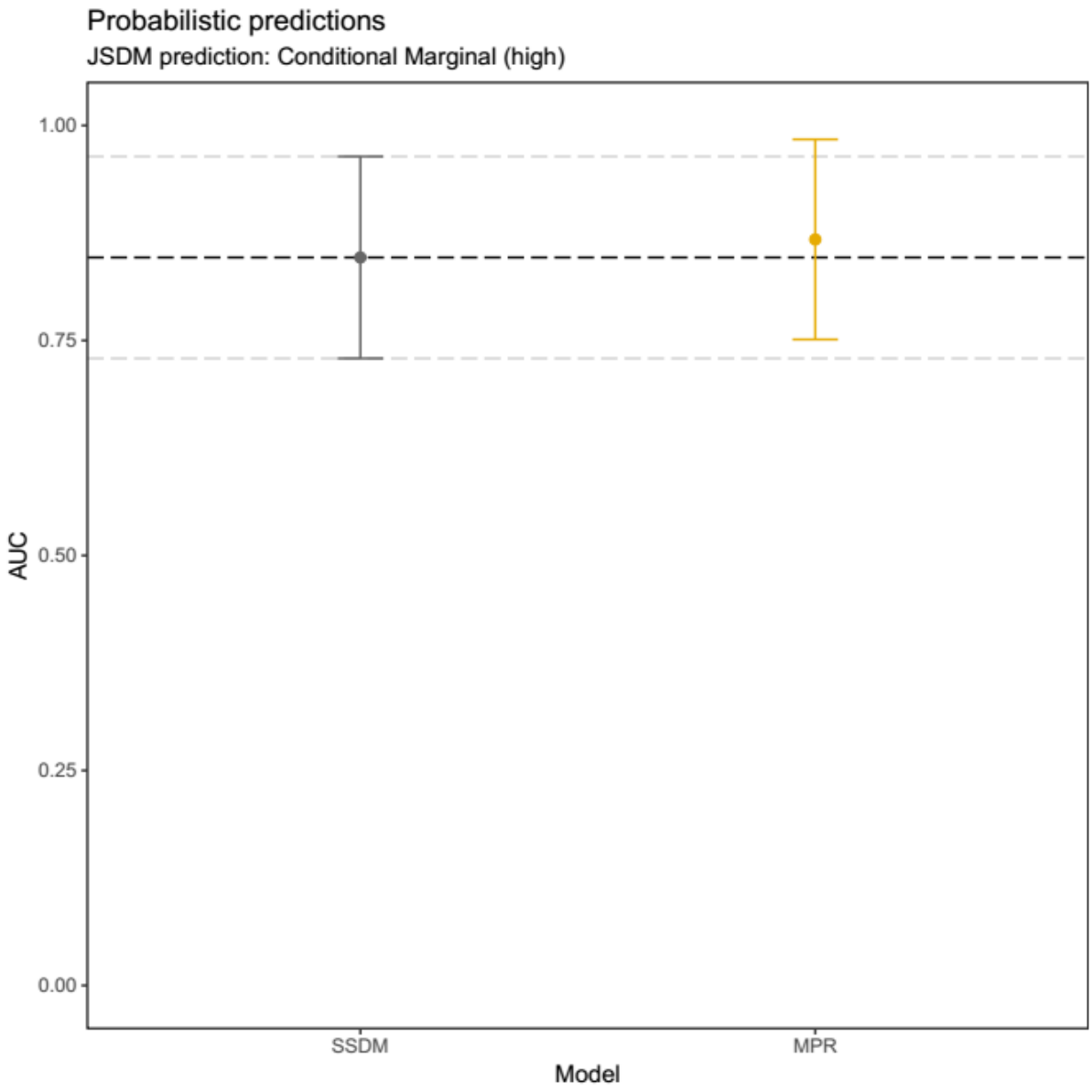


Figure 209: Performance of the SSDM and JSDM for the AUC evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

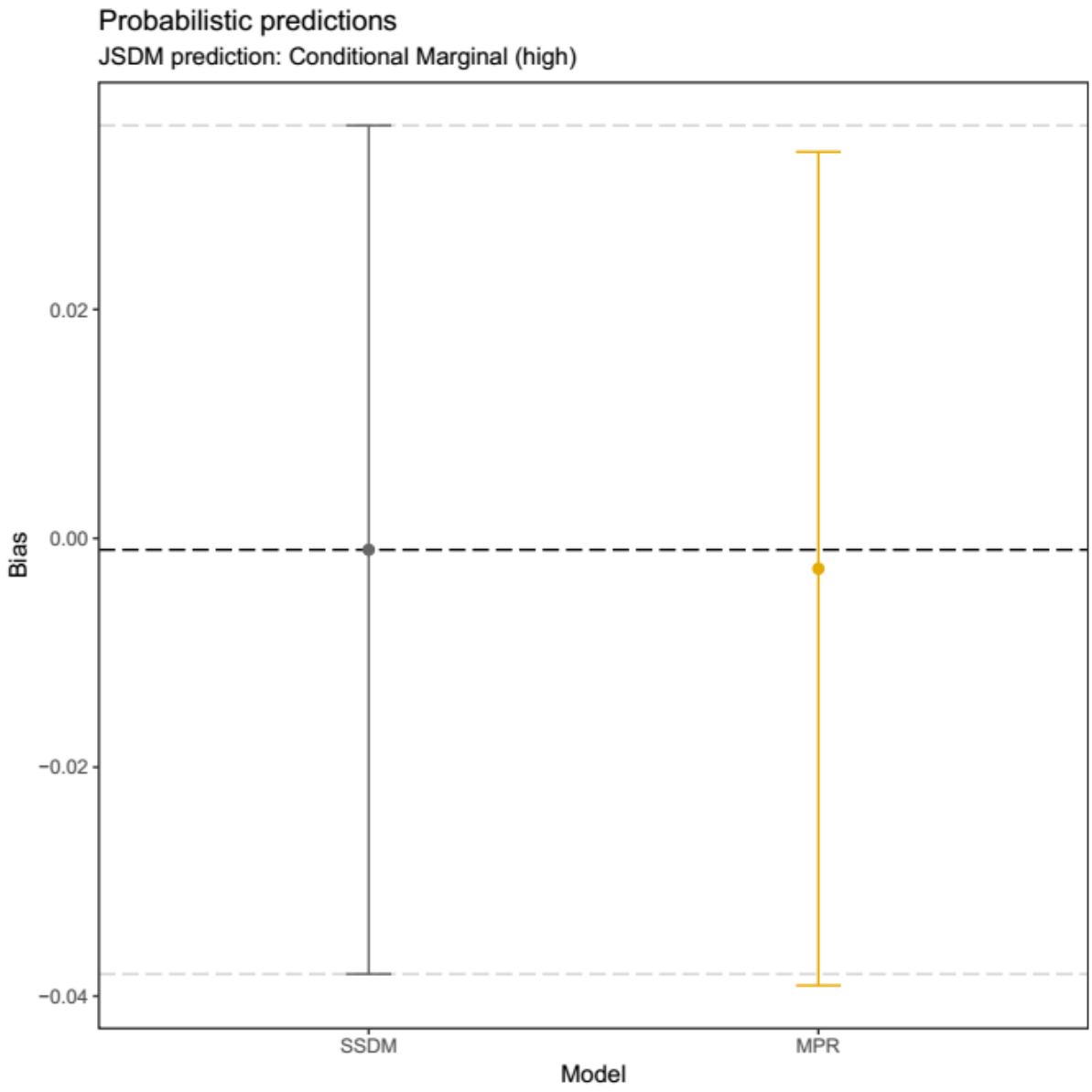


Figure 210: Performance of the SSDM and JSDM for the mean error/bias evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

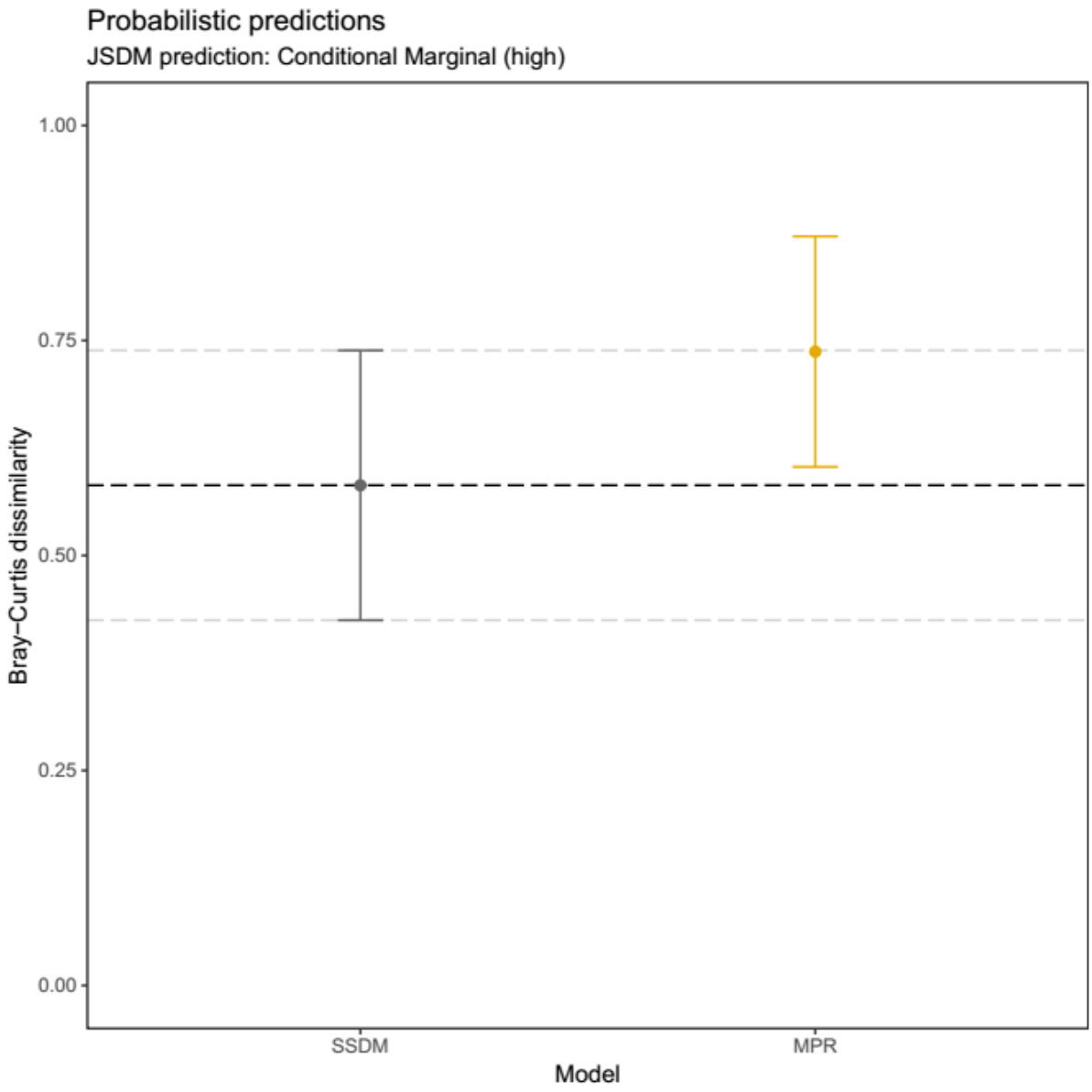


Figure 211: Performance of the SSDM and JSDM for the Bray-Curtis dissimilarity evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

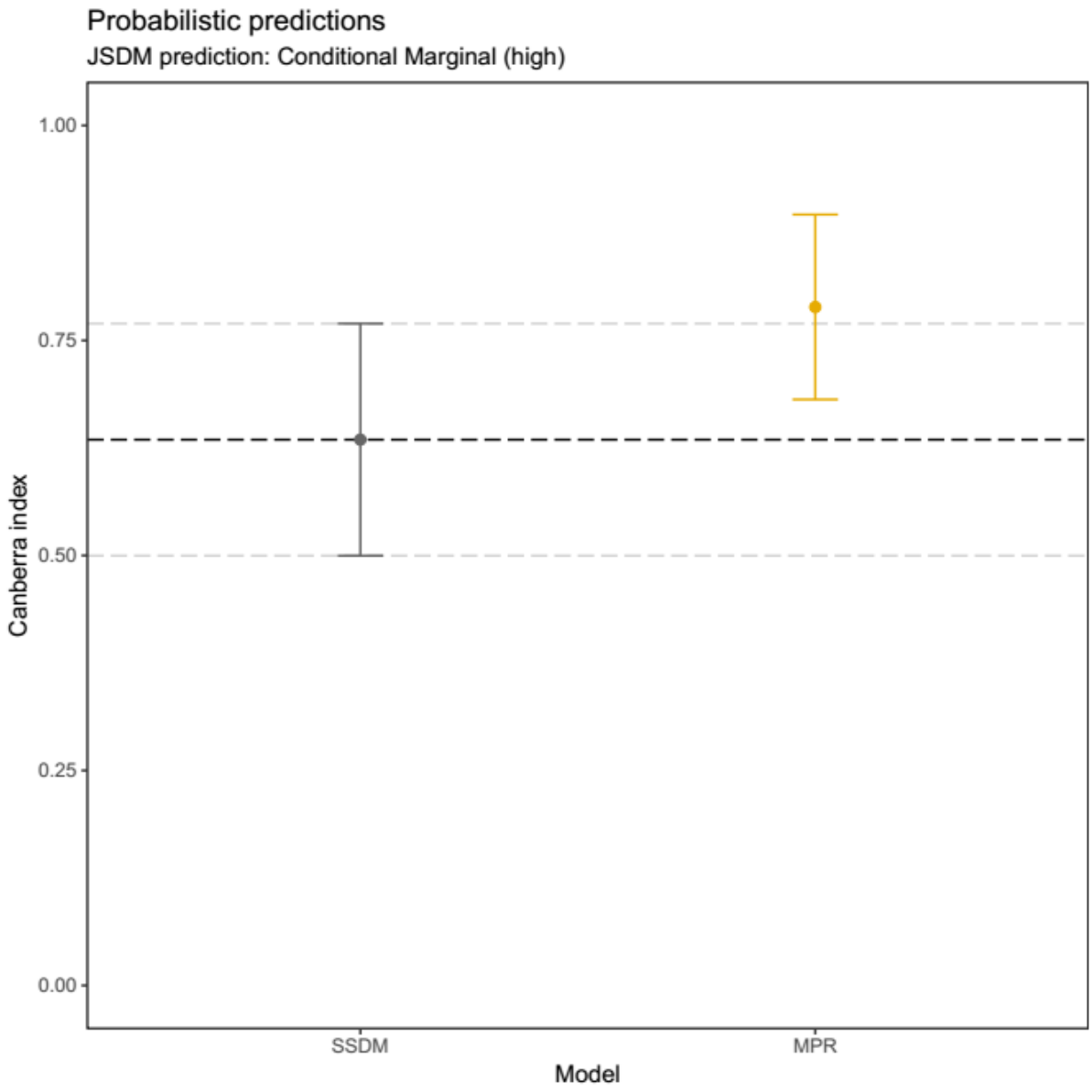


Figure 212: Performance of the SSDM and JSDM for the Canberra index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

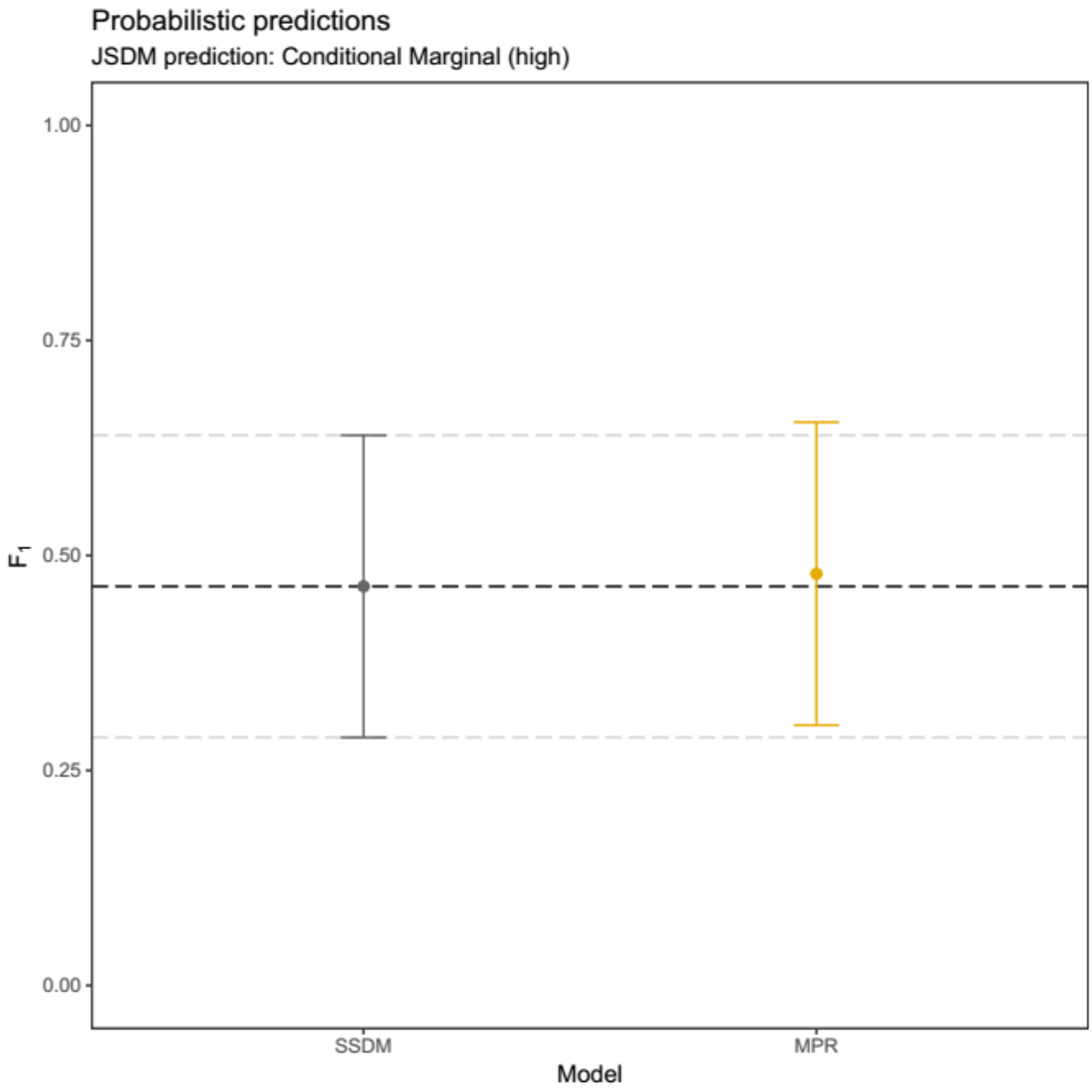


Figure 213: Performance of the SSDM and JSDM for the F_1 evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

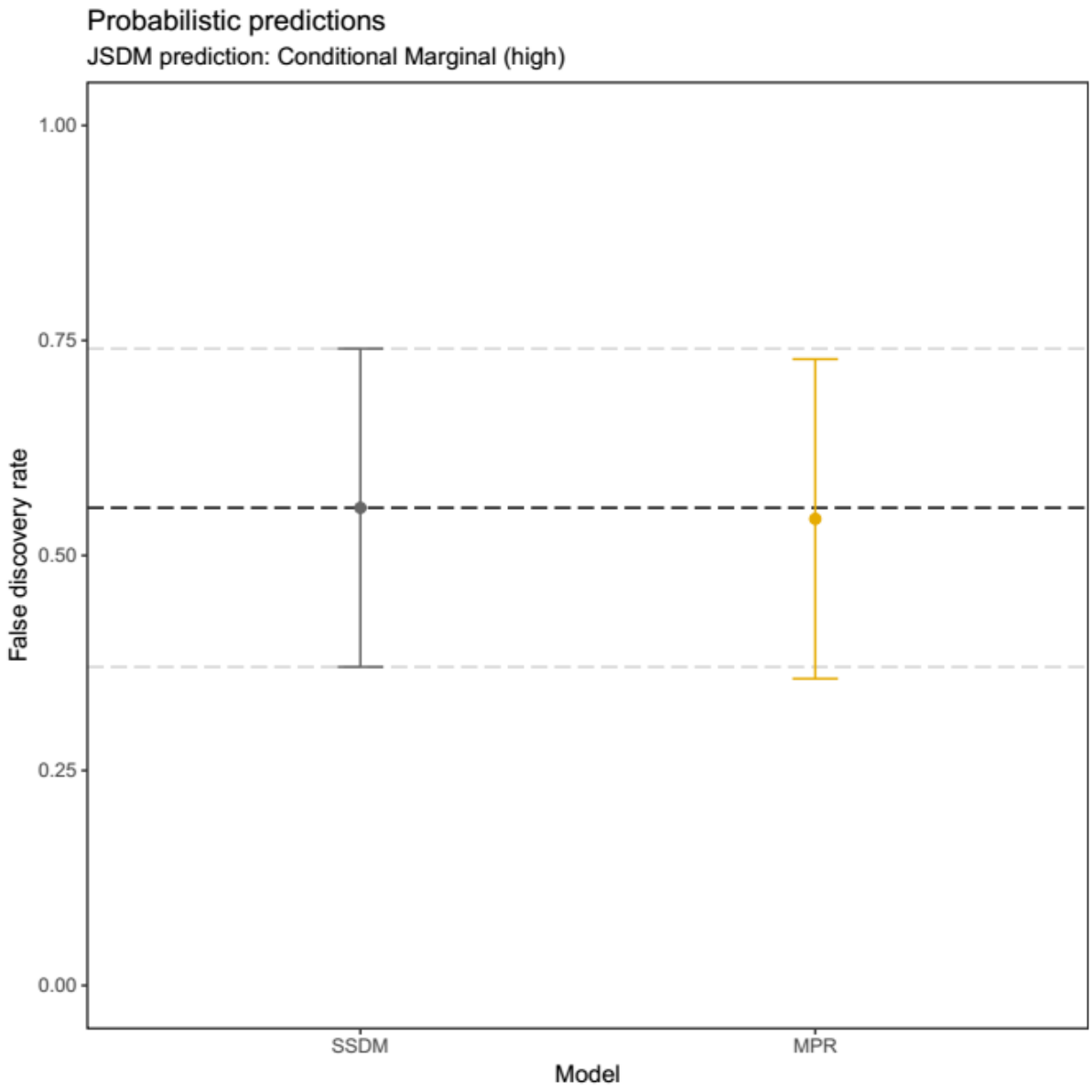


Figure 214: Performance of the SSDM and JSDM for the false discovery rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

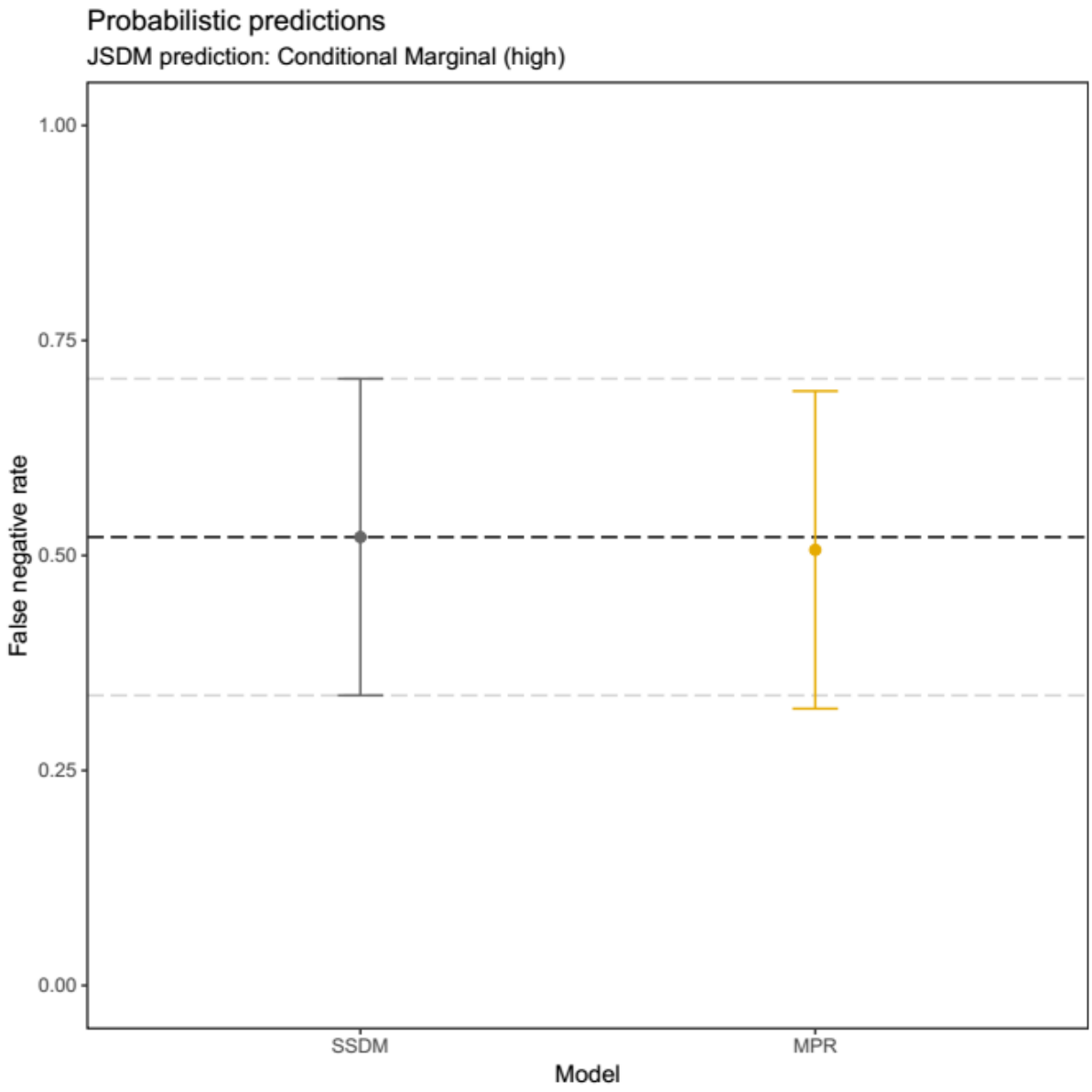


Figure 215: Performance of the SSDM and JSDM for the false negative rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

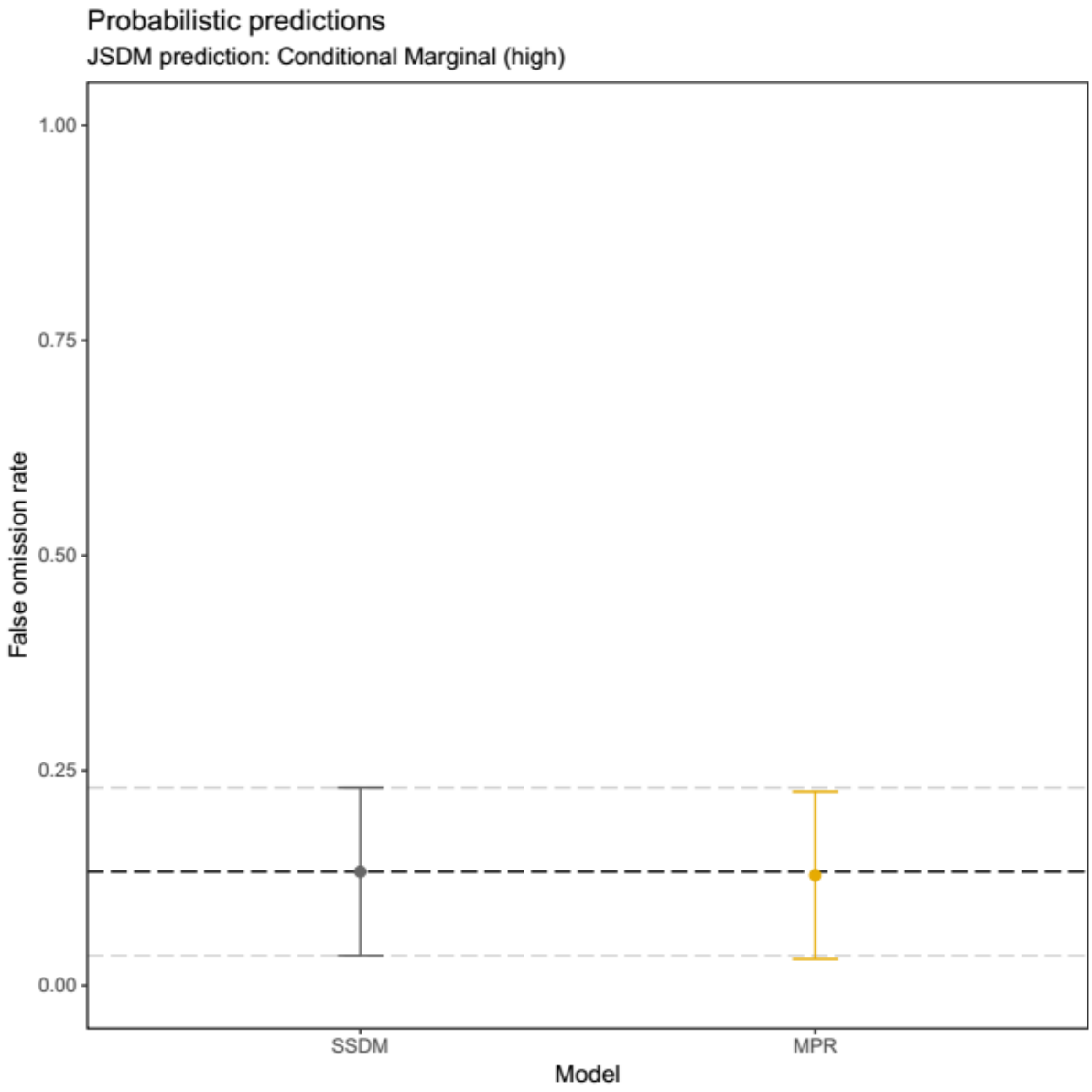


Figure 216: Performance of the SSDM and JSDM for the false omission rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

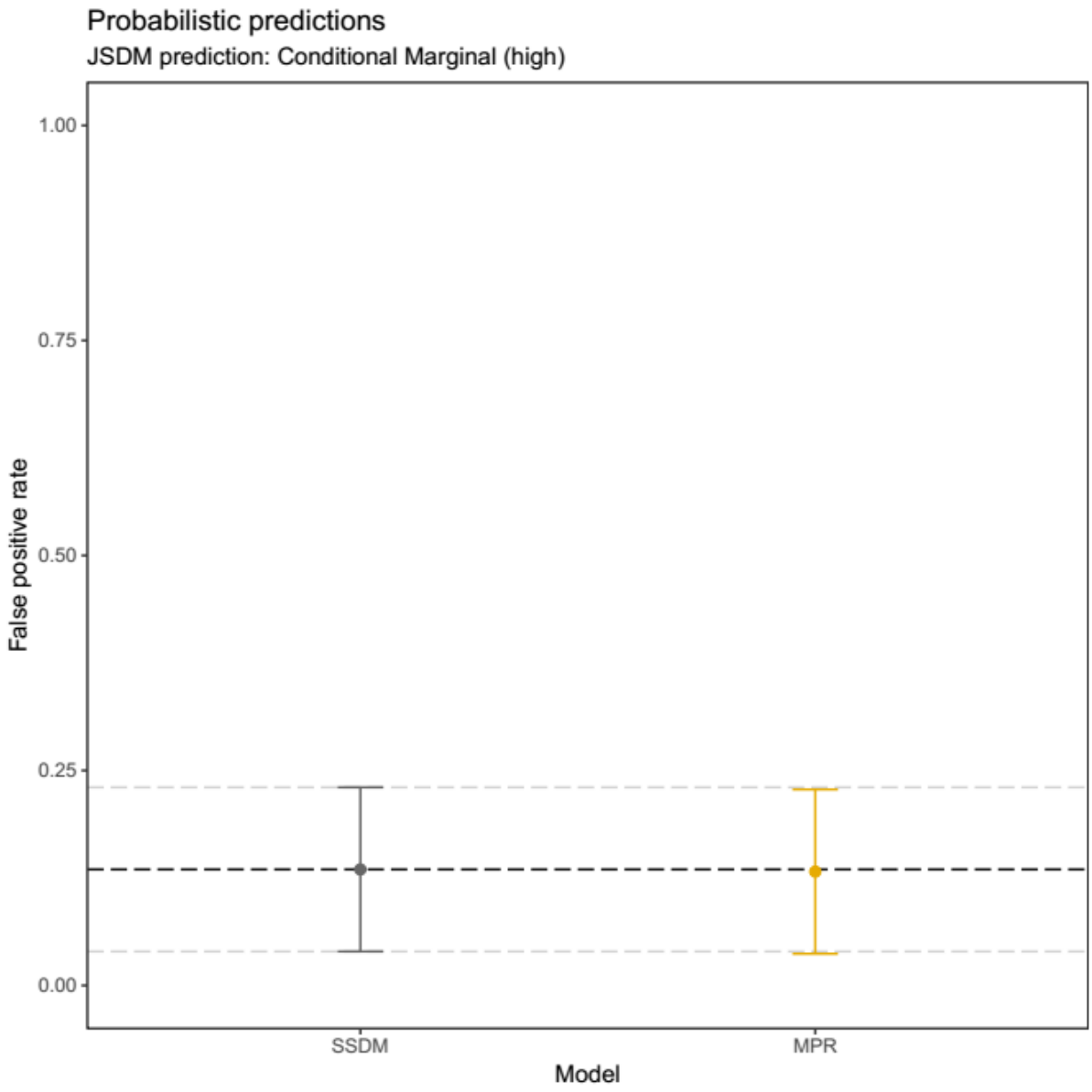


Figure 217: Performance of the SSDM and JSDM for the false positive rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

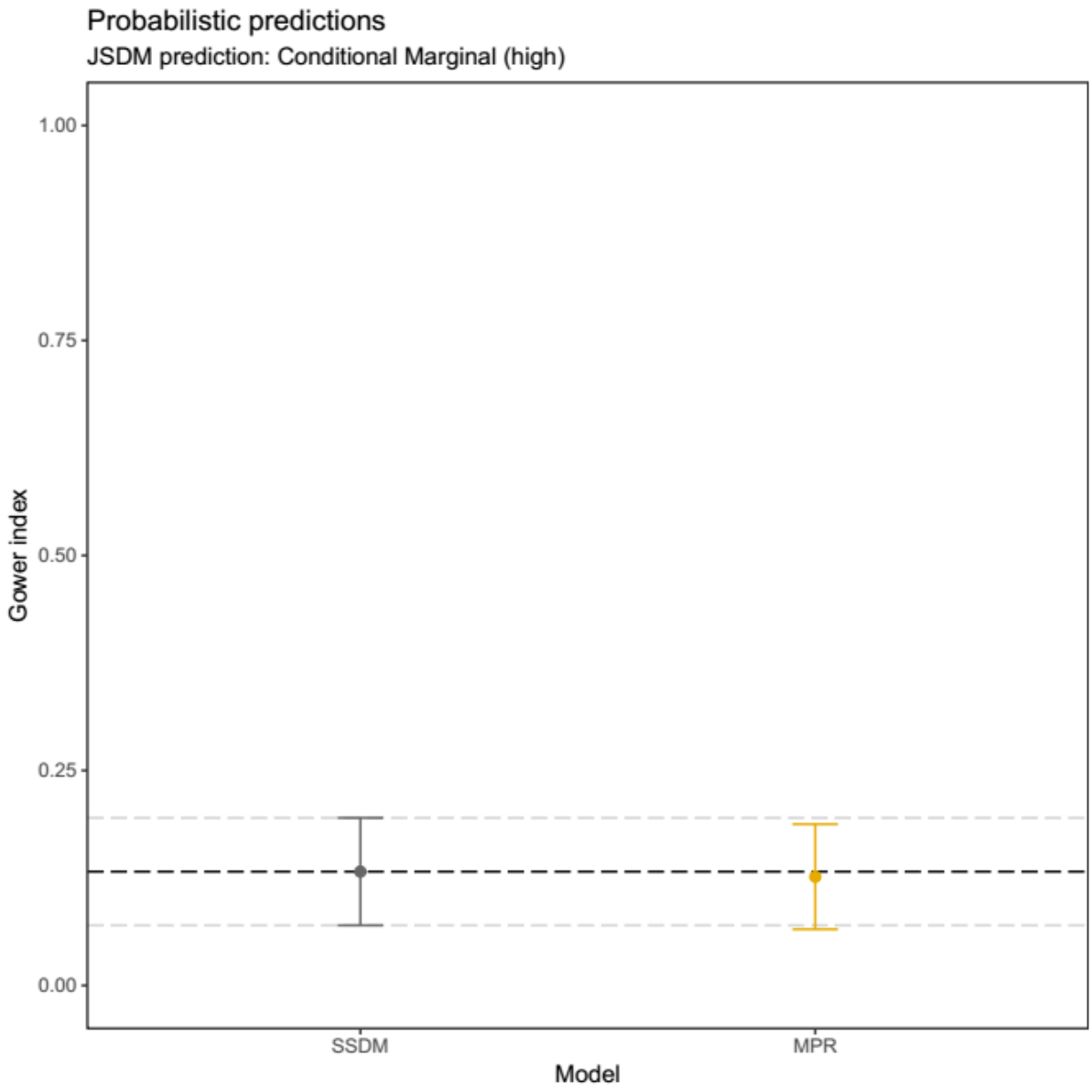


Figure 218: Performance of the SSDM and JSDM for the Gower index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

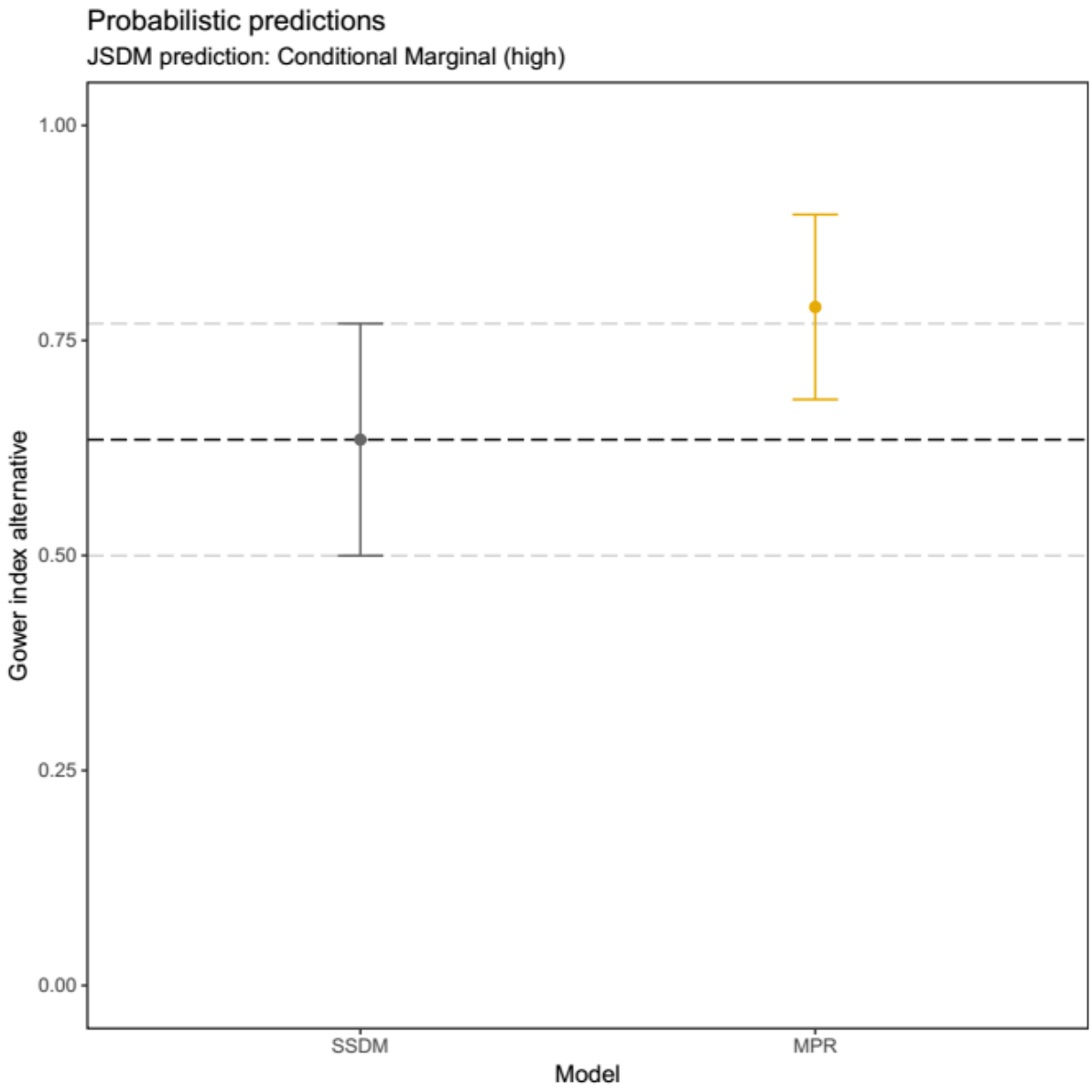


Figure 219: Performance of the SSDM and JSDM for the Gower index alternative evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

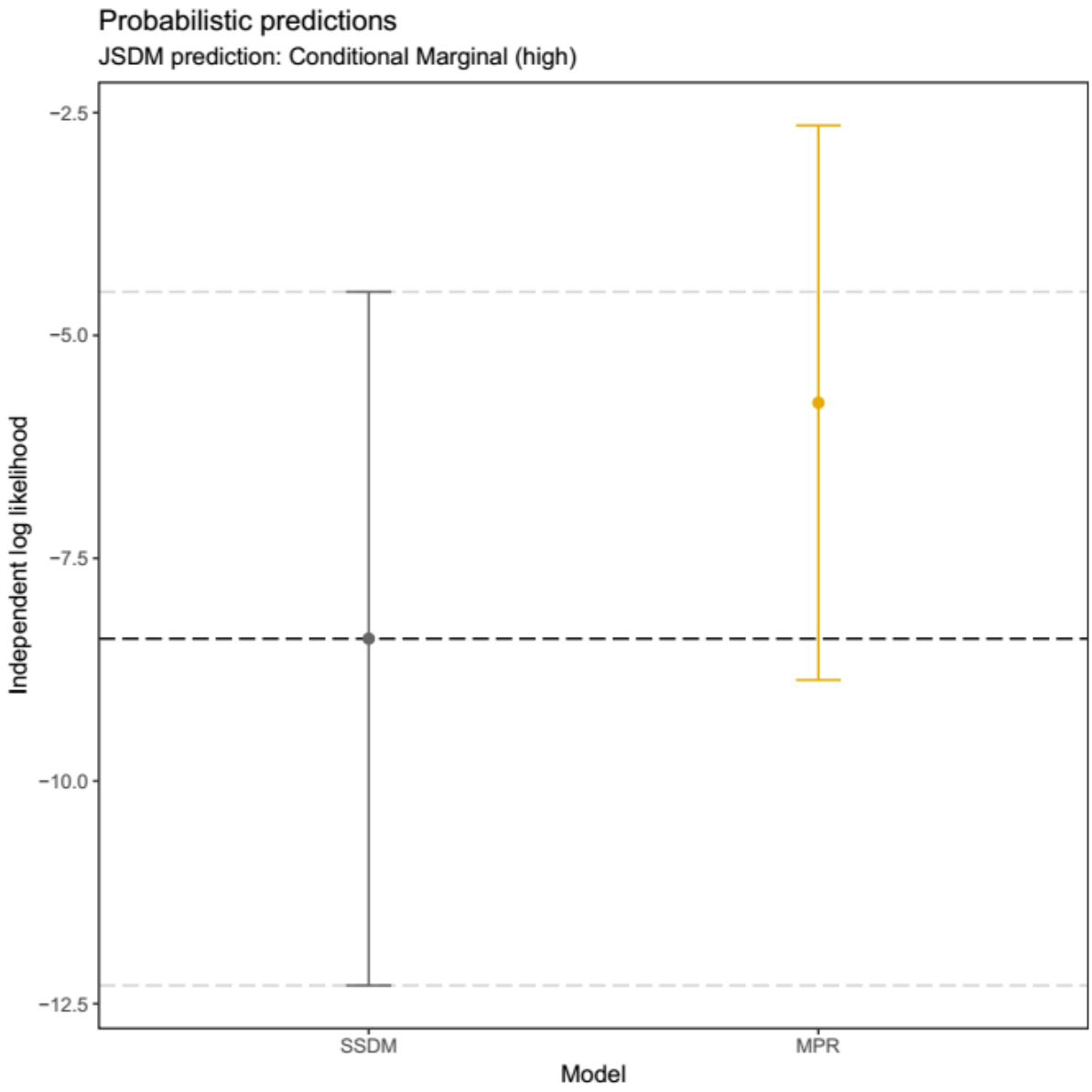


Figure 220: Performance of the SSDM and JSDM for the independent log likelihood evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

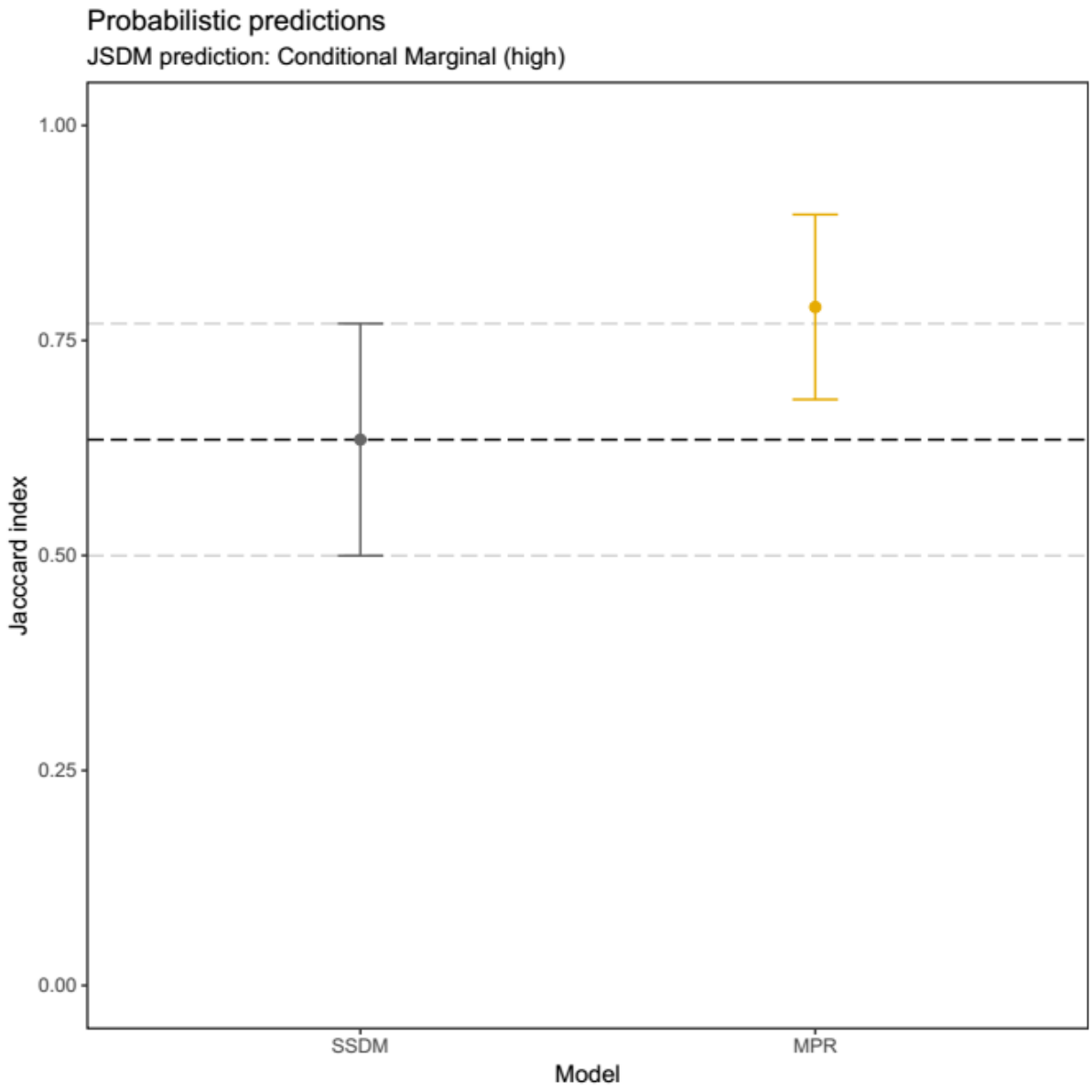


Figure 221: Performance of the SSDM and JSDM for the Jaccard index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

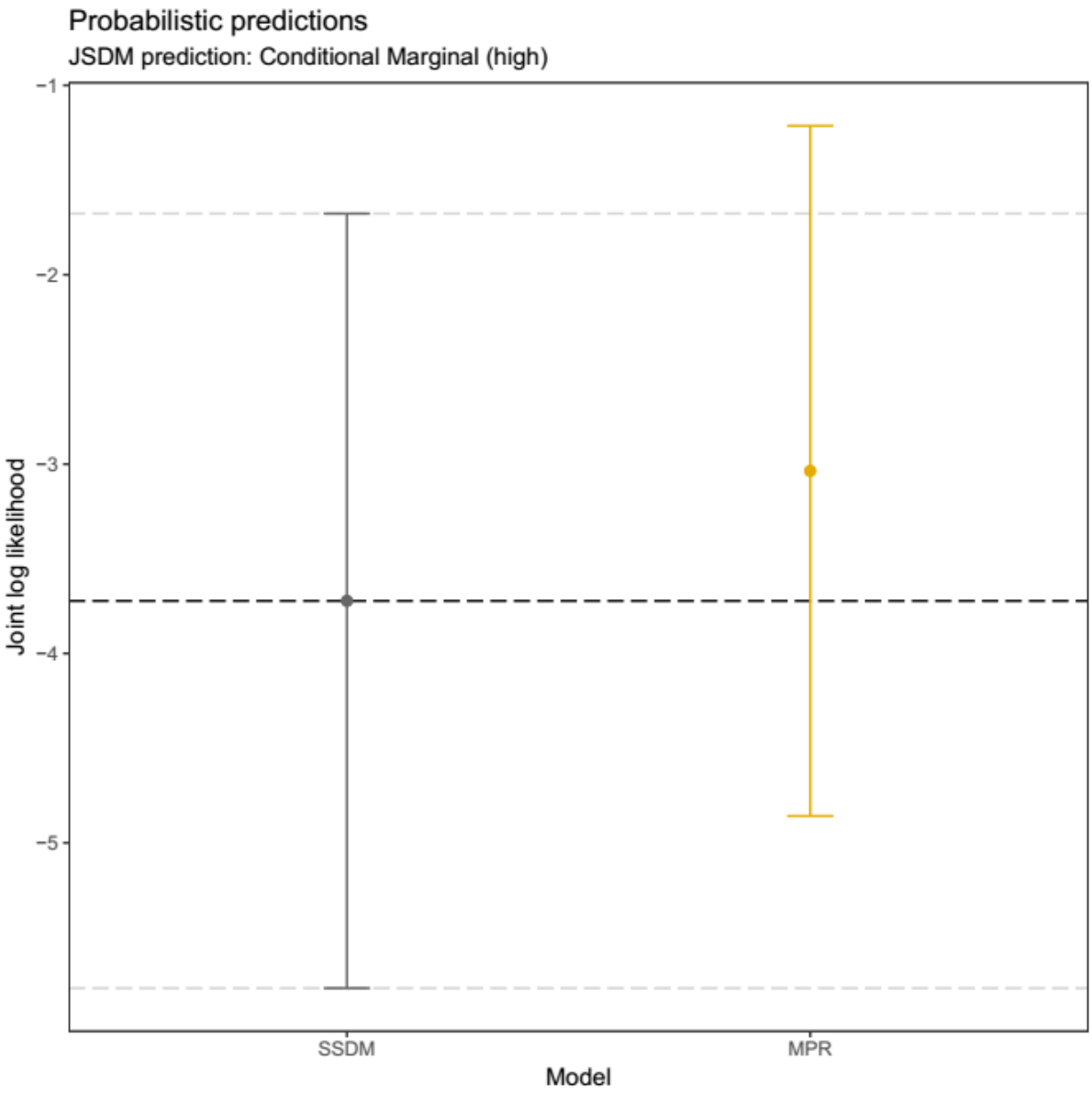


Figure 222: Performance of the SSDM and JSDM for the joint log likelihood evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

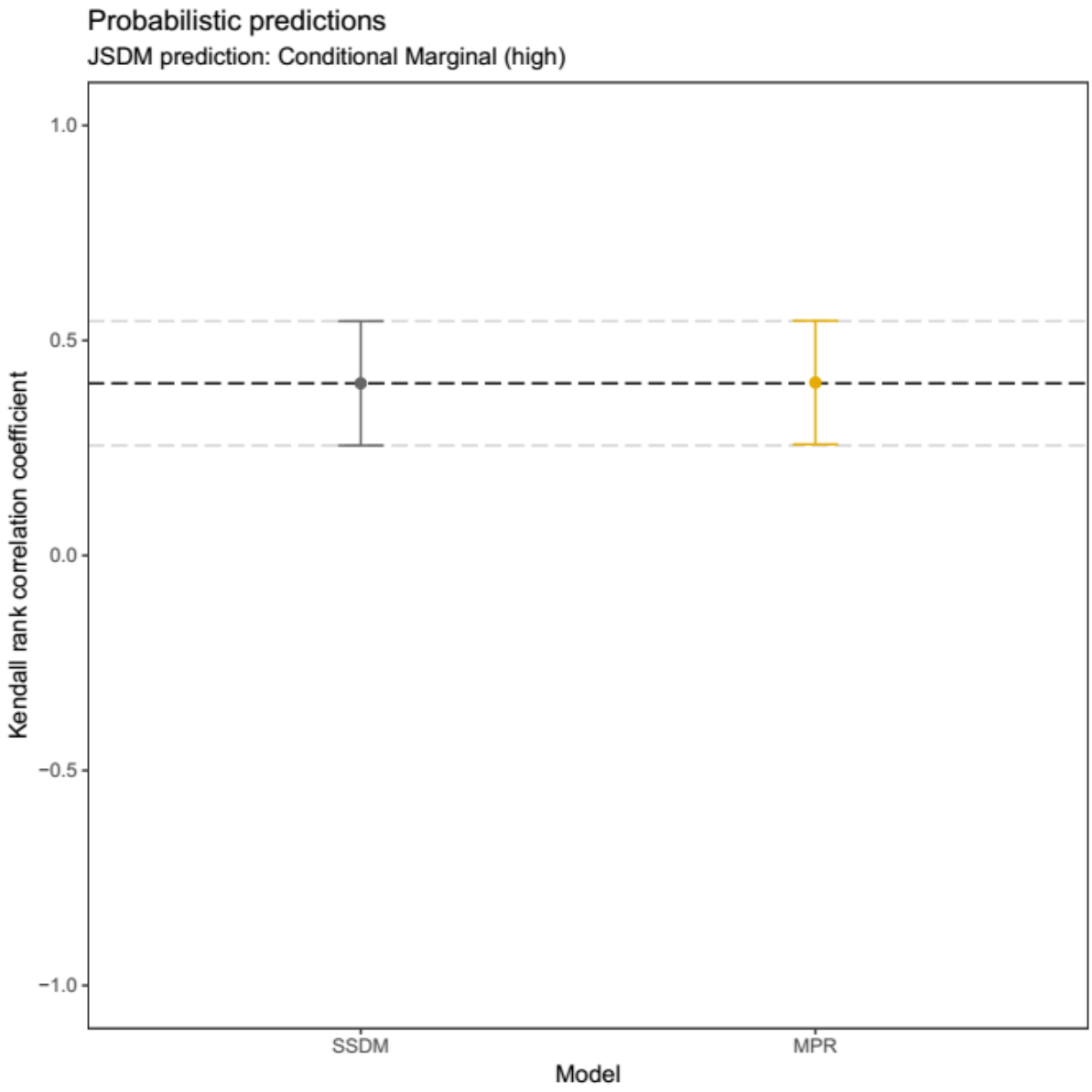


Figure 223: Performance of the SSDM and JSDM for the Kendall rank correlation coefficient evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

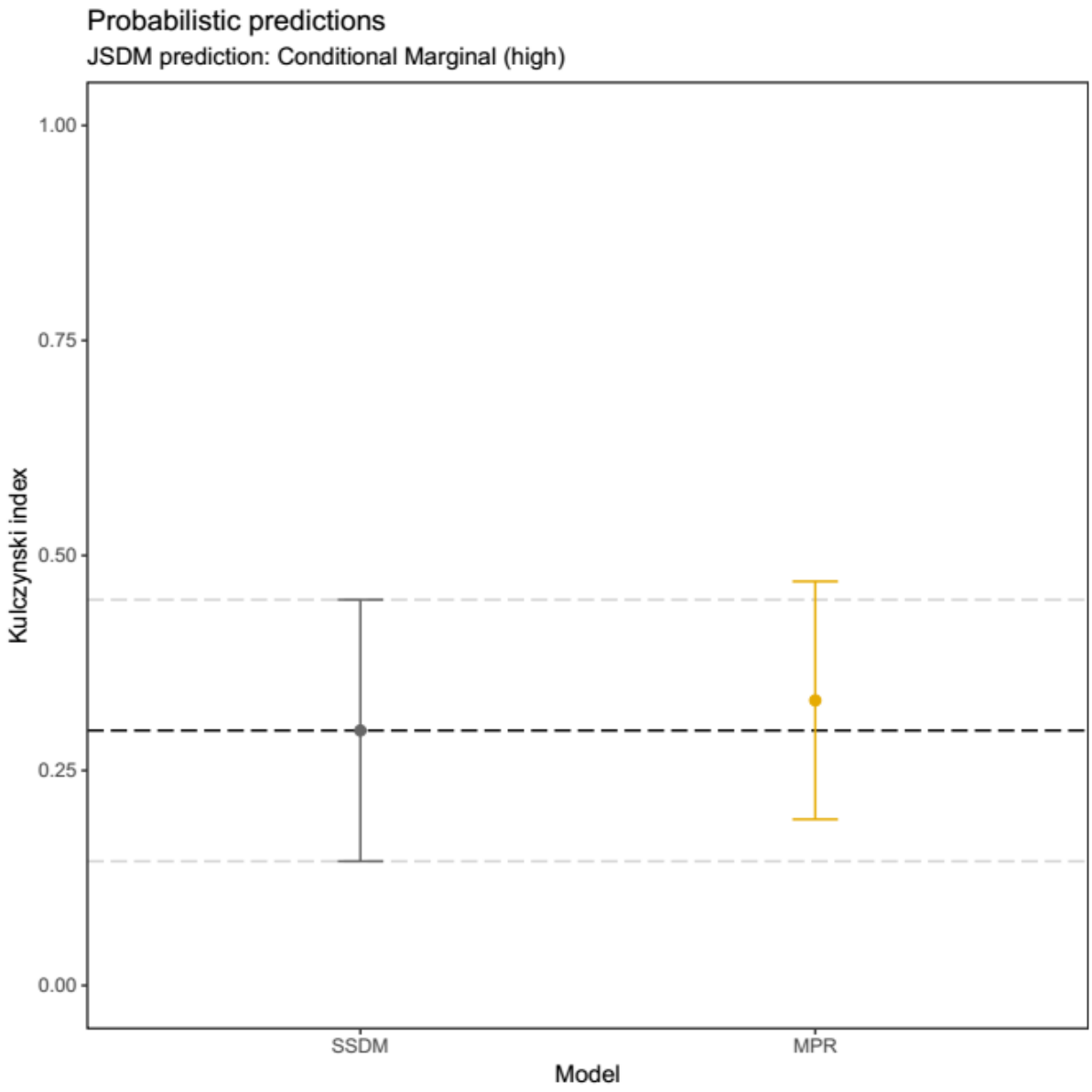


Figure 224: Performance of the SSDM and JSDM for the Kulczynski index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

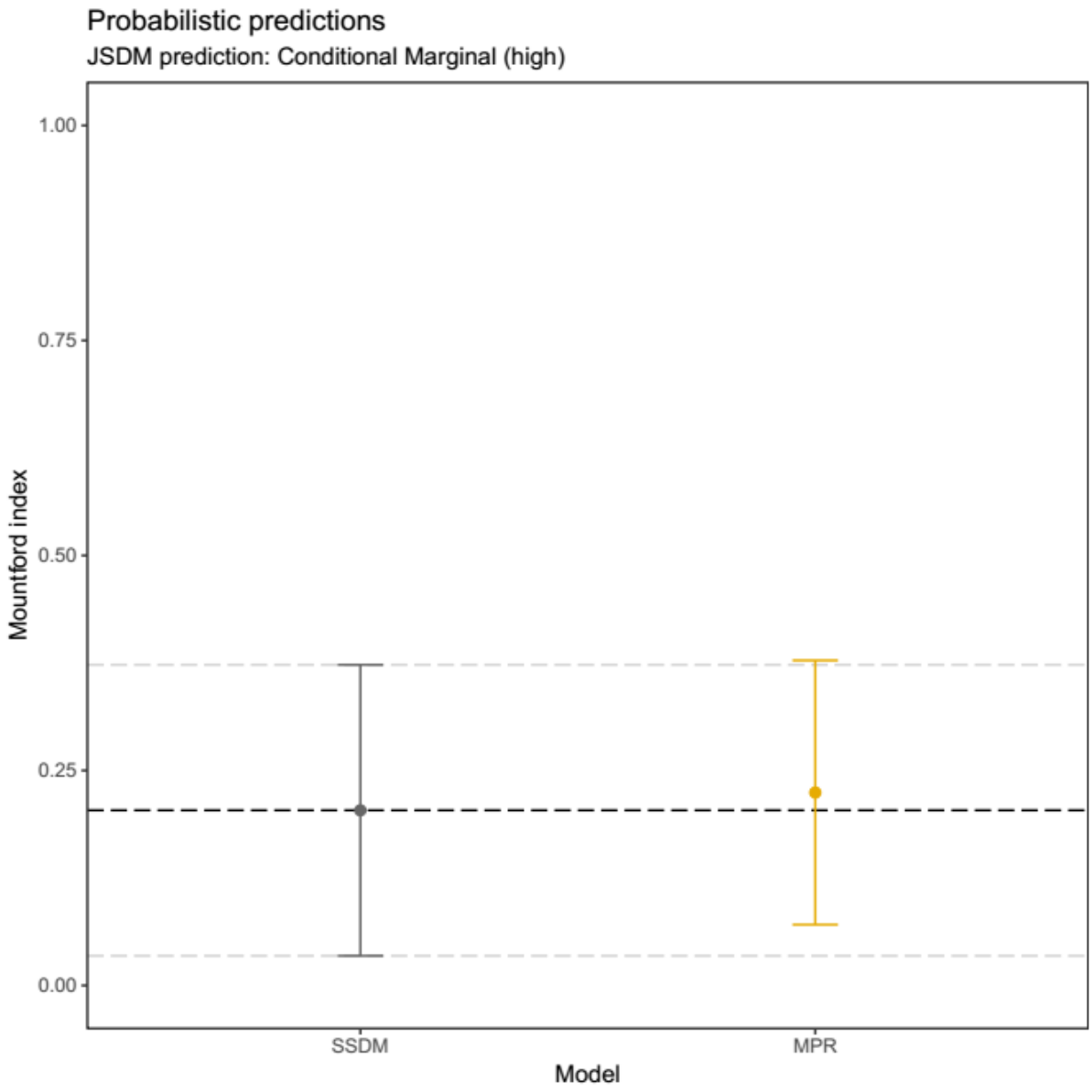


Figure 225: Performance of the SSDM and JSDM for the Mountford index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

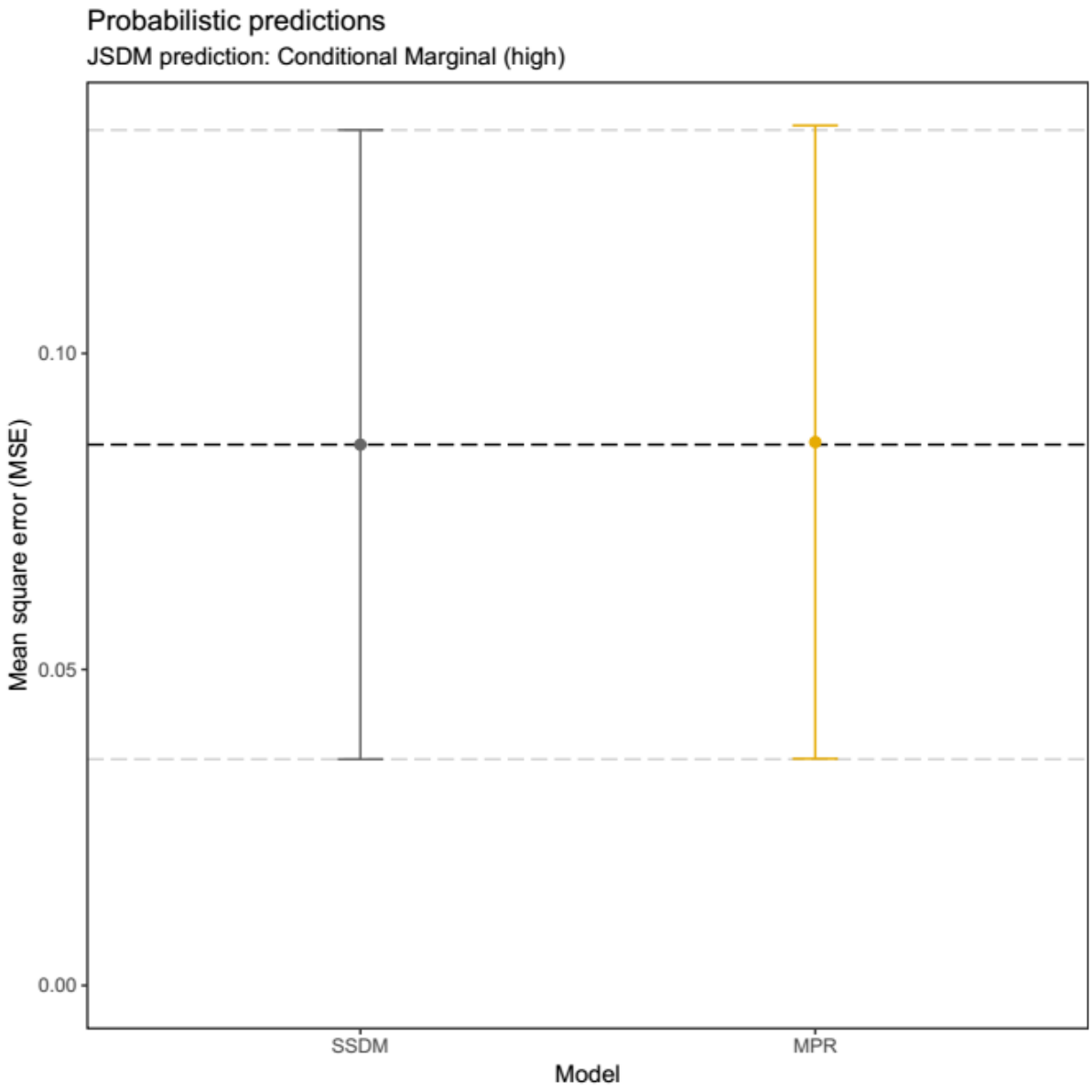


Figure 226: Performance of the SSDM and JSDM for the mean square error evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

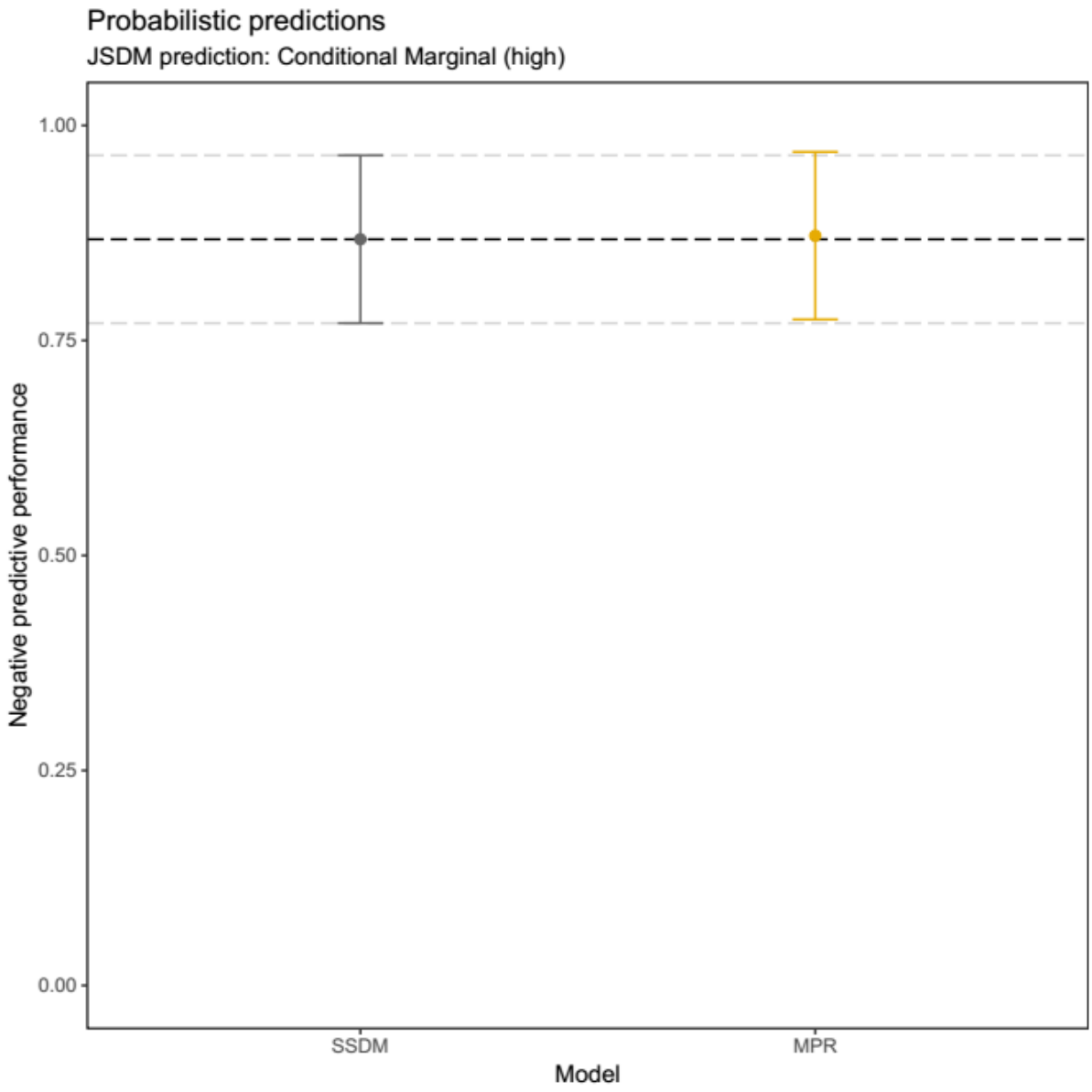


Figure 227: Performance of the SSDM and JSDM for the negative predictive performance evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

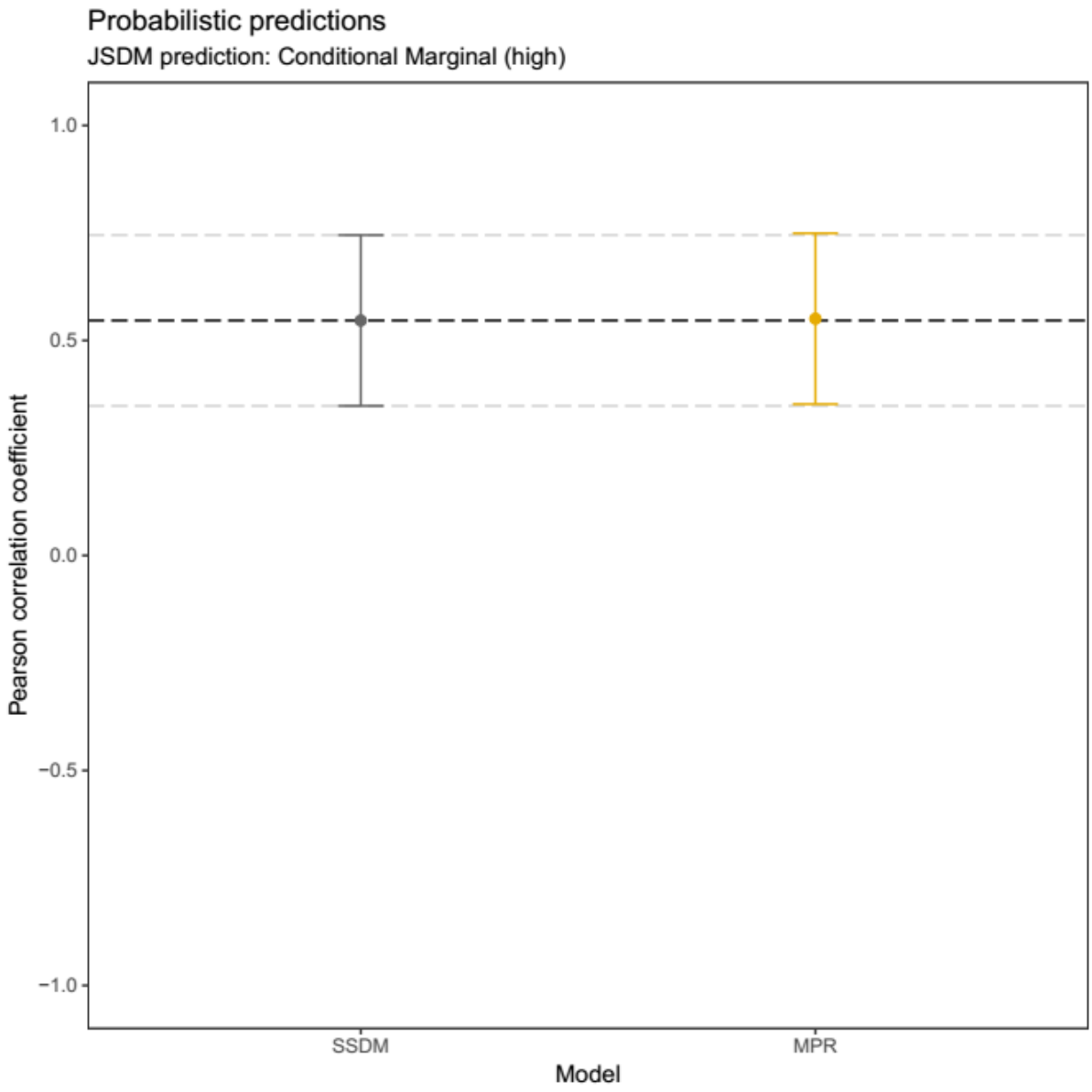


Figure 228: Performance of the SSDM and JSDM for the Pearson correlation coefficient evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

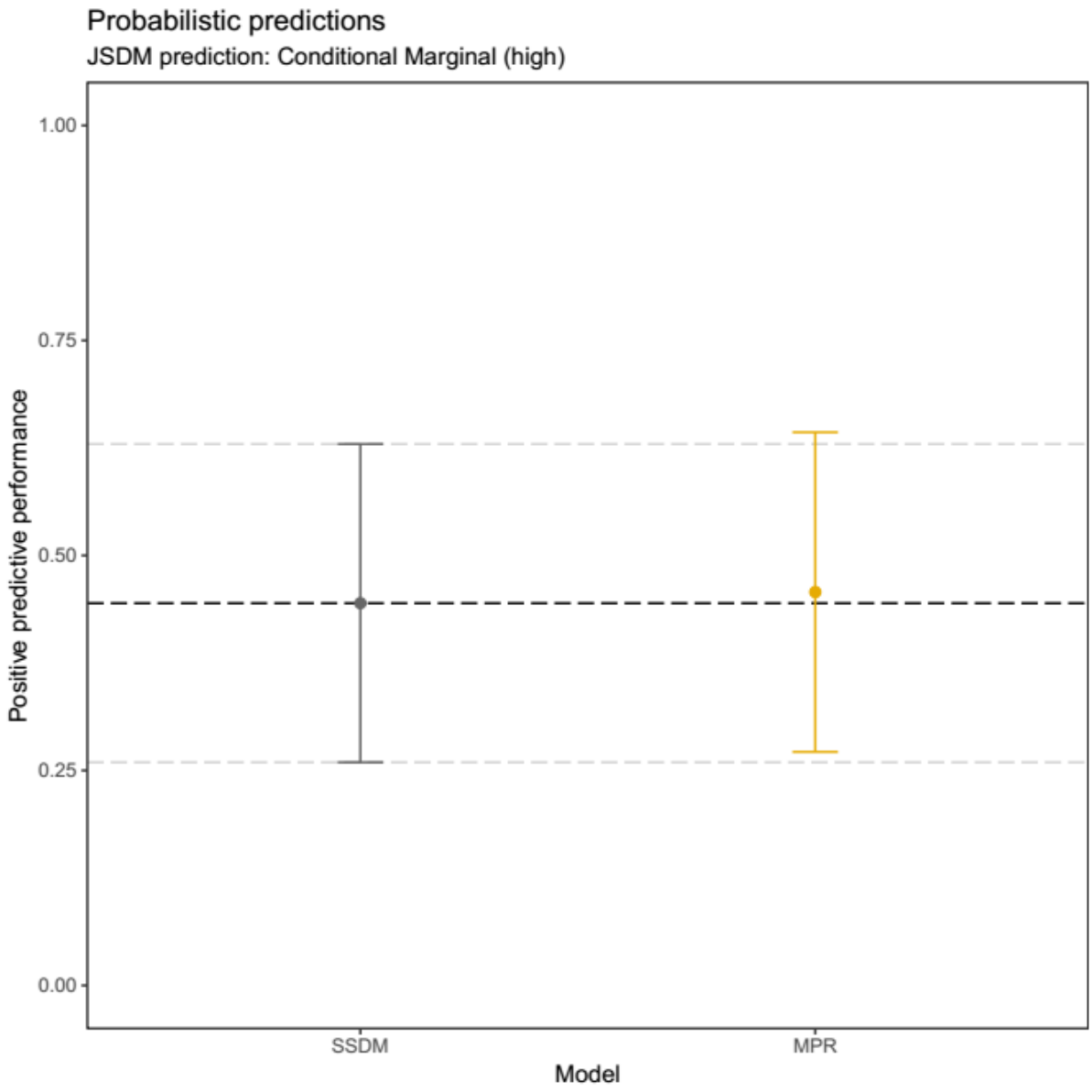


Figure 229: Performance of the SSDM and JSDM for the positive predictive performance evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

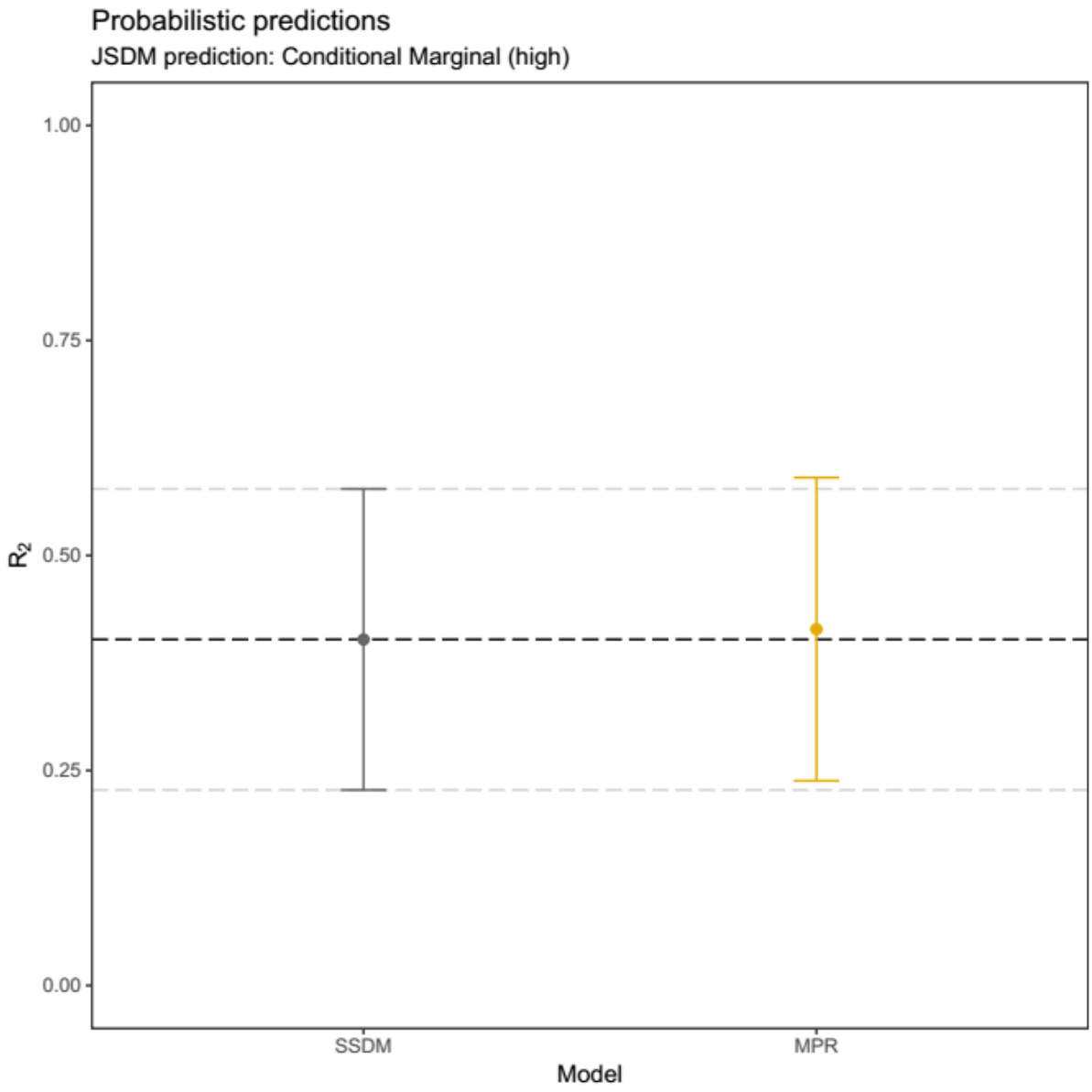


Figure 230: Performance of the SSDM and JSDM for the R^2 evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

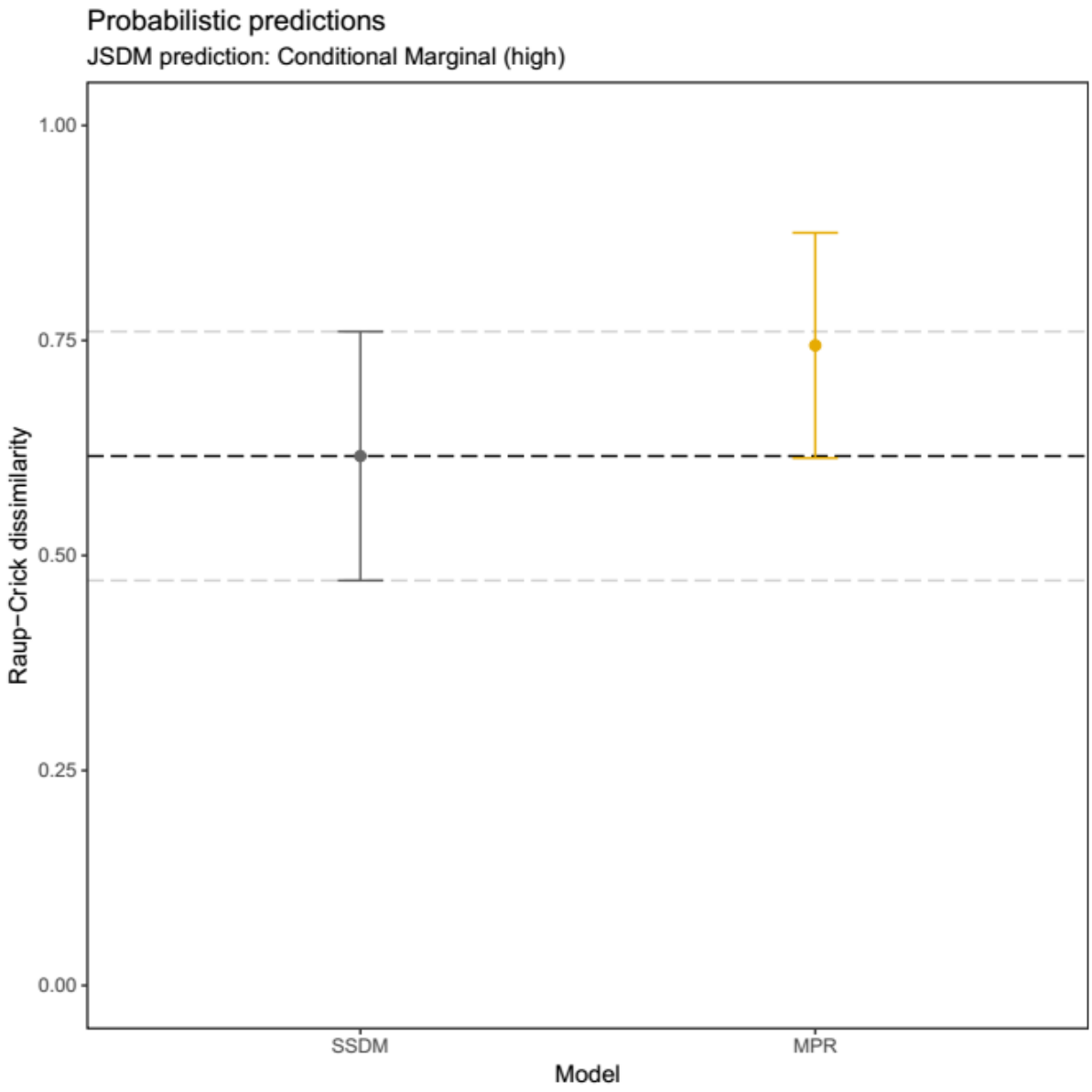


Figure 231: Performance of the SSDM and JSDM for the Raup-Crick dissimilarity evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

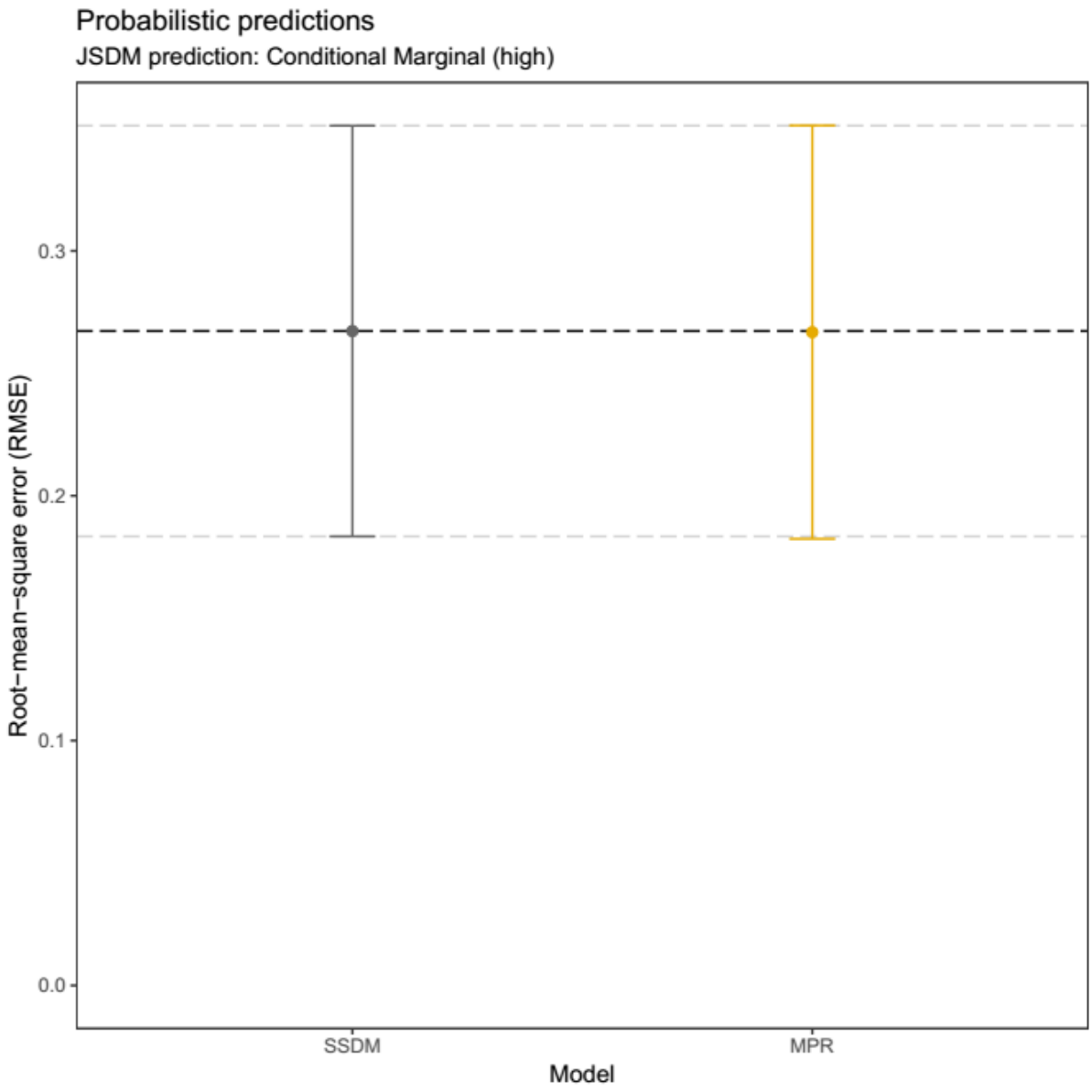


Figure 232: Performance of the SSDM and JSDM for the root-mean-square error evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

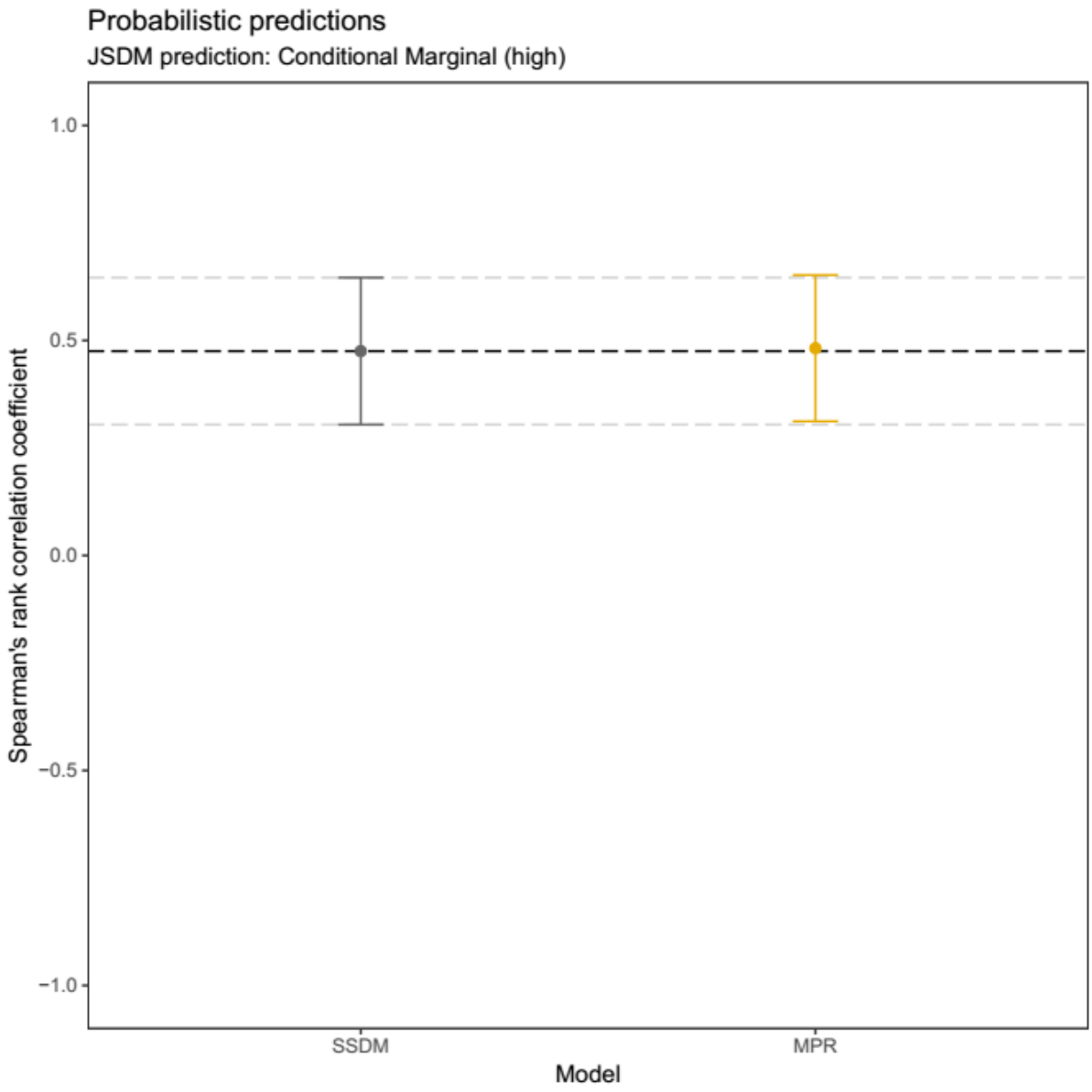


Figure 233: Performance of the SSDM and JSDM for the Spearman's rank correlation coefficient evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

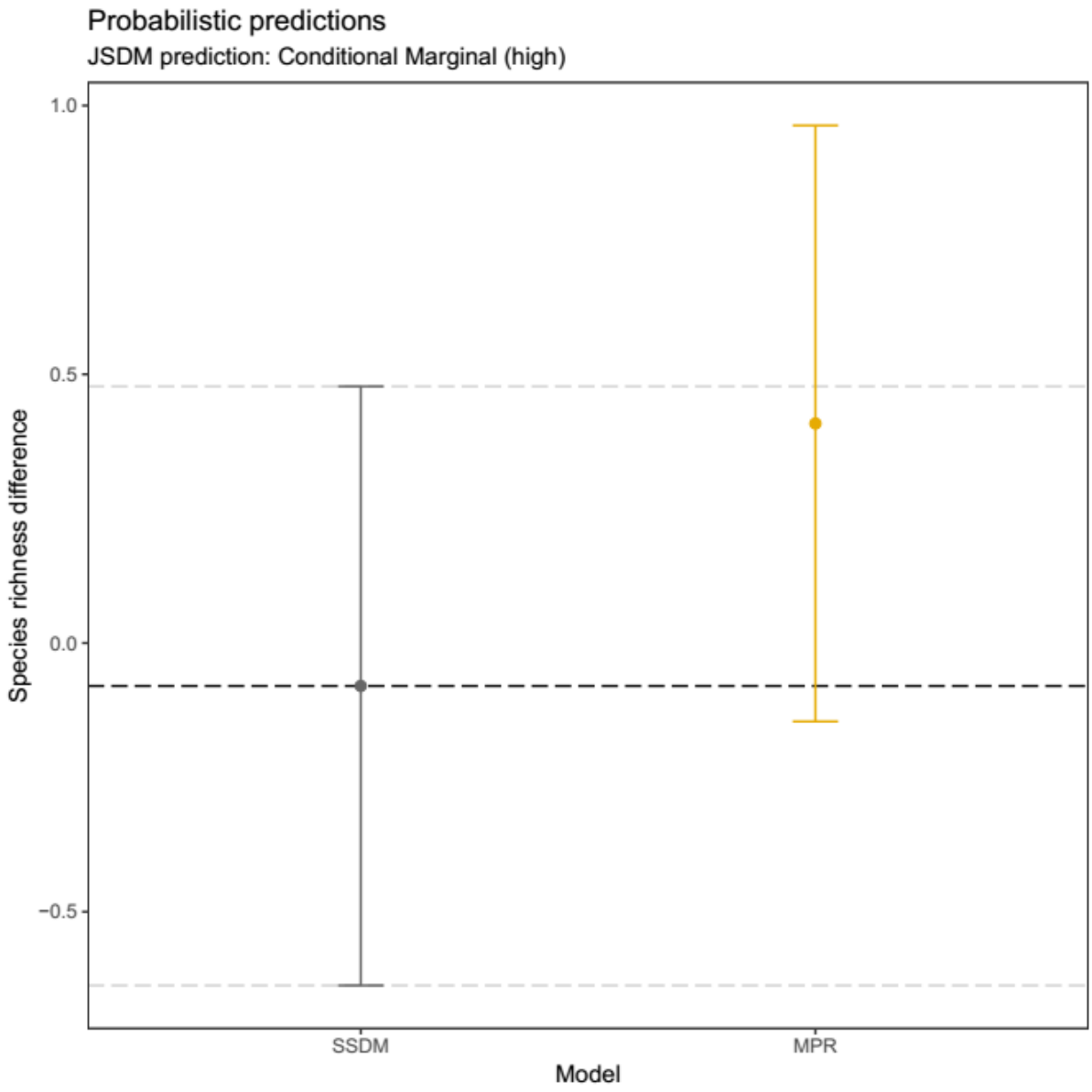


Figure 234: Performance of the SSDM and JSDM for the species richness difference evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

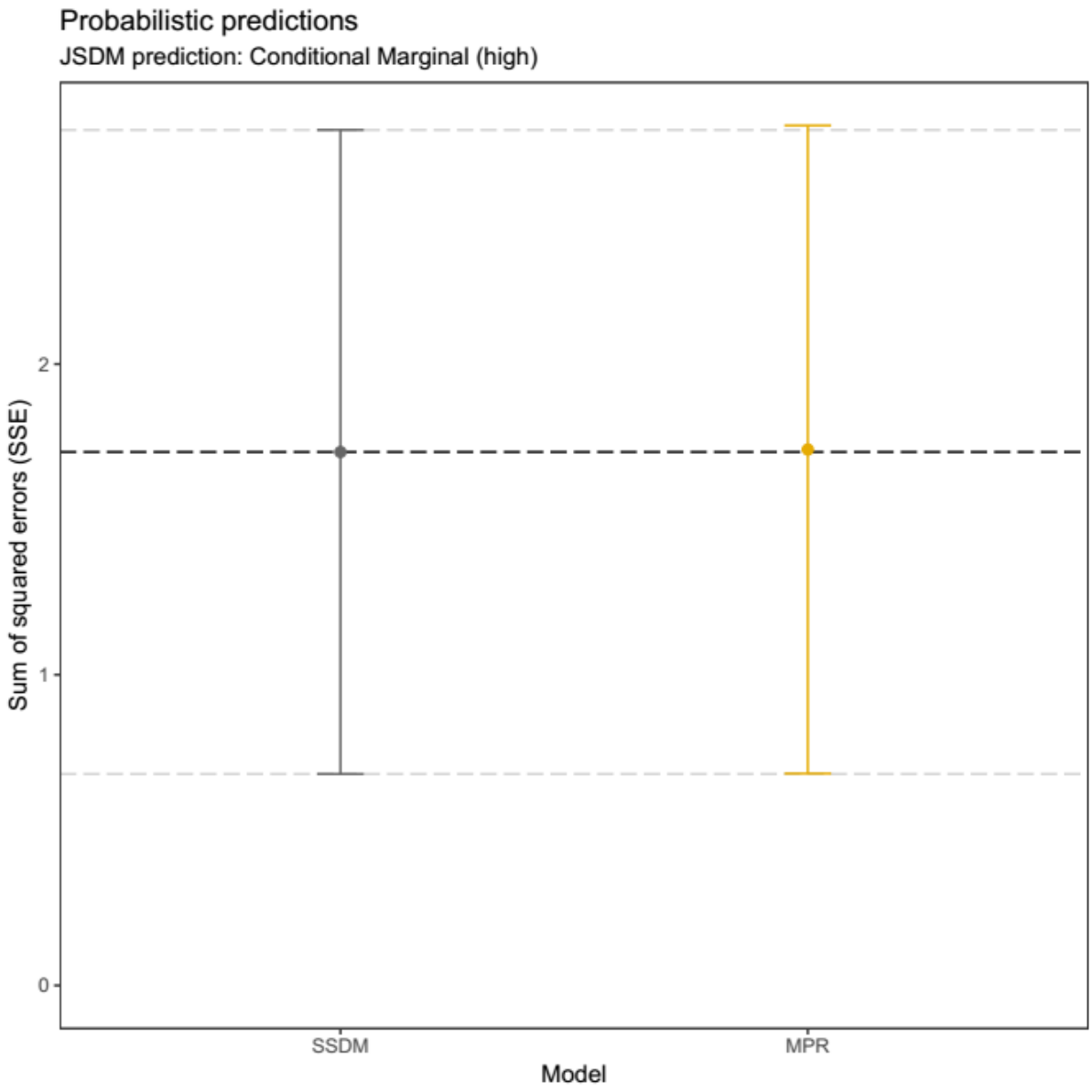


Figure 235: Performance of the SSDM and JSDM for the sum of squared errors evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

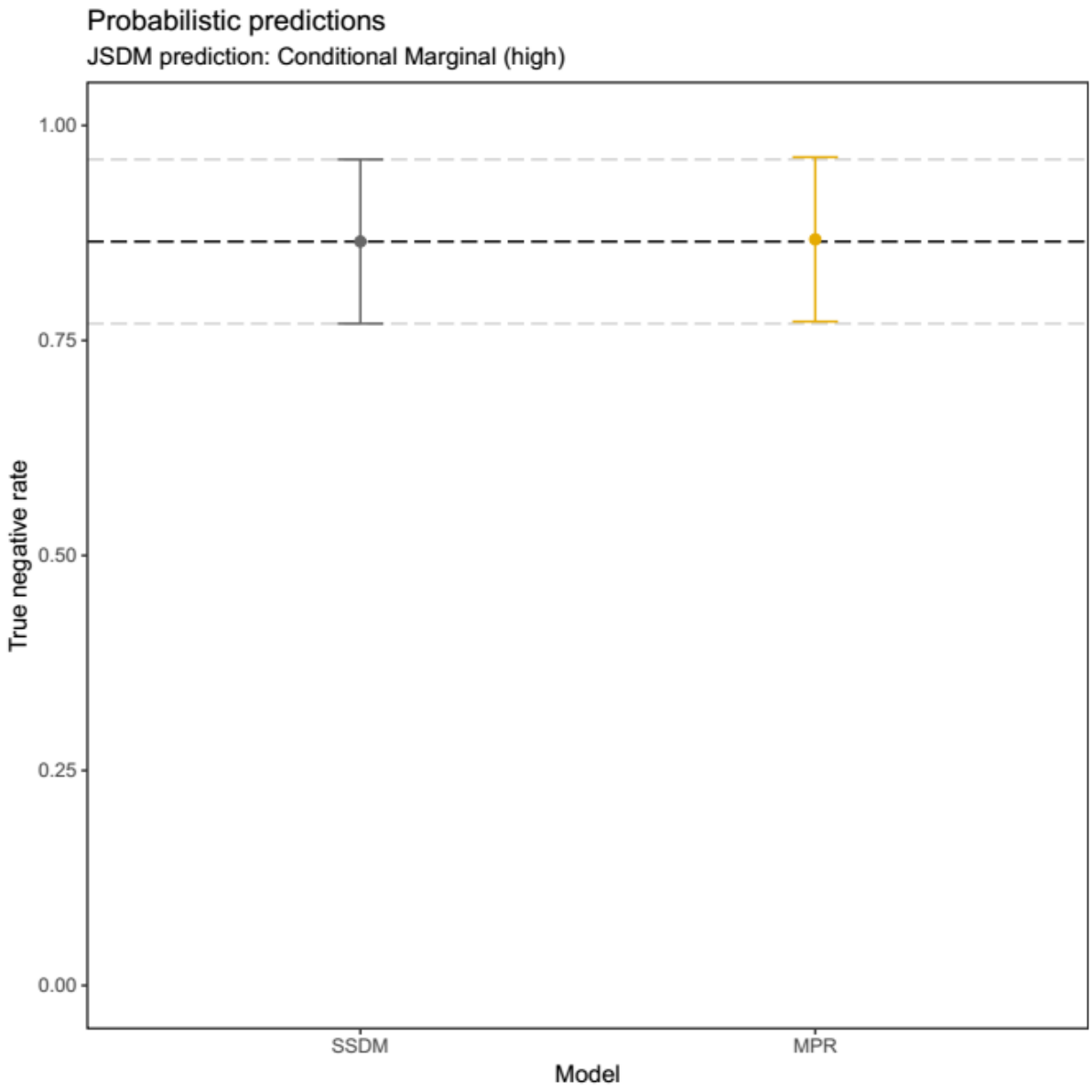


Figure 236: Performance of the SSDM and JSDM for the true negative rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

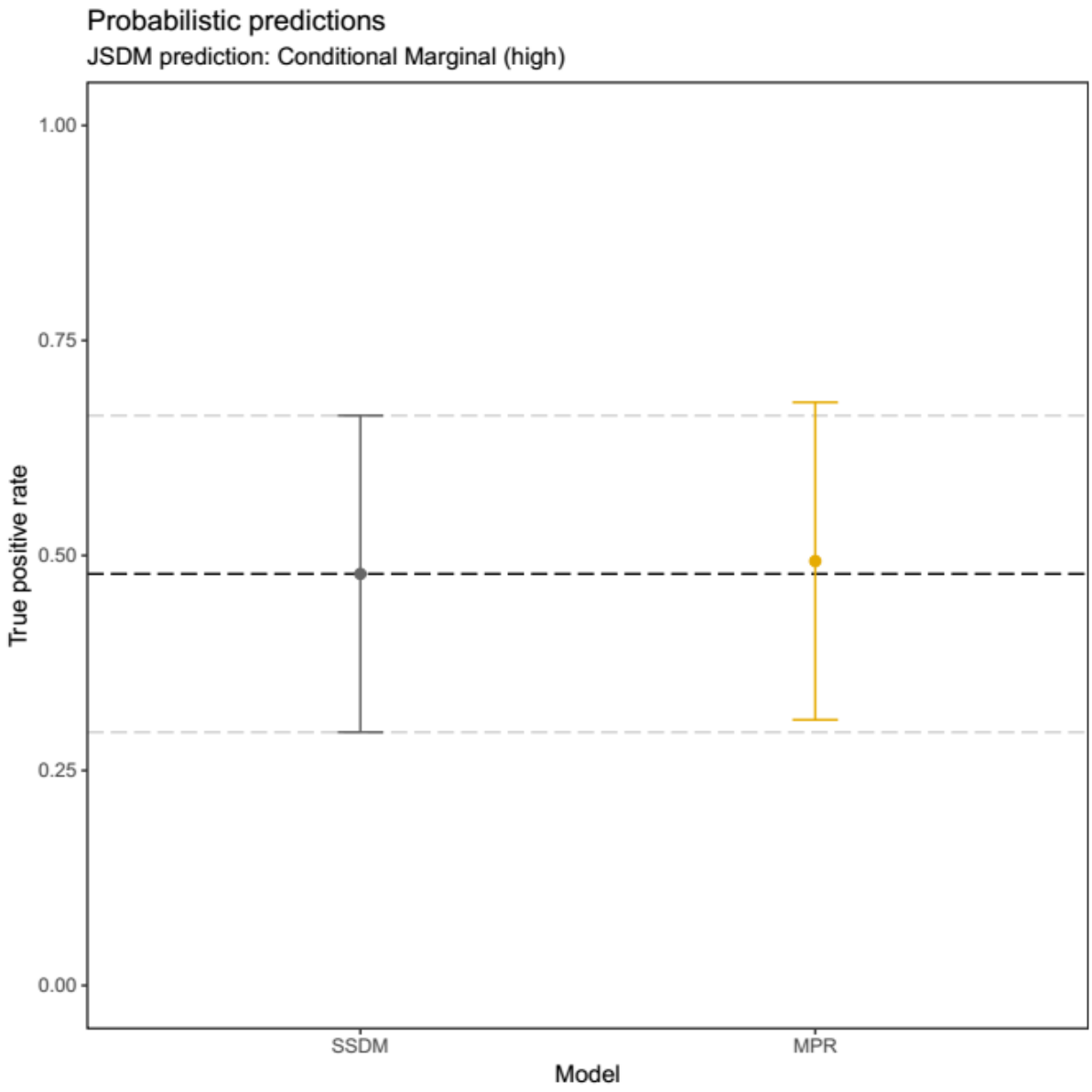


Figure 237: Performance of the SSDM and JSDM for the true positive rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

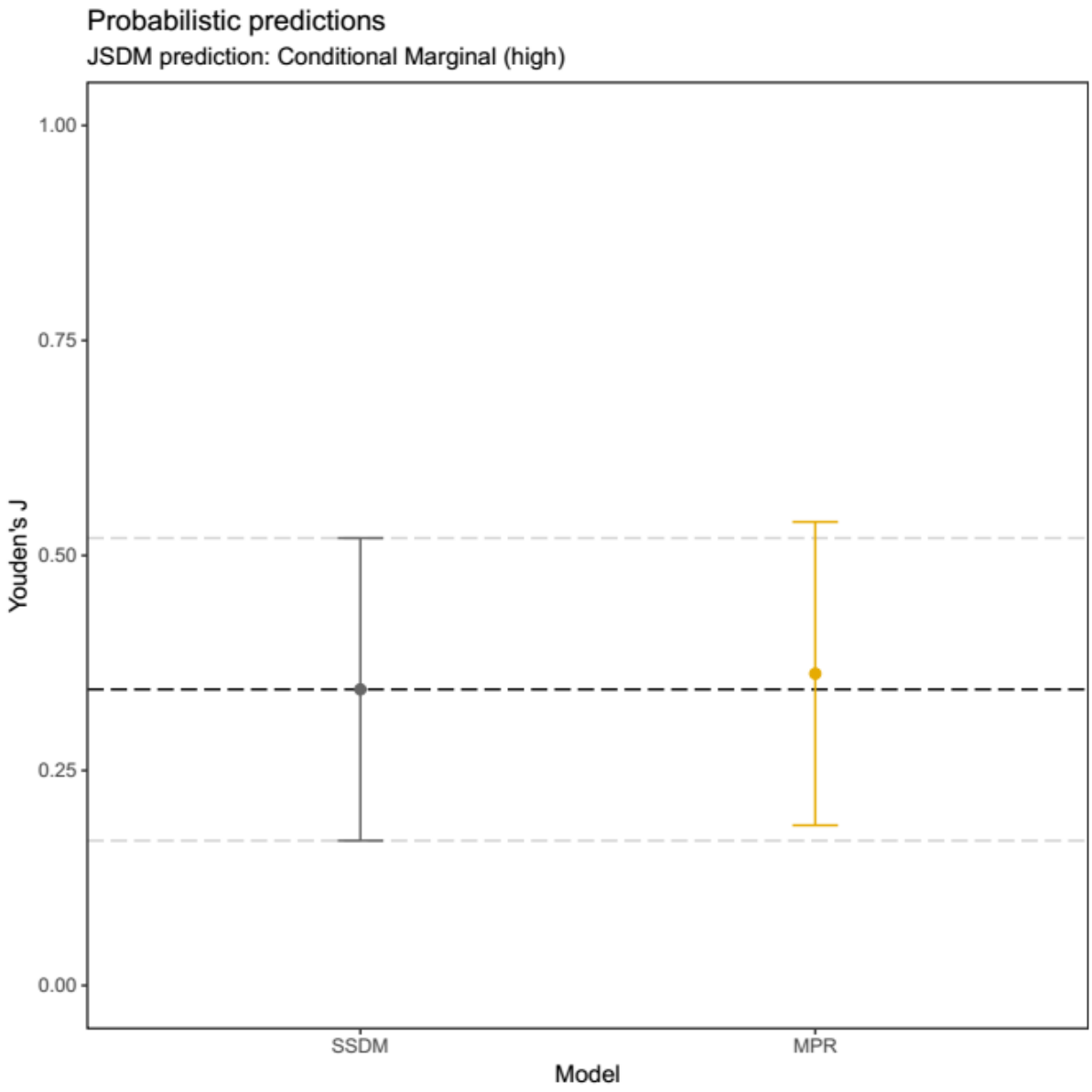


Figure 238: Performance of the SSDM and JSDM for the Youden's J evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval.

APPENDIX I

FOREST PLOTS OF THE RELATIVE PERFORMANCE DIFFERENCE BETWEEN THE JSDM AND SSDM

This appendix contains forest plots for each combination of prediction type and metric evaluation range. Rather than clustering metrics by the five groupings presented in the main article, here we have grouped them by the evaluation range and direction of the metrics. Threshold-independent metrics, for example, can potentially be evaluated between 0-1 and have either 1 or 0 be the optimal value. For convenience of navigation the figures are presented in subsections corresponding to prediction type.

1. **Marginal (probabilistic) predictions**

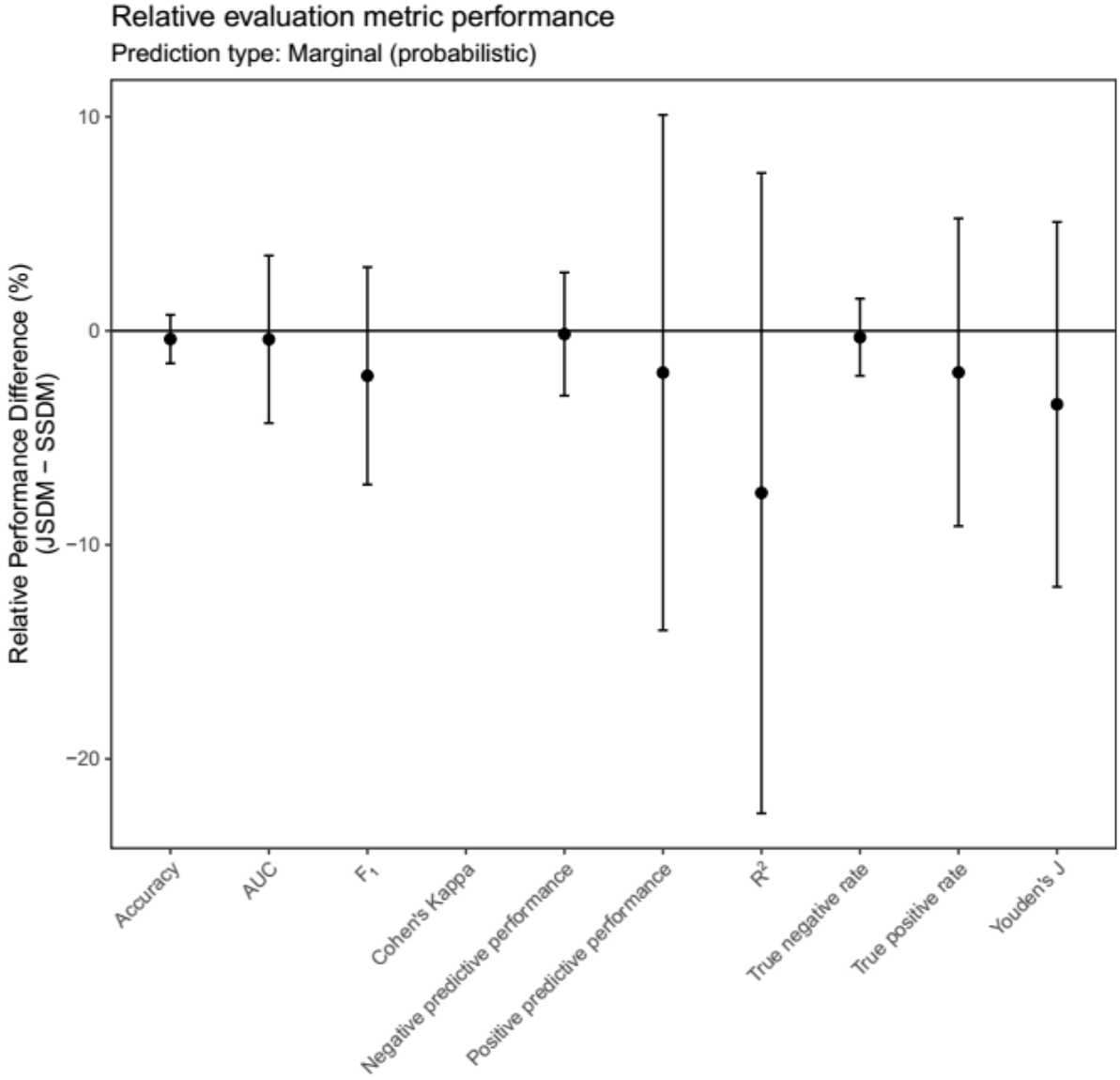


Figure 1: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

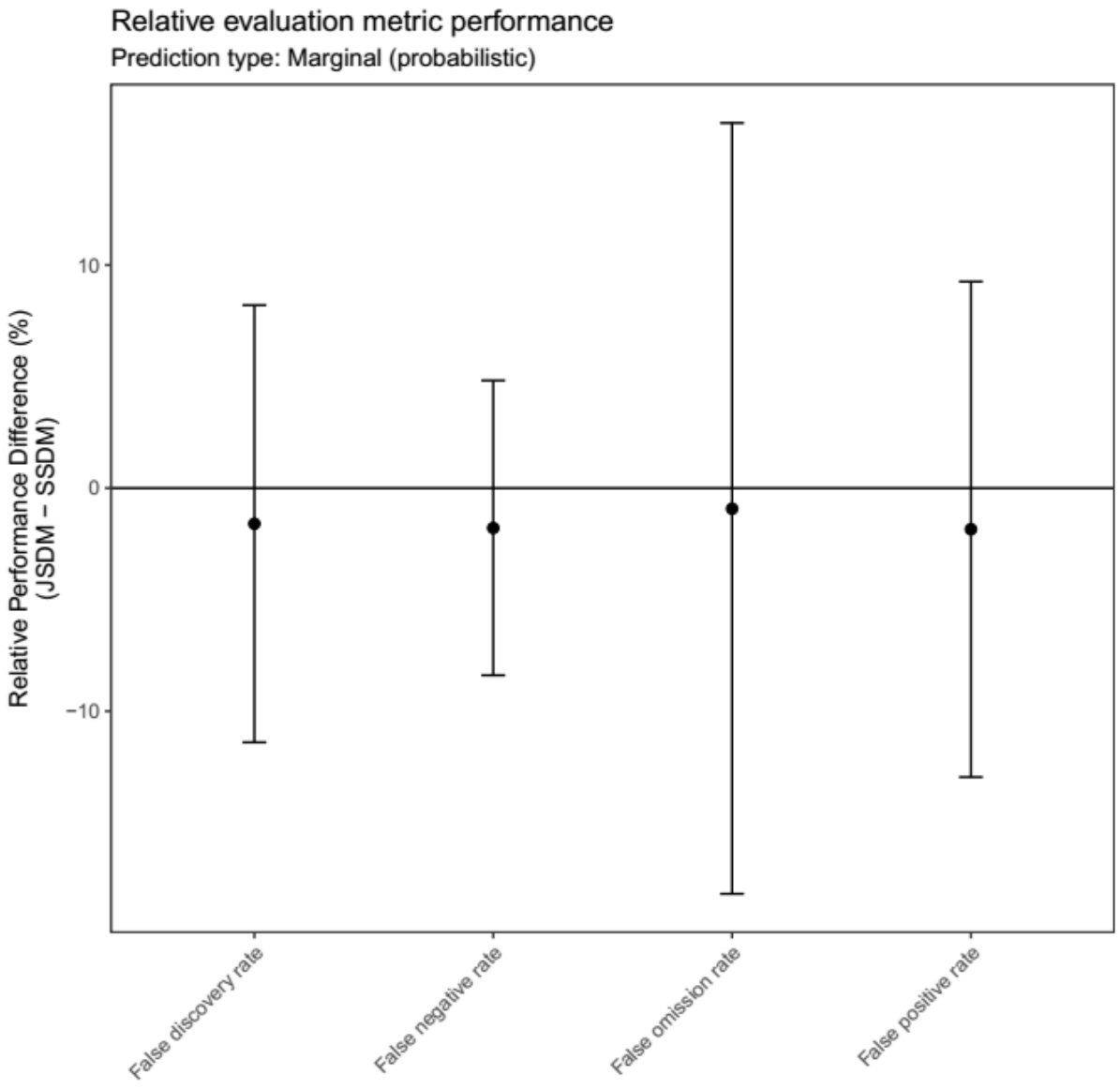


Figure 2: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

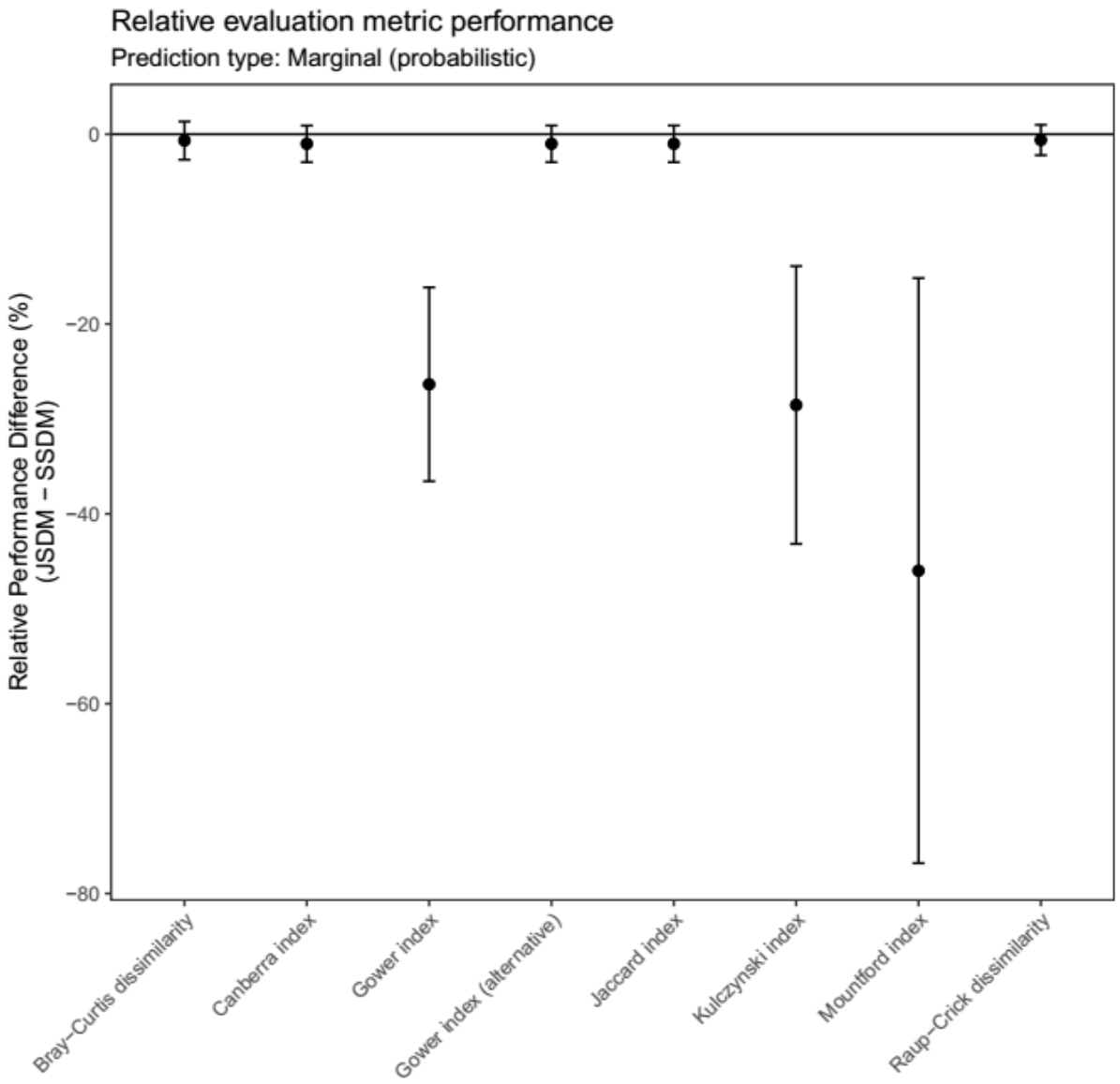


Figure 3: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

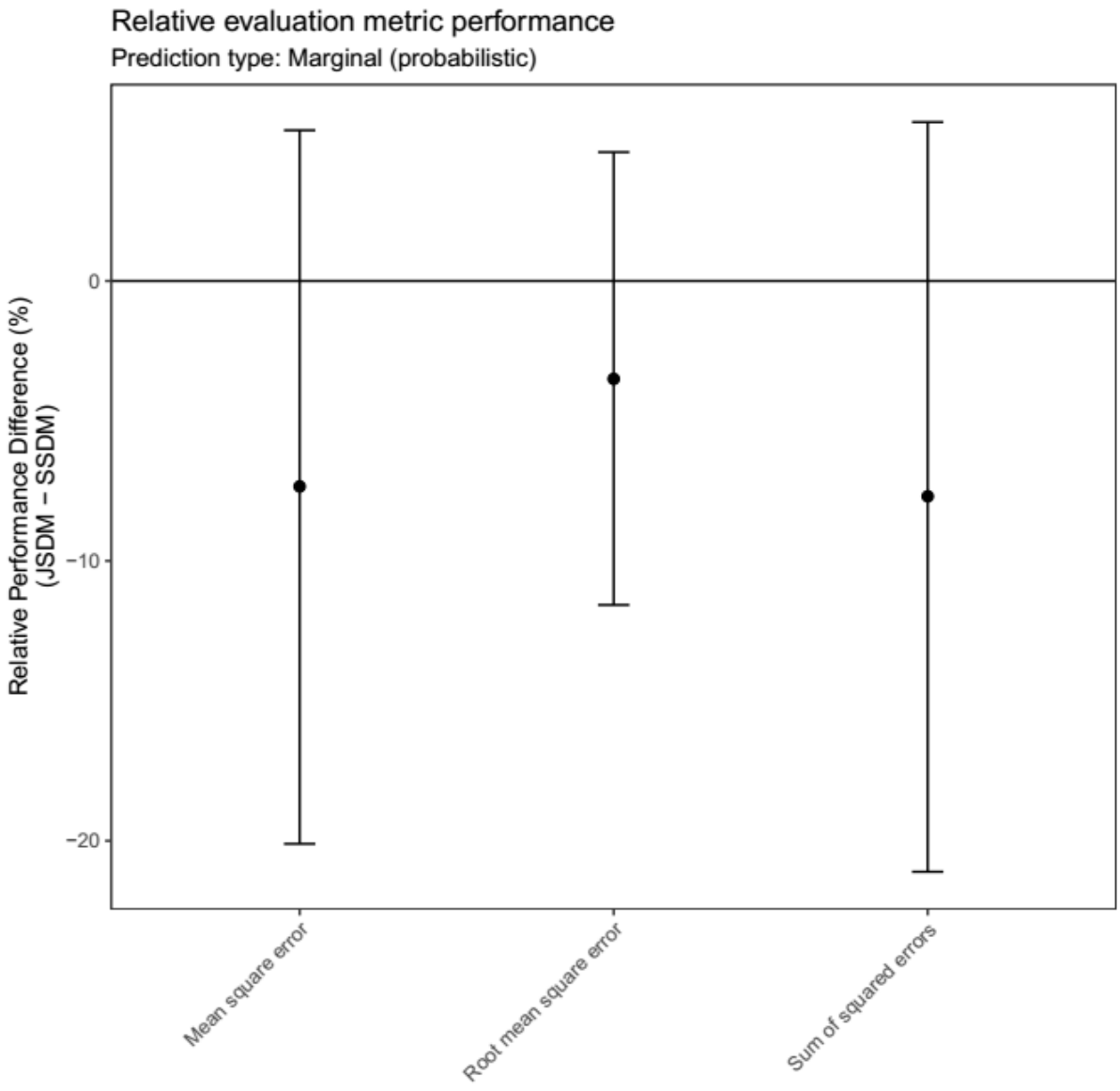


Figure 4: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – ∞, where lower values indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

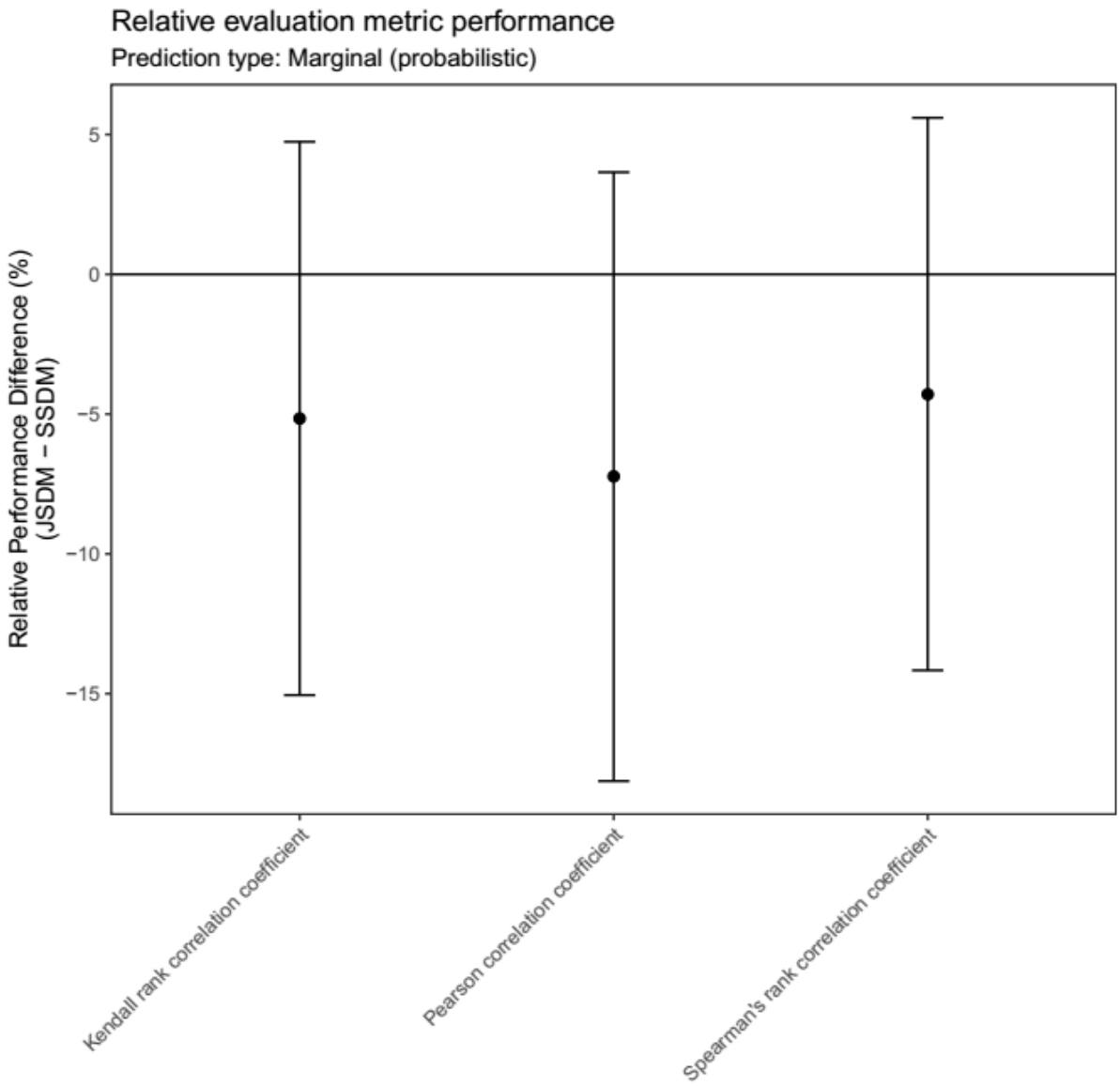


Figure 5: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -1 – 1, where higher values indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

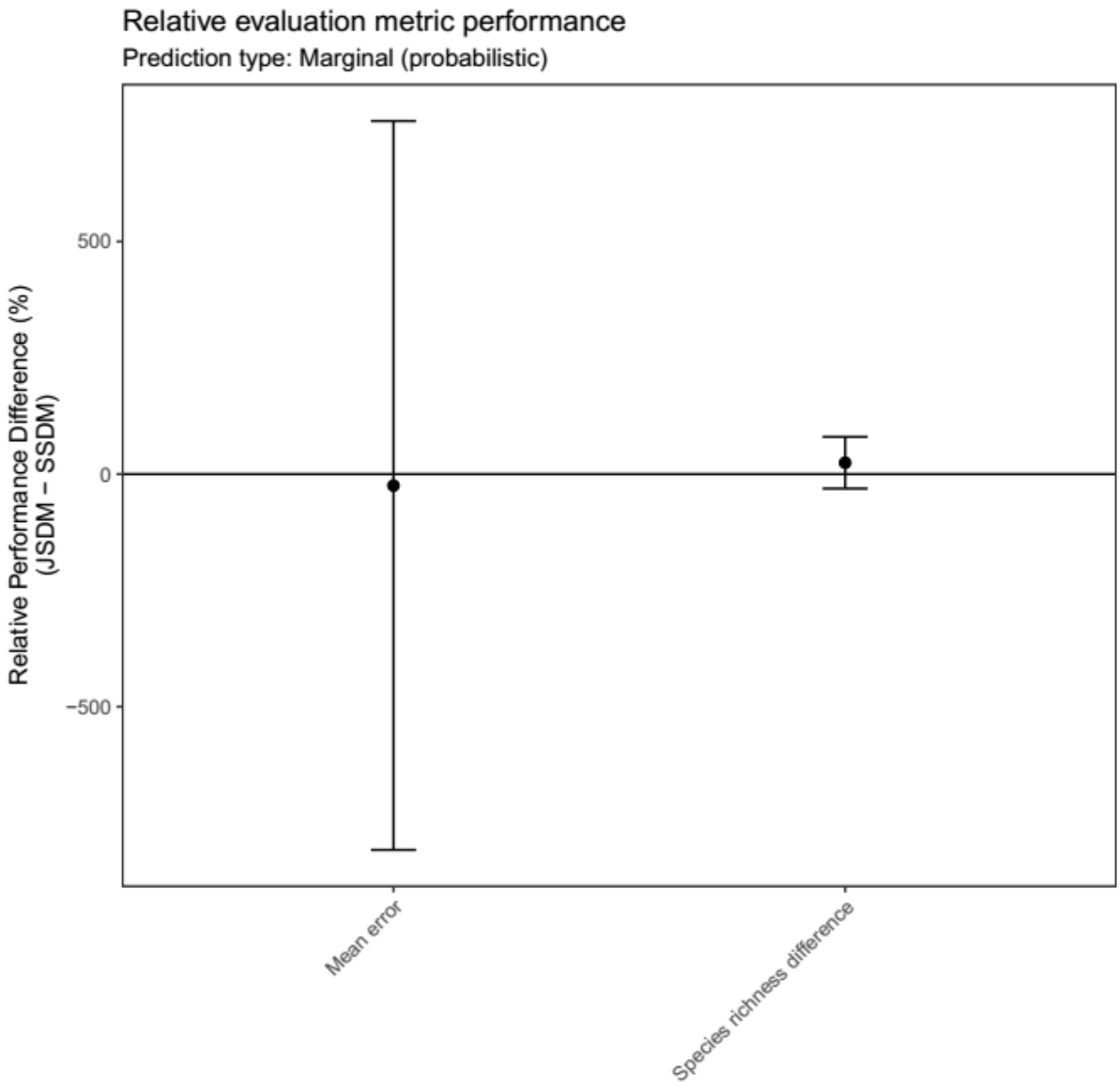


Figure 6: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

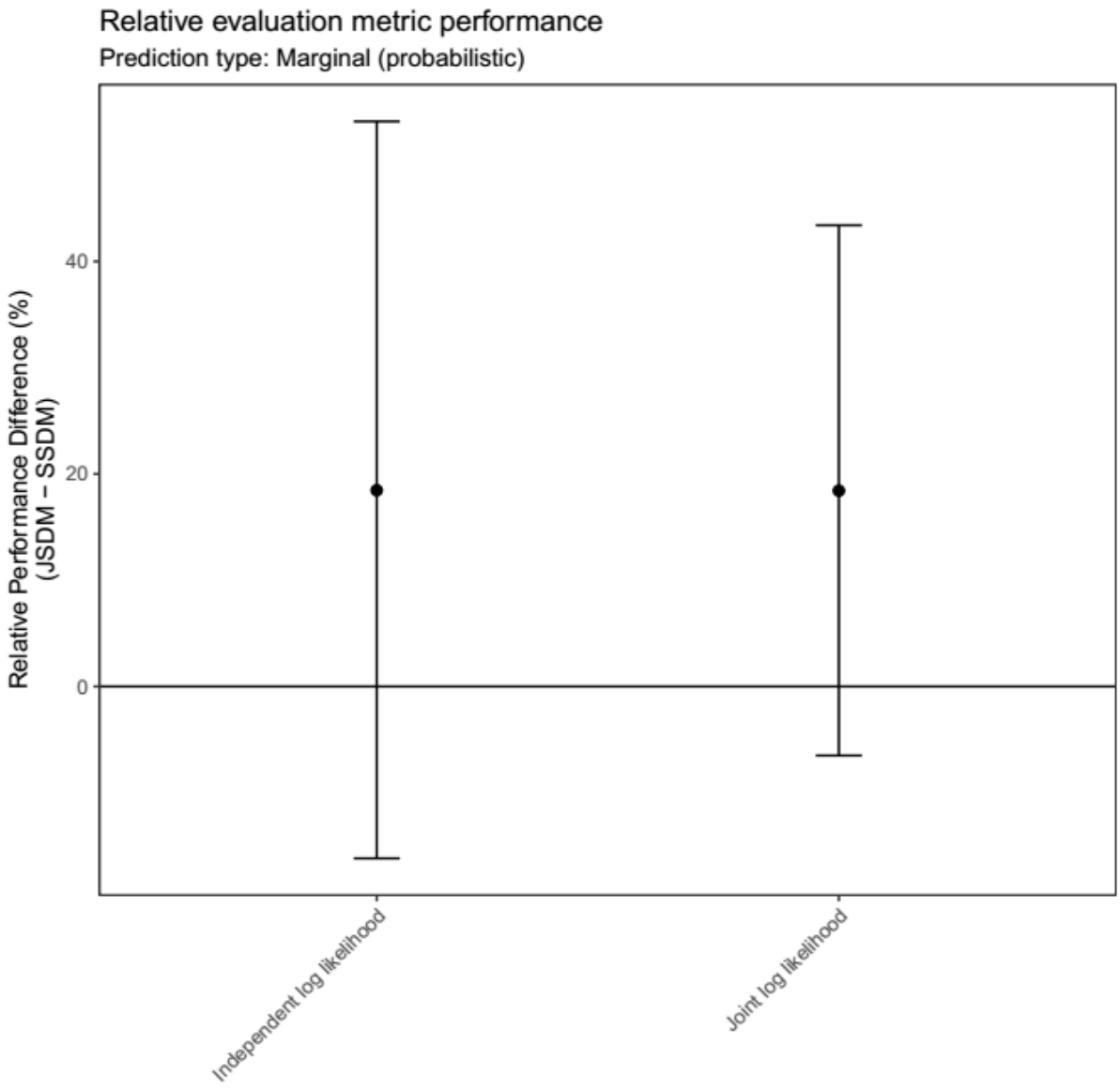


Figure 1: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where higher values indicate better performance, for marginal (probabilistic) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

2. Marginal (binary) predictions

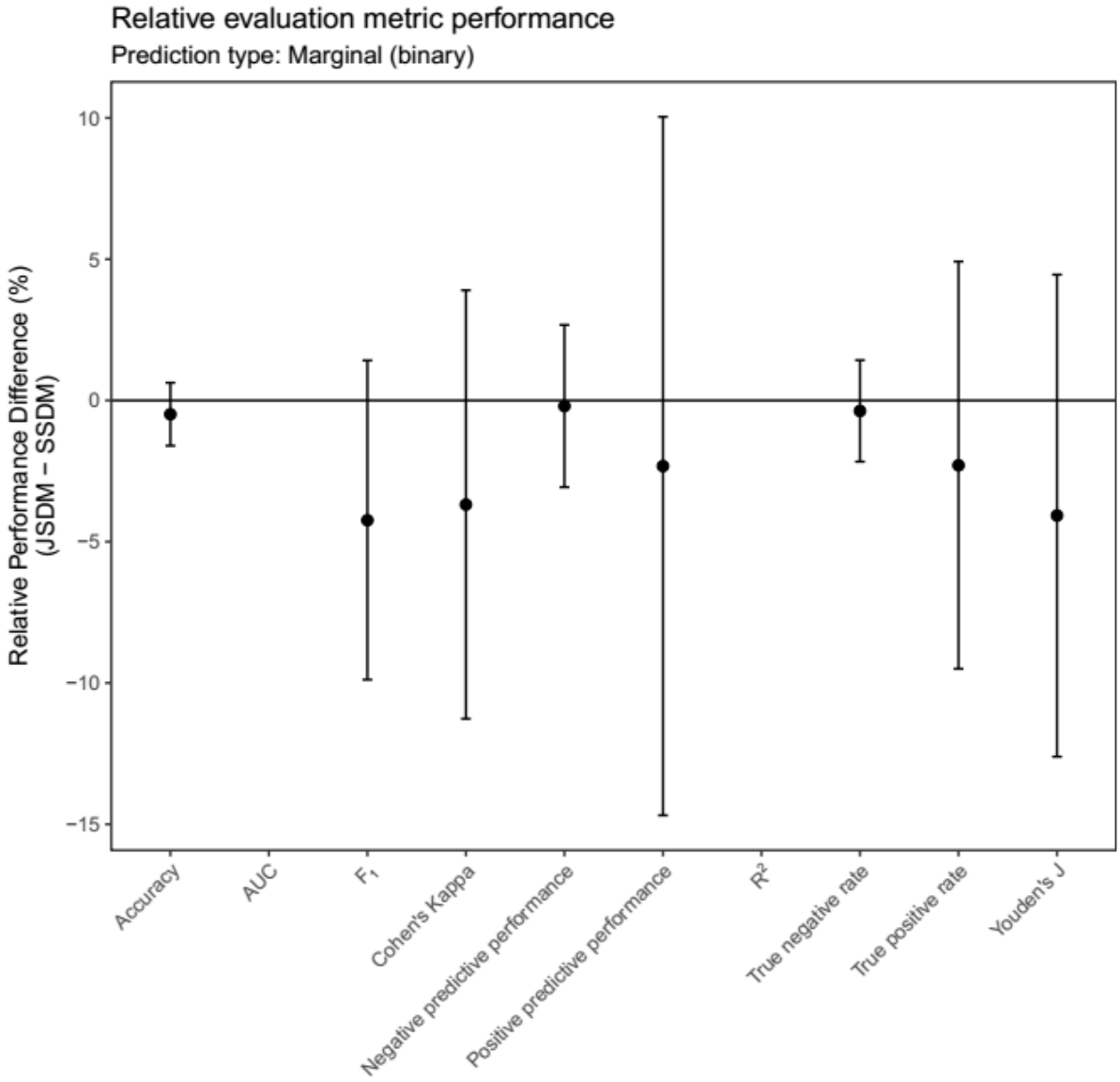


Figure 8: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for marginal (binary) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

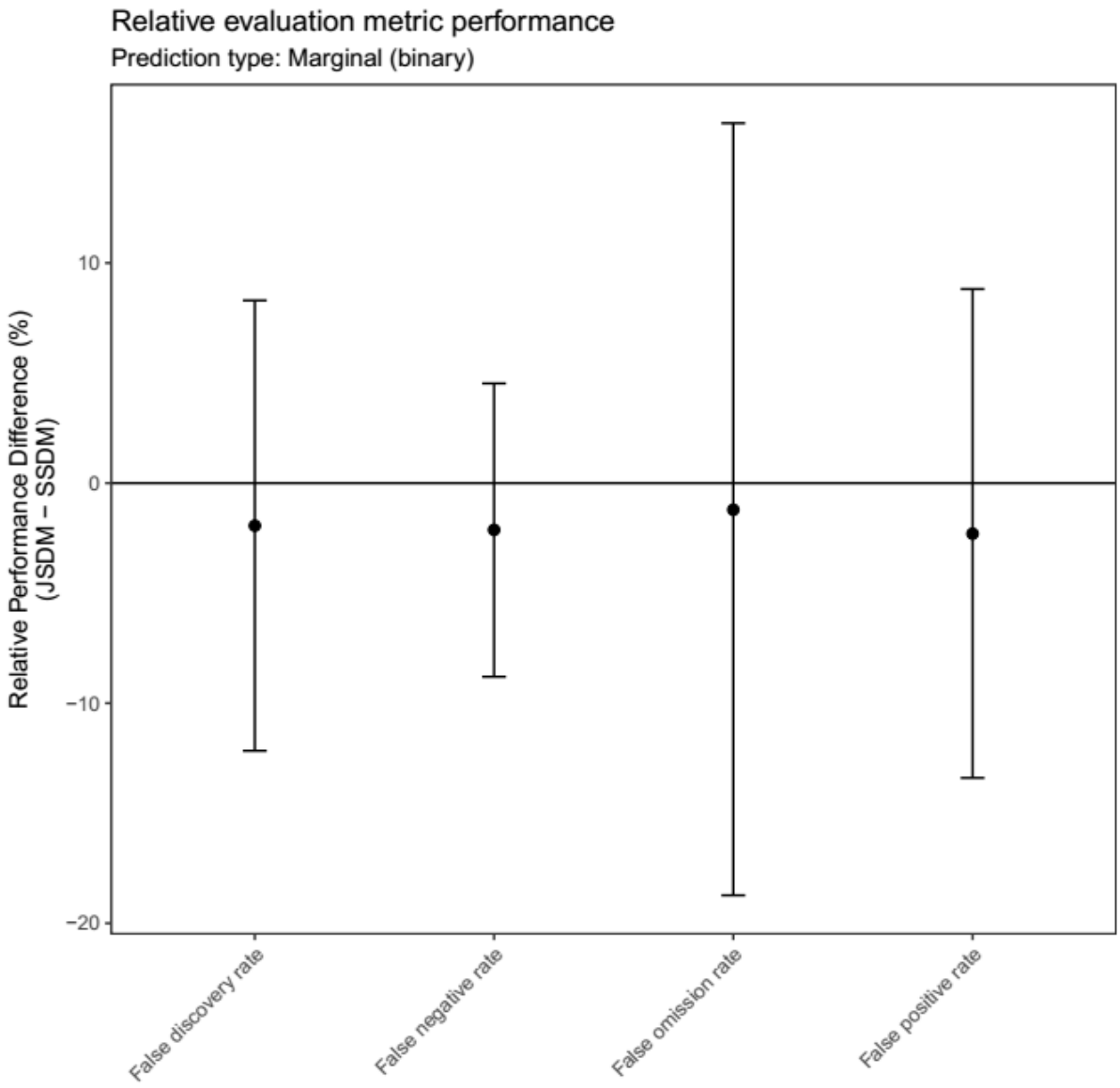


Figure 9: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for marginal (binary) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

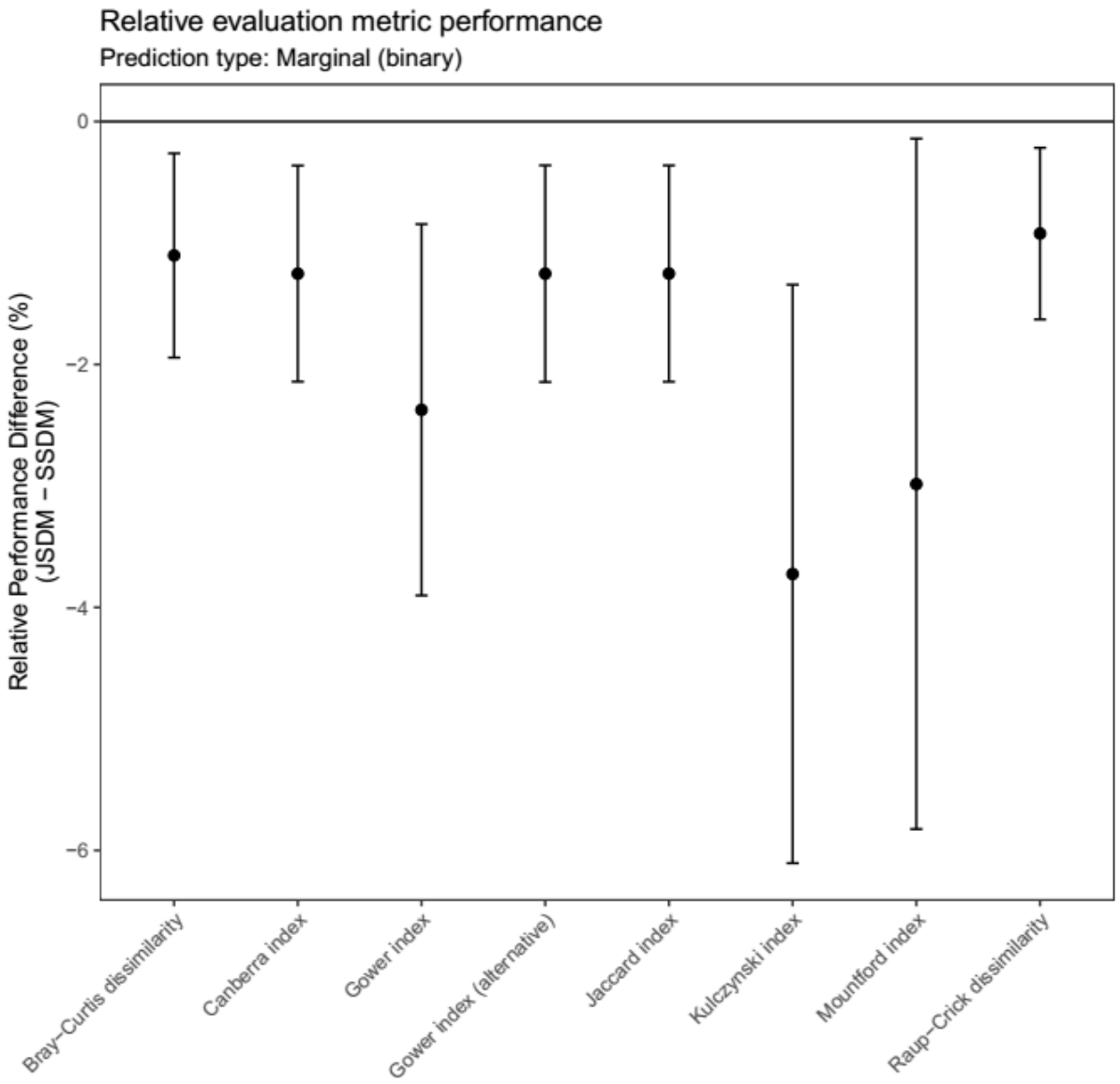


Figure 10: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for marginal (binary) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

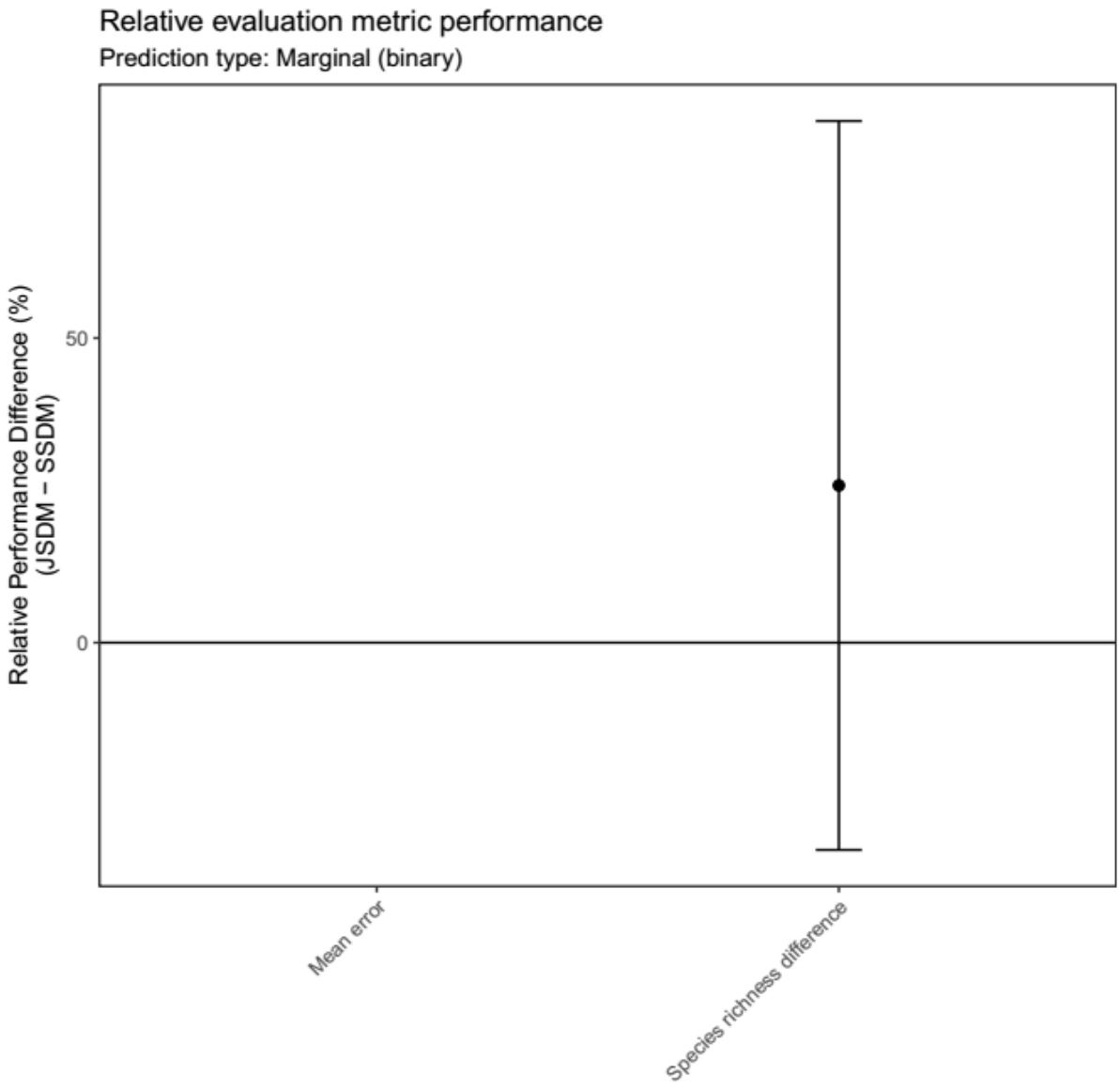


Figure 11: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for marginal (binary) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

3. Joint predictions

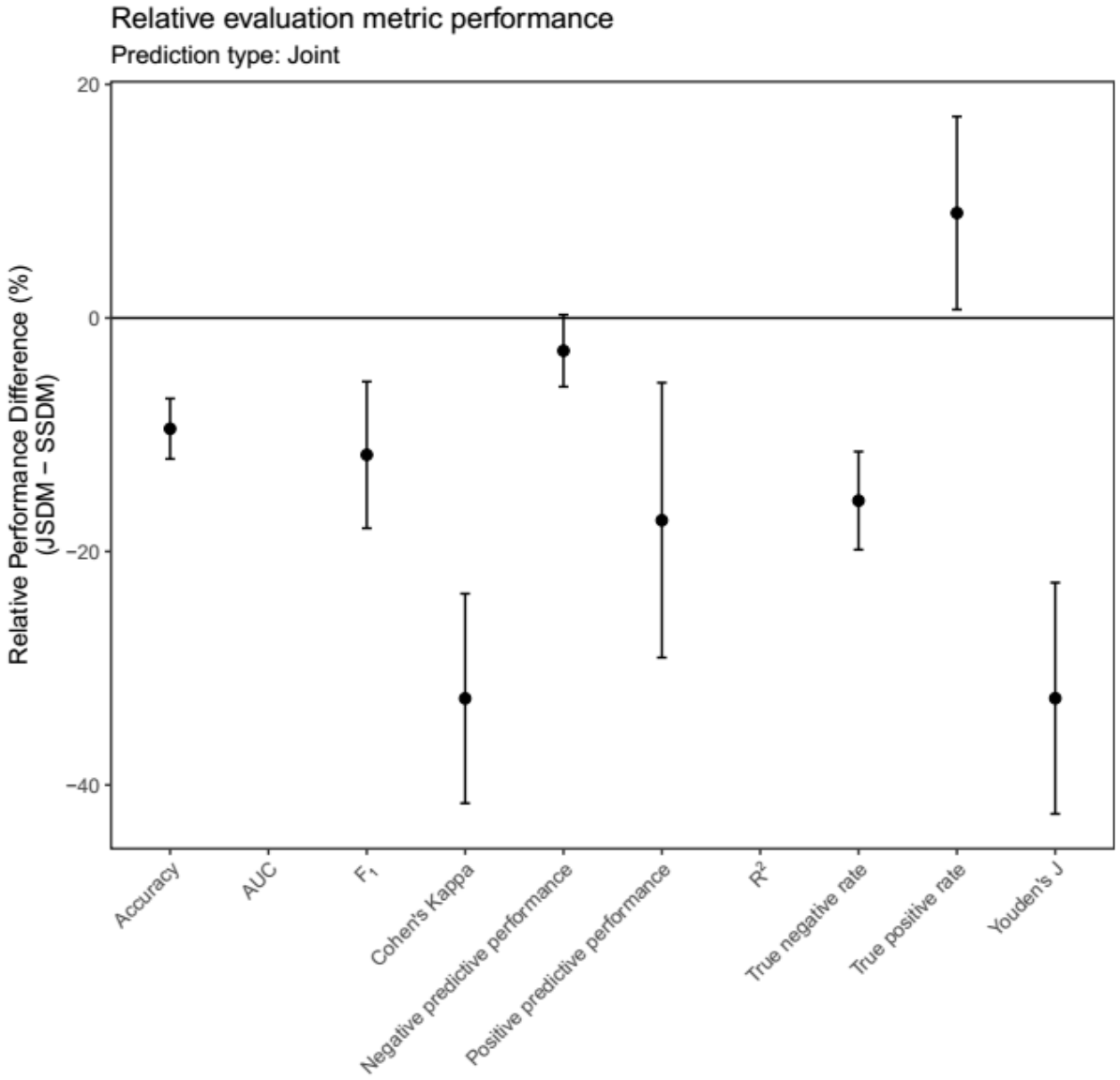


Figure 12: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for joint predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

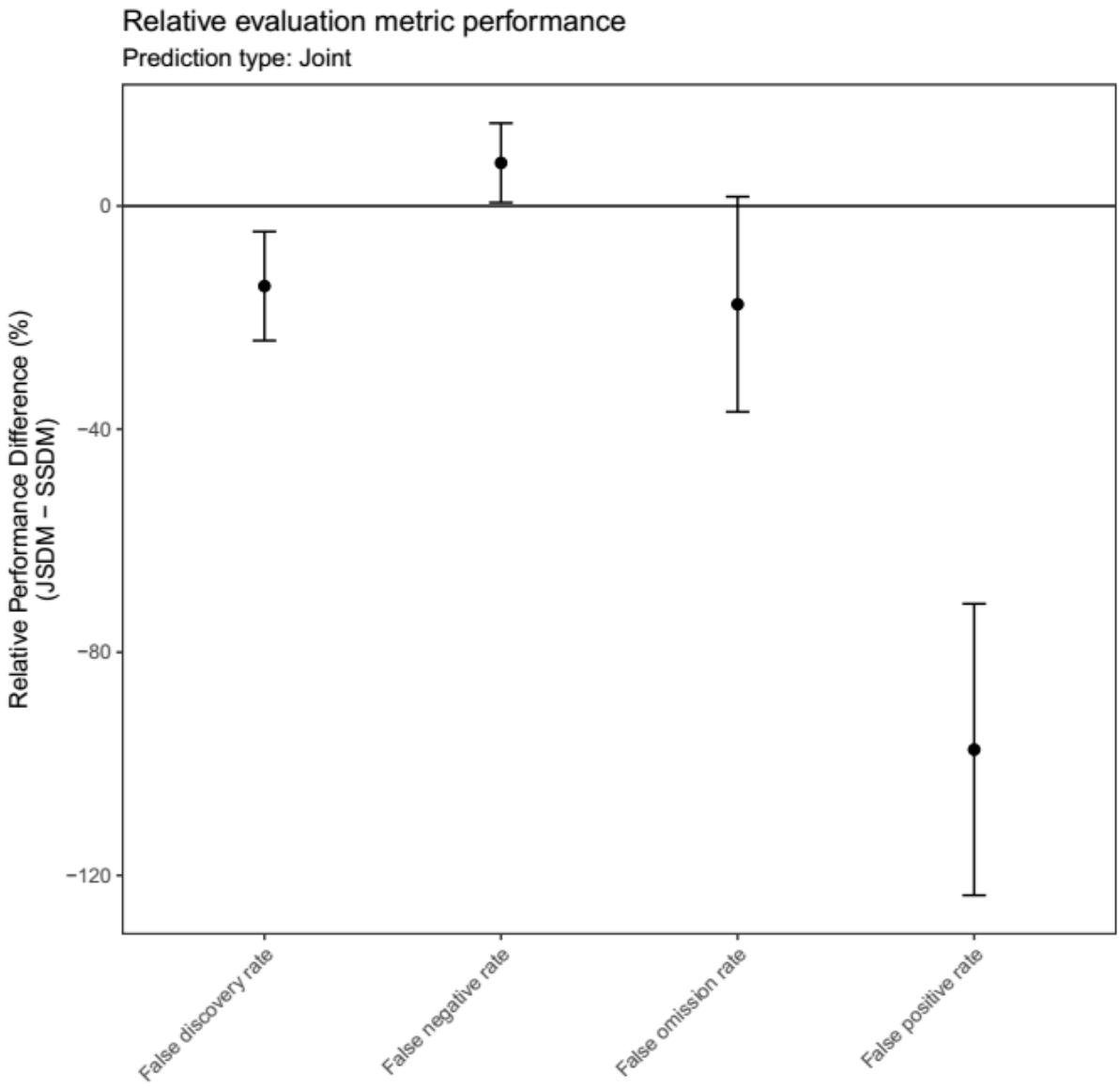


Figure 13: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for joint predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

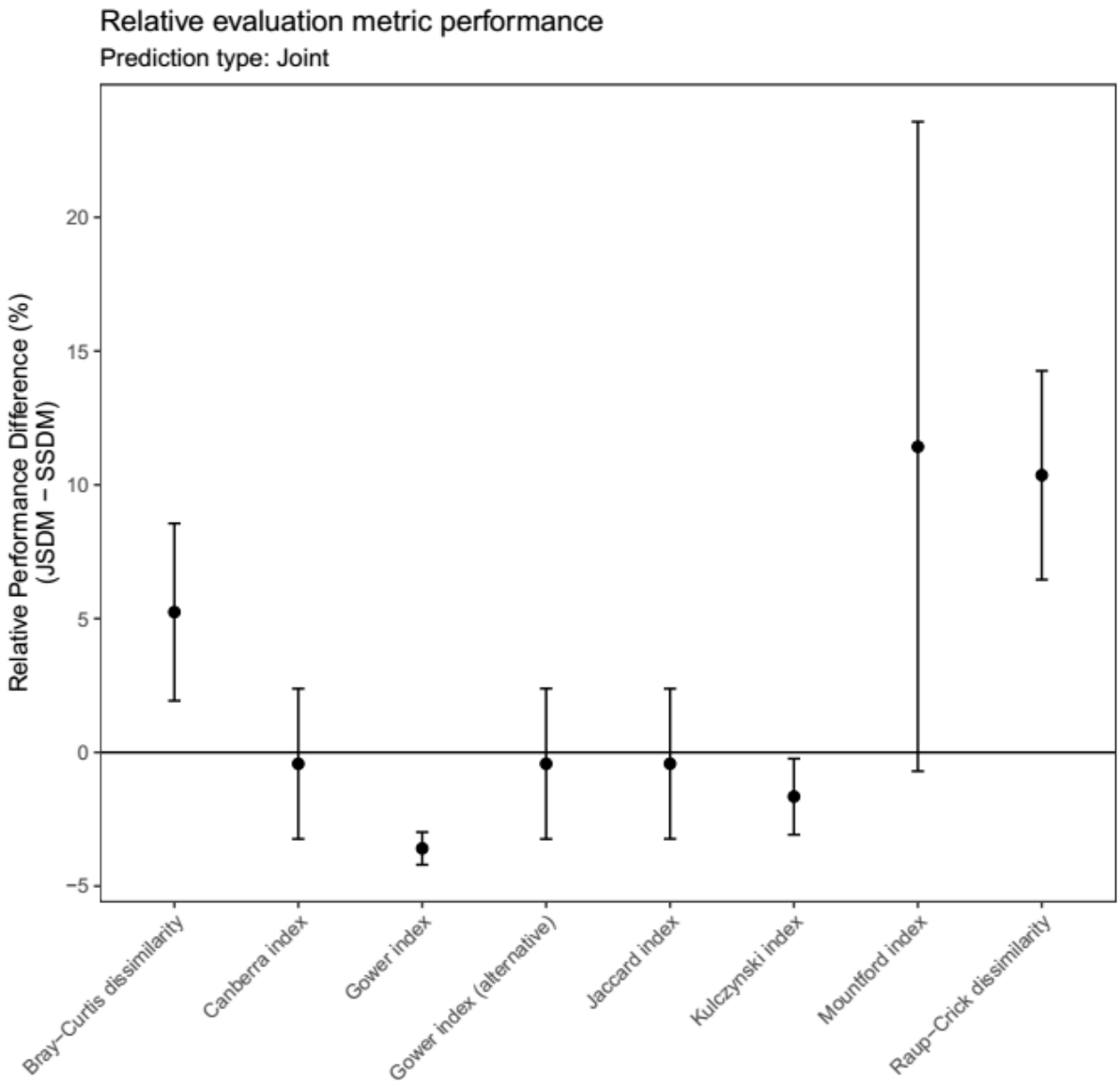


Figure 14: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for joint predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

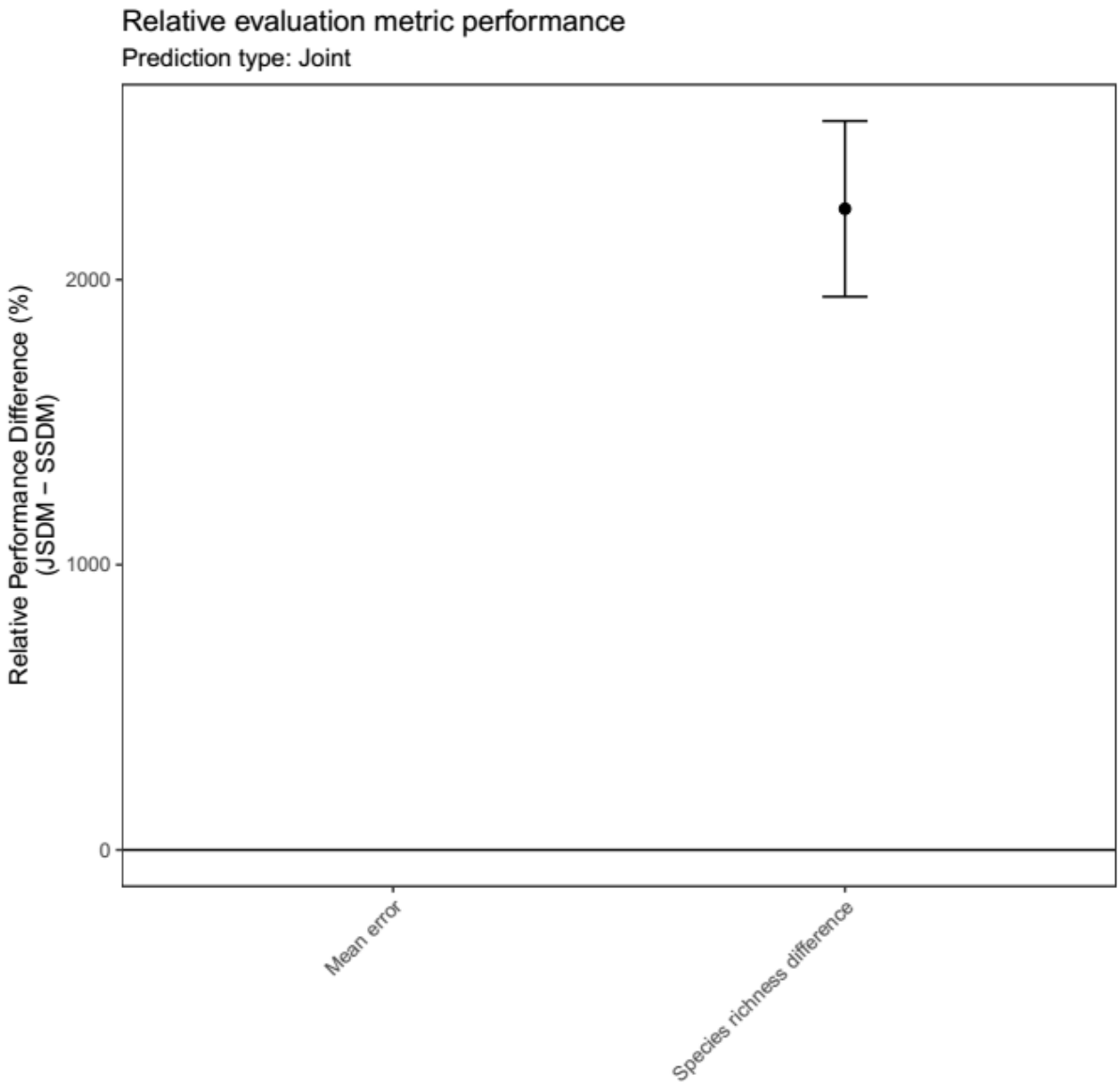


Figure 15: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for joint predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

4. Conditional (low-prevalence) predictions

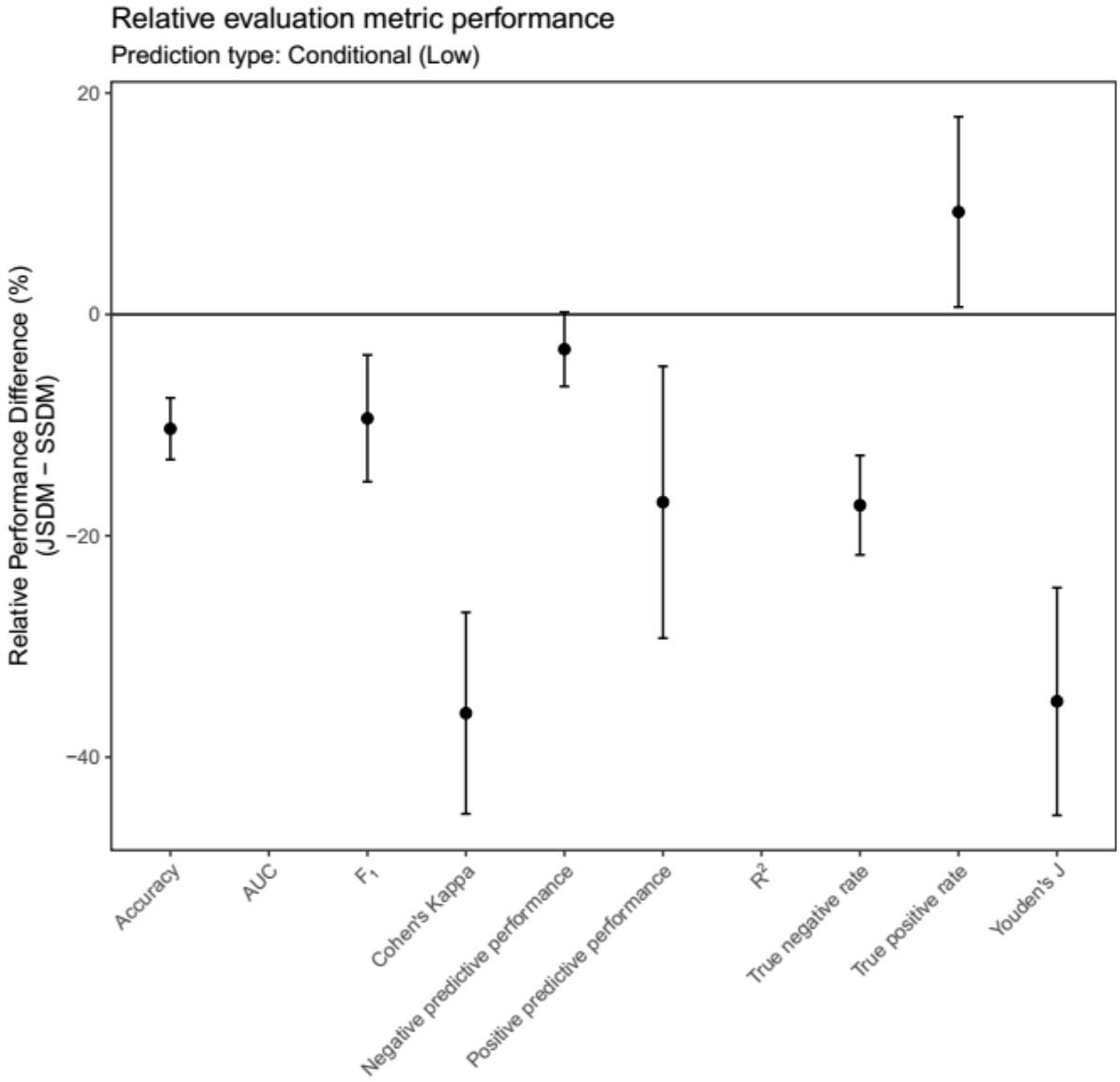


Figure 16: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for conditional (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

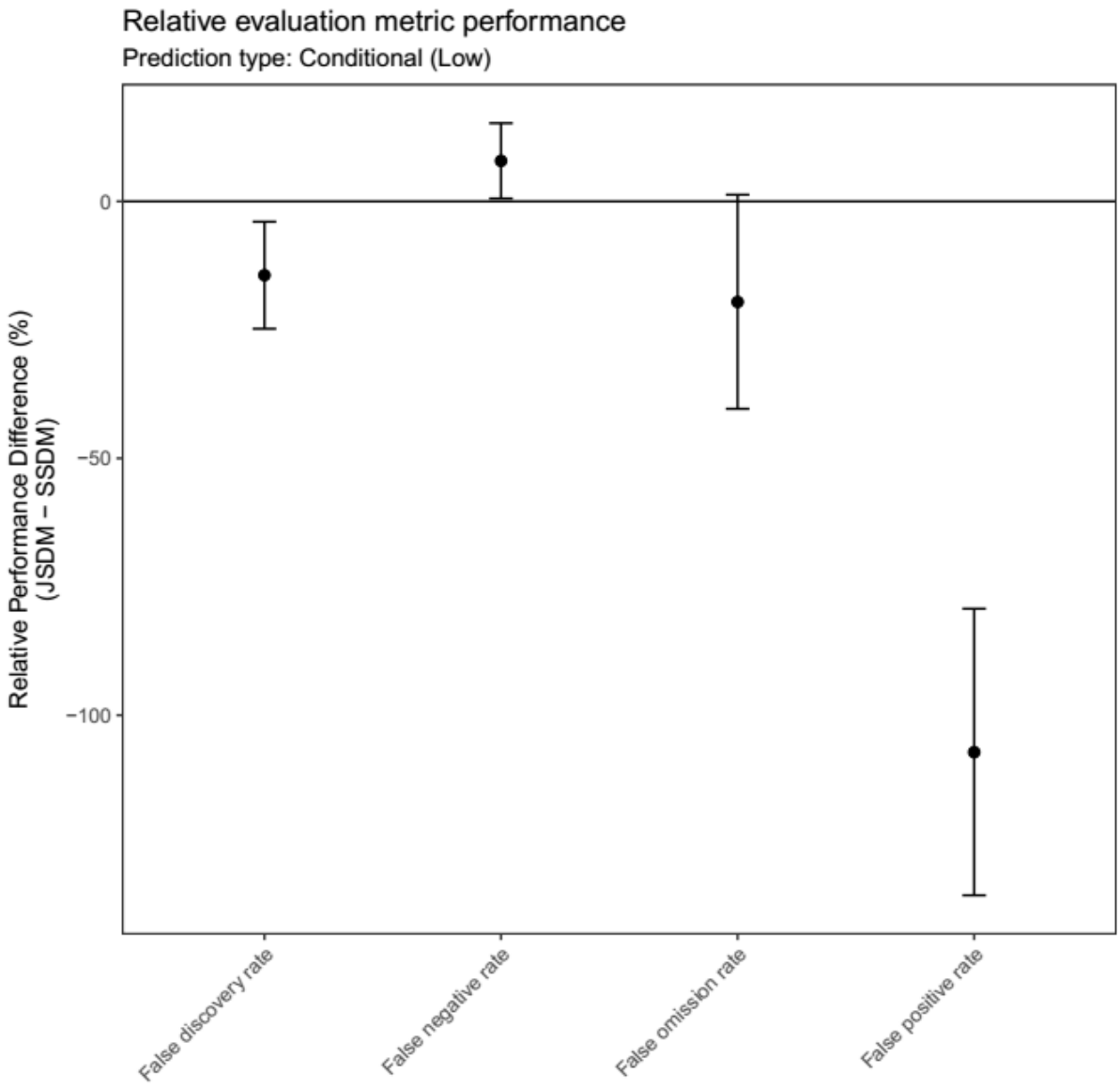


Figure 17: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

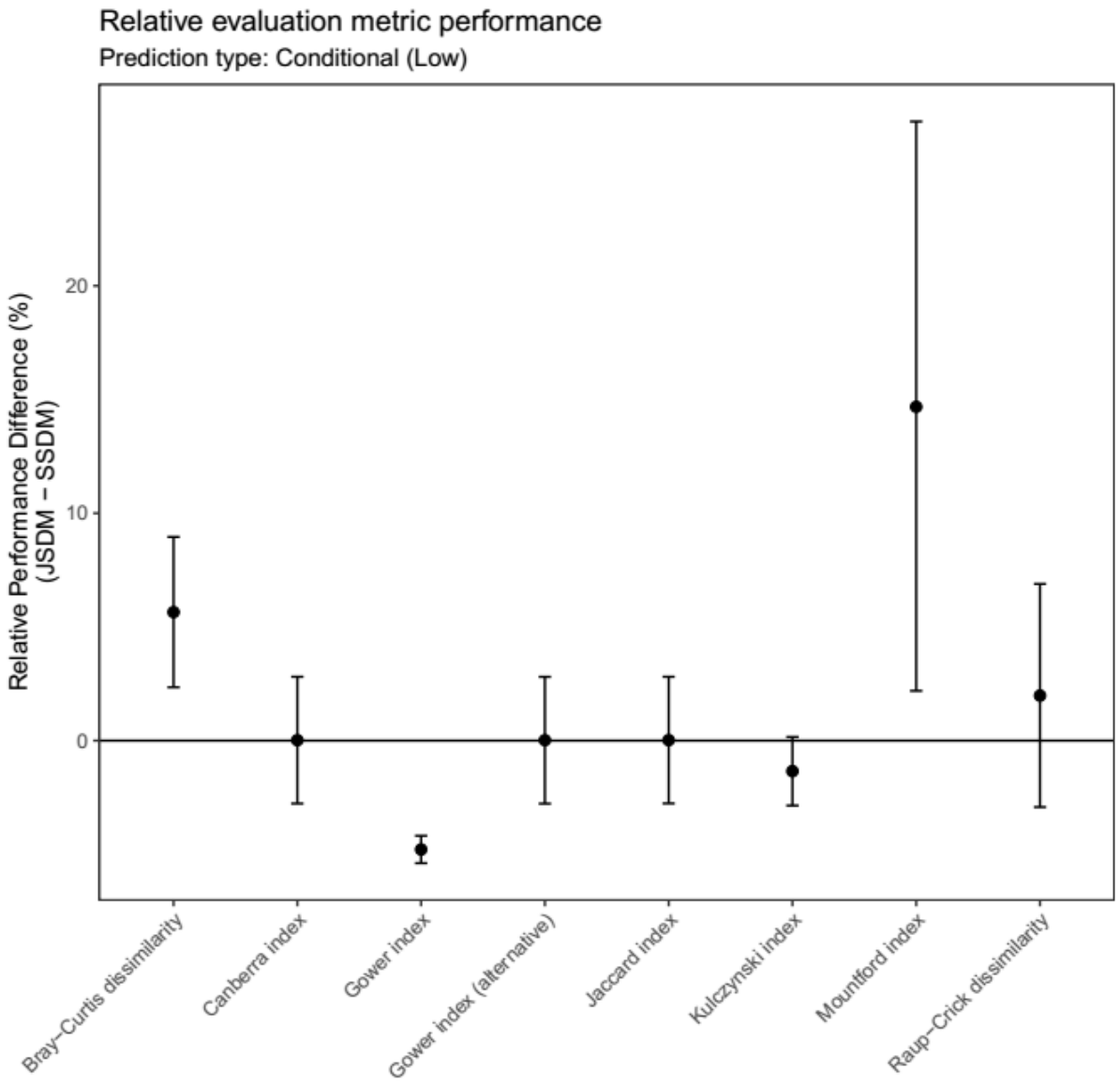


Figure 18: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

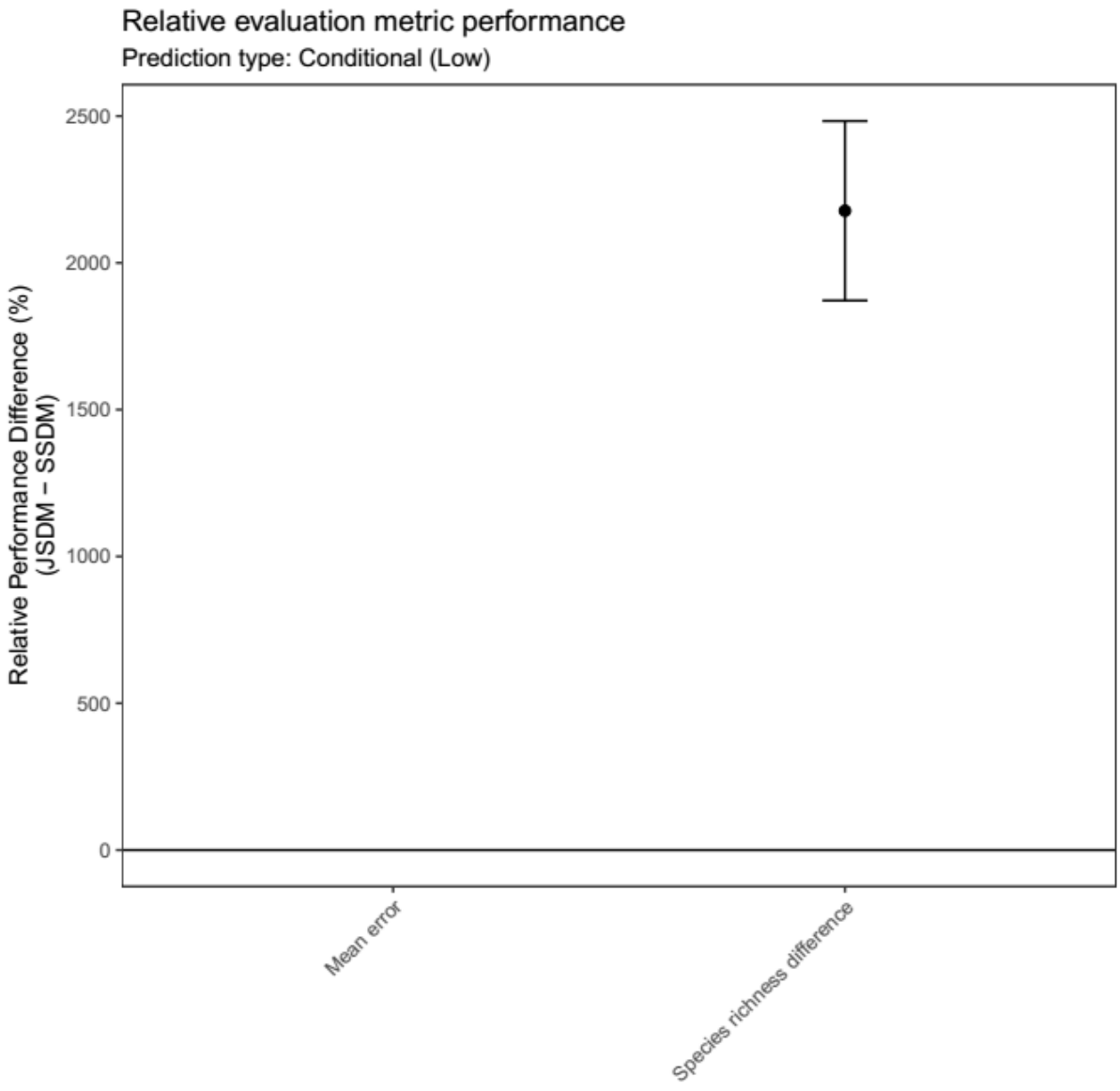


Figure 19: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for conditional (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

5. Conditional (medium-prevalence) predictions

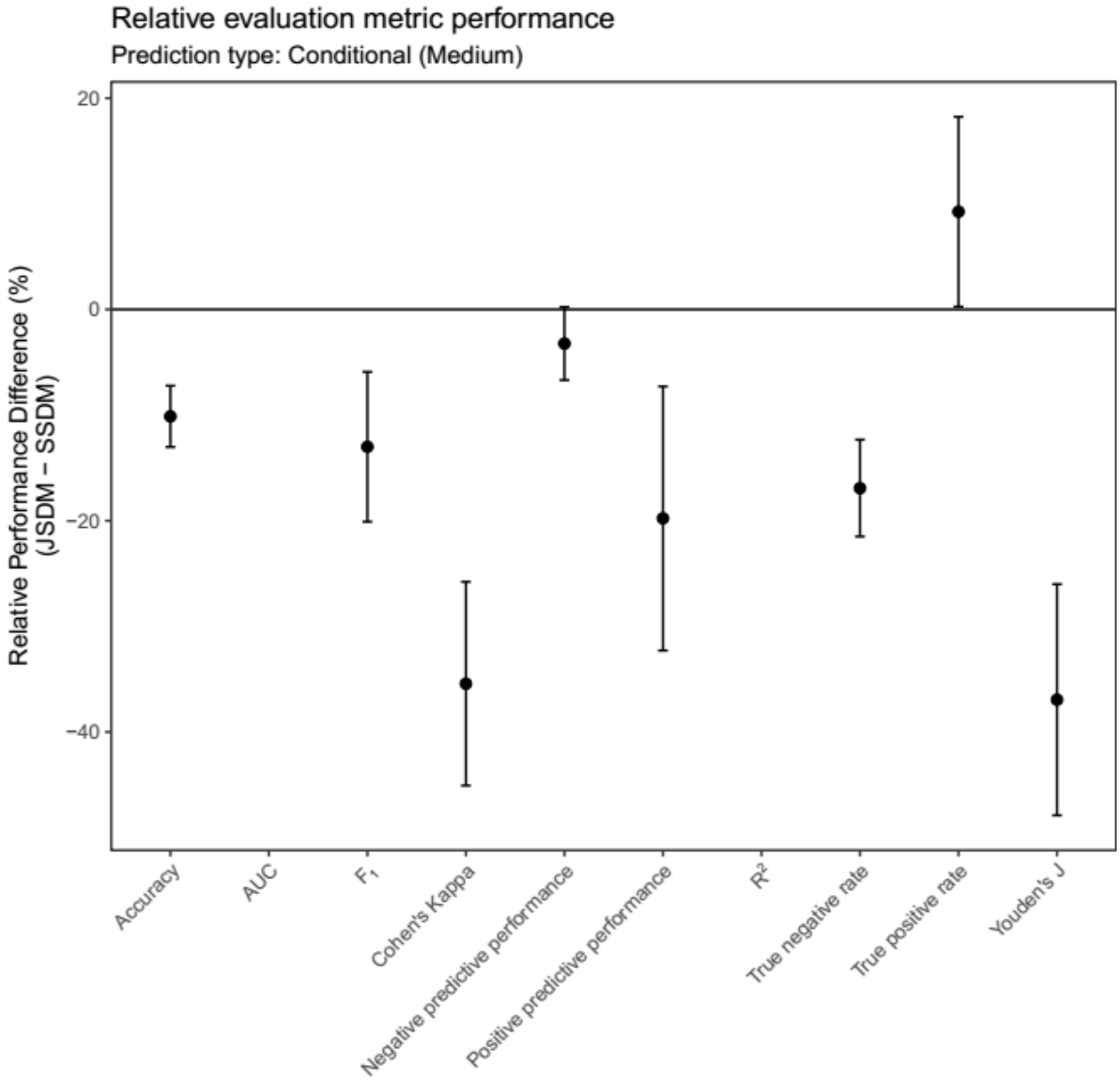


Figure 20: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for conditional (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

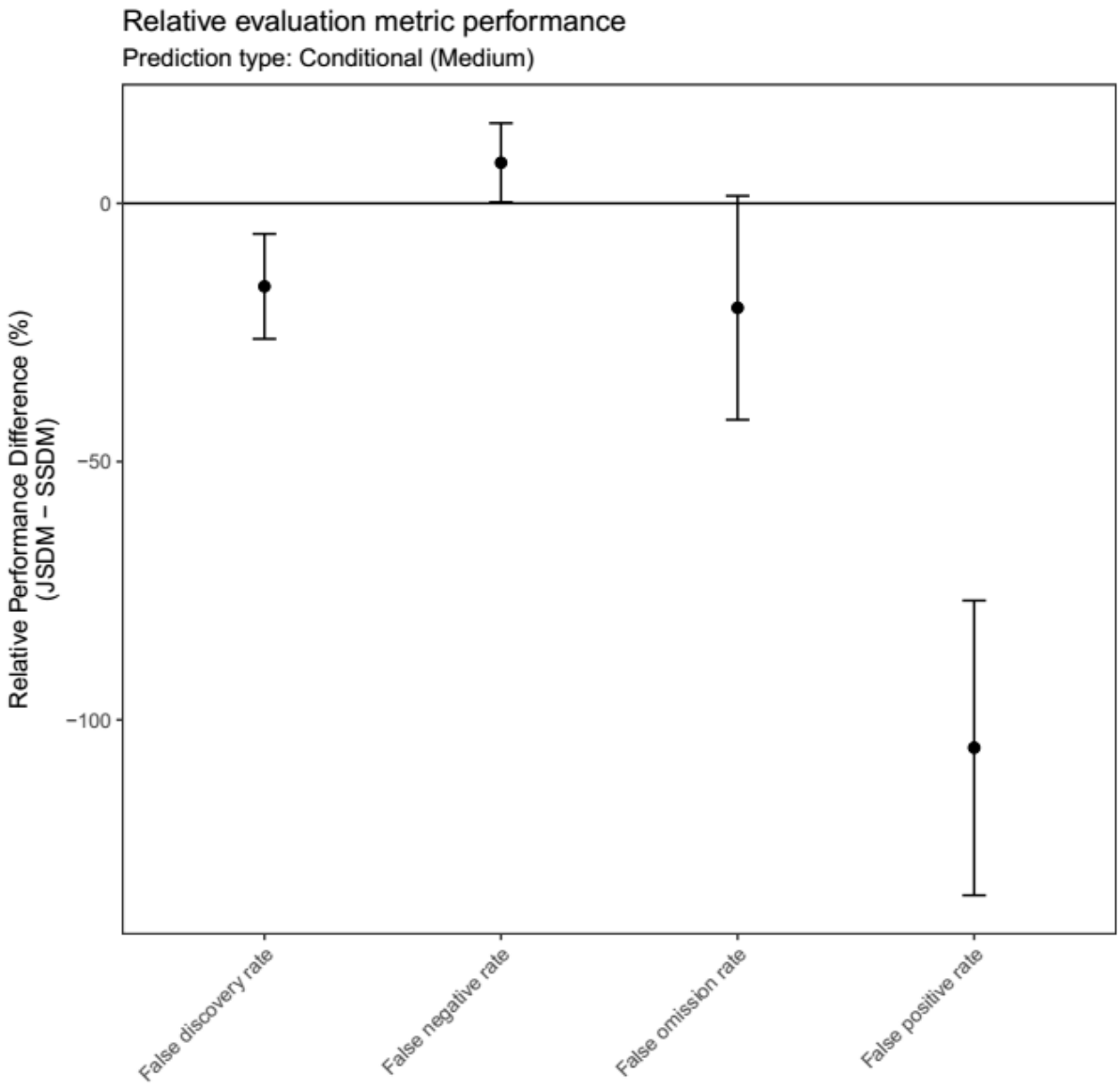


Figure 21: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

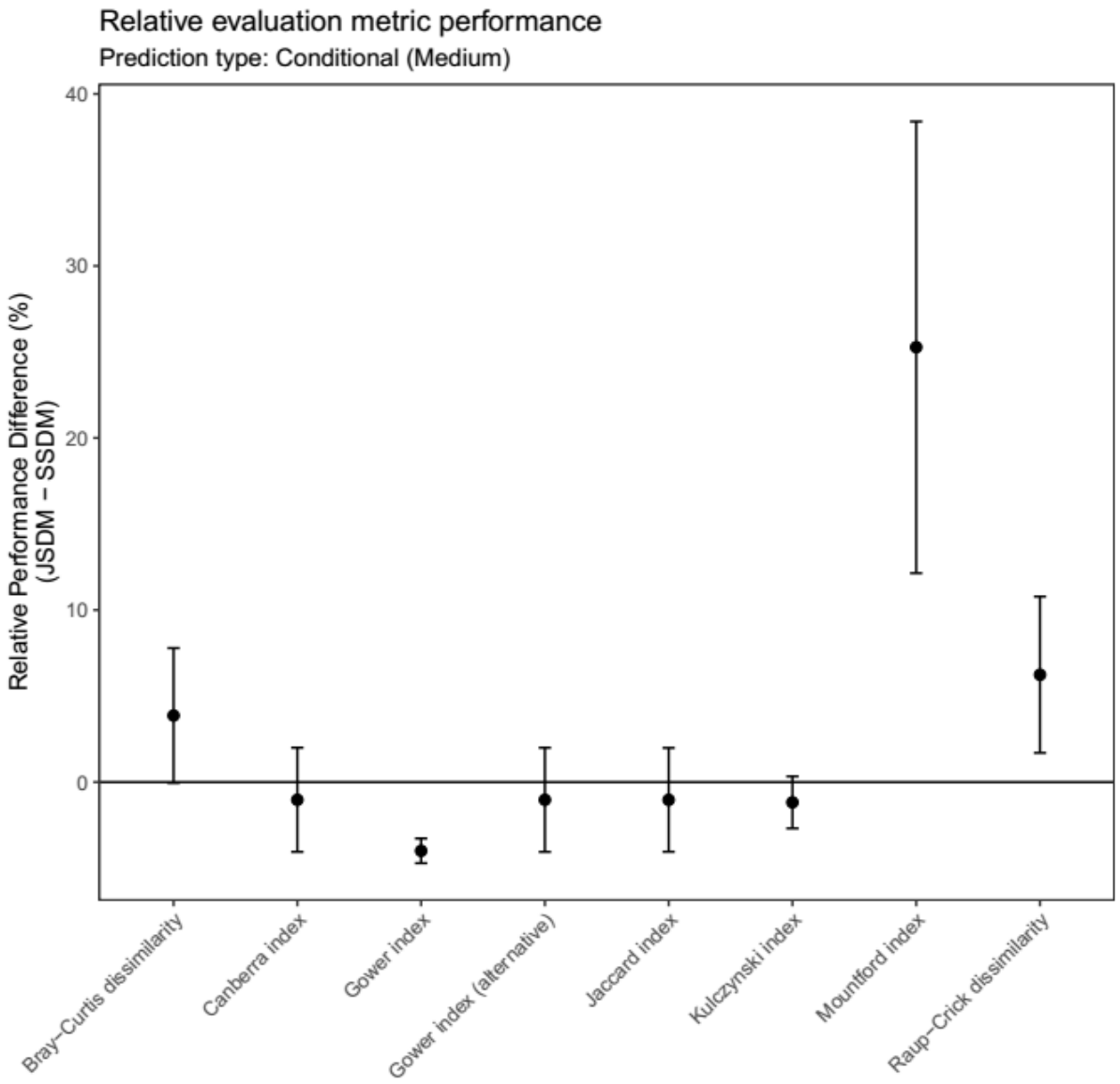


Figure 22: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

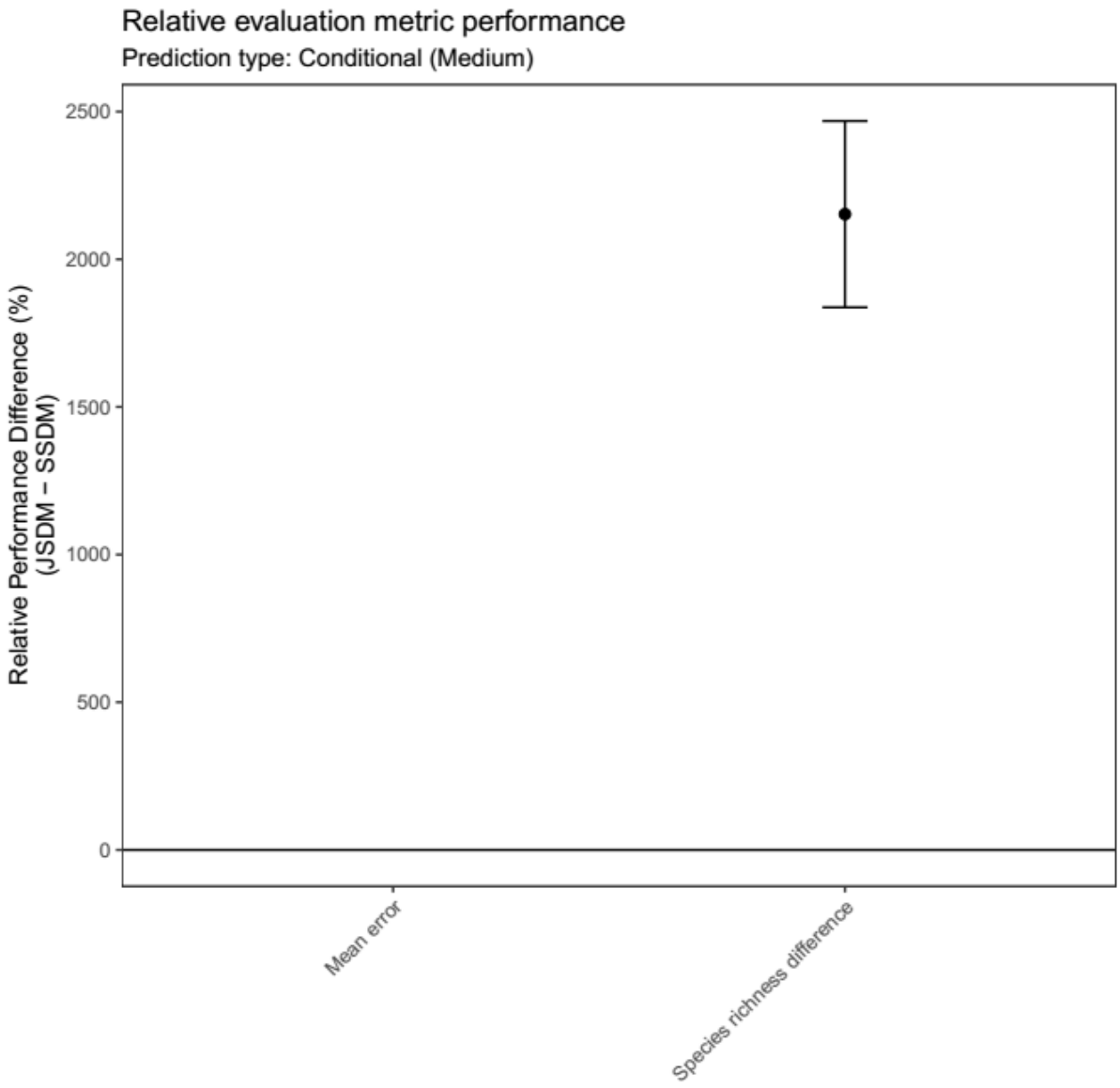


Figure 23: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for conditional (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

6. Conditional (high-prevalence) predictions

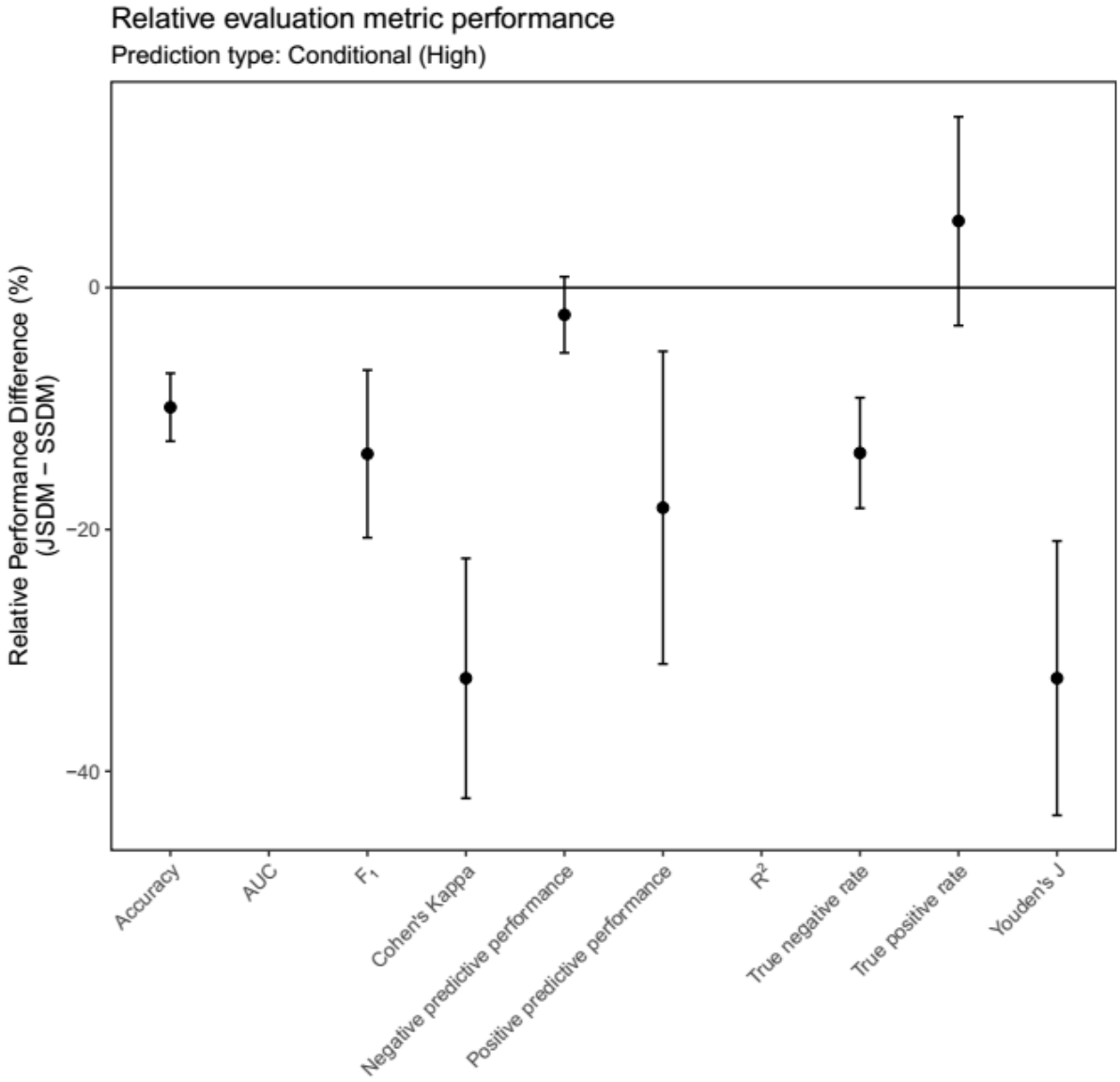


Figure 24: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for conditional (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

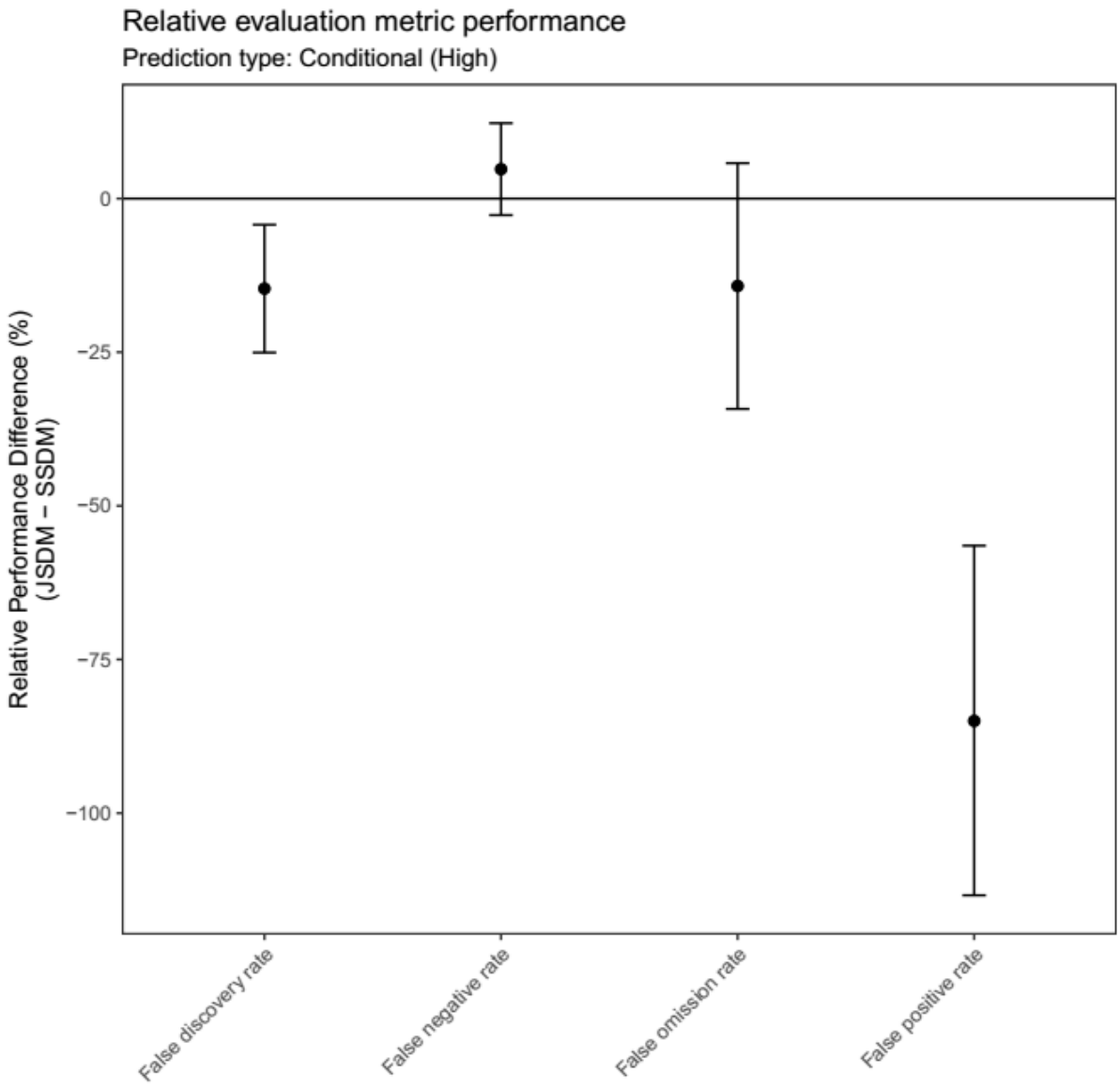


Figure 25: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

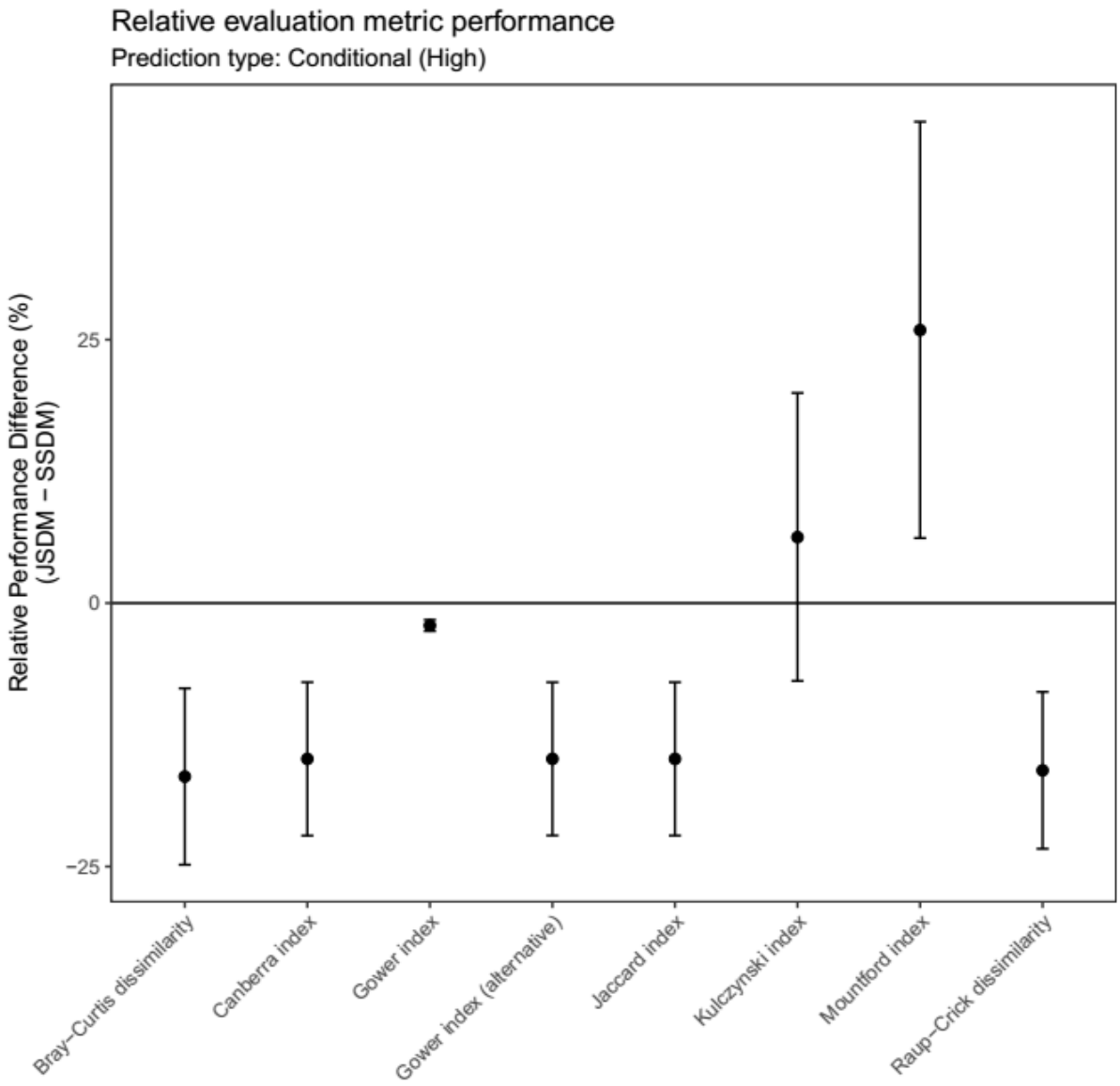


Figure 26: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

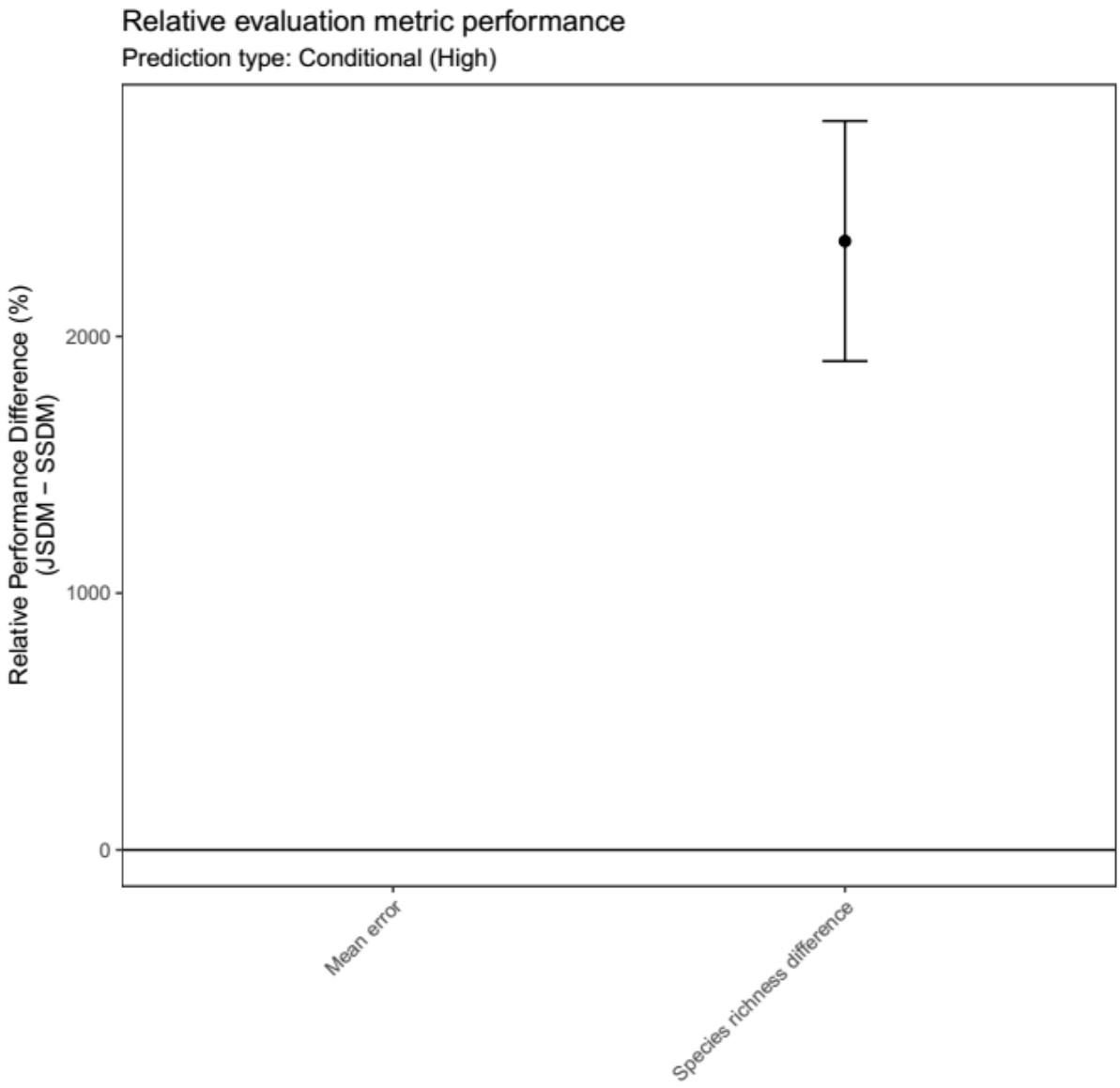


Figure 27: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for conditional (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

7. Conditional marginal (low-prevalence) predictions

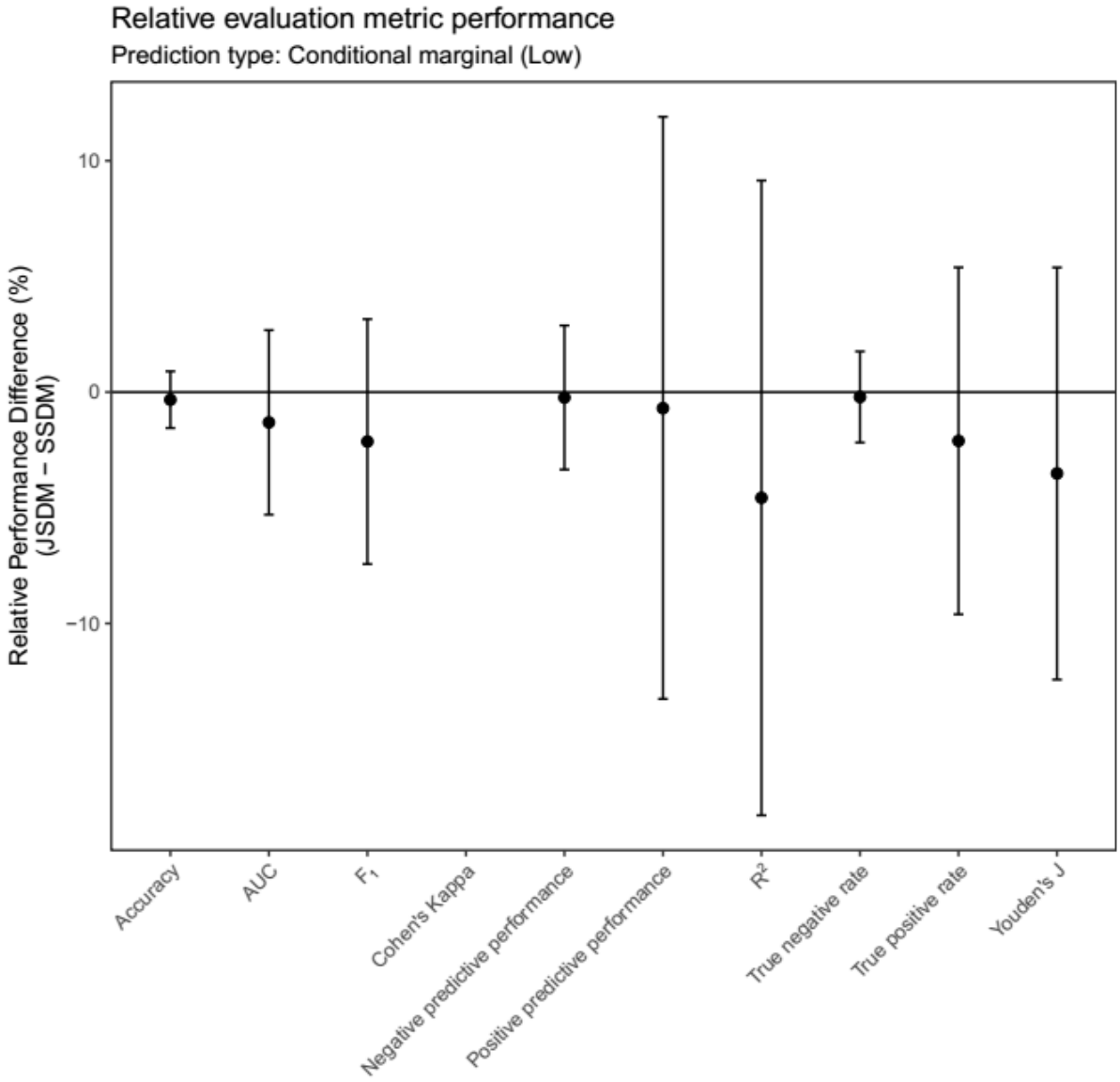


Figure 28: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

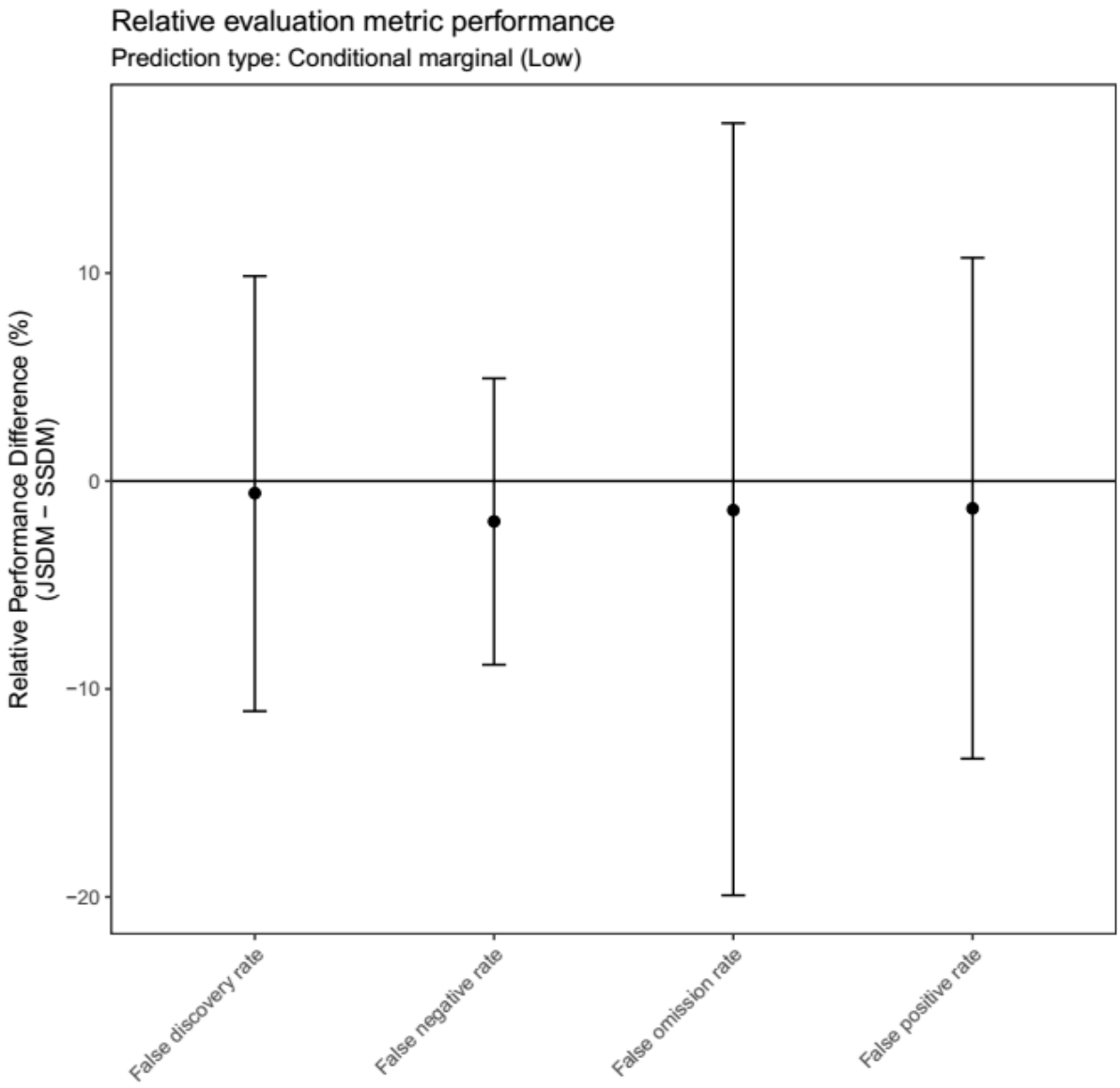


Figure 29: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

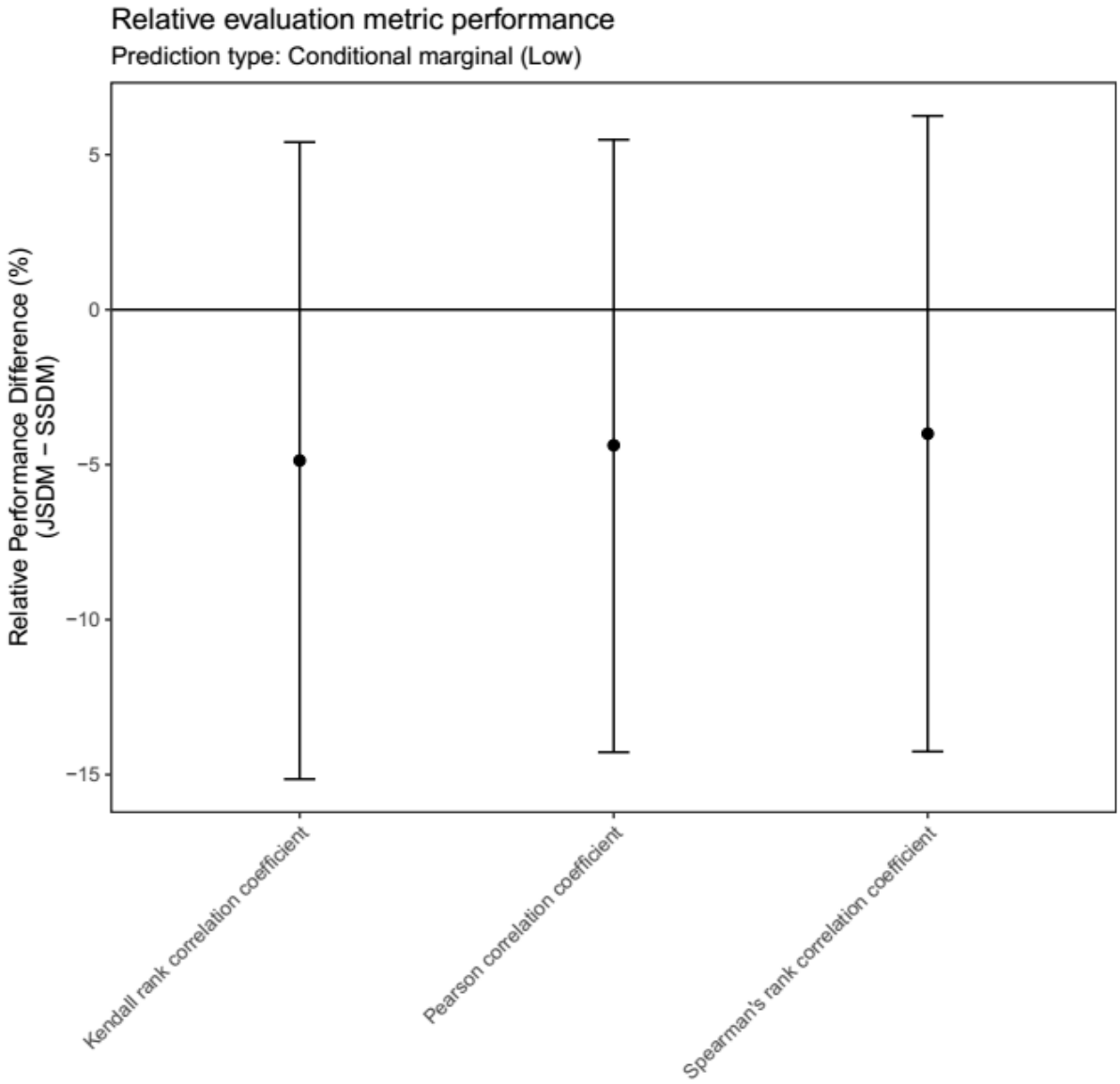


Figure 30: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -1 – 1, where higher values indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

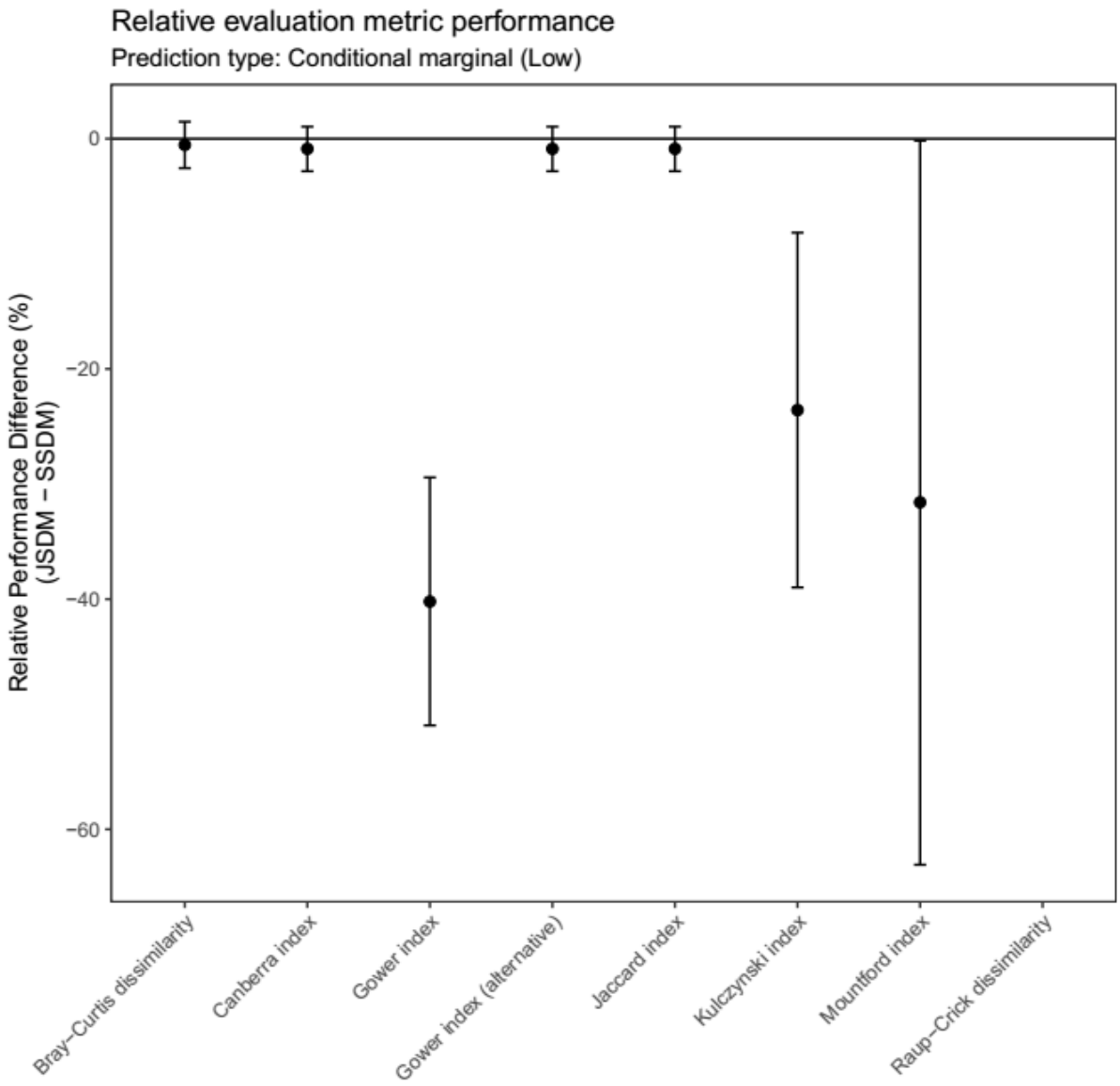


Figure 31: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

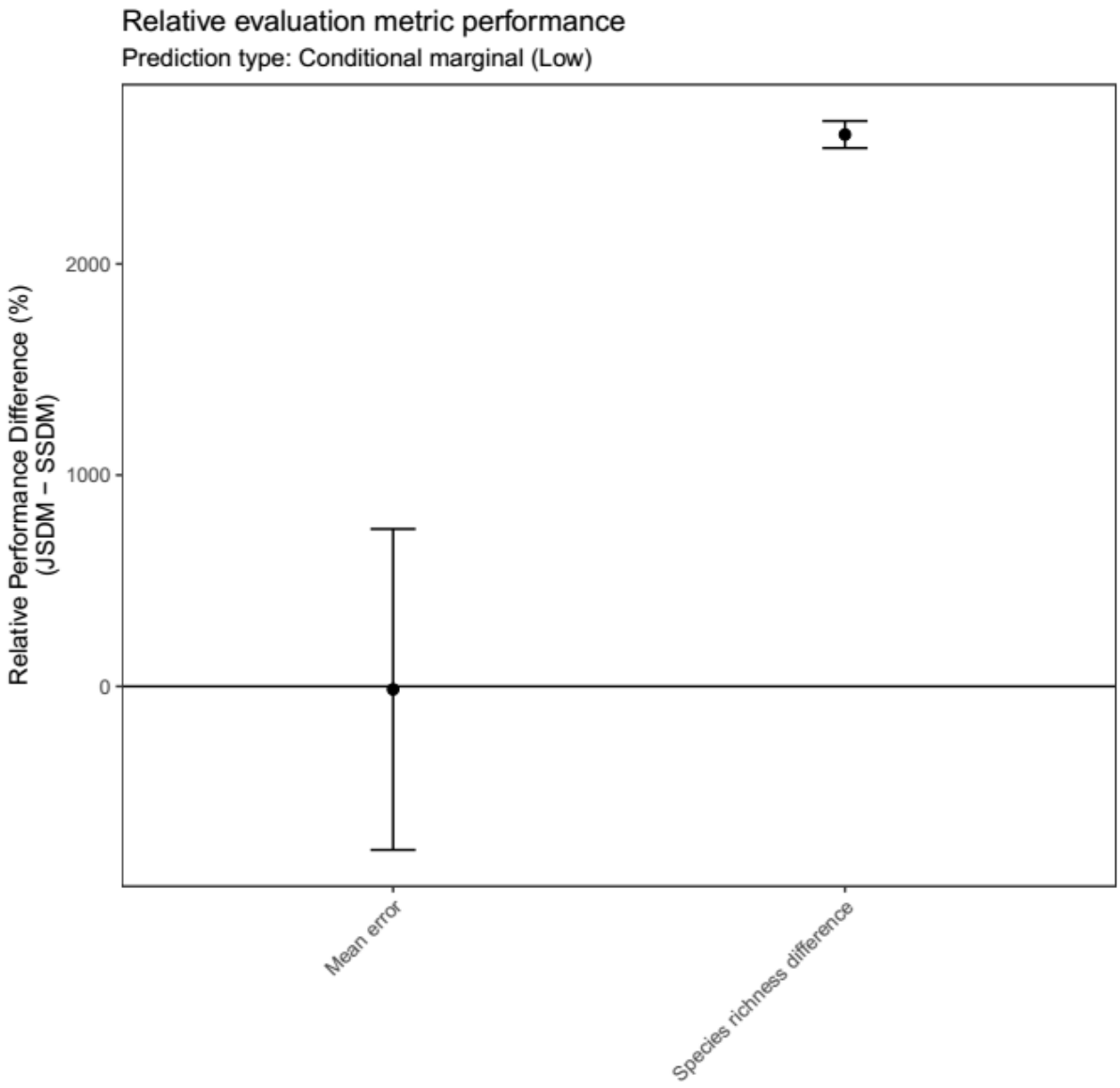


Figure 32: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

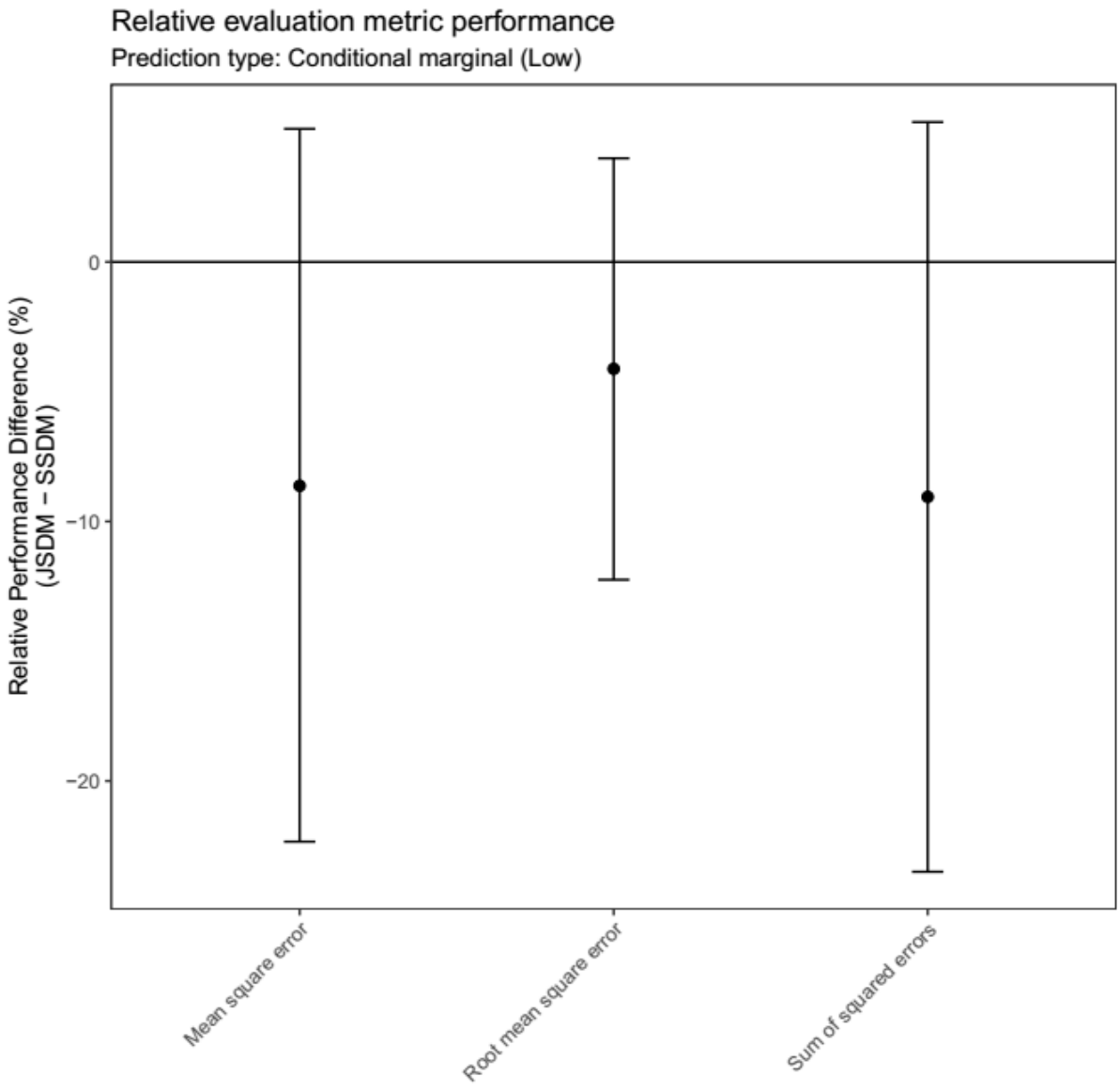


Figure 33: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – ∞, where lower values indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

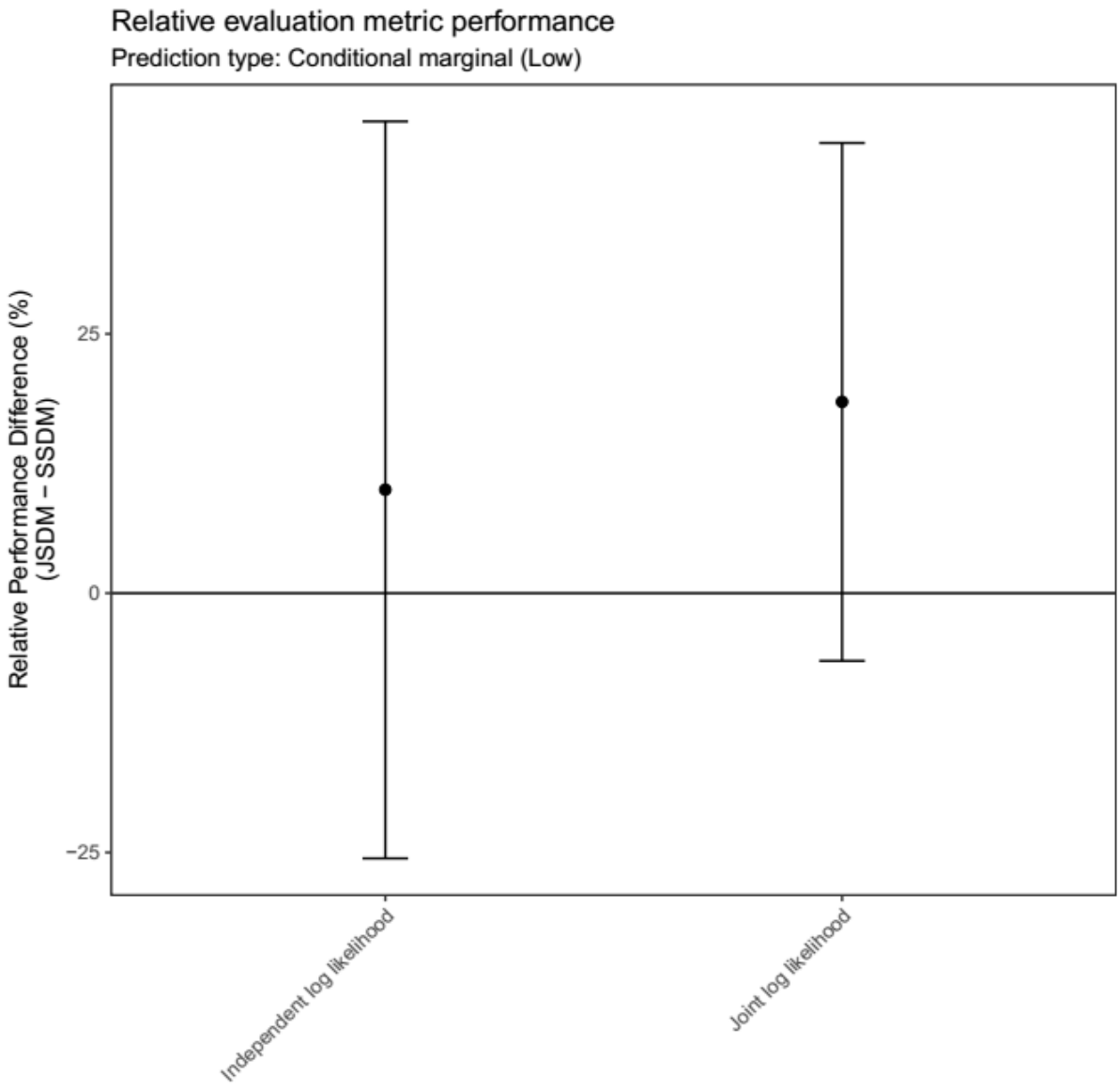


Figure 34: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where higher values indicate better performance, for conditional marginal (low-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

8. Conditional marginal (medium-prevalence) predictions

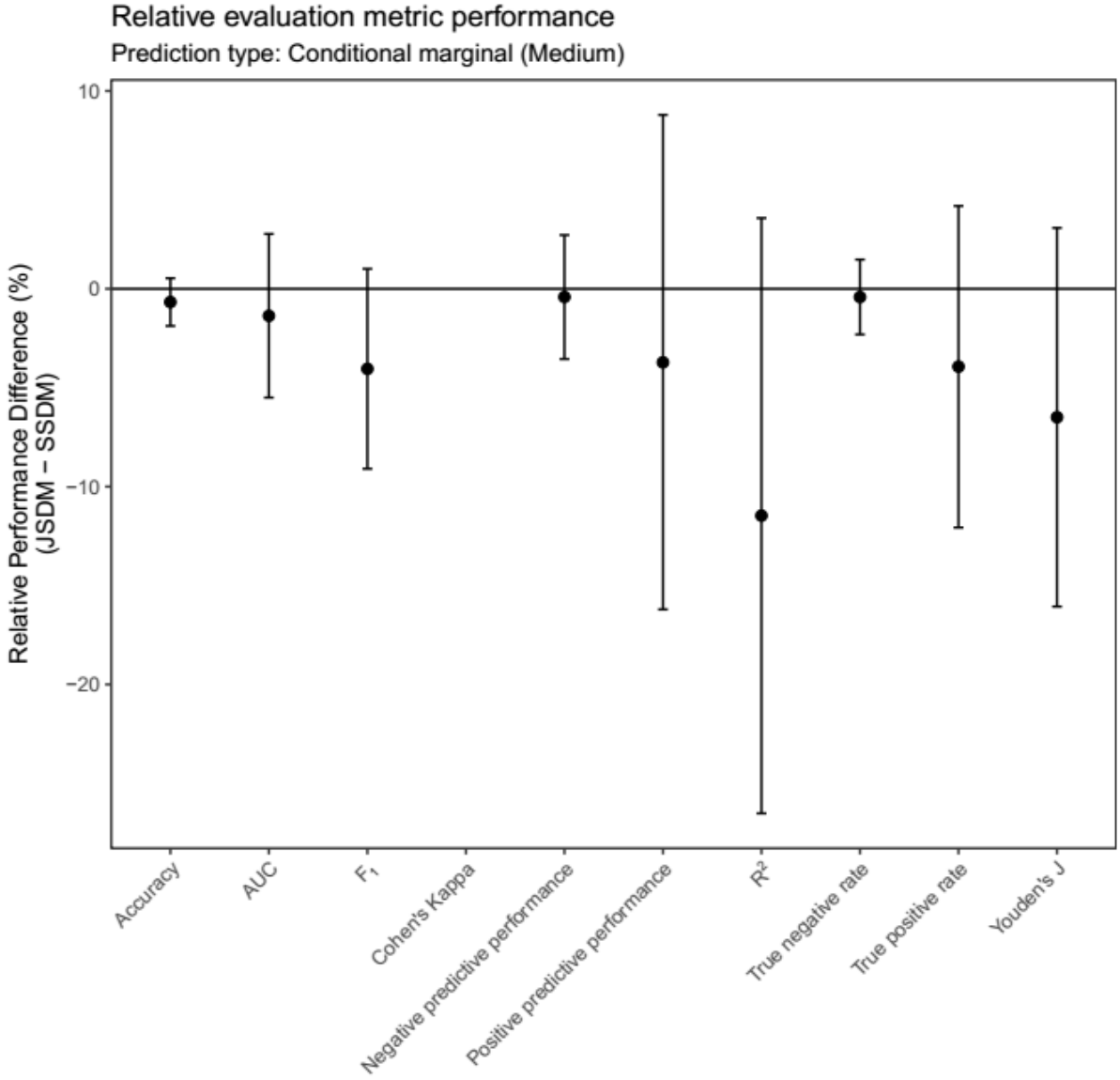


Figure 35: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

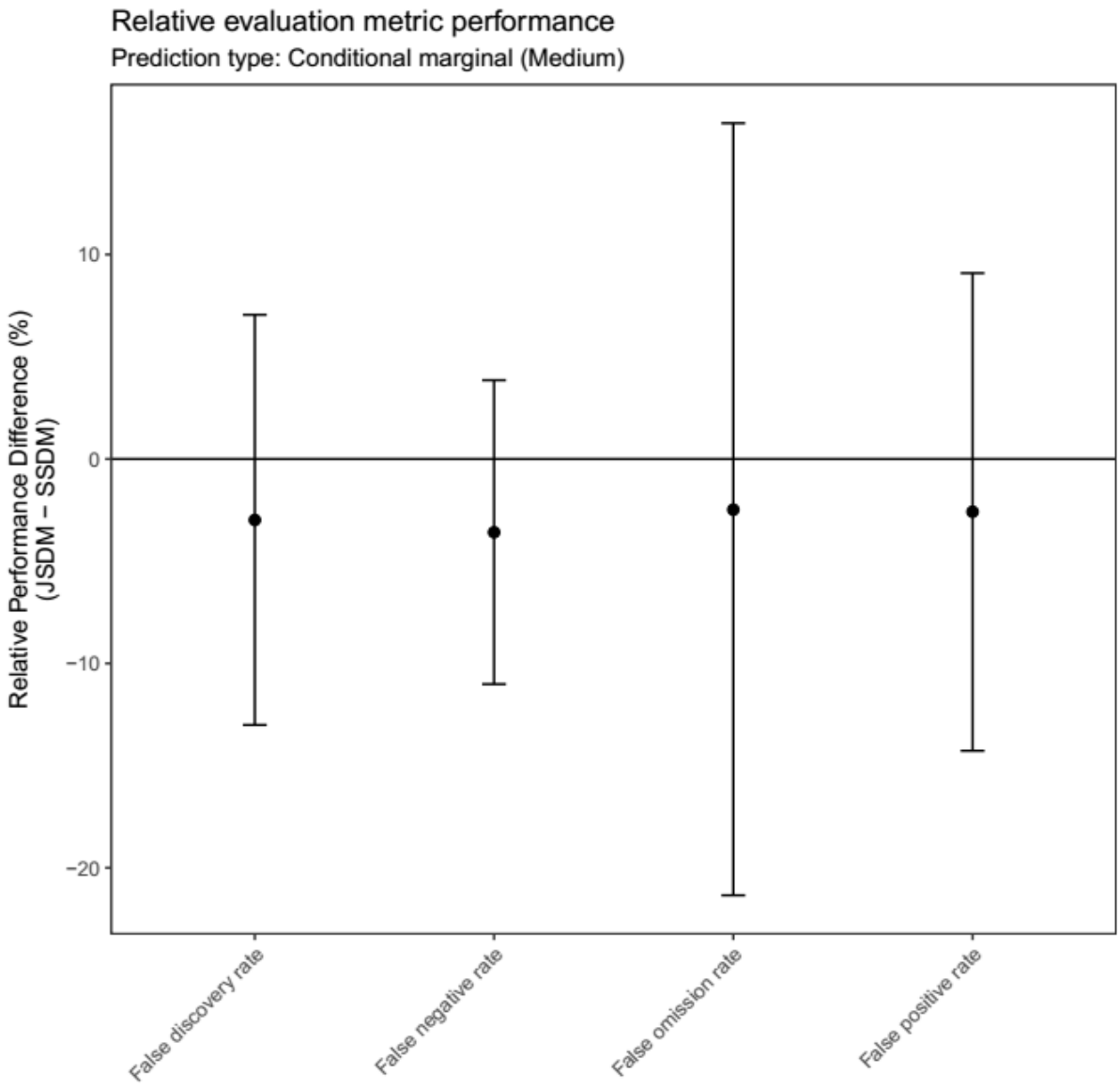


Figure 36: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

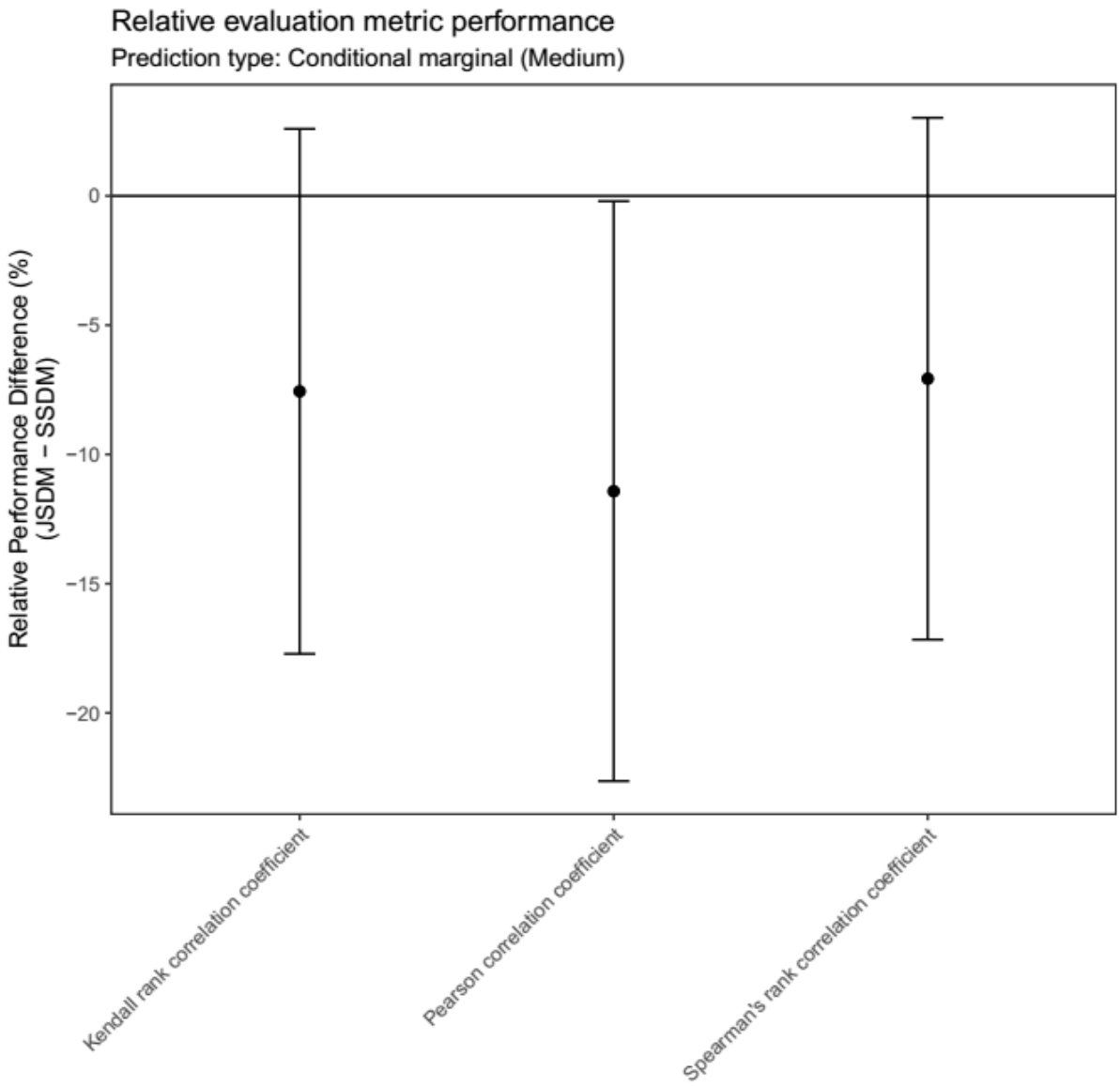


Figure 37: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -1 – 1, where higher values indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

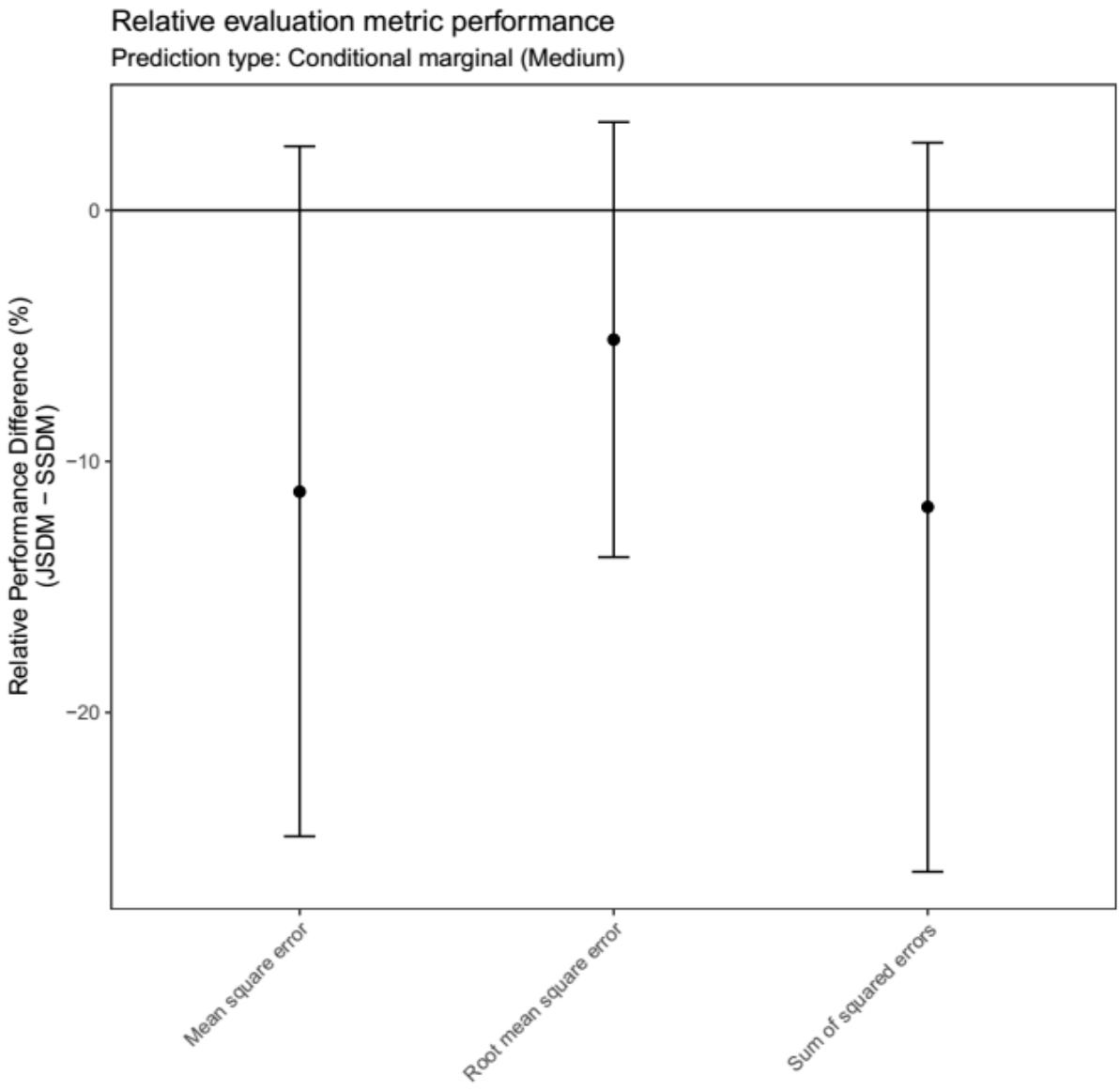


Figure 35: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – ∞, where lower values indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

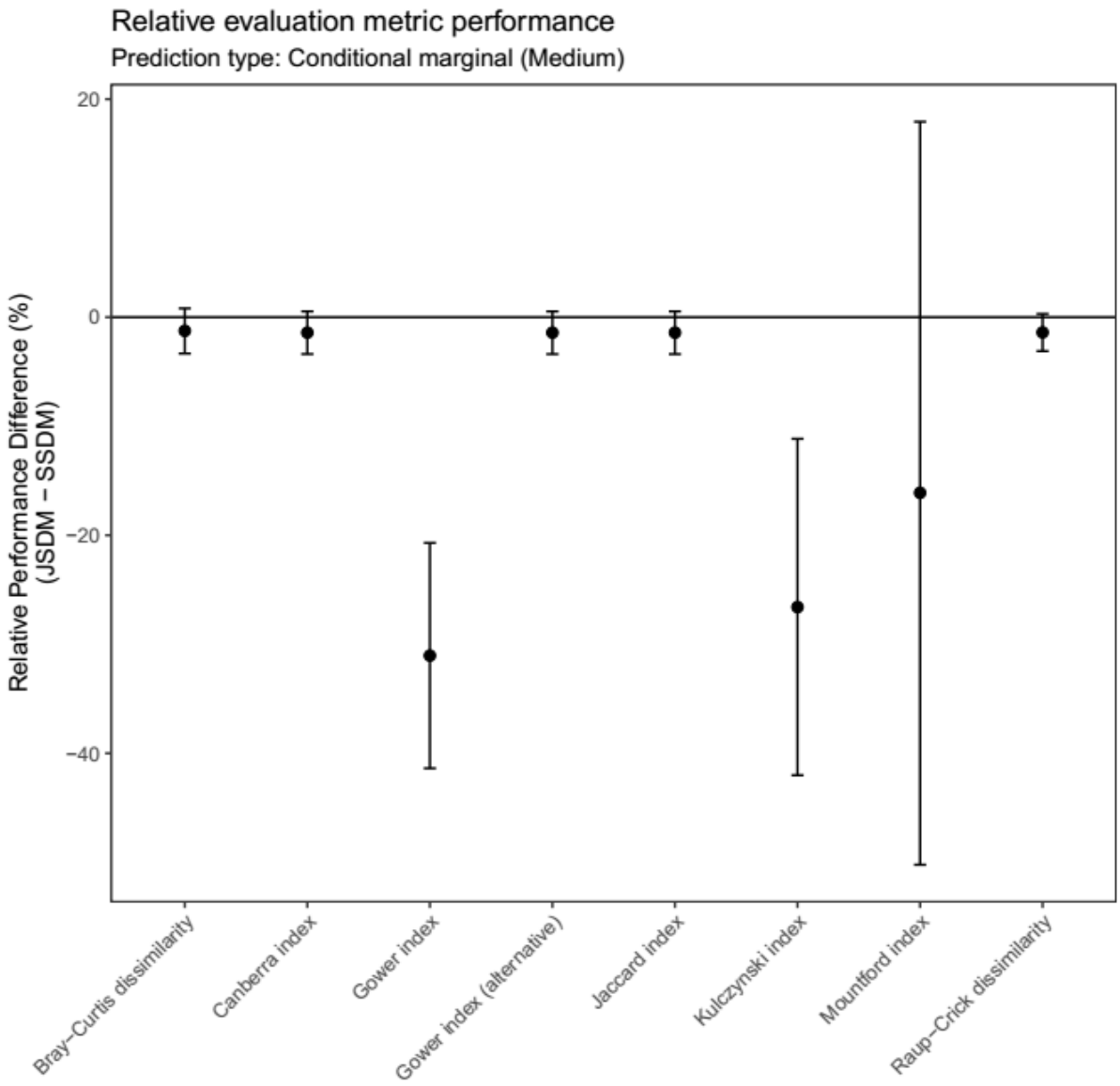


Figure 39: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

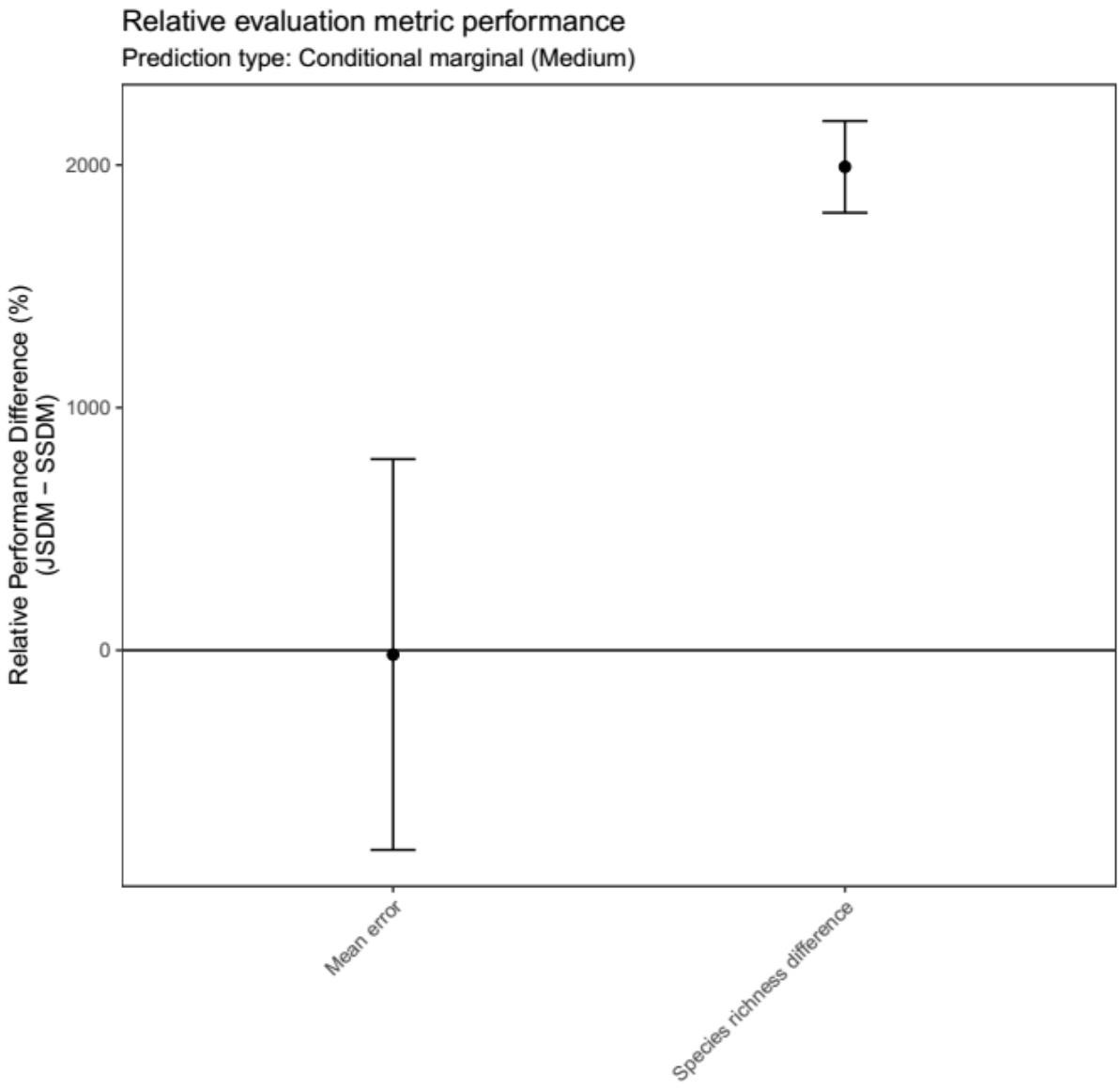


Figure 40: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

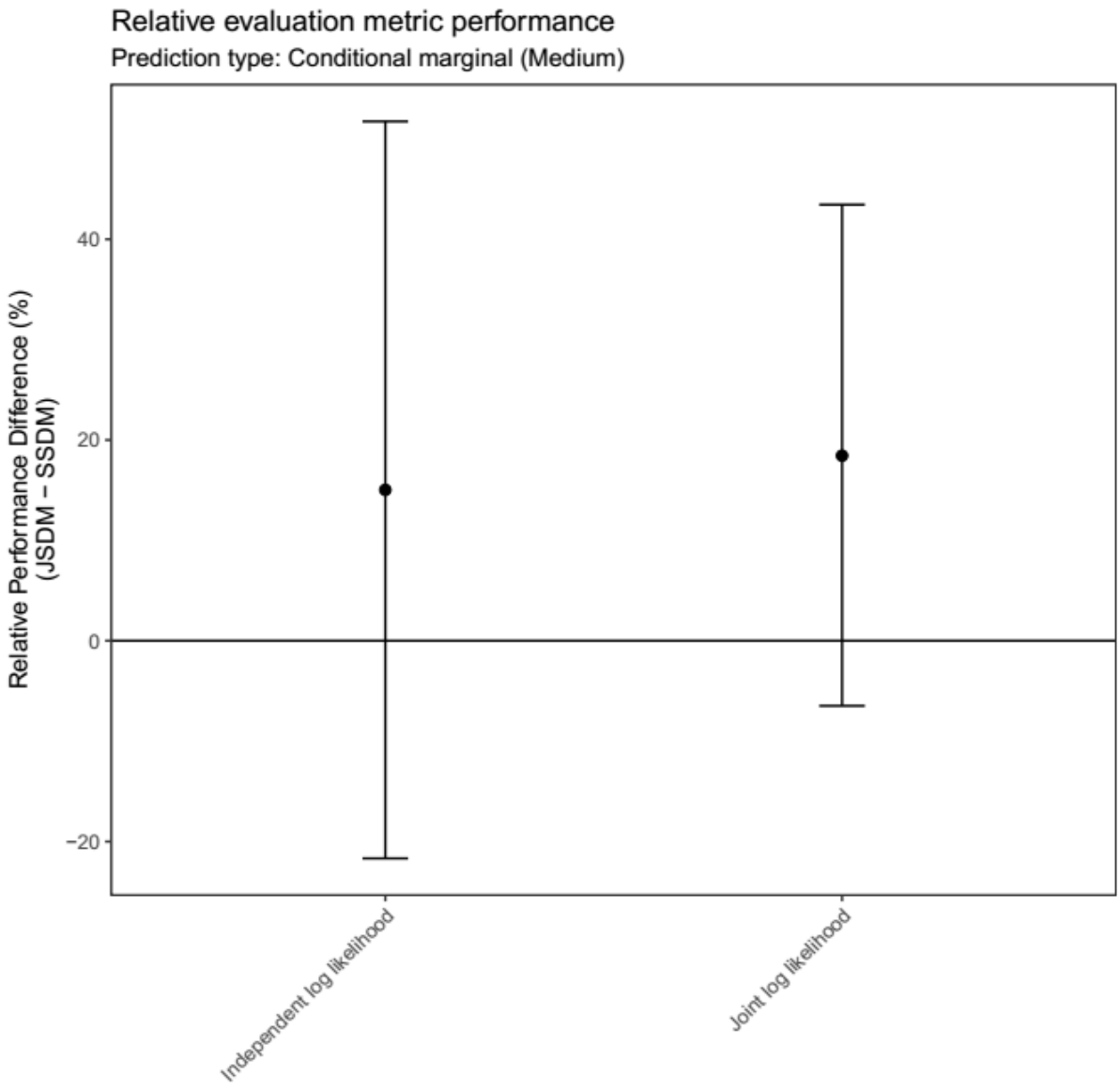


Figure 41: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where higher values indicate better performance, for conditional marginal (medium-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

9. Conditional marginal (high-prevalence) predictions

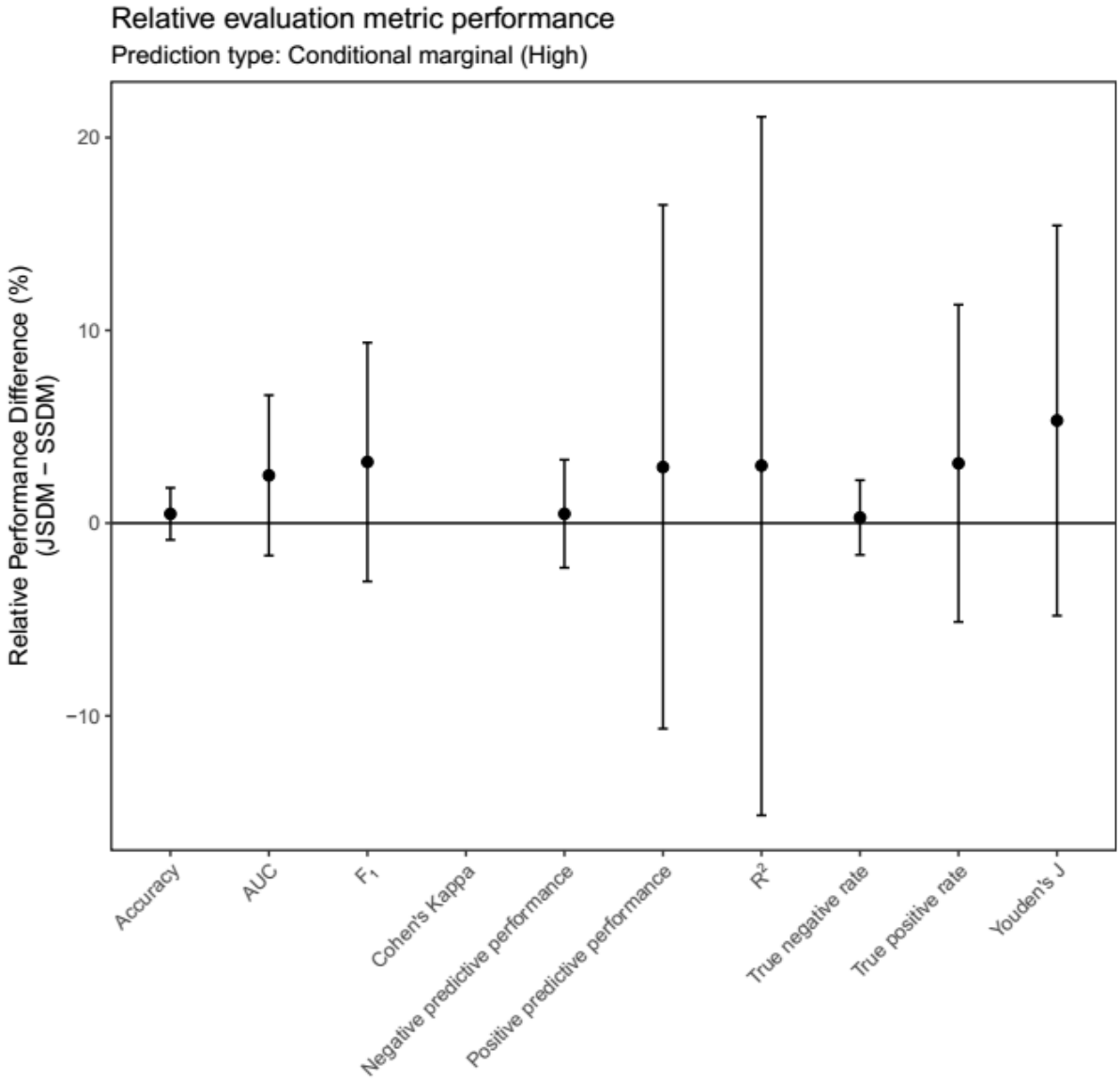


Figure 42: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where higher values indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

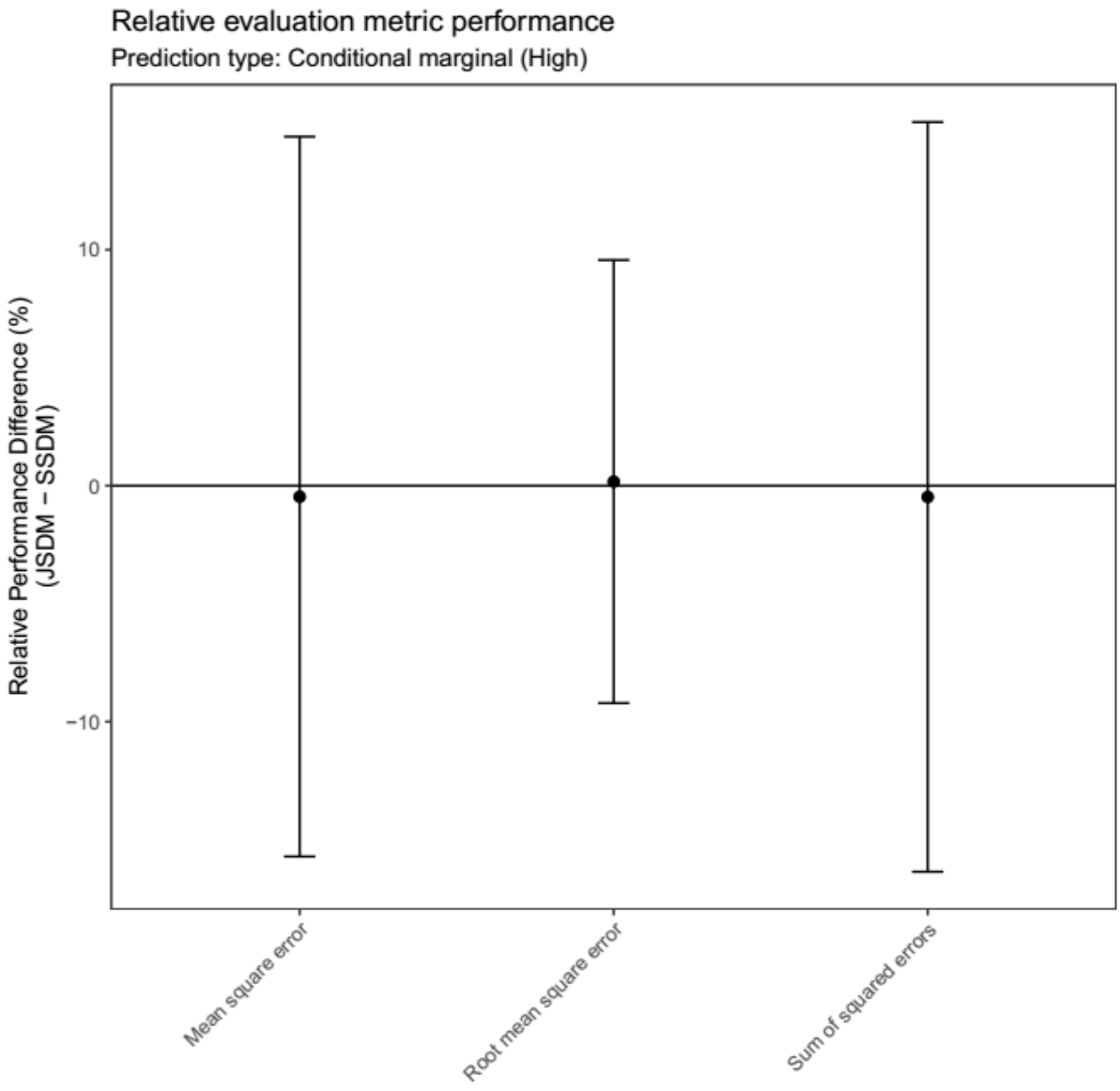


Figure 43: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – ∞, where lower values indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

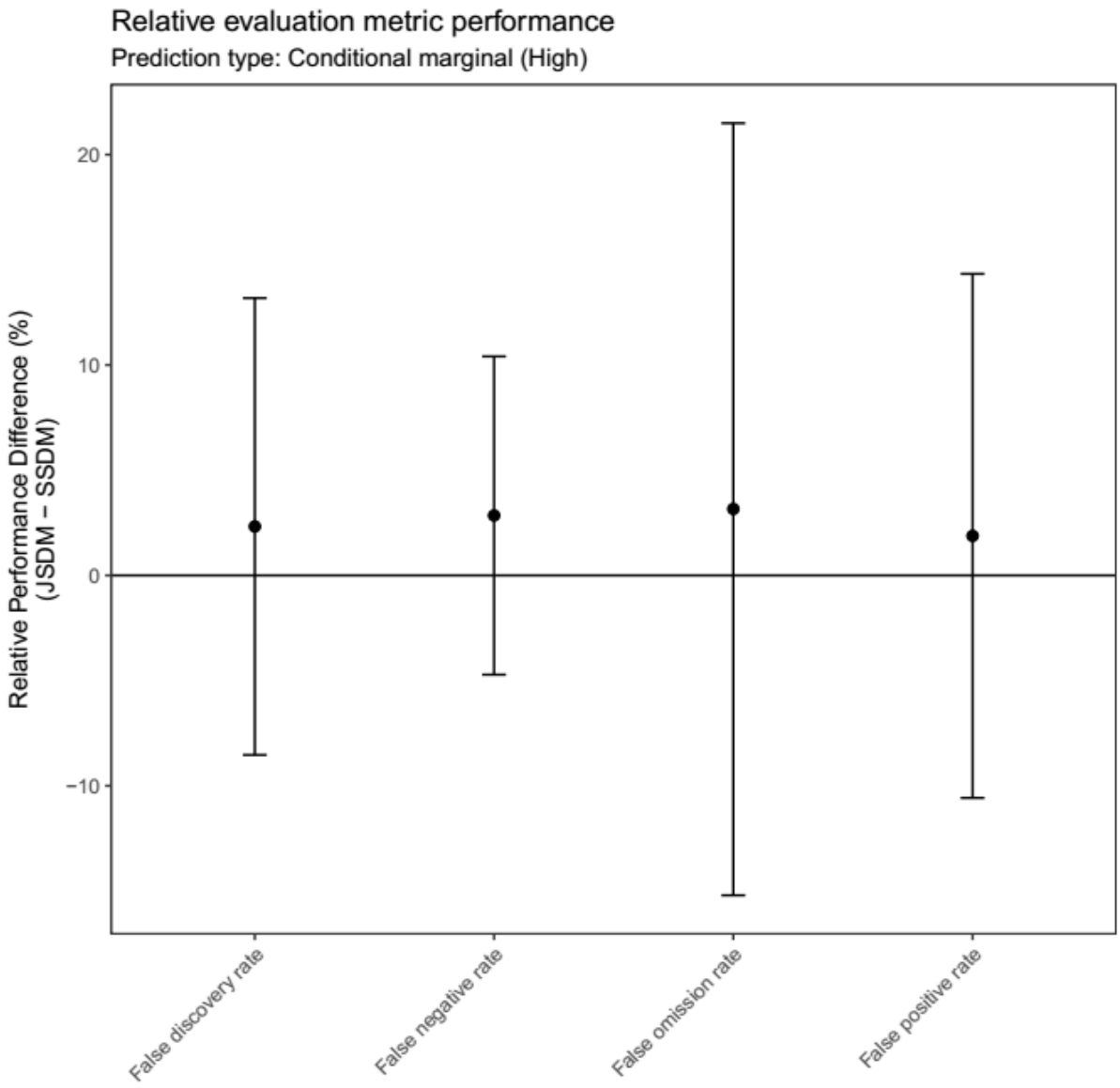


Figure 44: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

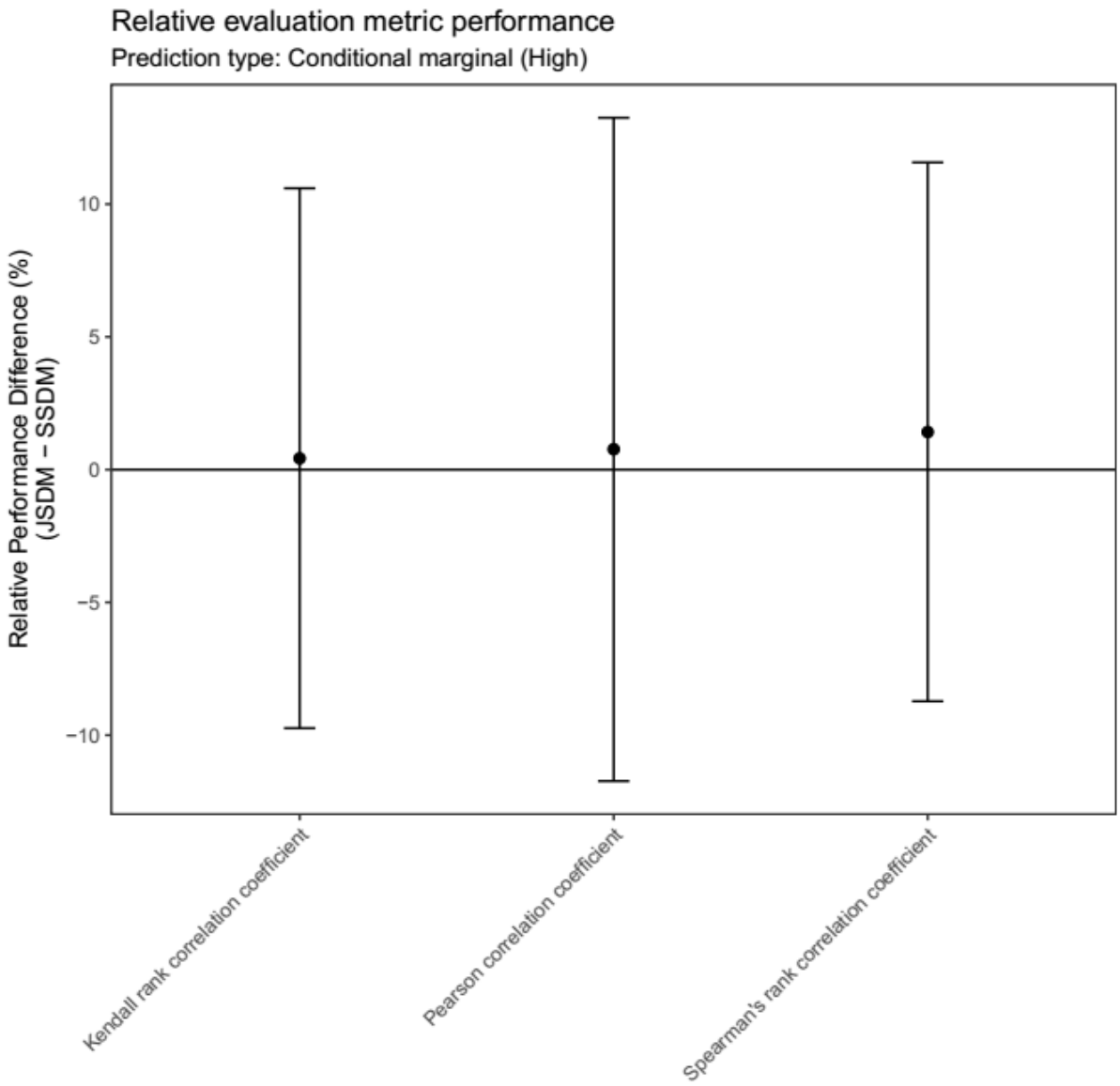


Figure 45: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -1 – 1, where higher values indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

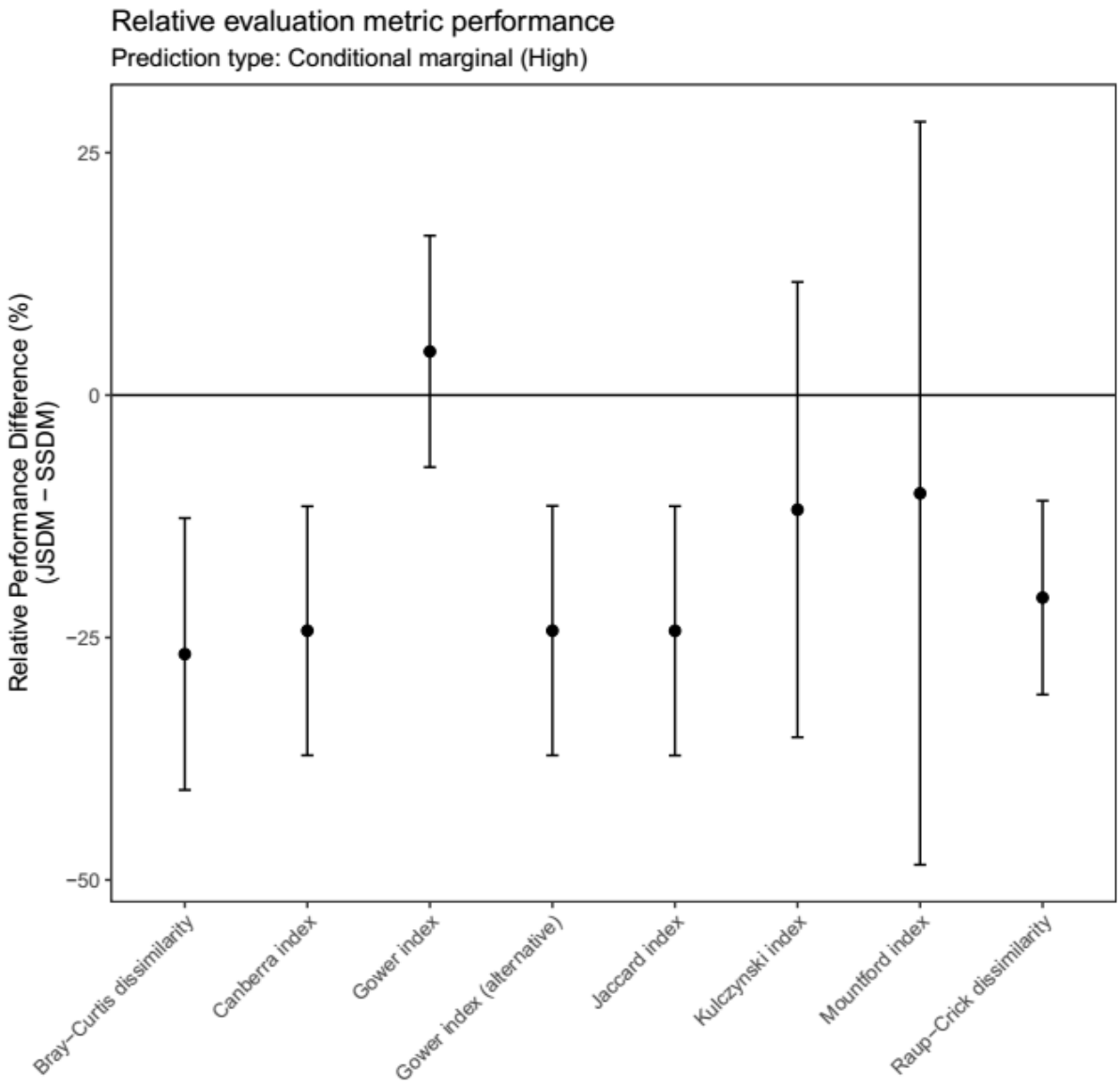


Figure 46: Relative performance of the JSDM compared to the SSDM for community dissimilarity metrics with an evaluation range of 0 – 1, where lower values indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

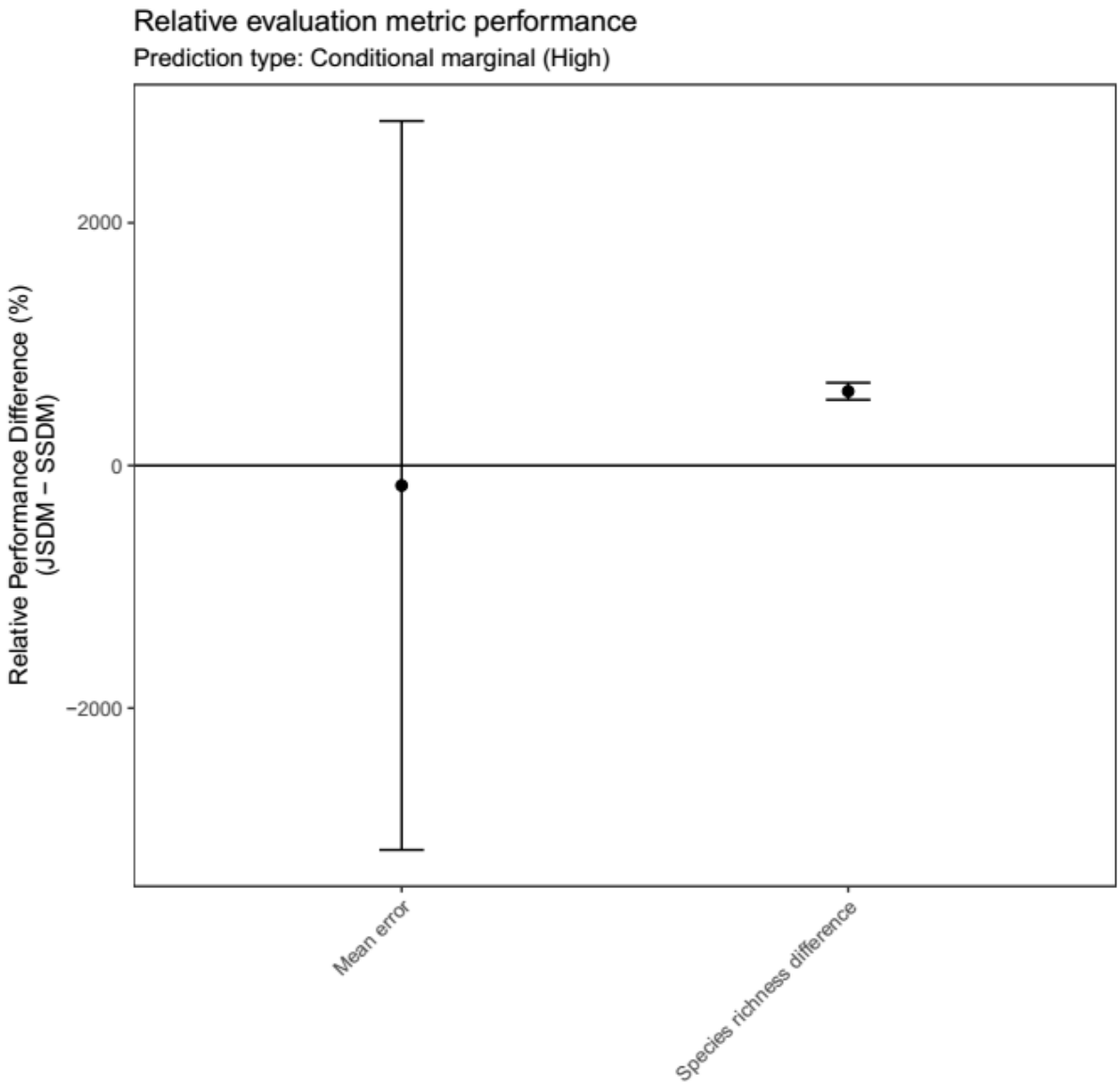


Figure 47: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where values closer to 0 indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

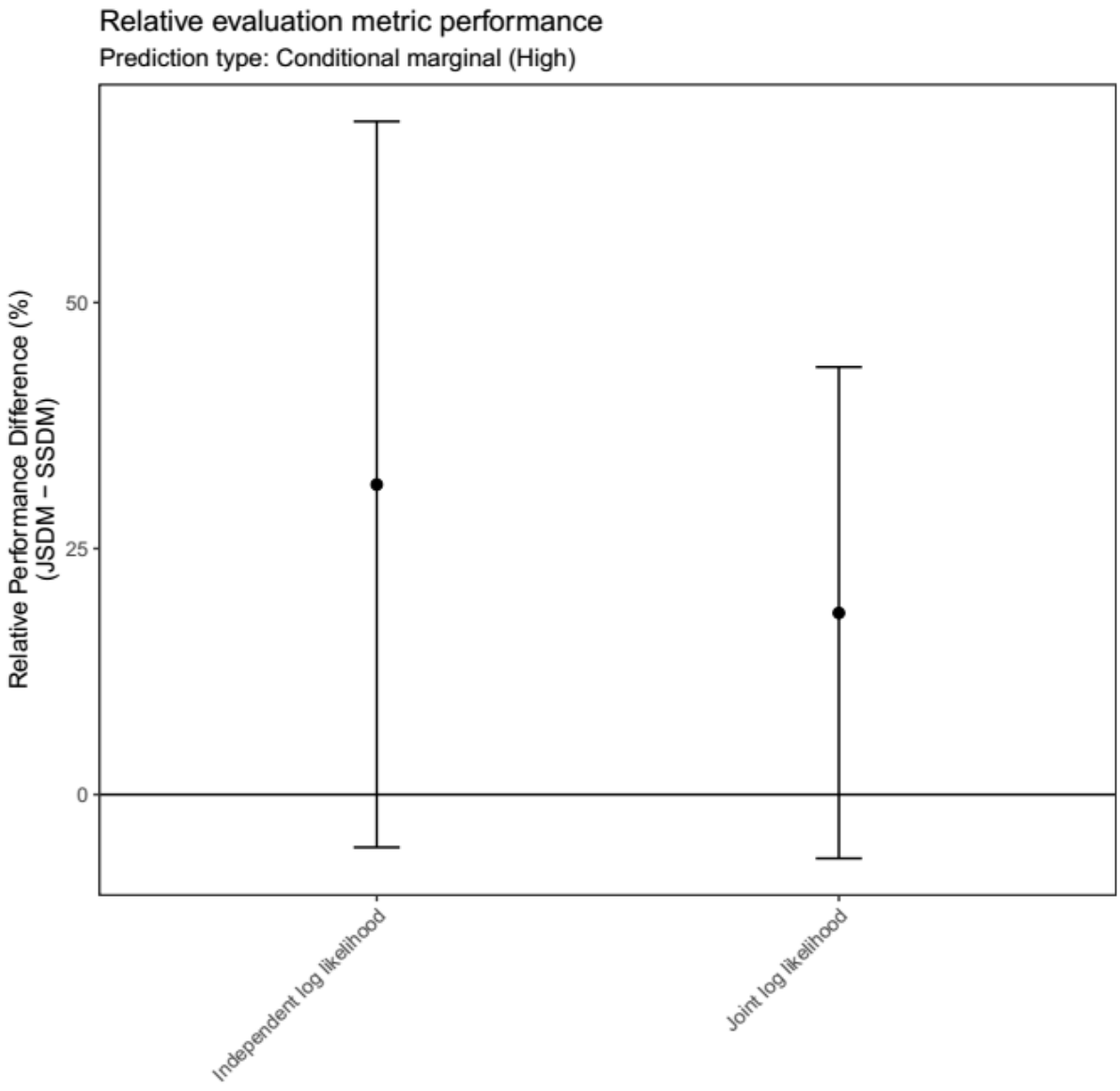


Figure 48: Relative performance of the JSDM compared to the SSDM for metrics with an evaluation range of -∞ – ∞, where higher values indicate better performance, for conditional marginal (high-prevalence) predictions. Positive relative performance difference indicates the JSDM outperforms the SSDM, negative relative performance difference indicates the SSDM outperformed the JSDM. Differences are represented by the mean and 95% confidence intervals, and calculated as (JSDM – SSDM) / |SSDM| * 100 using the back-transformed mean and standard error estimates from the mixed effects models.

APPENDIX J

MODEL DEFINITIONS

In this appendix we describe the model structure, default priors, and Markov chain Monte Carlo (MCMC) regimes of the JSDMs used in the main article. All JSDMs have been fit using the default or suggested settings as outlined in their source articles or, if not mentioned, their R package documentation.

All of the considered JSDMs are extensions of the generalised linear modelling (GLM) framework, and more specifically built on the foundation of Chib and Greenberg's (1998) multivariate probit regression model (hereafter, the core model). We first describe this core model and then for each JSDM define how it has modified or extended it. The statistical models are defined below using a common notation. The following terms are consistent across all models: $\boldsymbol{y}$, the response variable; $1(.)$, an indicator function that returns 1 when the expression in brackets is true and 0 otherwise; $\boldsymbol{z}$, a normally-distributed latent variable; $\boldsymbol{\mu}$, the linear predictor for the measured covariates; $\mathbf{X}$, the matrix of measured covariates; $\boldsymbol{\beta}$, the matrix of regression coefficients; and $\mathbf{I}$, the identity matrix. Subscript notation for sites is $i = 1, \ldots, n$; for species $j = 1, \ldots, J$; and for predictors $k = 1, \ldots, K$.

The core model uses a latent variable parameterisation of a probit model rather than the probit link directly. The probability of species presence is modelled as the probability of a latent multivariate normally-distributed variable exceeding a threshold, such that $y_{ij} = 1$ if $z_{ij} > 0$, and $y_{ij} = 0$ otherwise. The probability of species presence at a site is represented by a one-dimensional latent variable, while the community present at a site is represented by a multi-dimensional latent variable (see Chapter Four figures). This core model is described as follows:

$$y_{ij} = 1(z_{ij} > 0)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\boldsymbol{e}_i \sim MVN(\mathbf{0}, \mathbf{R})$$

The probability of species j being present at site i is the probability that latent variable z_{ij} is greater than zero. The latent variable is the sum of the linear predictor μ_{ij} and the correlated residual error e_{ij}. The linear predictor is the product of the measured environmental variables $\boldsymbol{X}_{i,.}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{.,j}$, as in generalised linear models. Correlations in the residual error $\boldsymbol{e}_i$ are captured in **R**, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements are restricted between -1 and 1. Standard deviations are constrained to equal to 1 in probit regression, thus covariance and correlation matrices are equivalent. The elements of **R** reflect species co-occurrence patterns not described by the environmental predictors (i.e. species interactions, or missing predictors).

(1) Multivariate Probit Regression (MPR)

The multivariate probit regression (MPR) model used by Golding *et al* (2015), *BayesComm* v0.1-2, is identical to the core model. The regression coefficients have a normal prior, $\beta \sim N(0, 10)$, and the correlation coefficients an inverse Wishart prior with $n + 2J$ degrees of freedom and scale matrix **I**. MPR was fit in R (R Core Team 2018) by MCMC using a Gibbs sampler implemented in R and C++. A single MCMC chain of 11,000 samples, discarding the first 1,000 as burn-in, sampled the posterior distribution.

(2) Hierarchical Multivariate Probit Regression (HPR)

The hierarchical multivariate probit regression model of Pollock *et al* (2014) (hereby HPR) models the regression coefficients hierarchically such that β_{jk} is drawn from a normal distribution with mean ω_k and standard deviation σ_k.

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + e_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\beta_{jk} \sim N(\omega_k, \sigma_k)$$

$$e_i \sim \boldsymbol{MVN}(\mathbf{0}, \mathbf{R})$$

Regression coefficients have a normal prior, $N(0, 100)$, on ω and a uniform prior, $U(0{,}100)$, on σ. Correlation coefficients have an inverse Wishart prior with $J + 1$ degrees of freedom and an $\mathbf{I}$ scale matrix. HPR was fit in R, and the posterior distributions sampled via MCMC using Gibbs sampling in JAGS (Plummer 2003), with three chains of 1,000,000 samples (thinned to keep 1 every 1,000) with the first 15,000 discarded as burn in.

(3) Multivariate Probit Regression with Latent Factors (LPR)

The *boral* JSDM (Hui 2016) is a multivariate probit regression model with latent factors (hereafter LPR).

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$$

$$\varepsilon_i \sim N(\mathbf{0}, \mathbf{I})$$

This differs from the core model by using latent factors ("hypothetical" unmeasured variables) to explain any residual covariation that is not accounted for by the measured variables. This explained variation is the product of H unmeasured latent factors, $\boldsymbol{\eta}_{i,.}$ $(h = 1, \ldots, H)$, and the factor loadings, $\boldsymbol{\lambda}_{.,j}$, of species j to latent factor h. We can interpret $\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$ and $\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$ as linear predictors for measured and unmeasured variables respectively. Now z_{ij} is the sum of the linear predictor of measured covariates, μ_{ij}, the linear predictor of unmeasured covariates, ν_{ij}, and uncorrelated residual error ε_{ij}. This model can be defined equivalently as an extension of MPR where the correlation matrix has low-rank structure, modelled as: $\mathbf{R} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{I}$ (Warton *et al.* 2015). Species interactions can be estimated by converting the factor loadings to correlation coefficients using this expression.

The regression coefficients in the LPR implementation have a normal prior, $\beta \sim N(0, 10)$, and the latent factors a normal prior, $\eta \sim N(0, 4.47)$, constrained such that upper diagonal elements are 0, and the diagonal elements are positive, $U(0, 20)$. The default model uses two latent factors. LPR was fit in R using the *boral* v1.7 package (Hui 2018). The posterior distribution was sampled via MCMC, with a Gibbs sampler using JAGS, in a single chain of 60,000 samples (thinned to keep 1 every 50) discarding the first 10,000 as burn in.

(4) Dimension Reduction Model (DPR)

The multivariate generalised regression model of Clark *et al* (2017), *gjam* (hereby DPR), fits various types of response data. For presence-absence data it is a multivariate probit regression model that takes on two different forms depending on the size of the dataset. The small dataset form is equivalent to the core model. For datasets above a size threshold ($J > 100$ or $J > \frac{2}{3} * n - 1$) , it is similar to LPR, but with an

additional dimension reduction step (with respect to species) using the Dirichlet process (Taylor-Rodríguez *et al.* 2017):

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,\cdot}\boldsymbol{\beta}_{\cdot,j}$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,\cdot}\boldsymbol{\lambda}_{\cdot,j}$$

$$\boldsymbol{\lambda}_{\cdot,j} = \mathbf{A}_{\boldsymbol{l}_j}$$

$$\varepsilon_i \sim N(\mathbf{0}, \mathbf{I})$$

The Dirichlet process assumes that groups of species will have the same response to the unmeasured variables, so *a priori* reduces $\boldsymbol{\lambda}$ along the j-dimension to a matrix, $\mathbf{A}$, for a fixed number, L, of species archetypes (rather than J species). The Dirichlet process can be seen as a distribution of distributions, and is used to cluster species into archetypes, via an index variable $\boldsymbol{l}$. Using $\mathbf{A}$ as a look-up table for the low-rank $\boldsymbol{\lambda}$ allows conversion back to full-rank $\boldsymbol{\lambda}$, and then estimate correlation coefficients via the factor loadings.

The regression coefficients in the DPR implementation have an improper uniform prior, $U(-Inf, Inf)$, correlation coefficients use an inverse-Wishart prior with $n - K + J - 1$ degrees of freedom and an $\mathbf{I}$ scale matrix, and latent factors are normally distributed as $\eta_{ih} \sim N(0,1)$. DPR was fit in R using the *gjam* package v2.2.5 (Clark & Taylor-Rodríguez 2018). The posterior distribution was sampled via MCMC using a Gibbs sampler in R of a single chain of 60,000 samples (thinned to keep 1 every 50), discarding the first 10,000 as burn in.

(5) Hierarchical Multivariate Probit Regression with Latent Factors (HLR-S and HLR-NS)

The last two JSDMs are spatially-explicit (HLR-S) and non-spatially-explicit (HLR-NS) versions of the hierarchical multivariate probit regression with latent factors used by Ovaskainen *et al*. (2016b):

$$y_{ij} = 1\left(z_{ij} > 0\right)$$

$$z_{ij} = \mu_{ij} + \nu_{ij} + \varepsilon_{ij}$$

$$\mu_{ij} = \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j}$$

$$\beta_j \sim MVN\left(\boldsymbol{\omega}_{\boldsymbol{l}_j}, \boldsymbol{\sigma}\right)$$

$$\nu_{ij} = \boldsymbol{\eta}_{i,.}\boldsymbol{\lambda}_{.,j}$$

$$\varepsilon_i \sim N(0,1)$$

The HLR models make use of latent factors like LPR and DPR, but also use hierarchical regression coefficients like HPR. However, instead of independent normal distributions unique to the K covariates, $\beta_{.,j}$ is drawn from a multivariate normal distribution, allowing for correlation among coefficients. The mean of this distribution, $\boldsymbol{\omega}_{\boldsymbol{l}_j}$, is unique to species' archetype, $\boldsymbol{l}$. Archetype is specified *a priori* in HLR models using trait data, assuming that species with similar traits exhibit similar responses. Here, $\boldsymbol{l}_j$ is set to 1 as trait data was unavailable for most datasets. Thus, each species' $\beta_{.,j}$ are independently drawn from a shared distribution. The hierarchical regression coefficients have a normal prior, $N(0,1)$ for $\boldsymbol{\omega}$, and an inverse Wishart prior with five degrees of freedom and a **I** scale matrix for $\boldsymbol{\sigma}$. The latent factors are normally distributed as

$\eta_{ih} \sim N(0,1)$ in HLR-NS, and as independent spatially homogenous Gaussian processes with exponential covariance in HLR-S:

$$Cov(\eta_{ih}, \eta_{i'h}) = exp(-d_{ii'}/\alpha_h)$$

Here, d is the spatial distance between sampling units, and α is a positive parameter controlling decay in correlation with distance. At zero distance the covariance function is 1, so the latent factors have zero mean and unit variance like in the non-spatial model, and this declines towards zero with increasing α. The model uses a multiplicative gamma process shrinkage prior on the factor loadings to estimate an appropriate number of latent factors during the burn-in stage of MCMC before fixing that number for the remaining samples (Bhattacharya & Dunson 2011; Ovaskainen *et al.* 2016b).

These JSDMs were fit in R using the *HMSC* package v2.2-0 (Blanchet, Tikhonov & Norberg 2019). The posterior distribution was sampled via MCMC using a Gibbs sampler in a single chain of 50,000 samples (thinned to keep 1 every 50) with the first 10,000 discarded as burn in.

REFERENCE LIST

Araújo, M.B. & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, **16**, 743–753.

Barner, A.K., Coblentz, K.E., Hacker, S.D. & Menge, B.A. (2018). Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, **99**, 557–566.

Beissinger, S.R., Iknayan, K.J., Guillera-Arroita, G., Zipkin, E.F., Dorazio, R.M., Royle, J.A. & Kéry, M. (2016). Incorporating Imperfect Detection into Joint Models of Communities: A response to Warton et al. *Trends in Ecology and Evolution*, **31**, 736–737.

Bhattacharya, A. & Dunson, D.B. (2011). Sparse Bayesian infinite factor models. *Biometrika*, **98**, 291–306.

Blanchet, F.G., Tikhonov, G. & Norberg, A. (2019). HMSC: Hierarchical Modelling of Species Community.

Calabrese, J.M., Certain, G., Kraan, C. & Dormann, C.F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, **23**, 99–112.

Chib, S. & Greenberg, E. (1998). Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.

Clark, J.S., Gelfand, A.E., Woodall, C.W. & Zhu, K. (2014a). More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, **24**, 990–999.

Clark, J.S., Gelfand, A.E., Woodall, C.W. & Zhu, K. (2014b). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, **24**, 990–999.

Clark, J.S., Nemergut, D., Seyednasrollah, B., Turner, P.J. & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: median-zero, multivariate, multifarious data. *Ecological Monographs*, **87**, 34–56.

Clark, J.S. & Taylor-Rodríguez, D. (2018). gjam: Generalised Joint Attribute Modelling.

Cornell, H. V. & Harrison, S.P. (2014). What Are Species Pools and When Are They Important? *Annual Review of Ecology, Evolution, and Systematics*, **45**, 45–67.

Dorazio, R.M., Connor, E.F. & Askins, R.A. (2015). Estimating the effects of habitat and biological interactions in an avian community. *PLoS ONE*, **10**, 1–16.

Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., Moretti, M.D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S.I., Sheppard, C.S., Steinbauer, M.J., Zeuss, D. & Kraan, C. (2018a). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*.

Dormann, C.F., Calabrese, J.M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C.M., Ciuti, S., Elith, J., Gerstner, K., Guelat, J., Keil, P., Lahoz-Monfort, J.J., Pollock, L.J., Reineking, B., Roberts, D.R., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Wood, S.N., Wüest, R.O. & Hartig, F. (2018b). Model averaging in ecology: a review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, **88**, 485–504.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F.,

Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.
Dunn, O.J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, **56**, 52–64.
Edwards, A.M. & Auger-Méthé, M. (2019). Some guidance on using mathematical notation in ecology (S. McMahon, Ed.). *Methods in Ecology and Evolution*, **10**, 92–99.
Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006a). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. & E. Zimmermann, N. (2006b). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
Elith, J. & Leathwick, J.J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, …*, **40**, 677–697.
Elith, J., Leathwick, J.R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, **77**, 802–813.
Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E. & Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, **17**, 43–57.
Fielding, A.H. & Bell, J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
Freeman, E.A. & Moisen, G.G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling*, **217**, 48–58.
Gelfand, A.E., Schmidt, A.M., Wu, S., Silander, J.A., Latimer, A. & Rebelo, A.G. (2005). Modelling species diversity through species level hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **54**, 1–20.
Gelfand, A.E., Silander, J.A., Wu, S., Latimer, A., Lewis, P.O., Rebelo, A.G. & Holder, M. (2006). Explaining species distribution patterns through hierarchical modeling. *Bayesian Analysis*, **1**, 41–92.
Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F. & Hothorn, T. (2019). mvtnorm: Multivariate Normal and t Distributions.
Golding, N. (2018). greta: Simple and Scalable Statistical Modelling in R.
Golding, N. & Harris, D.J. (2015). BayesComm: Bayesian Community Ecology Analysis.
Golding, N., Nunn, M.A. & Purse, B. V. (2015). Identifying biotic interactions which drive the spatial distribution of a mosquito community. *Parasites & Vectors*, **8**, 367.

Golding, N. & Purse, B. V. (2016). Fast and flexible Bayesian species distribution modelling using Gaussian processes. *Methods in Ecology and Evolution*, **7**, 598–608.
Götzenberger, L., de Bello, F., Bråthen, K.A., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. & Zobel, M. (2012). Ecological assembly rules in plant communities-approaches, patterns and prospects. *Biological Reviews*, **87**, 111–127.
Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, **40**, 281–295.
Guillera-Arroita, G., Kéry, M. & Lahoz-Monfort, J.J. (2019). Inferring species richness using multispecies occupancy modeling: Estimation performance and interpretation. *Ecology and Evolution*, **9**, 780–792.
Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Mccarthy, M.A., Tingley, R. & Wintle, B.A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, **24**, 276–292.
Guisan, A. & Rahbek, C. (2011). SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, **38**, 1433–1444.
Hamner, B. & Frasco, M. (2018). Metrics: Evaluation Metrics for Machine Learning.
Hanberry, B.B. & He, H.S. (2013). Prevalence, statistical thresholds, and accuracy assessment for species distribution models. *Web Ecol*, **13**, 13–19.
Hardy, O.J. (2008). Testing the spatial phylogenetic structure of local communities: Statistical performances of different null models and test statistics on a locally neutral community. *Journal of Ecology*, **96**, 914–926.
Harris, D.J. (2015). Generating realistic assemblages with a joint species distribution model (D. Warton, Ed.). *Methods in Ecology and Evolution*, **6**, 465–473.
Harris, D.J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, **97**, 3308–3314.
Hui, F.K.C. (2017). boral: Bayesian Ordination and Regression AnaLysis.
Hui, F.K.C. (2018). boral: Bayesian Ordination and Regression AnaLysis.
Hui, F.K.C. (2016). boral - Bayesian Ordination and Regression Analysis of Multivariate Abundance Data in r. *Methods in Ecology and Evolution*, **7**, 744–750.
Hutchinson, G.E. (1957). Concluding Remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415–427.
Johnson, D.S. & Sinclair, E.H. (2017). Modeling joint abundance of multiple species using Dirichlet process mixtures. *Environmetrics*, **28**, 1–13.
Kissling, W.D., Dormann, C.F., Groeneveld, J., Hickler, T., Kühn, I., McInerny, G.J., Montoya, J.M., Römermann, C., Schiffers, K., Schurr, F.M., Singer, A., Svenning, J.-C., Zimmermann, N.E. & O'Hara, R.B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, **39**, 2163–2178.
Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. & Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*, **19**, 1639–1645.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, L., Candan, C. & Hunt, T. (2019). caret: Classification and Regression Training.

Lahoz-Monfort, J.J., Guillera-Arroita, G. & Wintle, B.A. (2014). Imperfect detection impacts the performance of species distribution models. *Global Ecology and Biogeography*, **23**, 504–515.
Lawson, C.R., Hodgson, J.A., Wilson, R.J. & Richards, S.A. (2014). Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution*, **5**, 54–64.
Leathwick, J.R. & Austin, M.P. (2001). Competitive interactions between tree species in New Zealand old-growth indigenous forests. *Ecology*, **82**, 2560–2573.
Legendre, P. & De Cáceres, M. (2013). Beta diversity as the variance of community data: Dissimilarity coefficients and partitioning. *Ecology Letters*, **16**, 951–963.
Letten, A.D., Keith, D.A., Tozer, M.G. & Hui, F.K.C. (2015). Fine-scale hydrological niche differentiation through the lens of multi-species co-occurrence models. *Journal of Ecology*, **103**, 1264–1275.
Liu, C., Berry, P.M., Dawson, T.P. & Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, **28**, 385–393.
Liu, C., White, M. & Newell, G. (2009). Measuring the accuracy of species distribution models: A review. *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation: Interfacing Modelling and Simulation with Mathematical and Computational Sciences, Proceedings*, 4241–4247.
Lowry, W.P. (1959). On the Use of Mathematical Notation in Ecological Literature. *Ecology*, **40**, 492–492.
MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.
Massey Jr, F.J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, **46**, 68–78.
Mateo, R.G., Felicísimo, Á.M., Pottier, J., Guisan, A. & Muñoz, J. (2012). Do stacked species distribution models reflect altitudinal diversity patterns? *PLoS ONE*, **7**.
Meade, B., Lafayette, L., Sauter, G. & Tosello, D. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility.
Meier, E.S., Kienast, F., Pearman, P.B., Svenning, J.-C., Thuiller, W., Araújo, M.B., Guisan, A. & Zimmermann, N.E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, **33**, 1038–1048.
Nieto-Lugilde, D., Maguire, K.C., Blois, J.L., Williams, J.W. & Fitzpatrick, M.C. (2017). Multiresponse algorithms for community-level modeling: review of theory, applications, and comparison to species distribution models. *Methods in Ecology and Evolution*, **12**, 3218–3221.
Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., Araújo, M.B., Dallas, T., Dunson, D., Elith, J., Foster, S.D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N.A., Holt, R.D., Hui, F.K.C., Husby, M., Kålås, J.A., Lehikoinen, A., Luoto, M., Mod, H.K., Newell, G., Renner, I., Roslin, T., Soininen, J., Thuiller, W., Vanhatalo, J., Warton, D., White, M., Zimmermann, N.E., Gravel, D. & Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, **89**, 1–24.
O'Brien, S.M. & Dunson, D.B. (2004). Bayesian multivariate logistic regression. *Biometrics*, **60**, 739–746.
Oksanan, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. & Wagner, H. (2018). vegan: Community Ecology Package.

Ovaskainen, O., Abrego, N., Halme, P. & Dunson, D. (2016a). Using latent variable models to identify large networks of species-to-species associations at different spatial scales. *Methods in Ecology and Evolution*, **7**, 549–555.
Ovaskainen, O., Hottola, J. & Siitonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514–2521.
Ovaskainen, O., Roy, D.B., Fox, R. & Anderson, B.J. (2016b). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, **7**, 428–436.
Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T. & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software (J. Chave, Ed.). *Ecology Letters*, **20**, 561–576.
Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R.K. (2009). Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation*, **142**, 2501–2509.
Pellissier, L., Anne Bråthen, K., Pottier, J., Randin, C.F., Vittoz, P., Dubuis, A., Yoccoz, N.G., Alm, T., Zimmermann, N.E. & Guisan, A. (2010). Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography*, **33**, 1004–1014.
Pineda, E. & Lobo, J.M. (2009). Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology*, **78**, 182–190.
Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.
Pollock, L.J. (2014). Data from: Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Dryad Digital Repository*.
Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., Vesk, P.A. & McCarthy, M.A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM) (J. McPherson, Ed.). *Methods in Ecology and Evolution*, **5**, 397–406.
R Core Team. (2016). R: A Language and Environment for Statistical Computing.
R Core Team. (2018). R: A Language and Environment for Statistical Computing.
Rainey, C. (2016). Dealing with separation in logistic regression models. *Political Analysis*, **24**, 339–355.
Rasmussen, C.E. & Williams, C.K.I. (2006). *Gaussian Processes for Machine Learning*, Secondn. The MIT Press, Cambridge, Massachusetts.
Revelle, W. (2018). psych: Procedures for Personality and Psychological Research.
Richmond, O.M.W., Hines, J.E. & Beissinger, S.R. (2010). Two-species occupancy models: a new paramaterization applied to co-occurence of secretive rails. *Ecological Applications*, **20**, 2036–2046.
Rota, C.T., Ferreira, M.A.R., Kays, R.W., Forrester, T.D., Kalies, E.L., McShea, W.J., Parsons, A.W. & Millspaugh, J.J. (2016). A multispecies occupancy model for two or more interacting species (D. Warton, Ed.). *Methods in Ecology and Evolution*, **7**, 1164–1173.
Schweiger, O., Heikkinen, R.K., Harpke, A., Hickler, T., Klotz, S., Kudrna, O., Kühn, I., Pöyry, J. & Settele, J. (2012). Increasing range mismatching of interacting species under global change is related to their ecological characteristics. *Global*

Ecology and Biogeography, **21**, 88–99.
Stjernman, M., Sahlin, U., Olsson, O. & Smith, H.G. (2019). Estimating effects of arable land use intensity on farmland birds using joint species modeling. *Ecological Applications*, **29**, e01875.
Taylor-Rodríguez, D., Kaufeld, K., Schliep, E.M., Clark, J.S. & Gelfand, A.E. (2017). Joint Species Distribution Modeling: Dimension Reduction Using Dirichlet Processes. *Bayesian Analysis*, **12**, 939–967.
The MathWorks Inc. (2016). MATLAB.
Thorson, J.T., Ianelli, J.N., Larsen, E.A., Ries, L., Scheuerell, M.D., Szuwalski, C. & Zipkin, E.F. (2016). Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. *Global Ecology and Biogeography*, **25**, 1144–1158.
Thorson, J.T., Scheuerell, M.D., Shelton, A.O., See, K.E., Skaug, H.J. & Kristensen, K. (2015). Spatial factor analysis: A new tool for estimating joint species distributions and correlations in species range. *Methods in Ecology and Evolution*, **6**, 627–637.
Thuiller, W., Pollock, L.J., Gueguen, M. & Münkemüller, T. (2015). From species distributions to meta-communities. *Ecology Letters*, **18**, 1321–1328.
Tikhonov, G., Abrego, N., Dunson, D. & Ovaskainen, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context (D. Warton, Ed.). *Methods in Ecology and Evolution*, **8**, 443–452.
Tobler, M.W., Kéry, M., Hui, F.K.C., Guillera-Arroita, G., Knaus, P. & Sattler, T. (2019). Joint species distribution models with species correlations and imperfect detection. *Ecology*, **100**, e02754.
University of Melbourne. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility.
Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, **10**, 225–232.
Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2018). blockCV: Spatial and environmental blocking for k-fold cross-validation.
Waddle, J.H., Dorazio, R.M., Walls, S.C., Rice, K.G., Beauchamp, J., Schuman, M.J. & Mazzotti, F.J. (2010). A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, **20**, 1467–1475.
Warton, D.I., Blanchet, F.G., O'Hara, R., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2016). Extending Joint Models in Community Ecology: A Response to Beissinger et al. *Trends in Ecology and Evolution*, **31**, 737–738.
Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C. & Hui, F.K.C. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology and Evolution*, **30**, 766–779.
Wilhelm, S. & B G, M. (2015). tmvtnorm: Truncated Multivariate Normal and Student t Distribution.
Wilkinson, D.P. (2018). JSDM_Inference v0.1.6. Zenodo, https://dx.doi.org/10.5281/zenodo.1452066.
Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). A comparison of joint species distribution models for presence-absence data. *Methods in Ecology and Evolution*, **10**, 198–211.
Wisz, M.S., Pottier, J., Kissling, W.D., Pellissier, L., Lenoir, J., Damgaard, C.F., Dormann, C.F., Forchhammer, M.C., Grytnes, J.A., Guisan, A., Heikkinen, R.K.,

Høye, T.T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M.C., Normand, S., Öckinger, E., Schmidt, N.M., Termansen, M., Timmermann, A., Wardle, D.A., Aastrup, P. & Svenning, J.C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Wolfram Research Inc. (2016). Mathematica.

Zdravko, B.I. (2015). TruncatedNormal: Truncated Multivariate Normal.

Zhang, C., Chen, Y., Xu, B., Xue, Y. & Ren, Y. (2018). Comparing the prediction of joint species distribution models with respect to characteristics of sampling data. *Ecography*, **41**, 1876–1887.

Zurell, D., Pollock, L.J. & Thuiller, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co- occurrence data in homogenous environments? *Ecography*, 0–2.

Zurell, D., Zimmermann, N.E., Gross, H., Baltensweiler, A., Sattler, T. & Wüest, R.O. (2019). Testing species assemblage predictions from stacked and joint species distribution models. *Journal of Biogeography*, **00**, 1–13.

APPENDIX K

PREDICTION DEFINITIONS

In this appendix we describe the different prediction methods available to JSDMs. This material is reproduced in condensed form from Chapter Three.

Most occupancy-based JSDMs are built on the foundation of the Chib and Greenberg (1998) multivariate probit regression model that uses the latent variable parameterisation of a probit model, instead of using the probit link directly. We focus on this multivariate probit model here for simplicity, though the same approach can be applied to the multivariate logistic, and latent factor models that have also been used as JSDMs. For these models, the occurrence state, y, of a species, j, at a site, i, is modelled via a normally-distributed latent variable, z_{ij}. The latent occurrence state of a species (present or absent), y_{ij}, is equal to 1 when $z_{ij} > 0$, and 0 otherwise (see Figure S1 for a visual representation). This core model is as follows:

$$
\begin{aligned}
y_{ij} &= 1\left(z_{ij} > 0\right) \\
z_{ij} &= \mu_{ij} + e_{ij} \\
\mu_{ij} &= \boldsymbol{X}_{i,.}\boldsymbol{\beta}_{.,j} \\
\boldsymbol{e}_i &\sim MVN(\mathbf{0}, \mathbf{R})
\end{aligned}
\tag{1}
$$

where the latent variable, z_{ij}, is the sum of the linear predictor, μ_{ij}, and the correlated residual error, e_{ij}. The linear predictor is the product of the measured environmental variables $\boldsymbol{X}_{i,.}$, and their corresponding regression coefficients $\boldsymbol{\beta}_{.,j}$, as in standard generalised linear models. Correlations in the residual error $\boldsymbol{e}_i$ are captured in $\mathbf{R}$, a symmetric and positive-definite matrix; its diagonal elements are 1 and its off-diagonal elements – the residual correlations between species – are restricted between -1 and 1. The elements of $\mathbf{R}$ reflect species co-occurrence patterns not described by the environmental predictors (i.e. species interactions, or missing predictors). Standard

deviations, and in turn variances, are constrained to equal 1 in probit regression, thus covariance and correlation matrices are equivalent. Because the variance of the latent variable distribution remains constant, the probability of presence for a single species is controlled only by the mean value of the distribution, μ (compare Figures S1a and S1b).

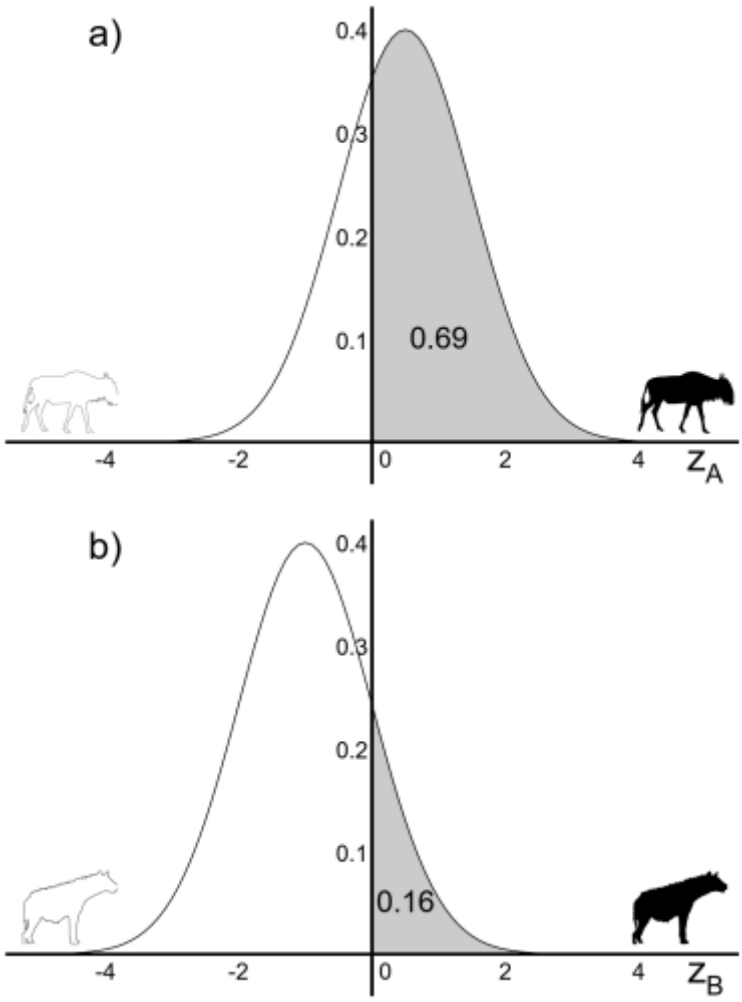


Figure S1: Visualisations of the normally-distributed latent variable, z_i, for two species (A and B). The mean of the distribution, μ_i, is 0.5 for species A and -1 for species B. The probability of occurrence of a species is equal to the area under the curve where $z_i > 0$, shown here in grey.

Since JSDMs simultaneously consider multiple species, the space of these latent variables is multivariate. We can consider each species as being represented by its own latent variable in a single dimension, and the whole species community at a site is represented by a multi-dimensional latent variable with as many dimensions as species. The JSDM therefore considers a joint probability distribution over these latent variables. With the latent variable parameterisation, the discrete response variables, e.g. $\Pr(A)$, correspond to the area under the curve of continuous response variables above or below a threshold. Thus, a species is present with probability $\Pr(A) = \Pr(z_A > 0)$ and absent with probability $\Pr(A') = \Pr(z_A \leq 0)$. This probability can be computed as the integral of the probability density function $f()$ of the latent variable z_A over all values greater than zero:$\Pr(A) = \Pr(z_A > 0) = \int_0^\infty f(z_A) dz_A$. In a two-species scenario (Figure S2), the joint probability $\Pr(A, B)$ that both species A and B are present can be calculated as:

$$\Pr(A, B) = \Pr(z_A > 0, z_B > 0) \quad (2)$$

$$\Pr(A, B) = \int_0^\infty \int_0^\infty f(z_A|z_B) \cdot f(z_B) dz_A dz_B$$

$$= \int_0^\infty \int_0^\infty f(z_B|z_A) \cdot f(z_A) dz_B dz_A$$

That is, this joint probability $Pr(A, B)$ is the integral of the product of two density functions, in only the region where both z_A and z_B are greater than zero. In this two-species case this can be visualised as the volume under a two-dimensional surface (as in Figure S2). In Eq. 2, the conditional probability distribution $f(z_A|z_B)$ reflects probable values of z_A (and therefore $\Pr(A)$) *given* a fixed value of z_B. The marginal probability distribution $f(z_B)$ reflects probable values of z_B (and therefore $\Pr(B)$) *independent* of

the latent occurrence state of species A. The order of items in the joint probability is inconsequential, so species can be marginalised out in any order (lines 2 and 3 in equation 4 are equivalent). Since the total area (or volume) under a probability distribution is always 1, the integral under a part of that distribution yields the probability that the value of the variable falls in that part of the space. This integration can be performed for any number of species, to compute joint probabilities for a community of any size.

Integration over the relevant proportion of the multivariate latent variable space in a JSDM can therefore be used to compute the *joint* probability of observing a given configuration of presence/absence of all species at a site. These integrals can also be used to compute the *marginal* probability of species presence by averaging over (marginalising) presence of the other species. Multiple joint probabilities can be combined to compute the *conditional* probability of presence of a species given the presence (or absence) of another species, by restricting the space of the integration to those indicating presence (or absence) of the other species. These predictive probability distributions can be combined in different ways to define a number of prediction types for JSDMs. We consider five types of JSDM prediction, each mapping onto ecological questions. These are illustrated in Figure S3 and detailed below.

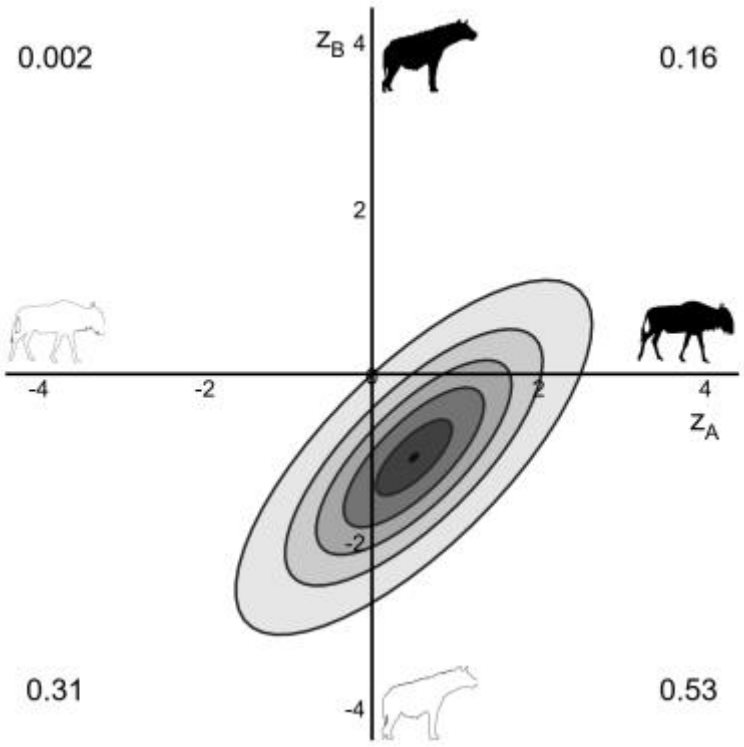


Figure S2: Visualisation of the multivariate, normally-distributed latent variable, z_{ij}, for a two species scenario. The mean of the distribution, μ_{ij}, on each species' respective axis is the same as their independent distributions from Figure 1 ($\mu_{iA} = 0.5$, $\mu_{iB} = -1$) and there is positive correlation of 0.75 between them. The contours of the probability distribution, the grey ellipses, indicate probability density values of 0.1, 0.3, 0.5, 0.7, and 0.9. The numbers in the four corners are the probabilities of the multivariate latent variable integrated in that quadrant, e.g. in the upper right quadrant; there is a probability of 0.16 that both species will occur at the site.

1.1.Prediction types

1.1.1. **Marginal prediction**

Similar to single-species SDM predictions, marginal predictions are based solely on environmental attributes. The predicted probability of occurrence of each species can be calculated using the JSDM regression coefficients and the covariate values found at each site, as in a standard generalised linear model. The sole difference between marginal predictions for JSDMs and those of single-species SDMs is that inter-species

correlations are accounted for in the estimation of JSDM regression coefficients. Since a species' marginal probability is independent of the occurrence state of other species (see Figure 3i), one can sum the different community assemblage probabilities (that contain the focal species) at a site to get the marginal probability. In Figure 2, we can see that species A, on the x-axis, is predicted to be present in the top-right and bottom-right quadrants. Thus:

$$Pr(A) = Pr(A|B) + Pr(A|B') \quad (5)$$

$$Pr(A) = 0.16 + 0.53$$

$$Pr(A) = 0.69$$

We can see that $Pr(A)$ calculated from the multivariate distribution in Figure 2 is the same as that from the univariate distribution in Figure 1a; the same can be calculated for species B. This is because the correlation between species affects their co-occurrence rates and not the unconditional probability of each species' occurrence at a site. Consequently, the marginal prediction from a JSDM can be computed without having to explicitly compute the integrals discussed in the previous section.

1.1.1. Joint prediction

We can also make predictions of species community composition that account for both the environmental covariates *and* species co-occurrence, by using the joint probability distribution. Whereas the marginal prediction of a community at a given site can be represented by a single vector of probabilities (one for each species), the joint prediction instead yields a probability value for each possible realisation of the community composition (a vector of 1s and 0s indicating presence/absence of each species; Figure 3ii).

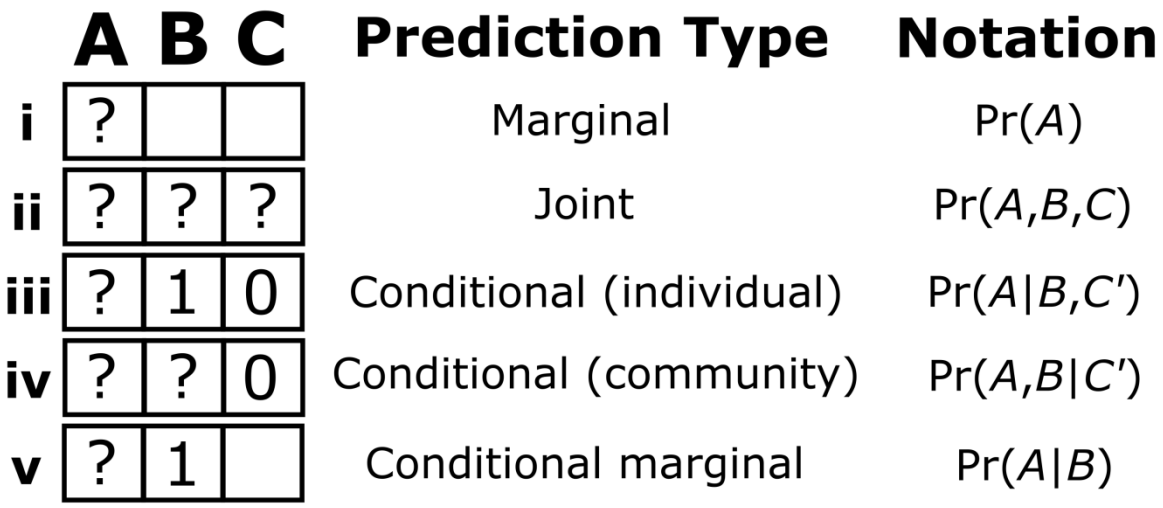


Figure S3: Five different prediction types possible with JSDMs. The boxes show the occurrence states of species A, B, and C. Question marks denote the species being predicted, empty boxes are species whose occurrence state is not being considered in that prediction type, and 1/0 a known presence/absence state of species being considered for prediction. Prediction type is the name of the prediction methodology, and Notation shows the corresponding probabilistic notation, following the definitions in the text.

However, with J species the number of possible community assemblages is 2^J. Whilst 3 species results in 8 probabilities (i.e. 8 possible species combinations), with 10 species that number rises to 1024, and with 20 species there are over a million possible assemblages. It is therefore generally infeasible to compute and store the probabilities of all assemblages, when using a JSDM for prediction in general.

Since the joint probability distribution is entirely defined by the JSDM parameters μ_{ij} and $\boldsymbol{R}$, these can instead be stored to represent the joint prediction. With these parameters, it then becomes straightforward to evaluate model predictions by computing the joint probability for the observed species assemblage in held-out data. These parameters can also be used to simulate plausible community assemblages by taking random draws from the joint probability distribution and these can then be summarised to compute other quantities of interest, such as a predictions of species richness.

Joint predictions cannot be represented via a single vector of probabilities of occurrence for individual species, since this discards crucial information about the jointness of the occurrence distribution. For example, computing an average over these simulated assemblages would simply result in (an approximation to) the marginal prediction. However, the joint prediction can be summarised in various ways that can be represented more concisely.

1.1.2. Conditional prediction

Another way to define prediction in multivariate space is by using the conditional probability distribution of species, in which we estimate a species' occurrence probability given a fixed occurrence state of other species. If we know the correlation between two species (estimated by the JSDM) and the occurrence state of one of those species, we can make a more informed prediction of the unknown occurrence state of the other species.

For each species with a known occurrence state, we can truncate the multivariate normal distribution over the latent variable in the dimension that represents that species, either to be positive if the species is present, or negative if it is absent (see Figure 4 for a visual representation). Because probability distributions integrate to one, truncating this distribution alters the probabilities of the remaining community assemblage possibilities. For instance, in our example, the marginal probability of species B, $\Pr(B)$, is 0.16 (Figure 1b), but if we know that species A is present, then the conditional probability of species B is $\Pr(B|A) = 0.23$ (Figure 4). This modification of the joint probability can be interpreted in another way: the conditional probability of species B being present is the joint probability of species A and B being present, normalised by

the marginal probability of species A: $\Pr(B|A) = \Pr(B,A)/\Pr(A) = 0.16/0.69 = 0.23$. This follows from rearranging Eq 1.

Depending on the amount of information available to condition our joint distribution, there are two possible ways of approaching conditional predictions. In the scenario in which we know the occurrence state of all but one species at a site, we can calculate the probability of occurrence of that individual species conditional on the rest (see Figure 3iii). Alternatively, if we know the occurrence state of only a single, or small subset of, other species, we can calculate the joint probability of the remaining species conditional on the occurrence state of the known species (see Figure 3iv). This joint prediction has the same characteristics as the joint prediction described above, but for a subset of the species.

1.1.3. Conditional marginal prediction

The final way of defining prediction using JSDMs is simultaneously conditional on and marginal to the occurrence states of the other species. As noted previously, conditional prediction can be defined in a variety of ways dependant on the number of species occurrence states that are known. This may result in the prediction of one species conditional on the fixed presence/absence of all other species, or of $J-1$ species, conditioned on the presence/absence of one species (a joint, conditional prediction). We can also combine the conditional and marginal approaches to predict the presence of one or more species, conditional on some species, but marginal to others (see Figure 3v). This prediction type is also simple to represent: like marginal predictions it can be stored as a vector of probabilities for each species. It also accounts for residual correlations, and can be matched onto ecological hypotheses. For example, predicting the marginal probability of presence of one set of species given the presence or absence

of some other set of species - say predators/competitors, or perhaps some easy-to-detect indicator species.

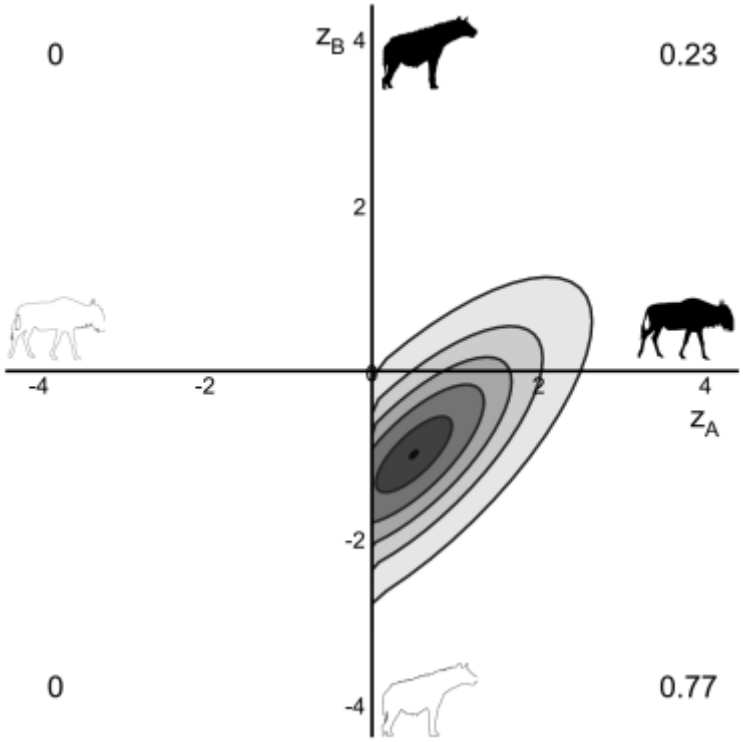


Figure 4: The conditional probability distribution of species B given species A is known to be present. This probability distribution is identical to that of Figure 2 except that it is truncated such that $z_A > 0$. The contours of the probability distribution, the grey ellipses, are the probability densities for 0.1, 0.3, 0.5, 0.7, and 0.9 proportions of the distribution's volume. The values in the four corners are the probabilities of the multivariate latent variable occurring in that quadrant. Since species A is known to be present, there is zero probability that the value of the latent variable falls in either of the left-hand quadrants. $Pr(z_B|z_A) = 0.23$, which is higher than the marginal distribution, $Pr(B) = 0.16$, of the un-truncated distributions seen in Figures 1b and 2.

APPENDIX L

EVALUATION METRICS

Name	Formula	Range	Binary predictions	Probabilistic predictions	Threshold-dependent	Threshold-independent	Community dissimilarity metrics	Species Richness	Likelihood	Interpretation
Accuracy / True Skill Statistic	$\frac{TP+TN}{TP+FP+TN+FN}$	0-1	X	X	X					Higher = more correct classifications
Area under the Receiver Operating Characteristic curve (AUC)		0-1		X		X				Higher = more correct classifications
Bray-Curtis dissimilarity	$\frac{A+B-2J}{A+B}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X			X			0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Canberra Index	$\frac{A+B-2J}{A+B-J}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X			X			0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Cohen's Kappa	$\frac{(TP+TN)-\frac{(TP+FN)*(TP+FP)+(TN+FP)*(TN+FN)}{TP+FP+TN+FN}}{(TP+FP+TN+FN)-\frac{(TP+FN)*(TP+FP)+(TN+FP)*(TN+FN)}{TP+FP+TN+FN}}$	0-1	X		X					Higher = more correct classifications
F_1 score	$\frac{2}{\frac{1}{Sensitivity}+\frac{1}{Precision}}$	0-1	X	X	X					Higher = more correct classifications
False discovery	$\frac{FP}{TP+FP}$	0-1	X	X	X					Higher = more predicted presences are

ratio								incorrect
False negative rate	$\frac{FN}{TP + FN}$	0-1	X	X	X			Higher = more observed presences are predicted incorrect
False omission rate	$\frac{FN}{TN + FN}$	0-1	X	X	X			Higher = more predicted absences are incorrect
False positive rate	$\frac{FP}{FP + TN}$	0-1	X	X	X			Higher = more observed absences are predicted incorrect
Gower Index	$\frac{A + B - 2J}{M}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences M = Total number of species	0-1	X	X		X		0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Gower Index (alternative)	$\frac{A + B - 2J}{A + B - J}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X		X		0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Jaccard distance	$\frac{2D}{1 + D}$ Where: D = Bray-Curtis	0-1	X	X		X		0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Kendall rank correlation coefficient	$\frac{(\#\ concordant\ pairs) - (\#\ discordant\ pairs)}{n(n-1)/2}$	-1 – 1		X	X			>0 = more correct predictions <0 = more incorrect predictions
Kulczynski Index	$1 - \frac{\frac{J}{A} + \frac{J}{B}}{2}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences	0-1	X	X		X		0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Log Likelihood – Independent		-∞ - ∞		X			X	Higher = Model fit to data better
Log Likelihood		-∞ - ∞		X			X	Higher = model fit to data better

– Joint							
Mountford Index	Positive root of the equation: $e^{A*E} + e^{B*E} = 1 + e^{(A+B-J)*E}$ Where: A = Number of observed presences B = Number of predicted presences J = Number of correctly predicted presences E = Mountford Index	0-1	X	X		X	0 = same composition 1 = opposite composition Lower = more correct assemblage prediction
Mean error (bias)	$\frac{1}{n}\sum_{i=1}^{n}(observed - predicted)$	-∞ - ∞		X		X	Increasing magnitude = more bias Positive value = more false absences (underprediction) Negative value = more false presences (overprediction)
Mean square error	$\frac{1}{n}\sum_{i=1}^{n}(observed - predicted)^2$	0-∞		X		X	Higher = increased error/bias Can't differentiate between over/underprediction
Negative predictive performance	$\frac{TN}{TN + FN}$	0-1	X	X	X		Higher = more observed absences are correctly identified
Pearson correlation coefficient	$\frac{cov(observed, predicted)}{\sigma_{observed}\sigma_{predicted}}$	-1 – 1		X		X	>0 = more correct predictions <0 = more incorrect predictions
Positive predictive performance / Precision	$\frac{TP}{TP + FP}$	0-1	X	X	X		Higher = more predicted presences are correct
R^2 / Coefficient of determination	$1 - \frac{\sum(observed - predicted)^2}{\sum(observed - \overline{observed})^2}$	0-1		X		X	Higher = predictions more closely match the observed values
Raup-Crick dissimilarity	$1 - \int_0^A hypergeometric(A, B, M, 1)$ Where: A = Number of observed presences B = Number of predicted presences M = Total number of species	0-1	X	X	X		0 = same composition 1 = opposite composition Lower = more correct assemblage prediction

Root mean square error	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(observed - predicted)^2}$	0-∞		X		X		Higher = increased error/bias Can't differentiate between over/underprediction
Spearman rank correlation coefficient	$\frac{cov(r_{obs}, r_{pred})}{\sigma_{r_{obs}}\sigma_{r_{pred}}}$	-1 – 1		X		X		>0 = more correct predictions <0 = more incorrect predictions
Species richness difference	$Predicted\ richness - observed\ richness$	-∞ - ∞	X	X			X	>0 = overpredicts species richness <0 = underpredicts species richness
Sum of squared errors	$\sum_{i=1}^{n}(observed - predicted)^2$	0-∞		X		X		Higher = increased error/bias Can't differentiate between over/underprediction
True negative rate / Specificity	$\frac{TN}{FP + TN}$	0-1	X	X	X			Higher = more correctly identified absences
True positive rate / Sensitivity	$\frac{TP}{TP + FN}$	0-1	X	X	X			Higher = more correctly identified presences
Youden's J statistic	$Sensitivity + Specificity - 1$	0-1	X	X	X			Higher = more correct classifications

APPENDIX M

RESULTS: ABSOLUTE VALUE FOREST PLOTS

1. Marginal (probabilistic) predictions

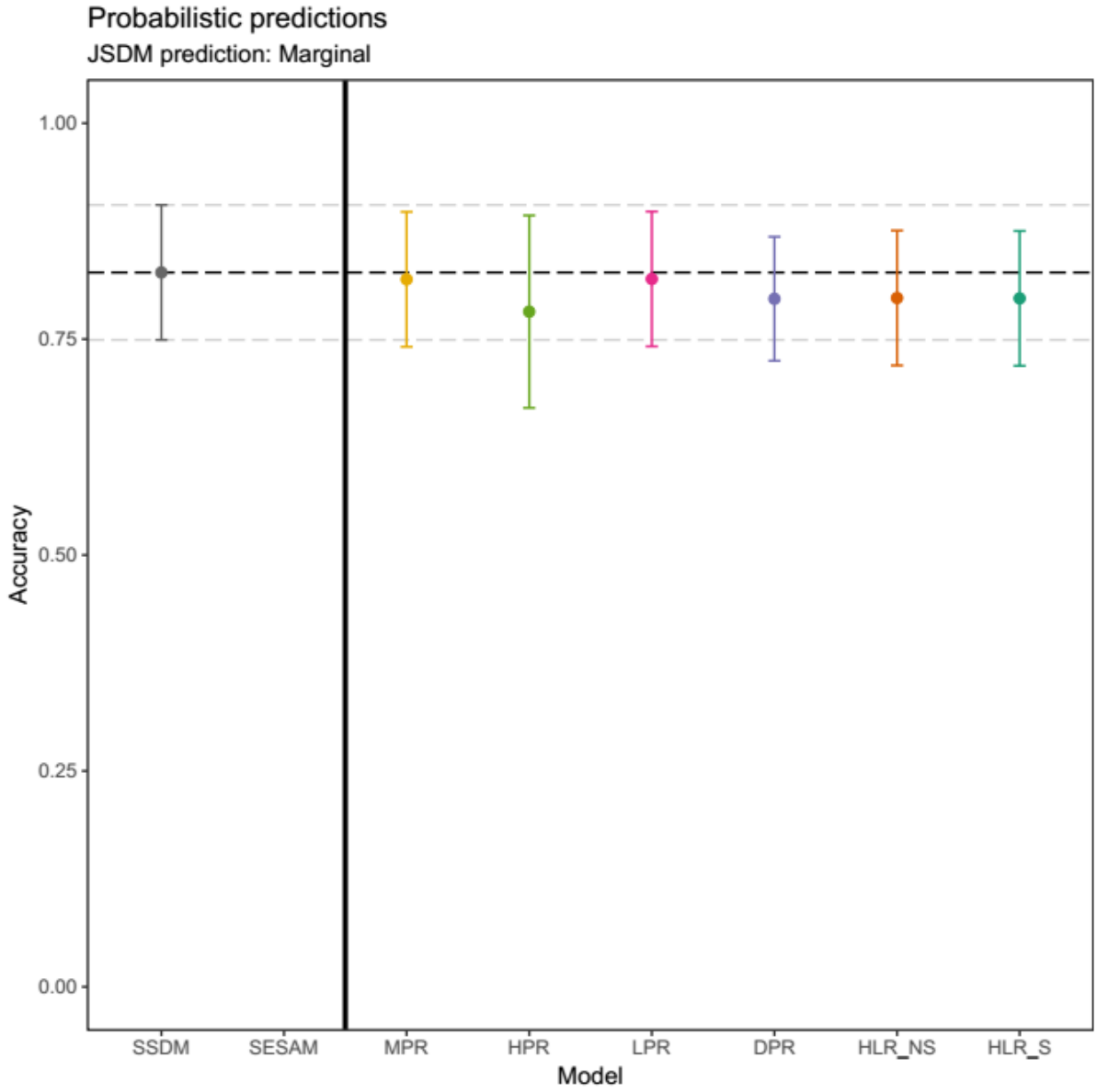


Figure S1: Model performance for the accuracy evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

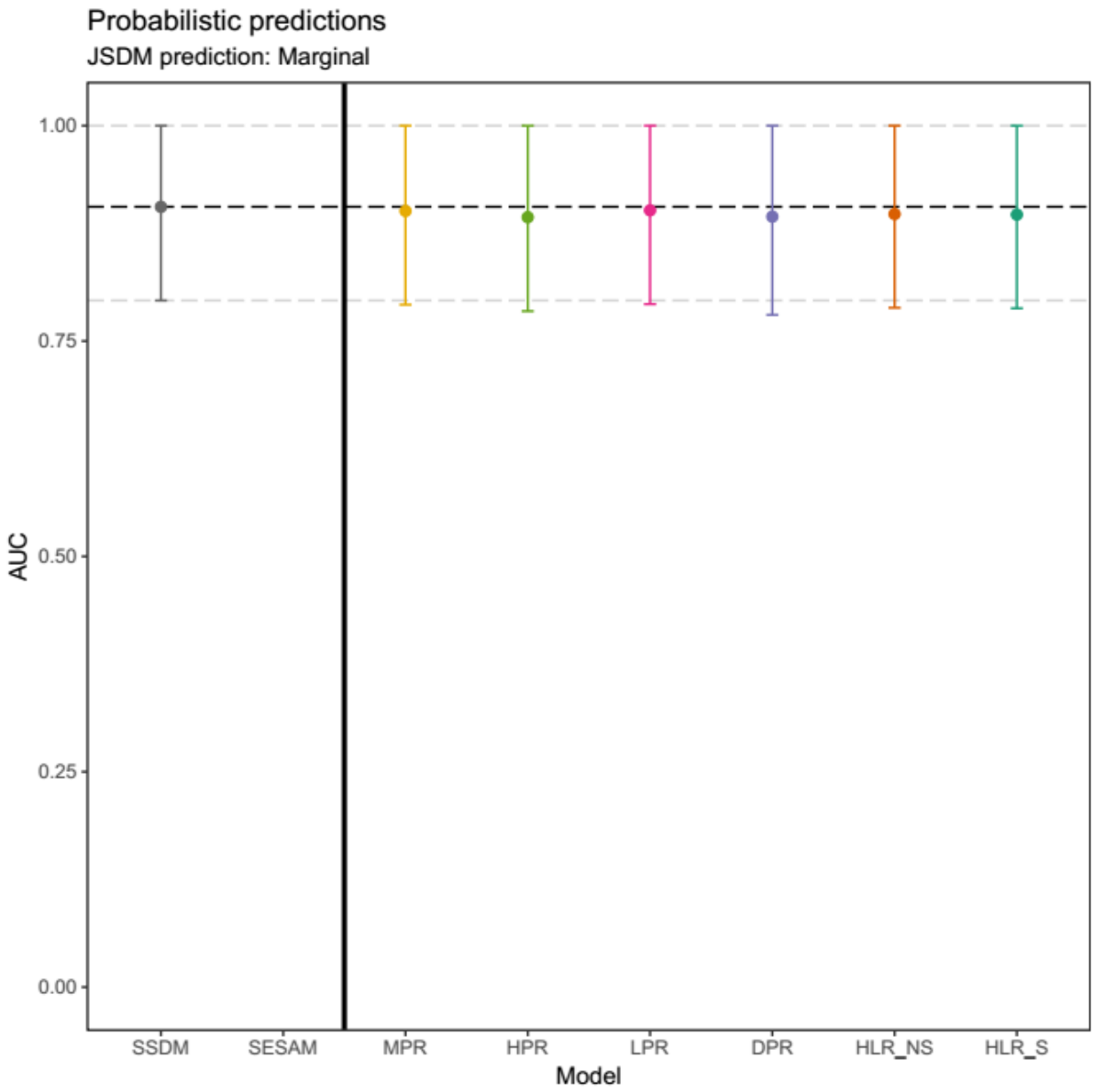


Figure S2: Model performance for the AUC evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

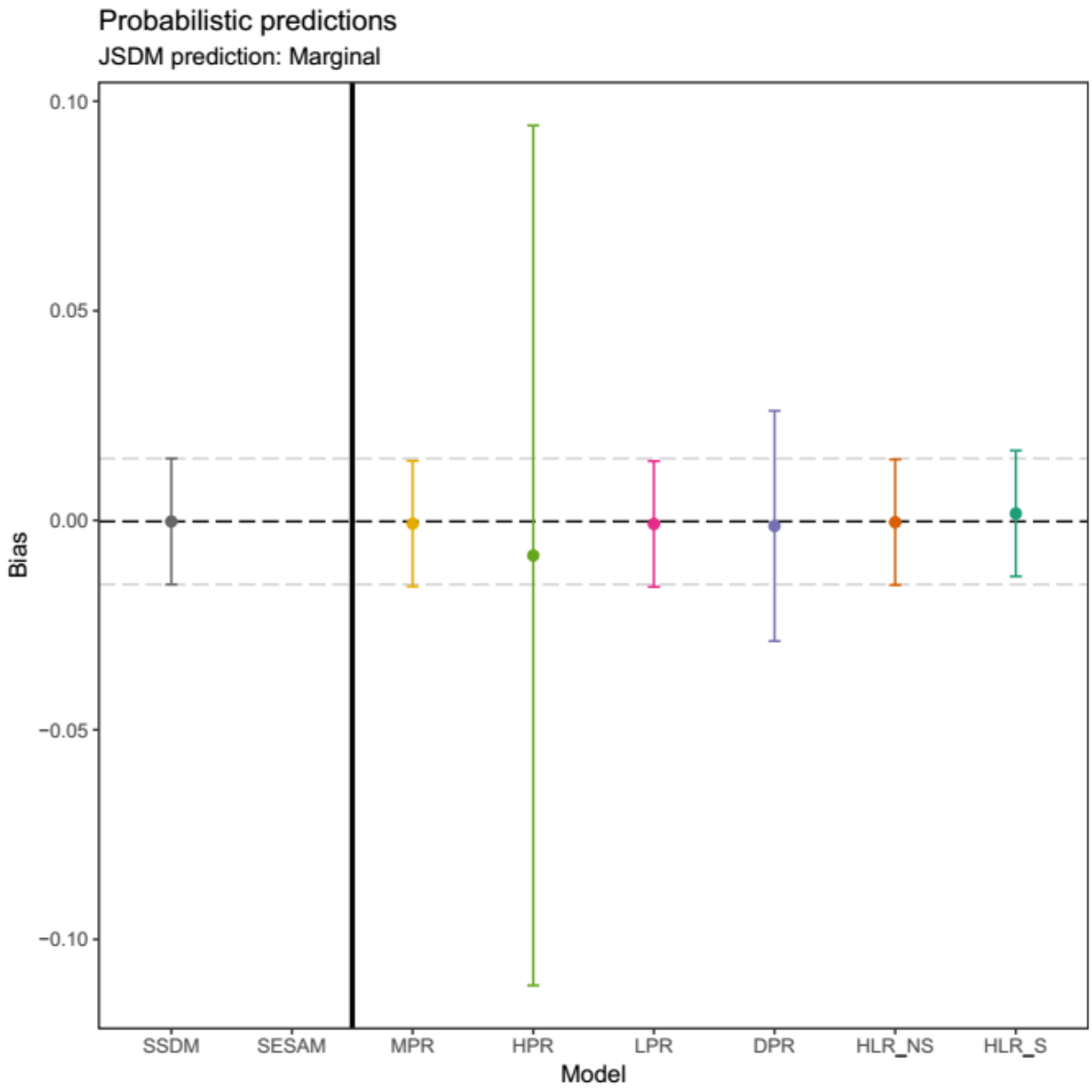


Figure S3: Model performance for the bias (mean error) evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

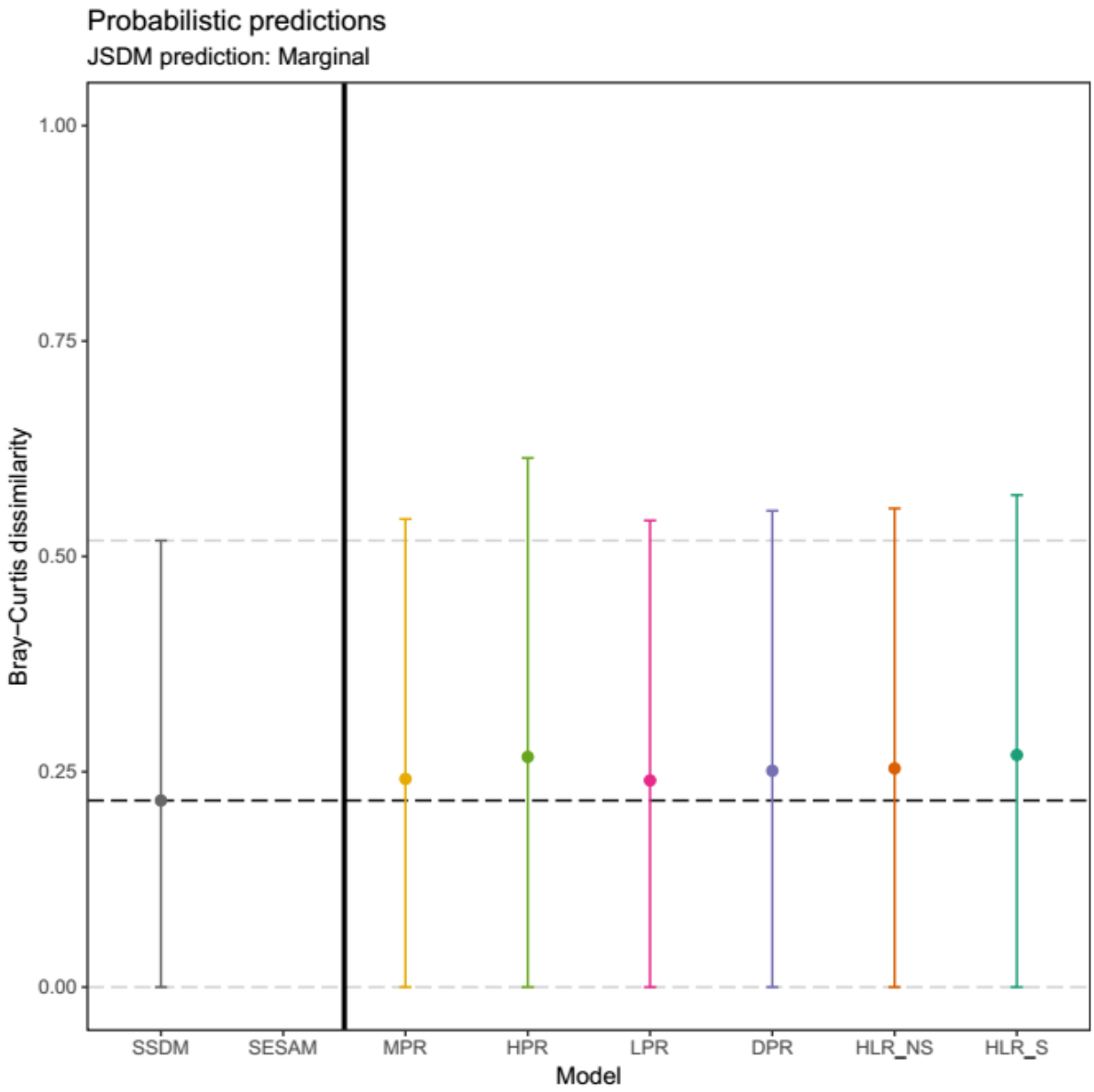


Figure S4: Model performance for the Bray-Curtis dissimilarity evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

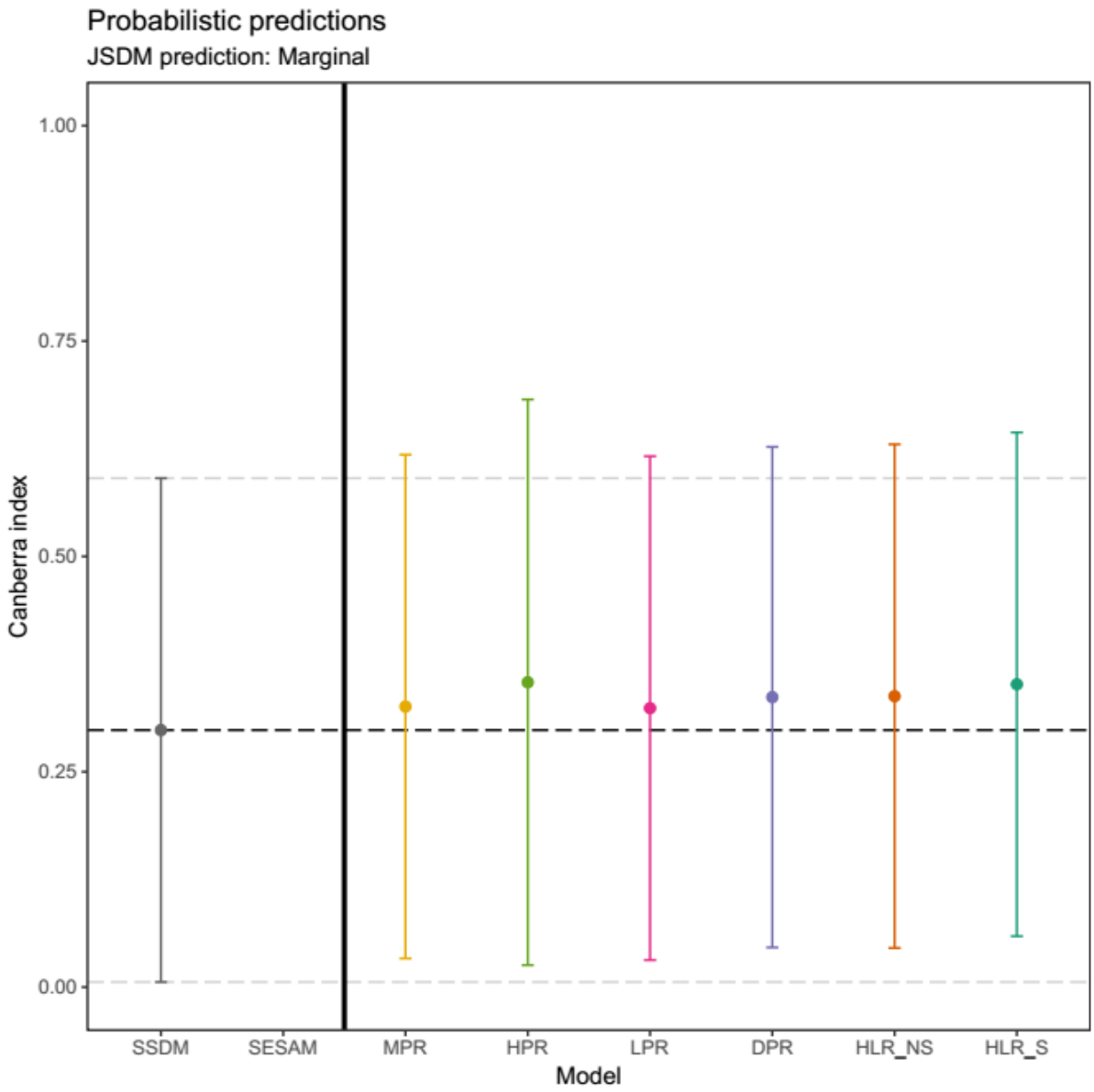


Figure S5: Model performance for the Canberra index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

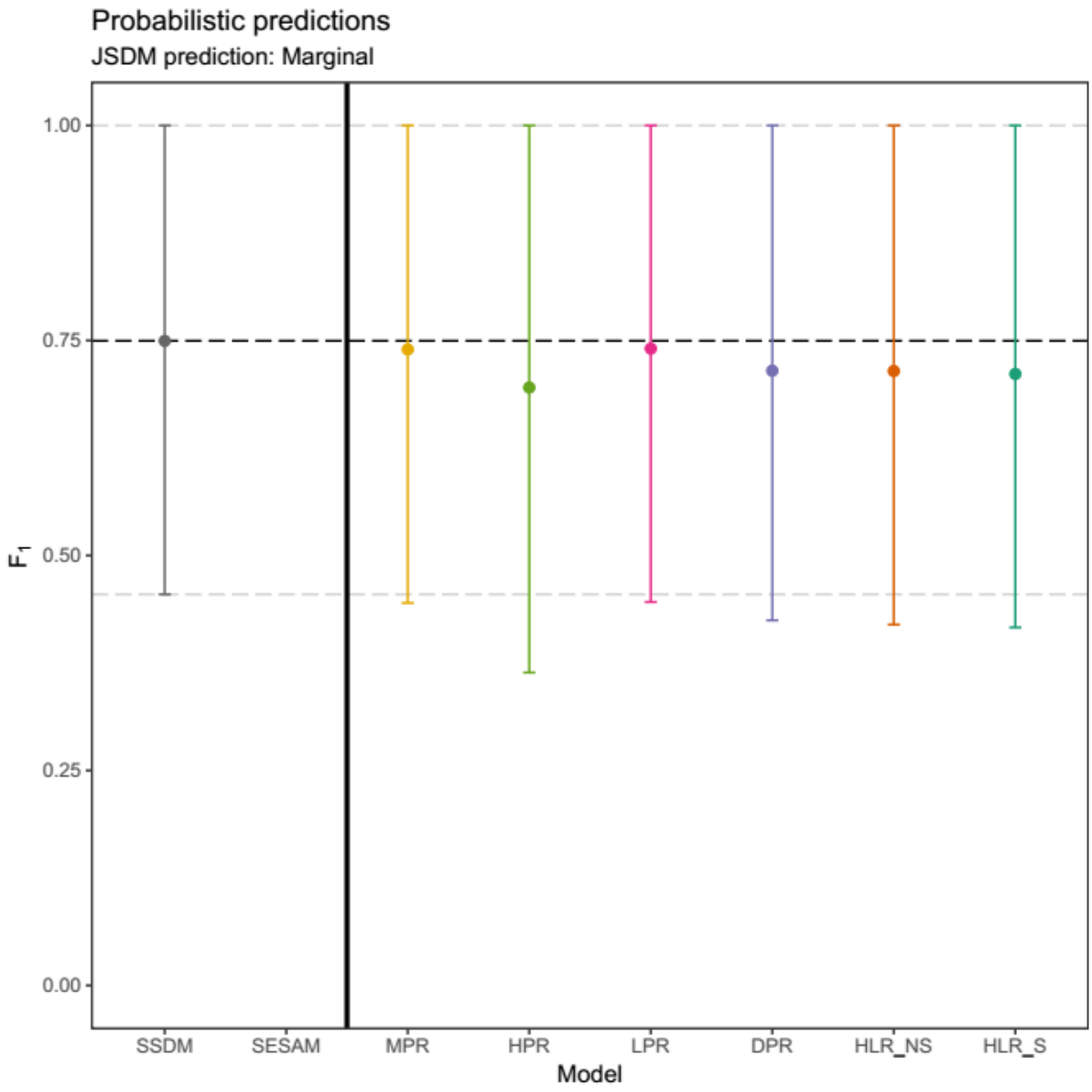


Figure S7: Model performance for the F_1 evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

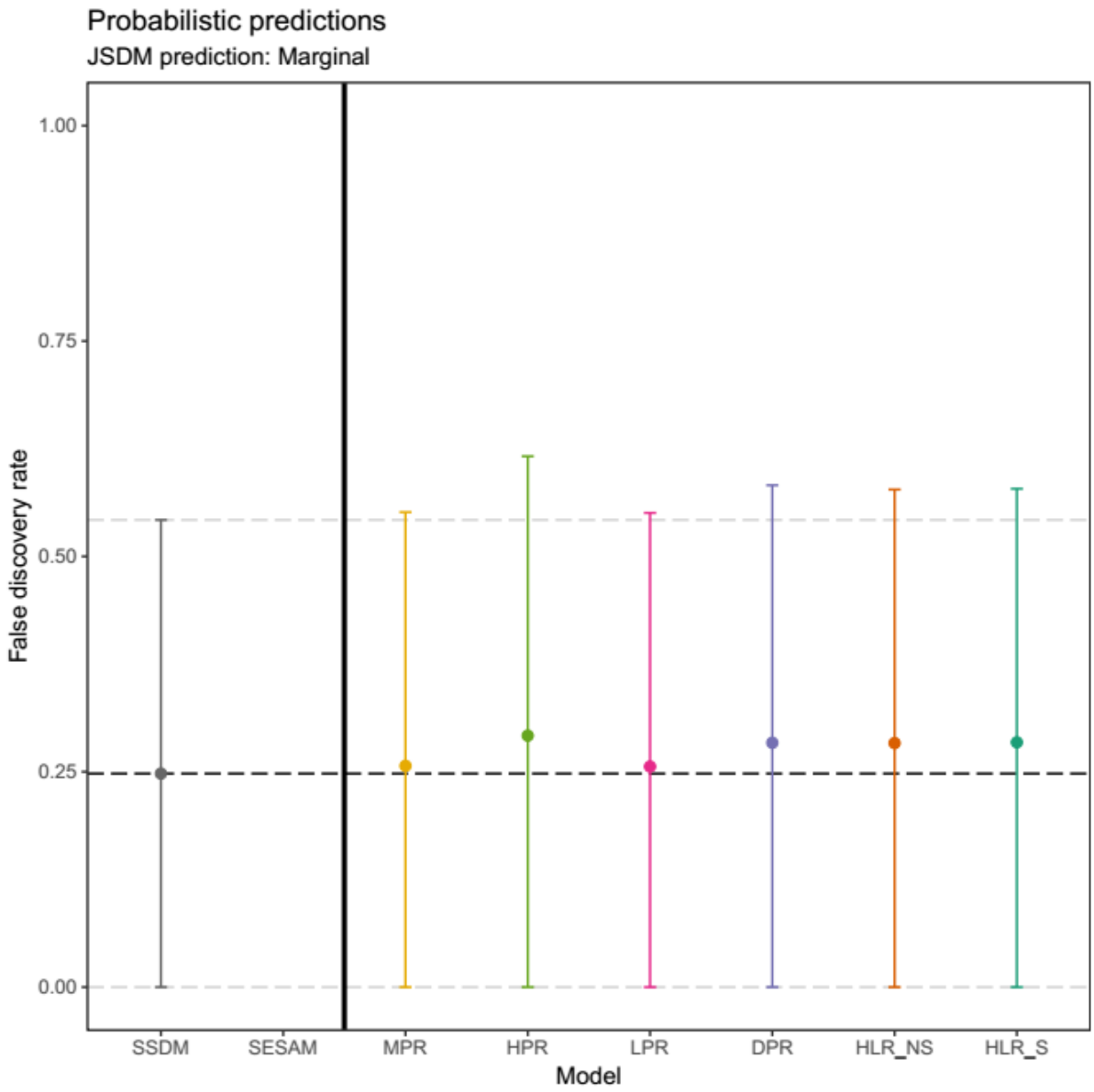


Figure S7: Model performance for the false discovery rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

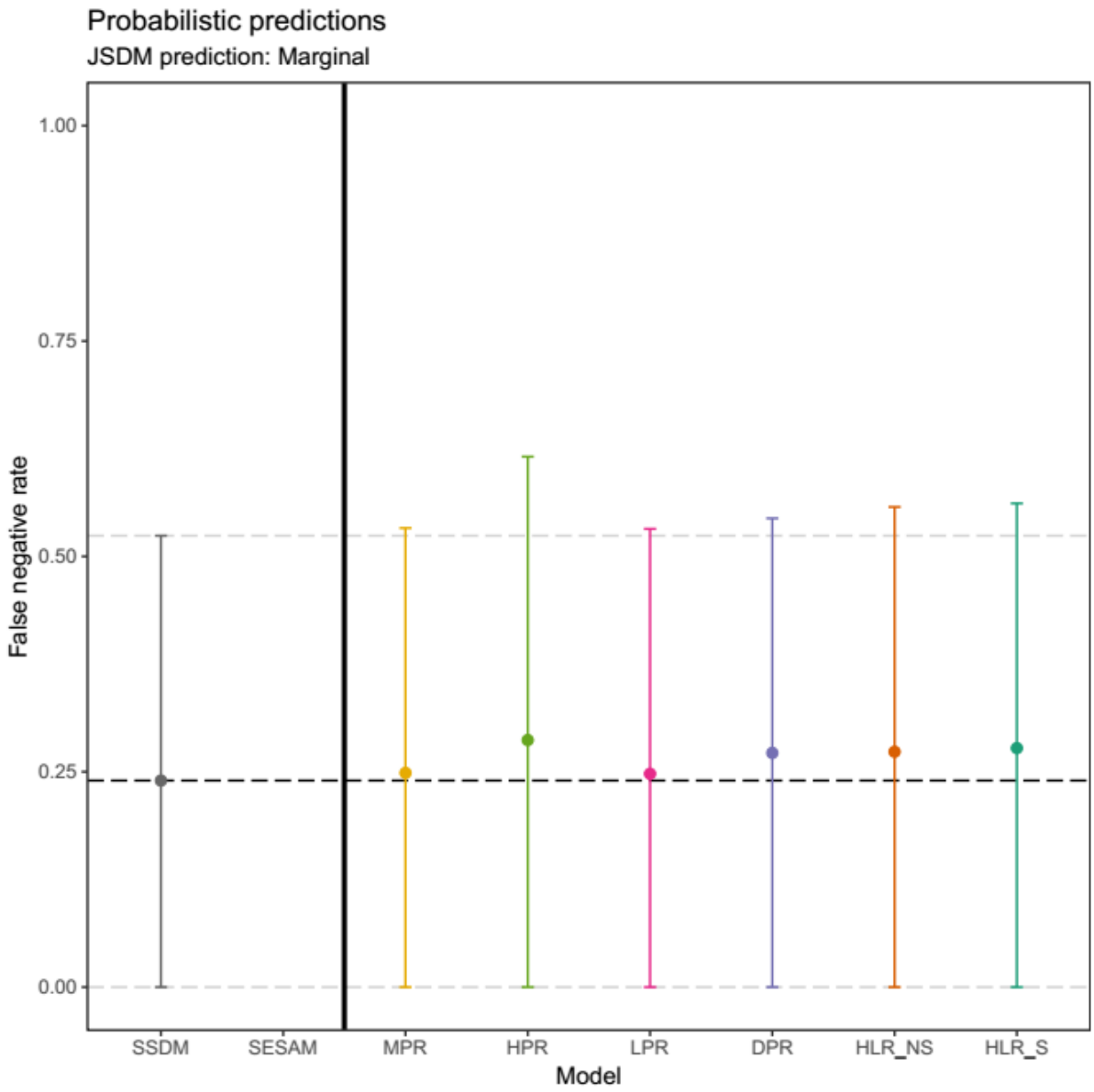


Figure S8: Model performance for the false negative rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

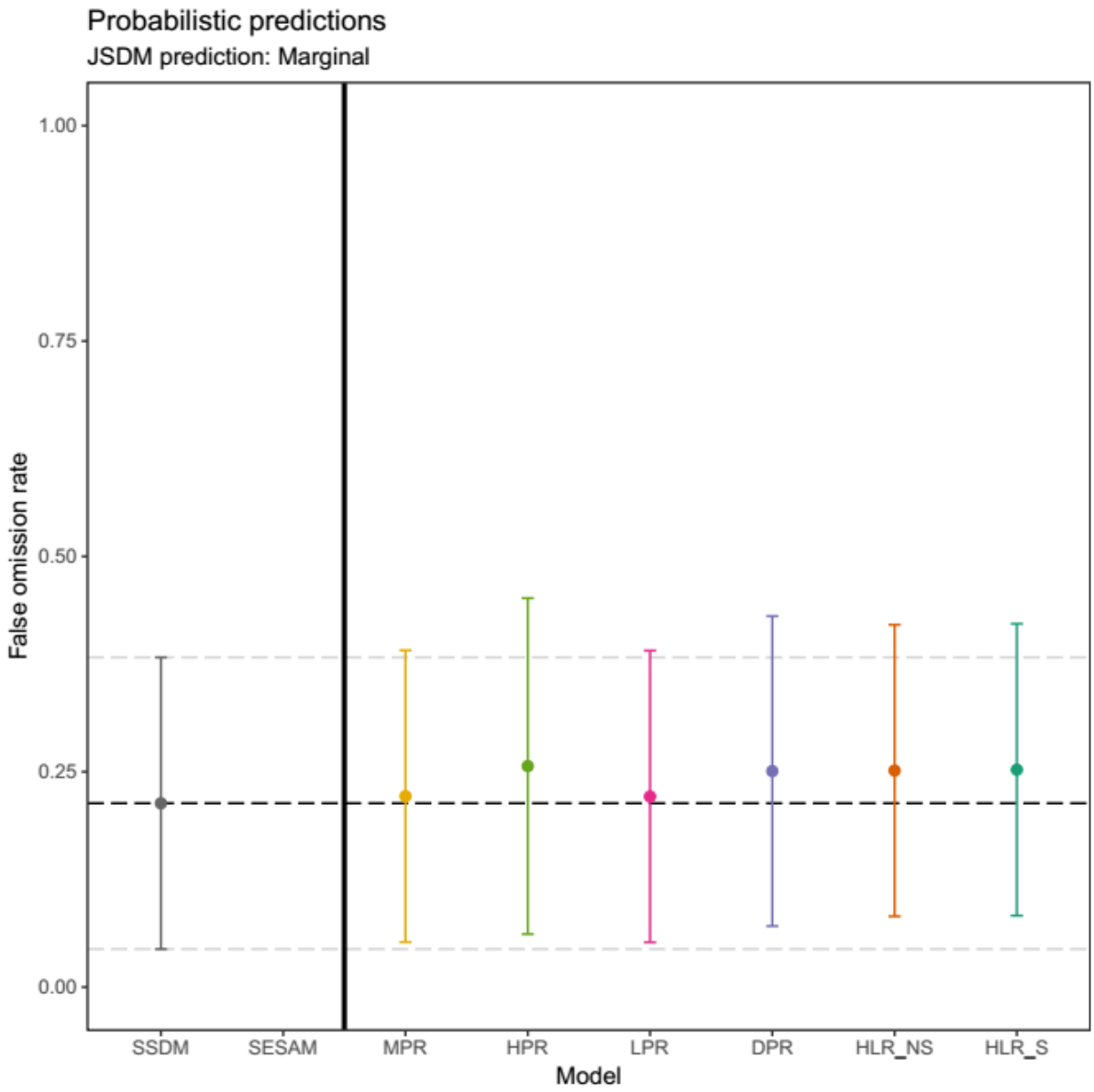


Figure S9: Model performance for the false omission rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

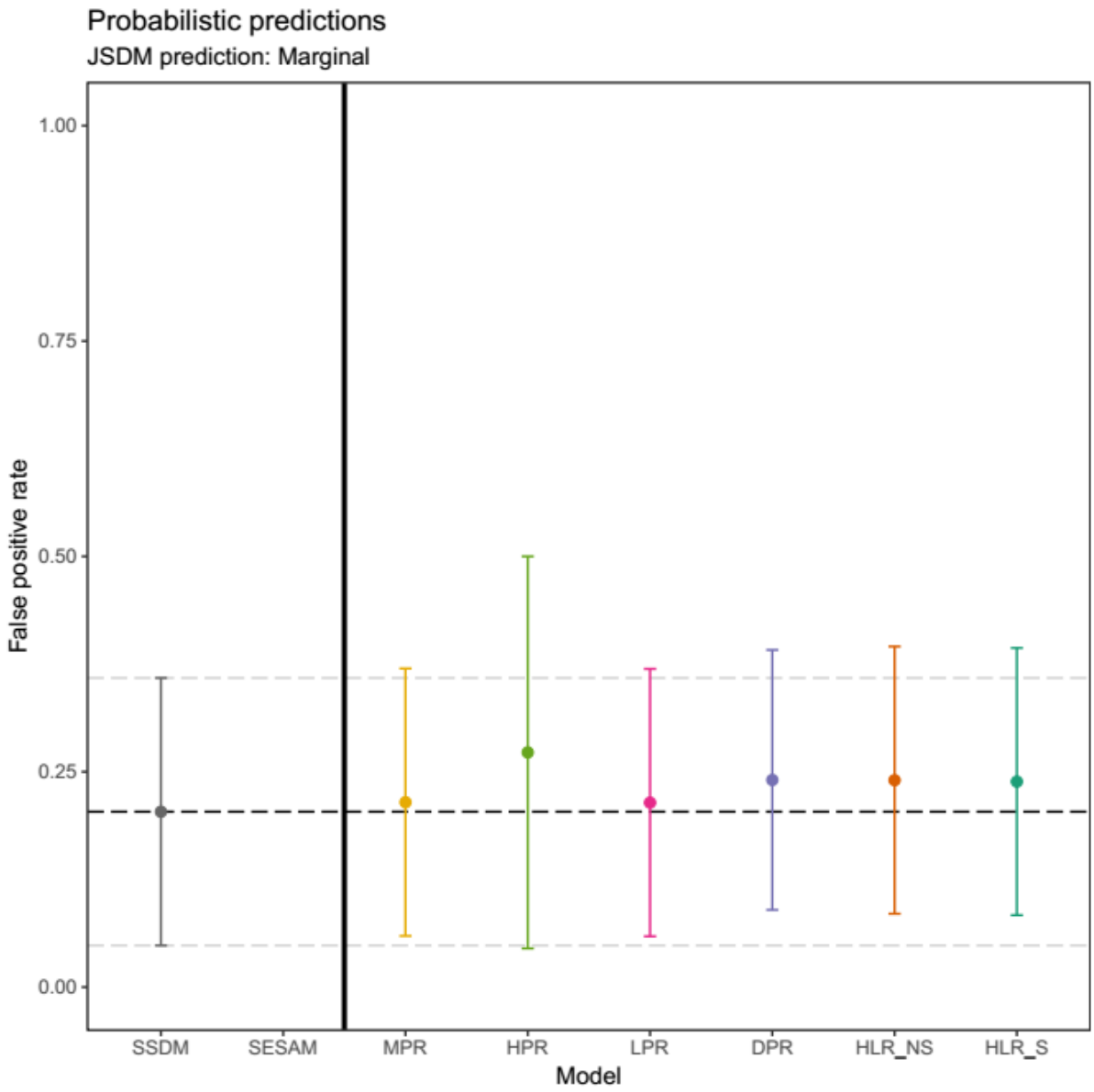


Figure S10: Model performance for the false positive rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

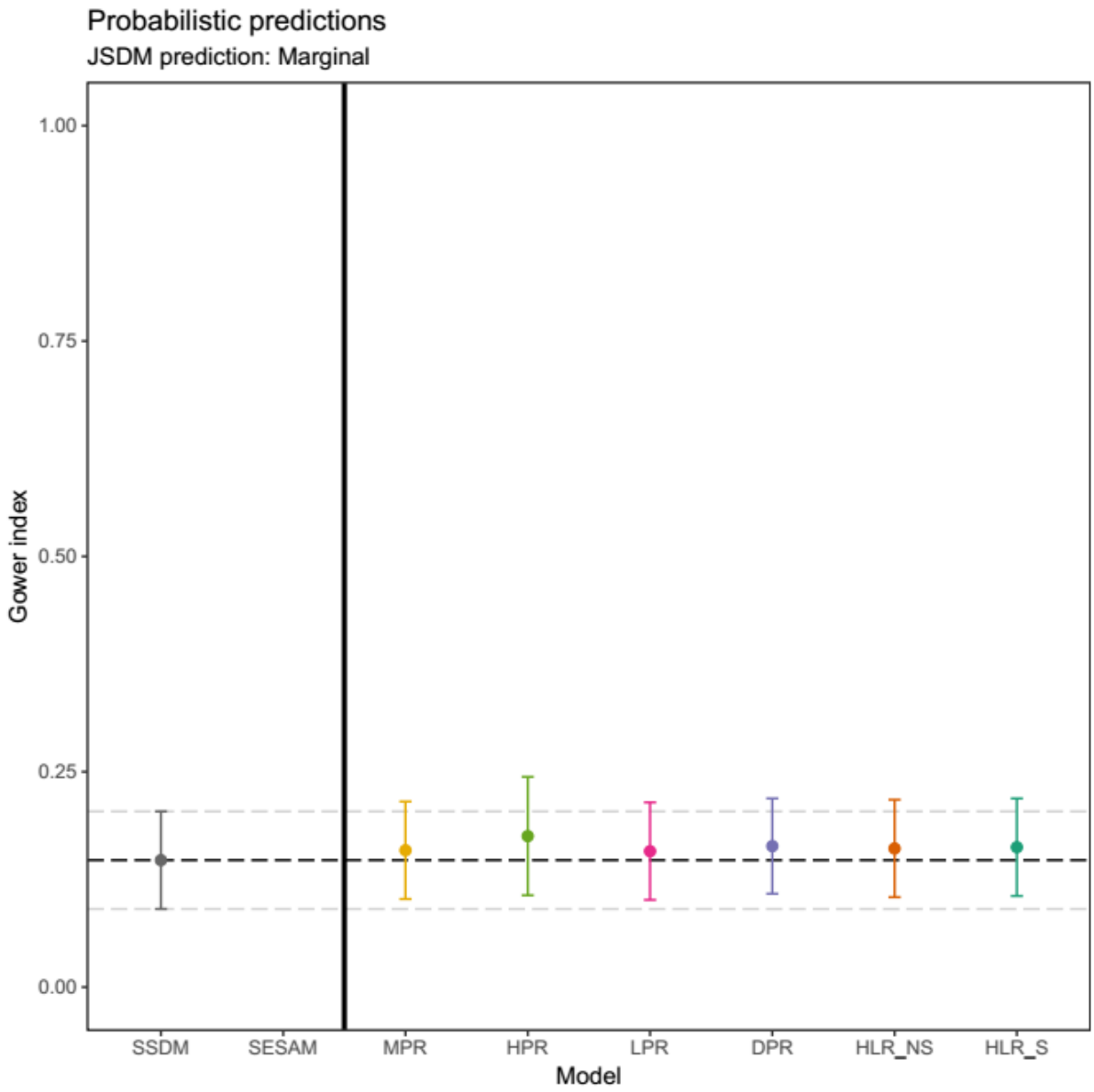


Figure S11: Model performance for the Gower index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

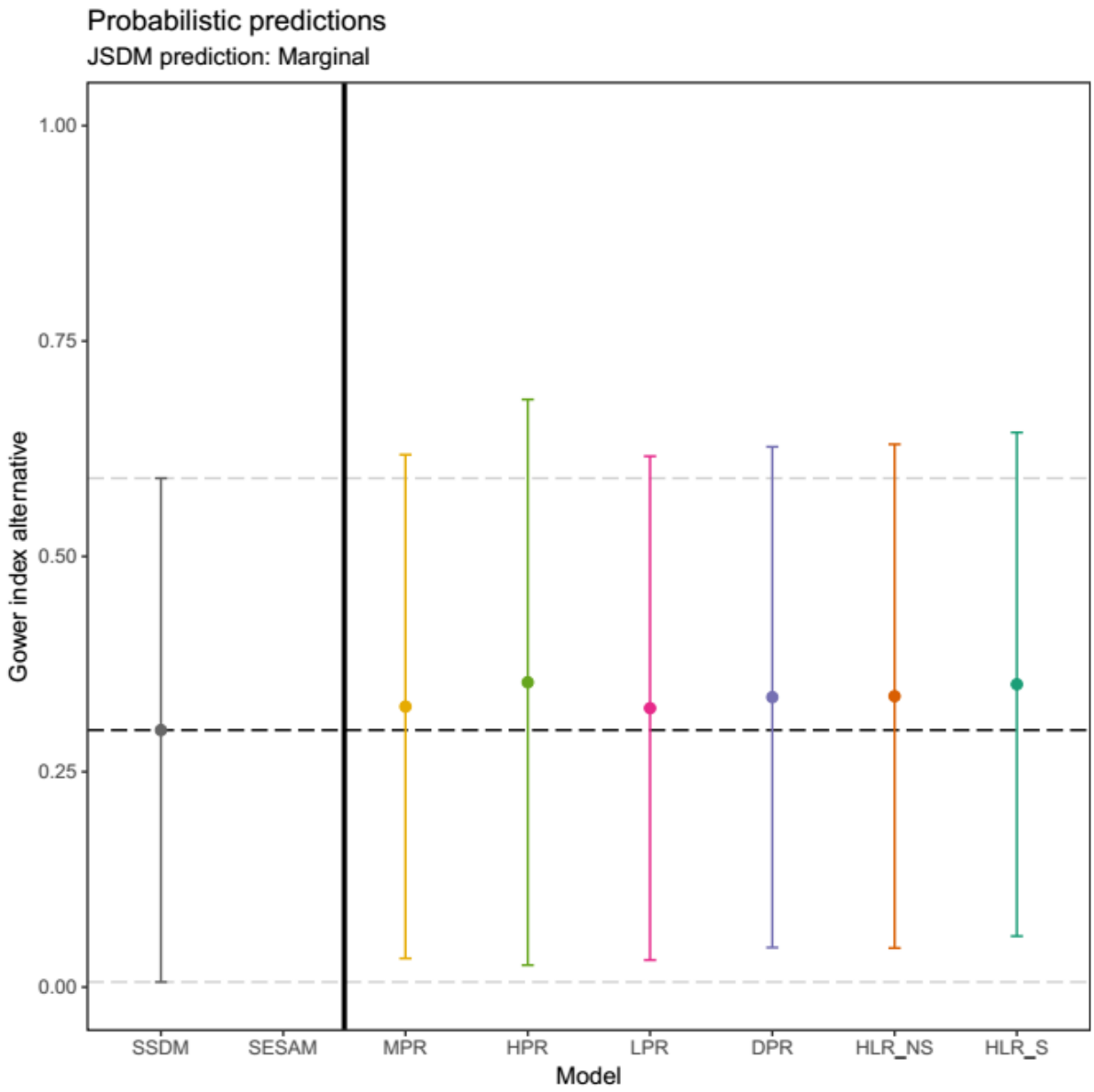


Figure S12: Model performance for the alternative Gower index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

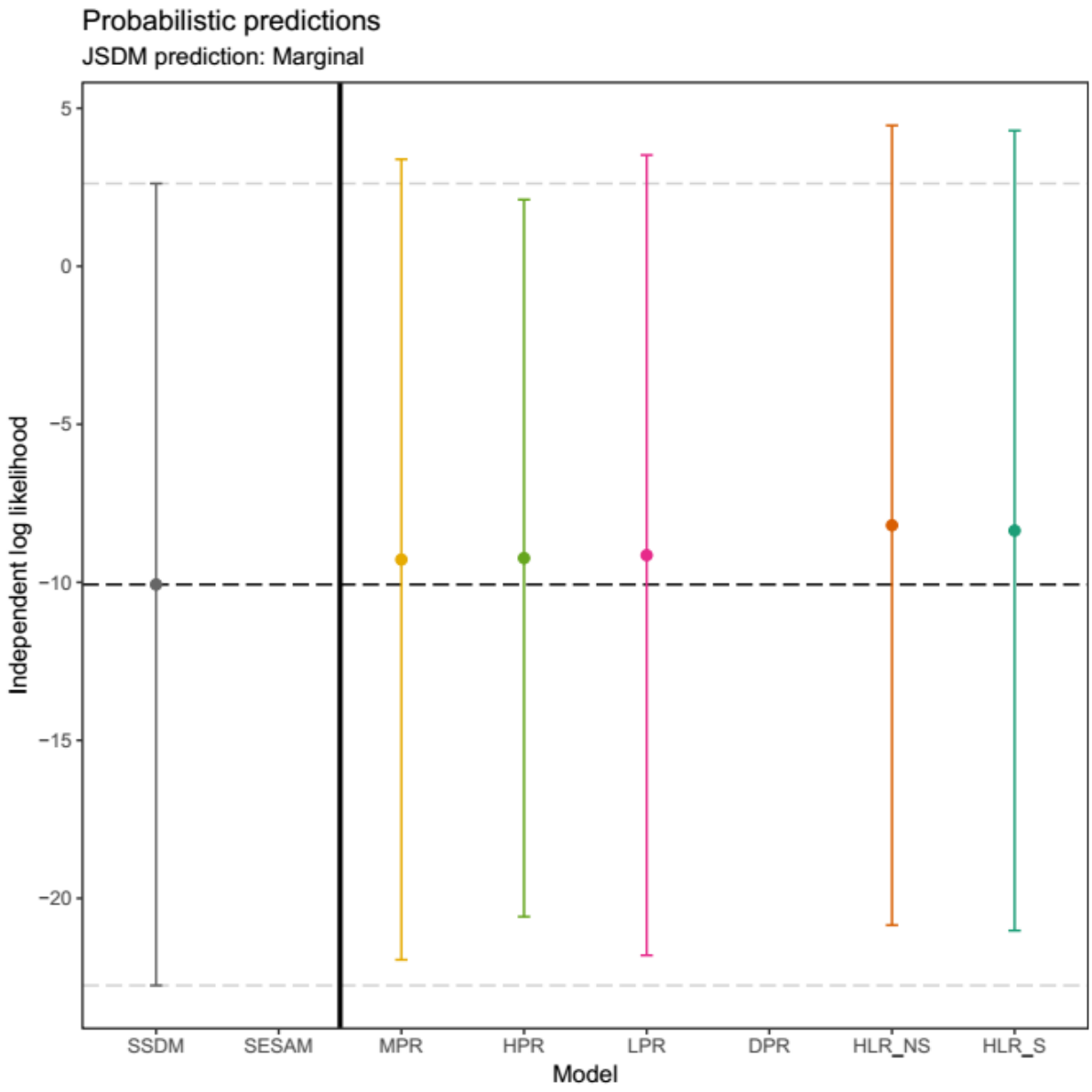


Figure S13: Model performance for the independent log-likelihood evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

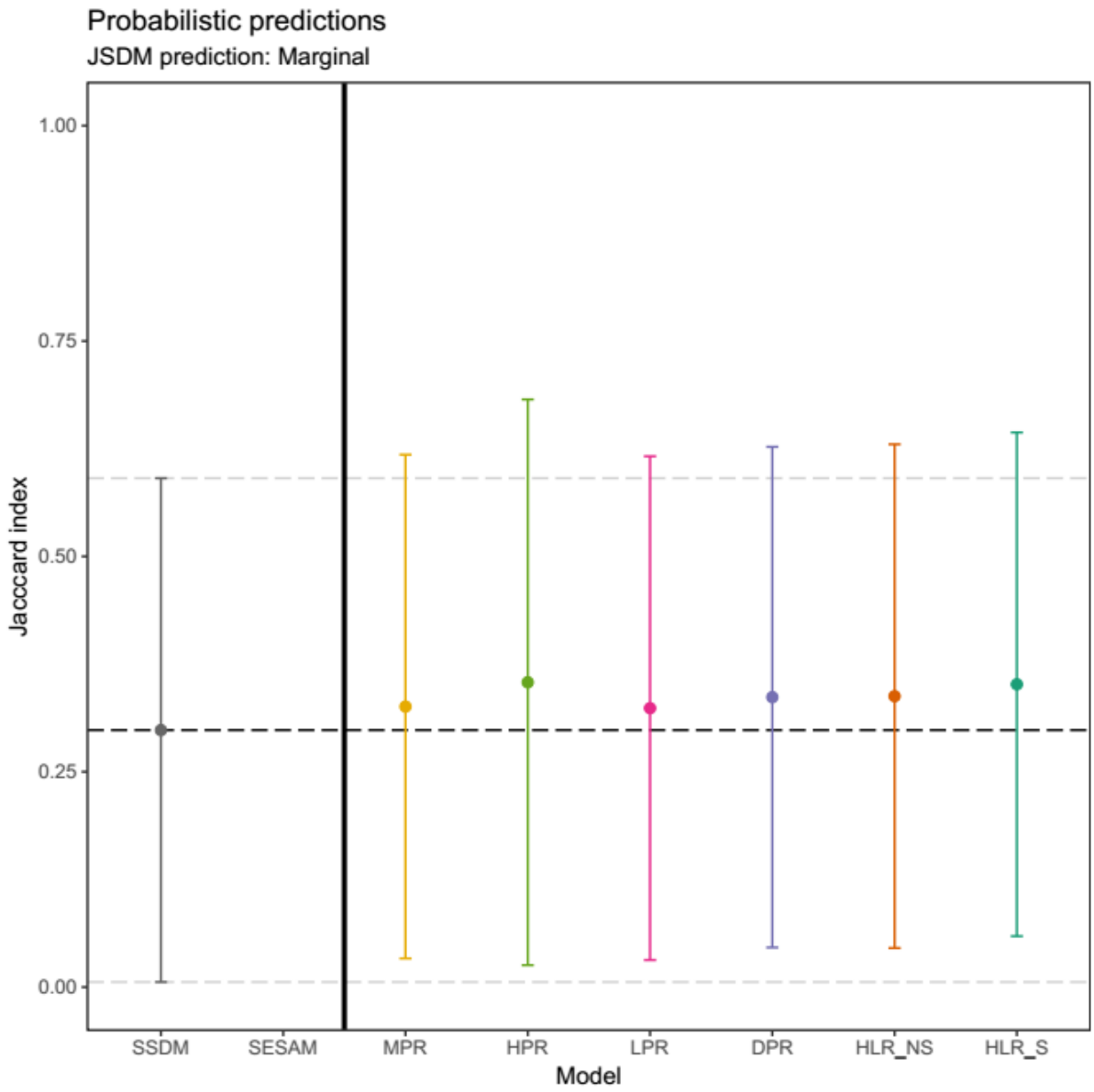


Figure S14: Model performance for the Jaccard index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

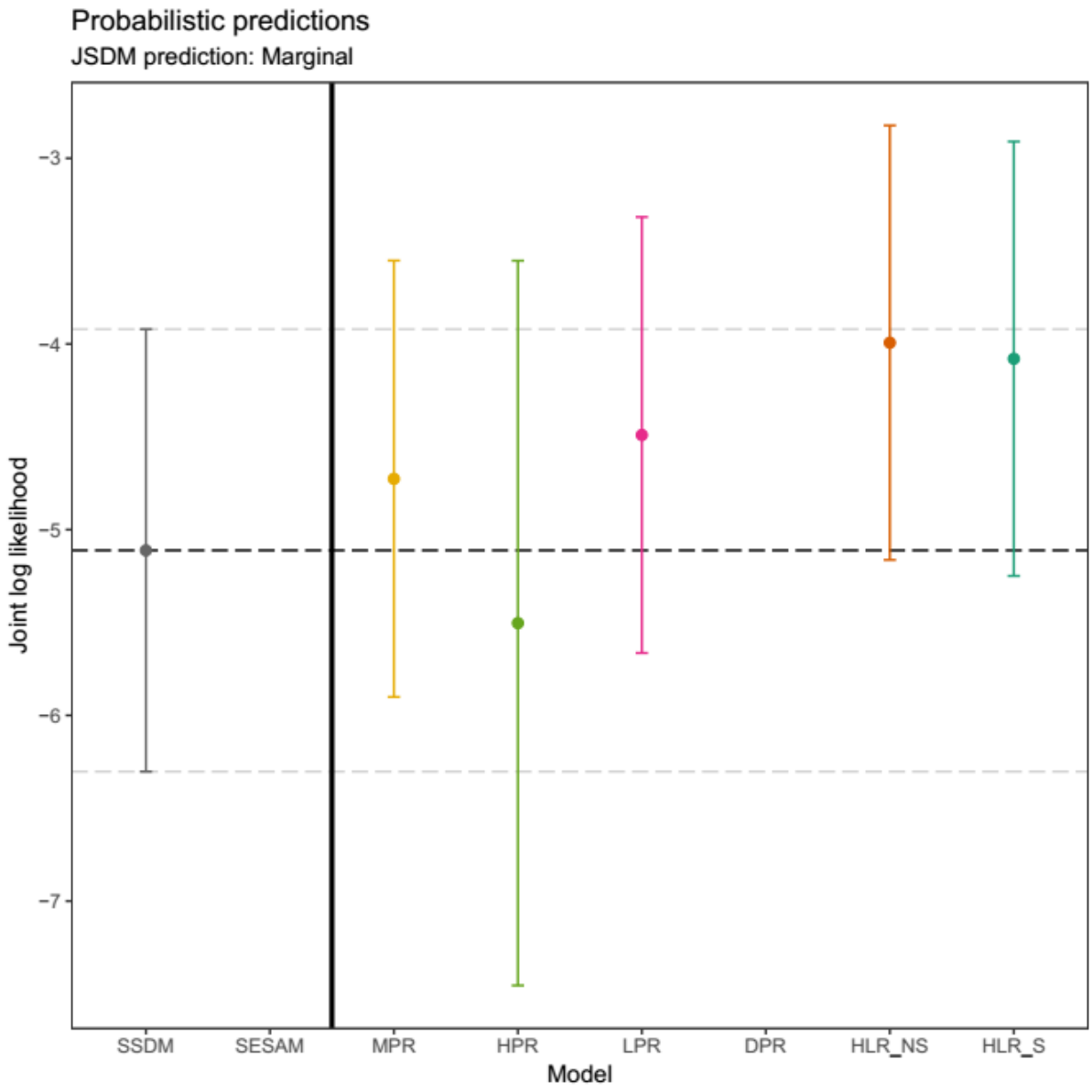


Figure S15: Model performance for the joint log-likelihood evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

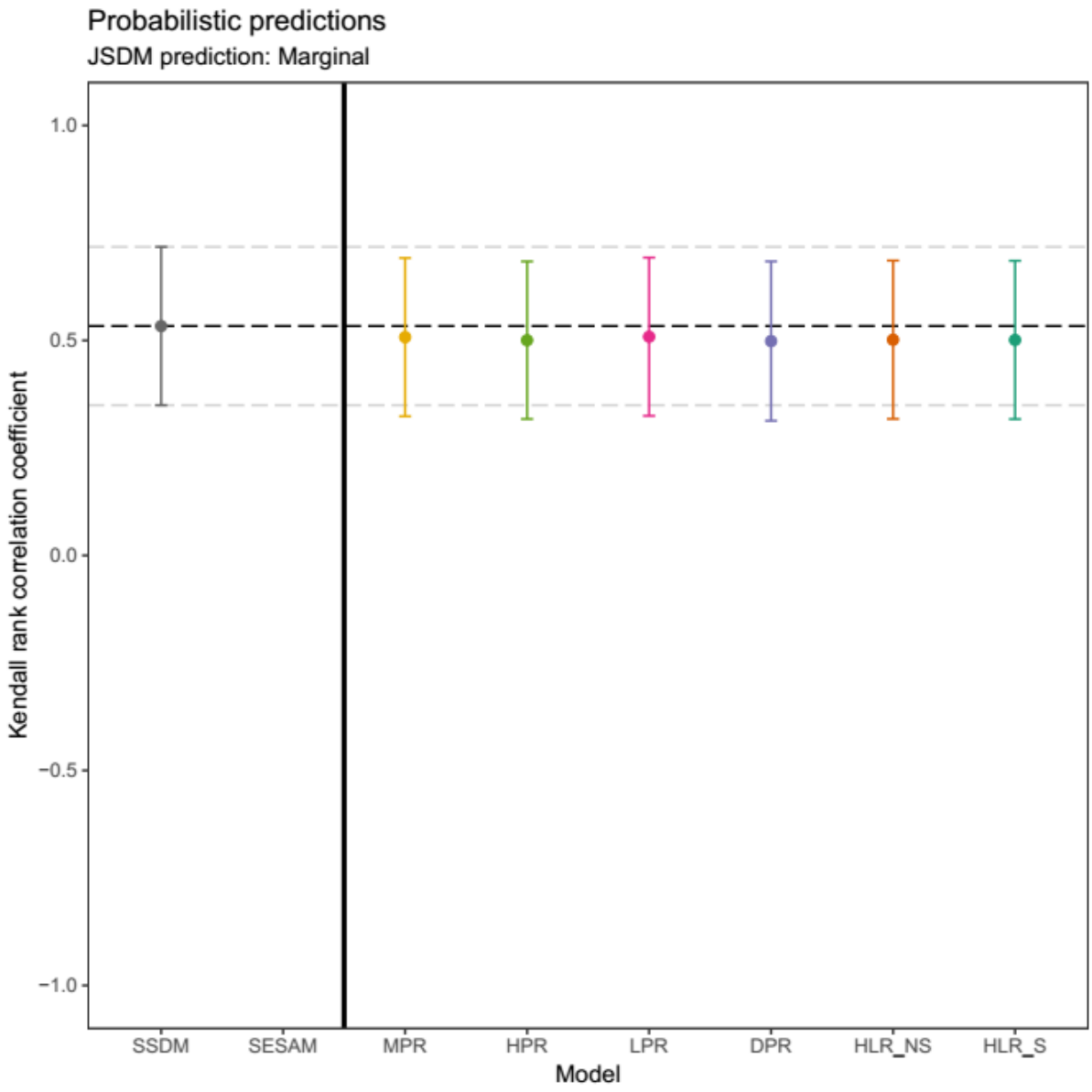


Figure S16: Model performance for the Kendall rank correlation coefficient evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

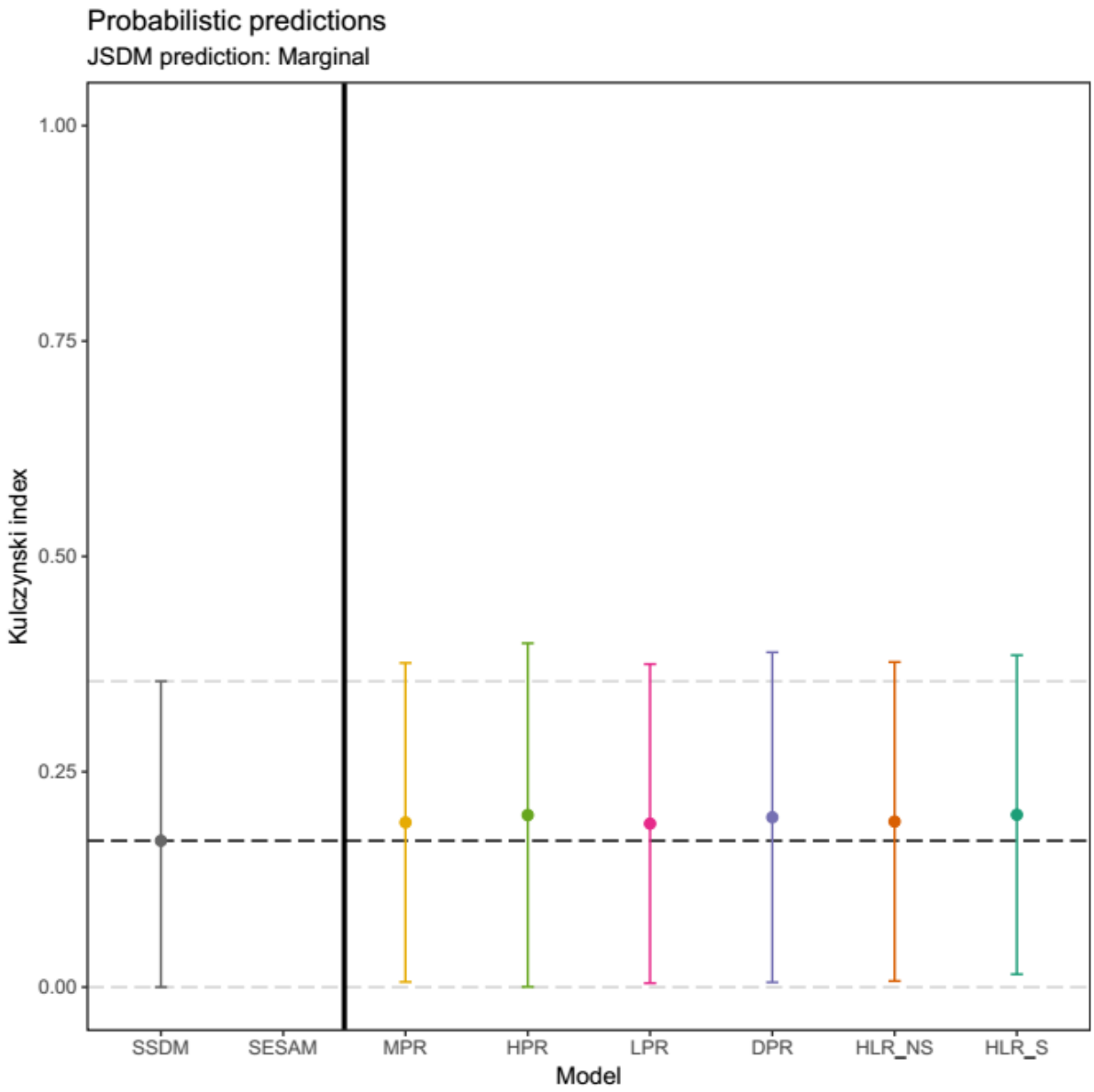


Figure S17: Model performance for the Kulczynski index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

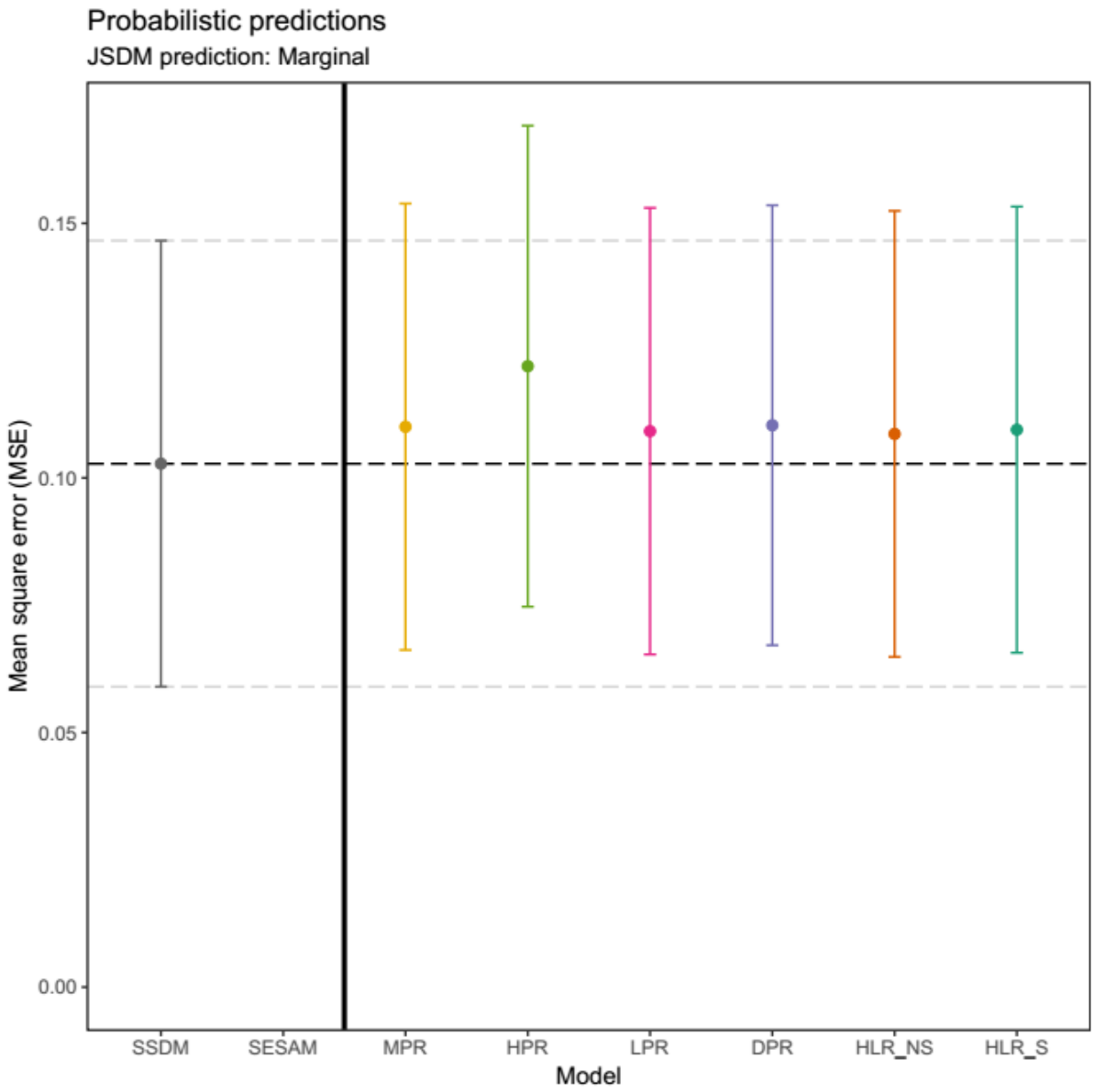


Figure S18: Model performance for the mean square error evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

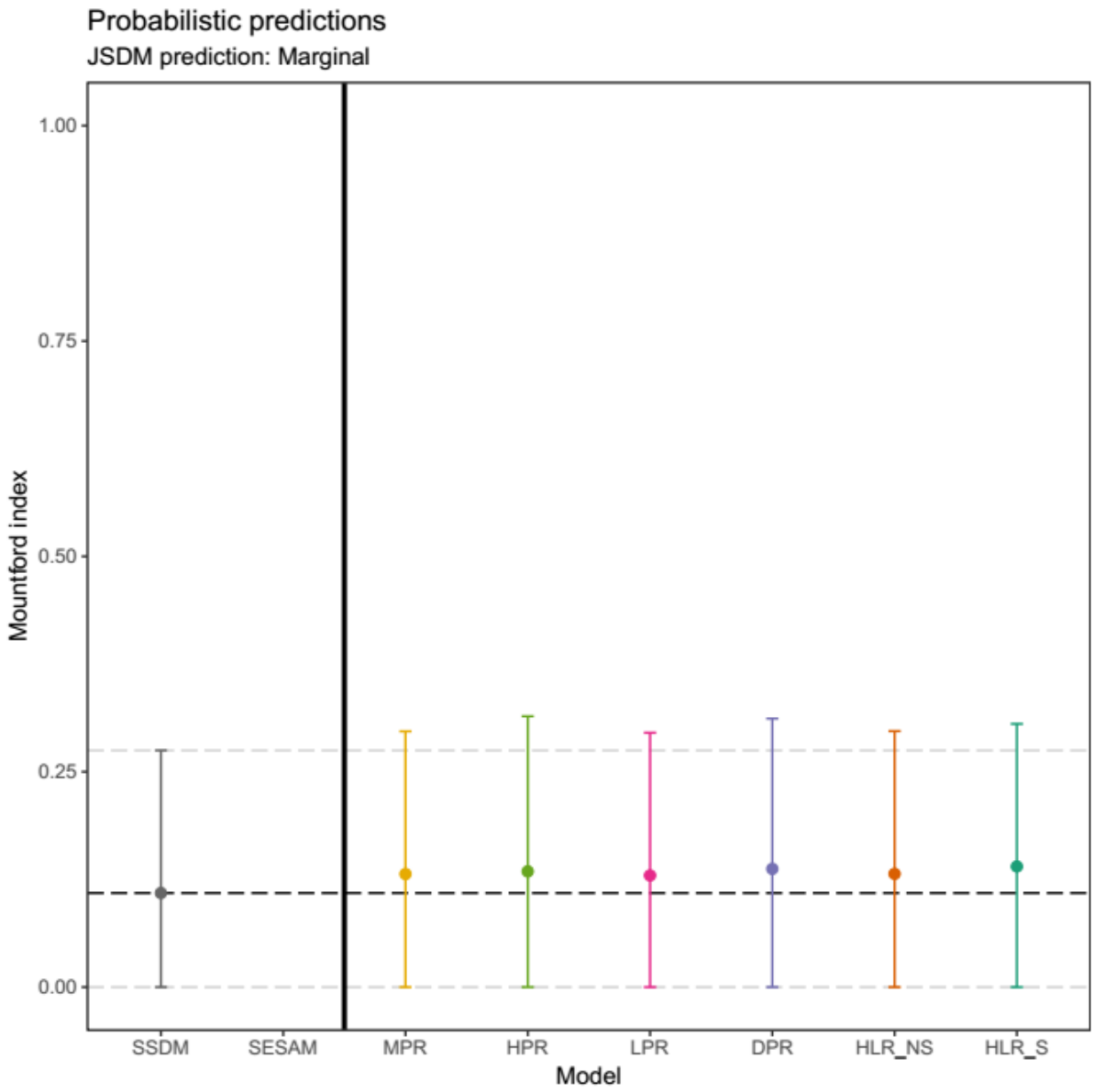


Figure S19: Model performance for the Mountford index evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

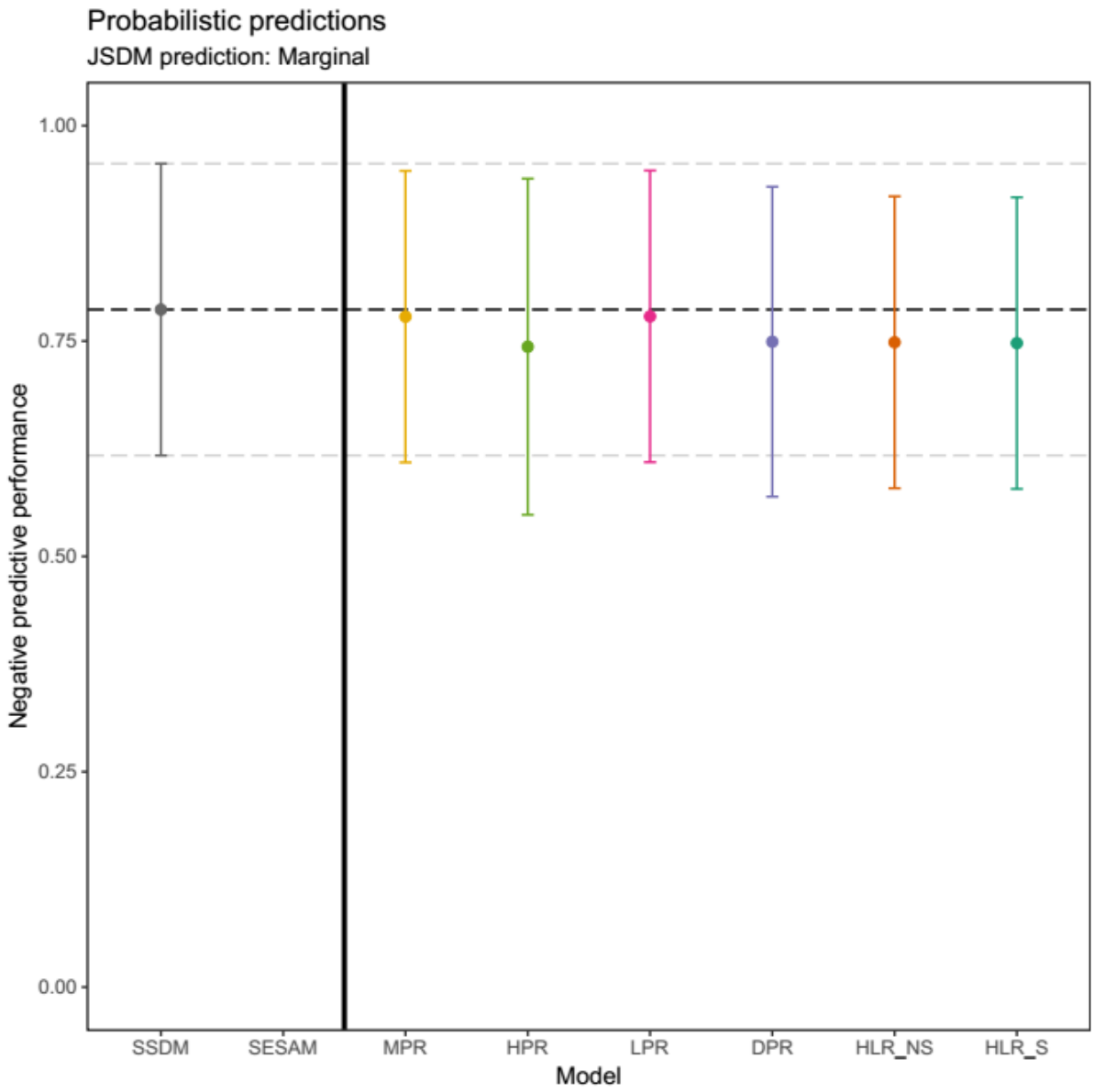


Figure S20: Model performance for the negative predictive performance evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

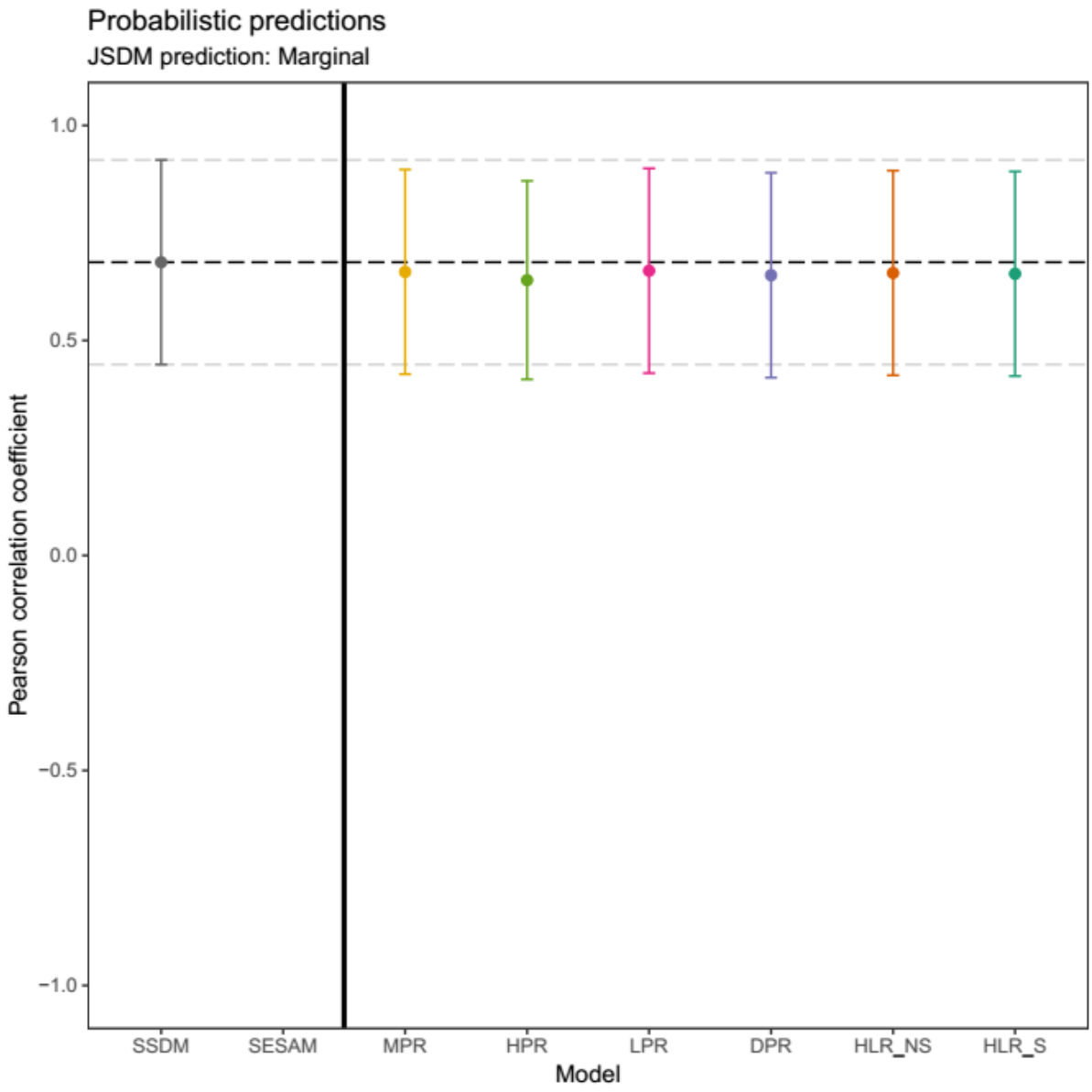


Figure S21: Model performance for the Pearson correlation coefficient evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

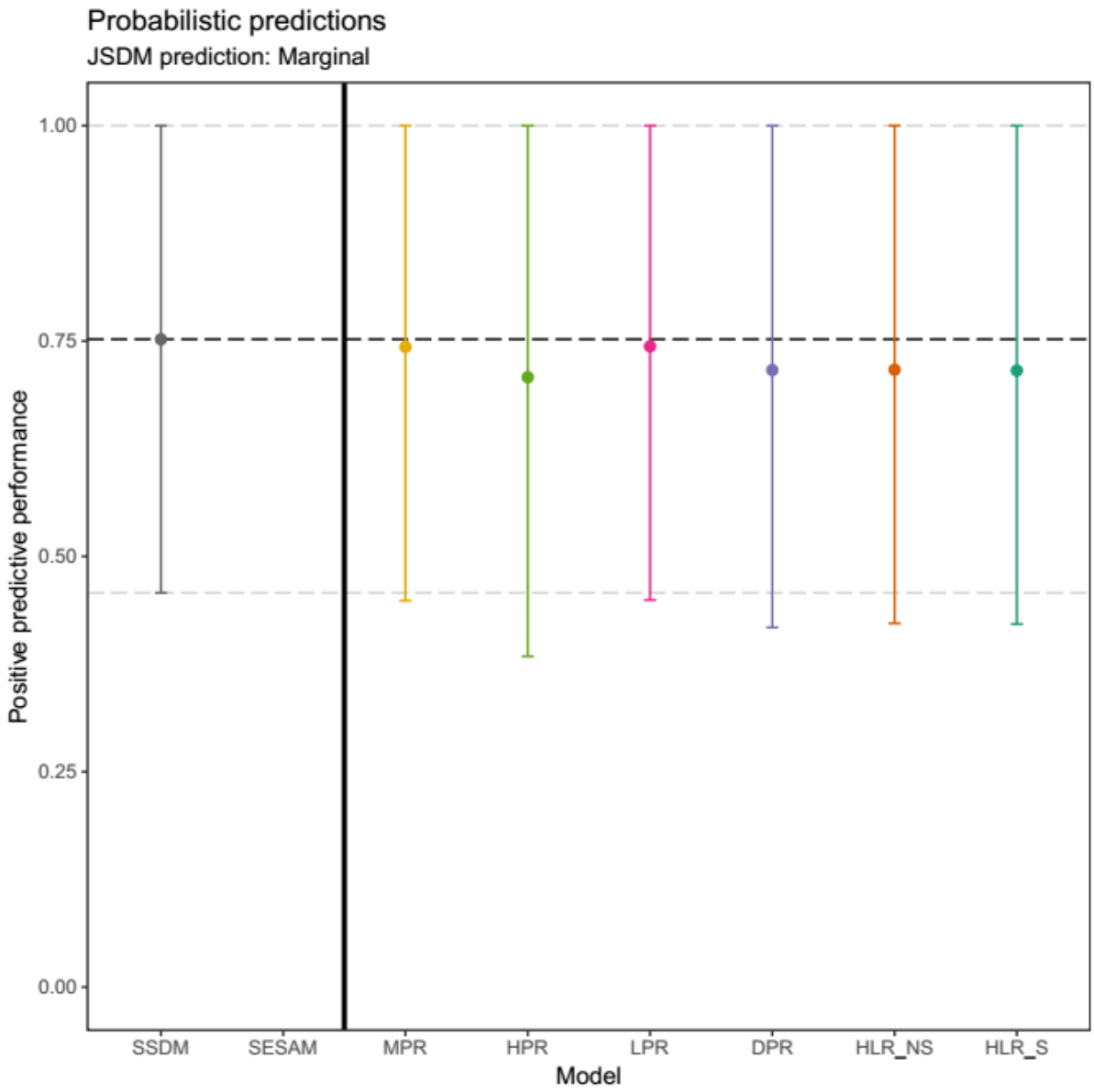


Figure S22: Model performance for the positive predictive performance evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

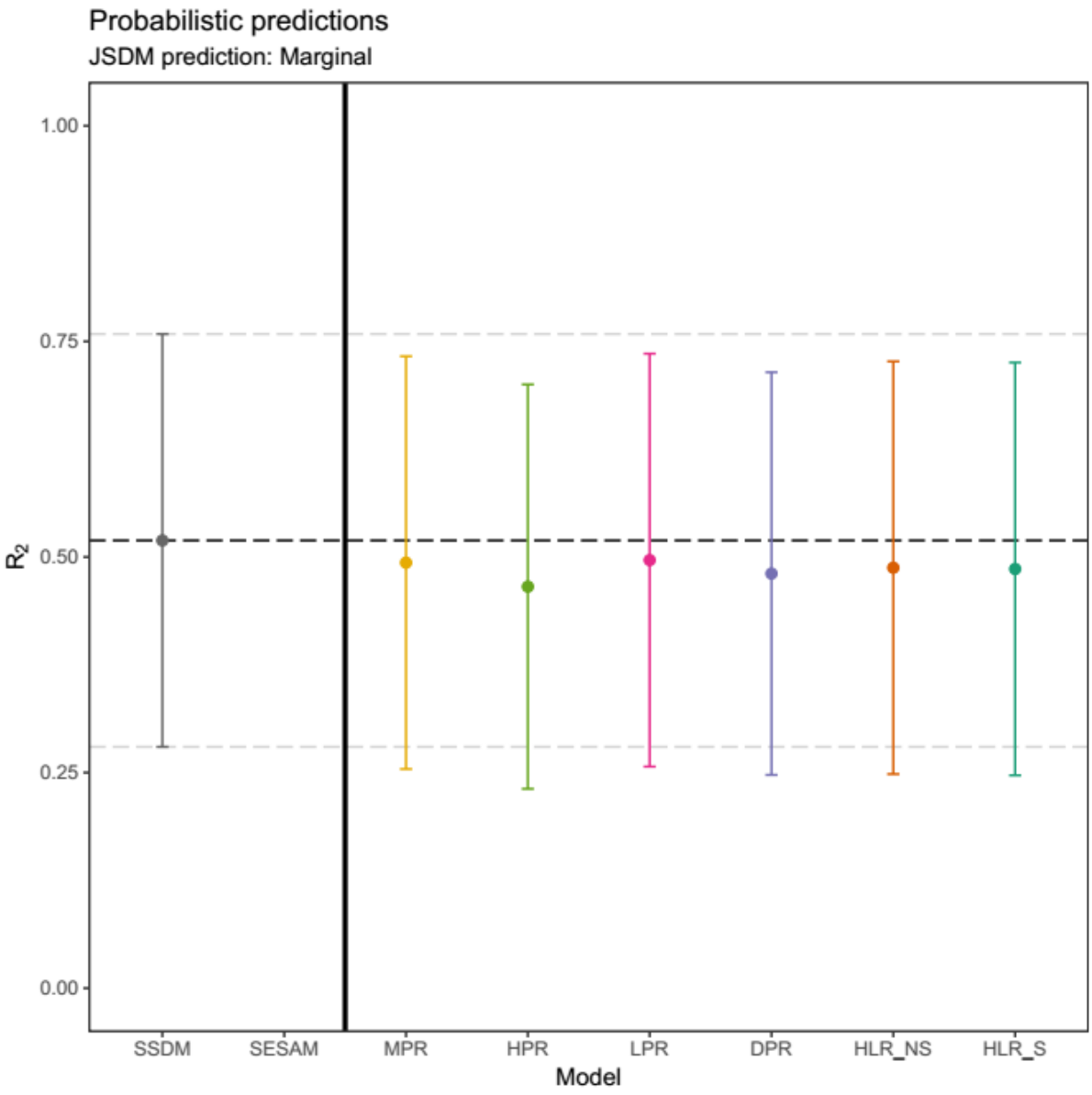


Figure S23: Model performance for the R^2 evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

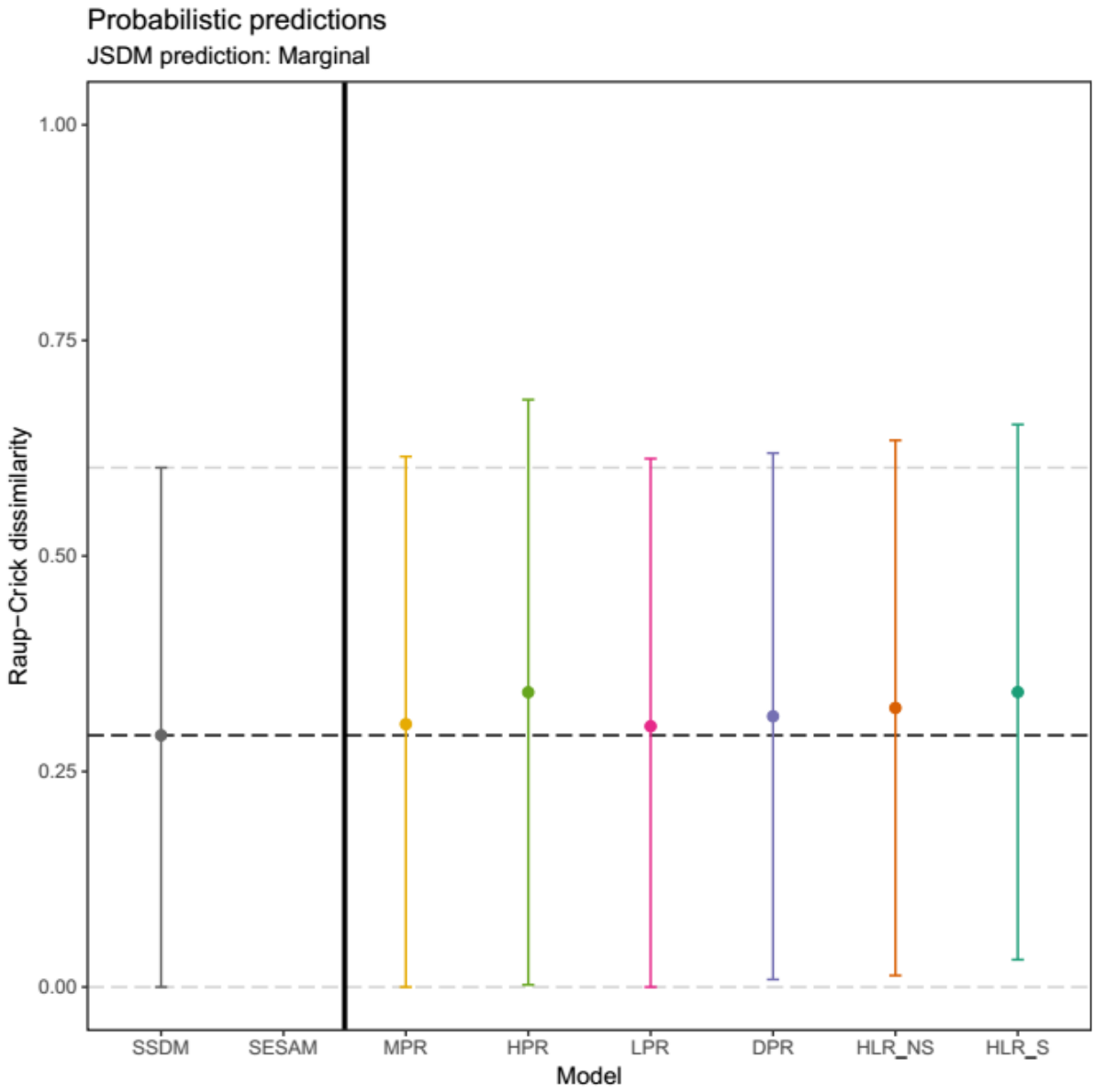


Figure S24: Model performance for the Raup-Crick dissimilarity evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

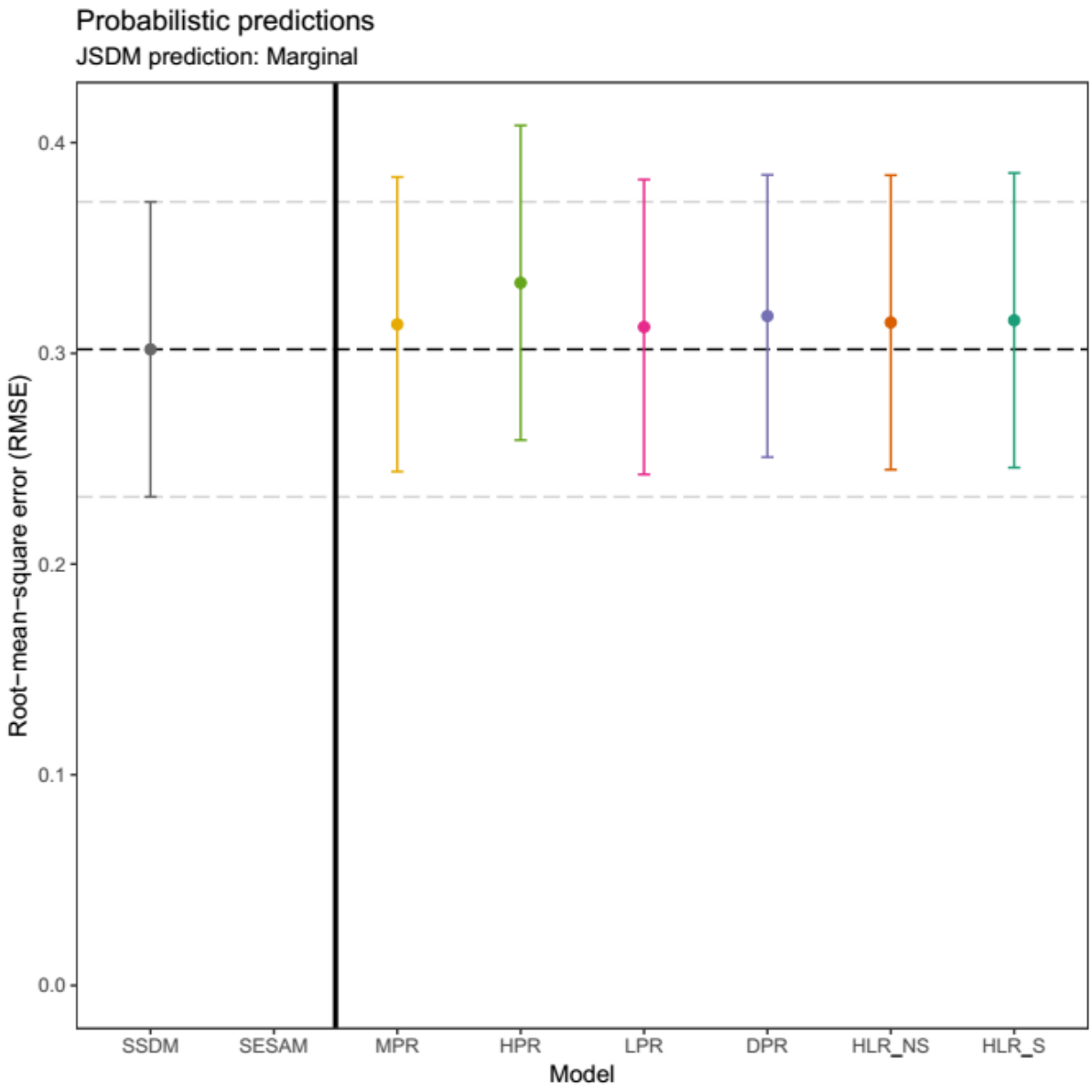


Figure S25: Model performance for the root-mean-square error evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

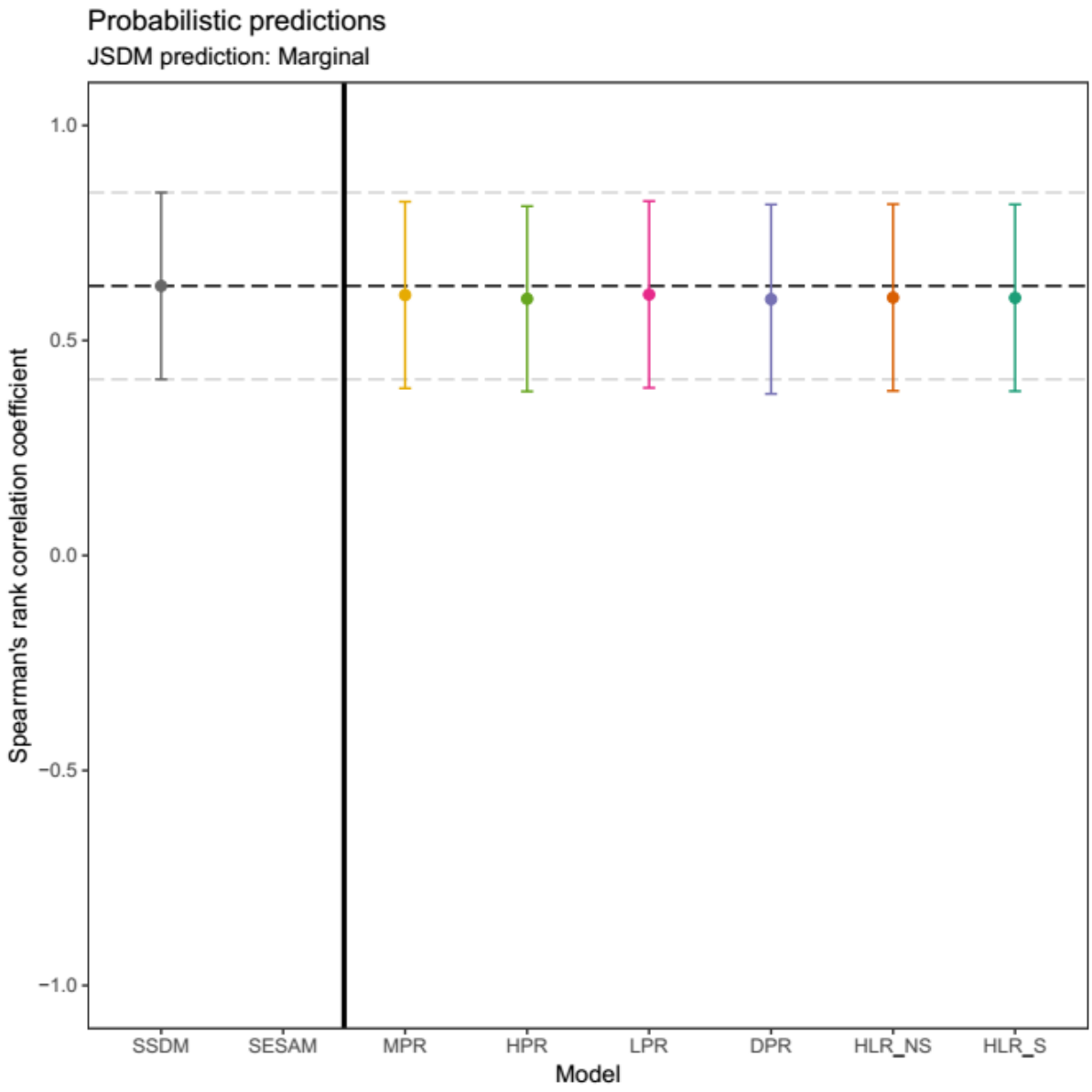


Figure S25: Model performance for Pearson's rank correlation coefficient evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

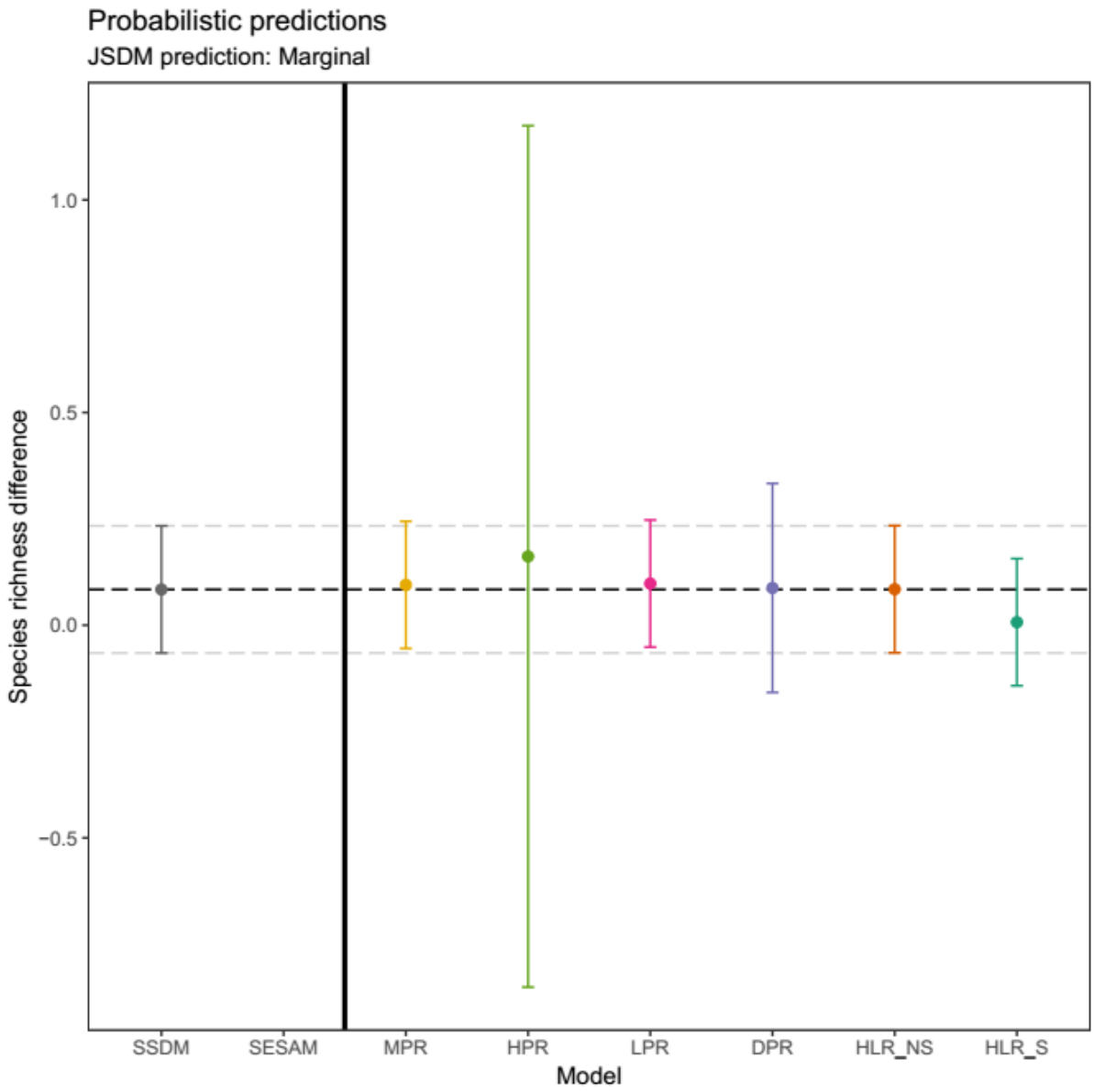


Figure S27: Model performance for the species richness difference evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

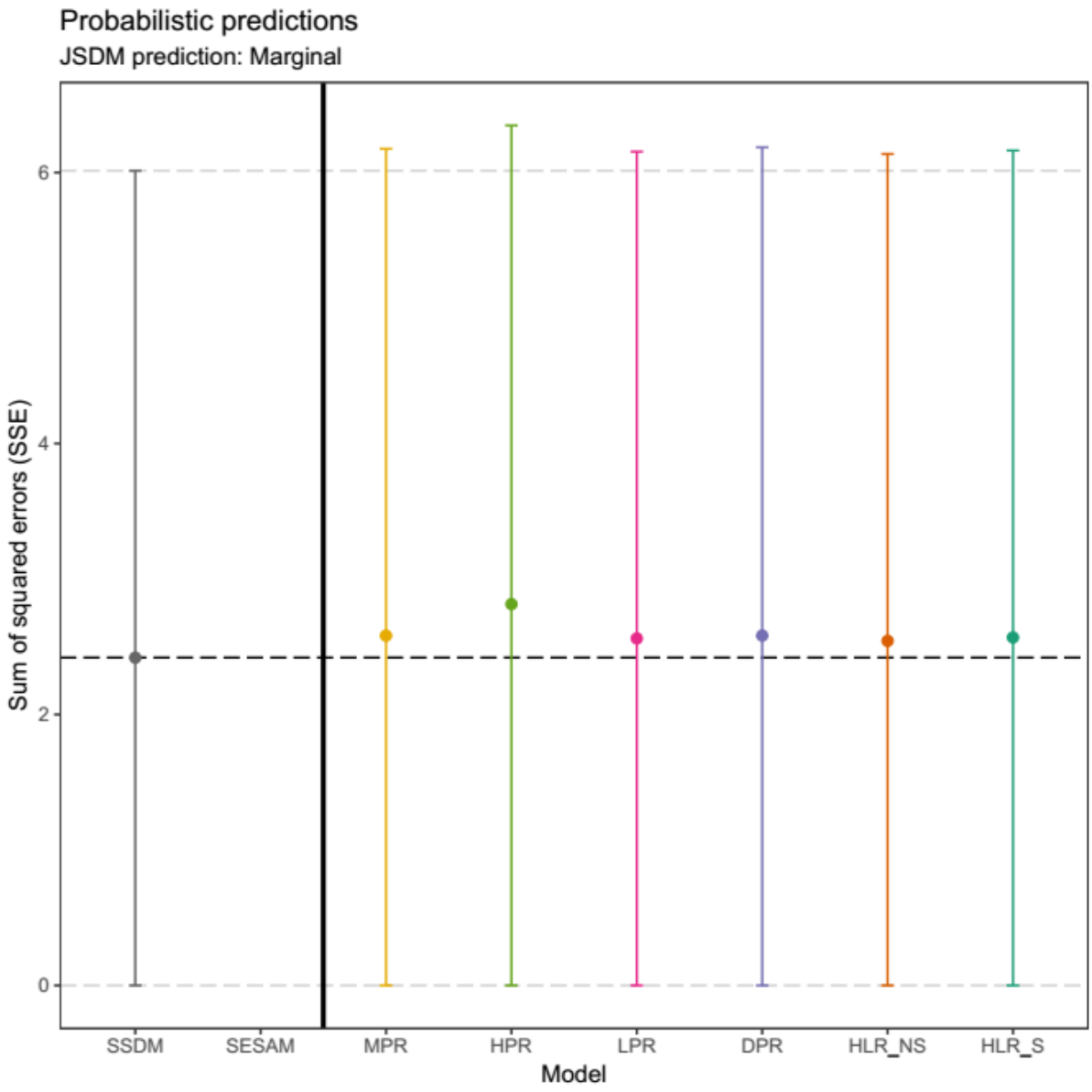


Figure S28: Model performance for the sum of square errors evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

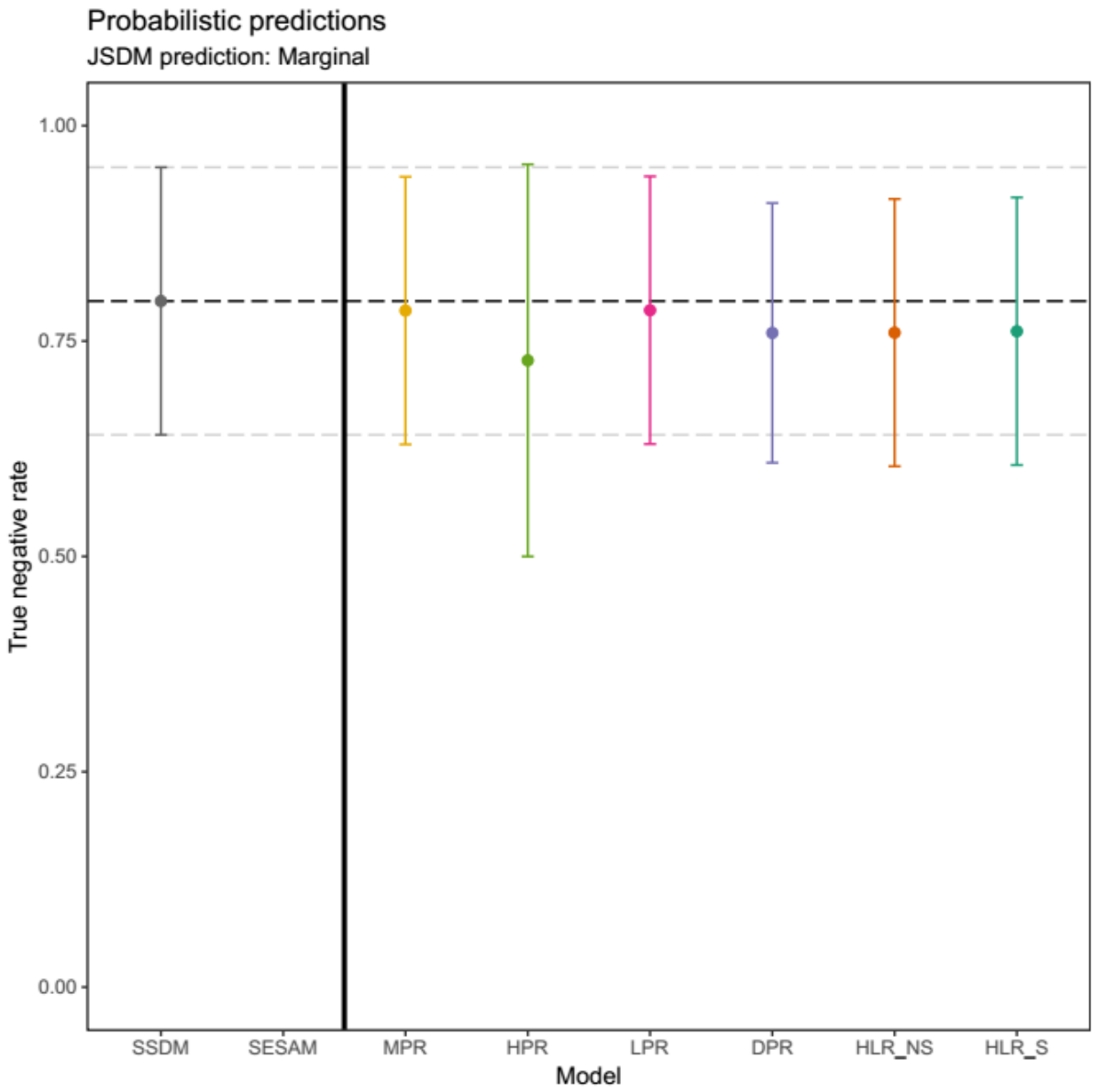


Figure S29: Model performance for the true negative rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

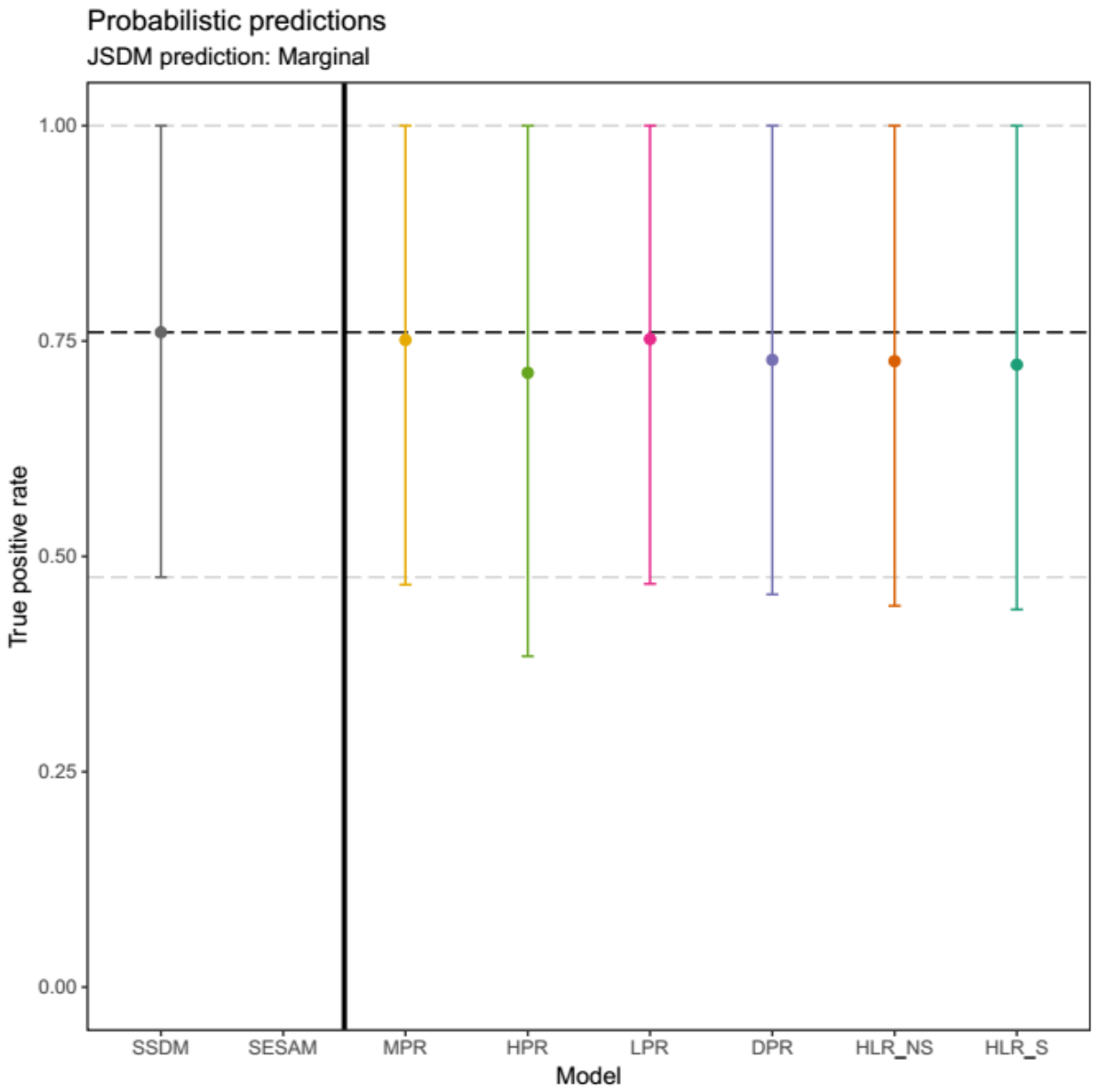


Figure S30: Model performance for the true positive rate evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

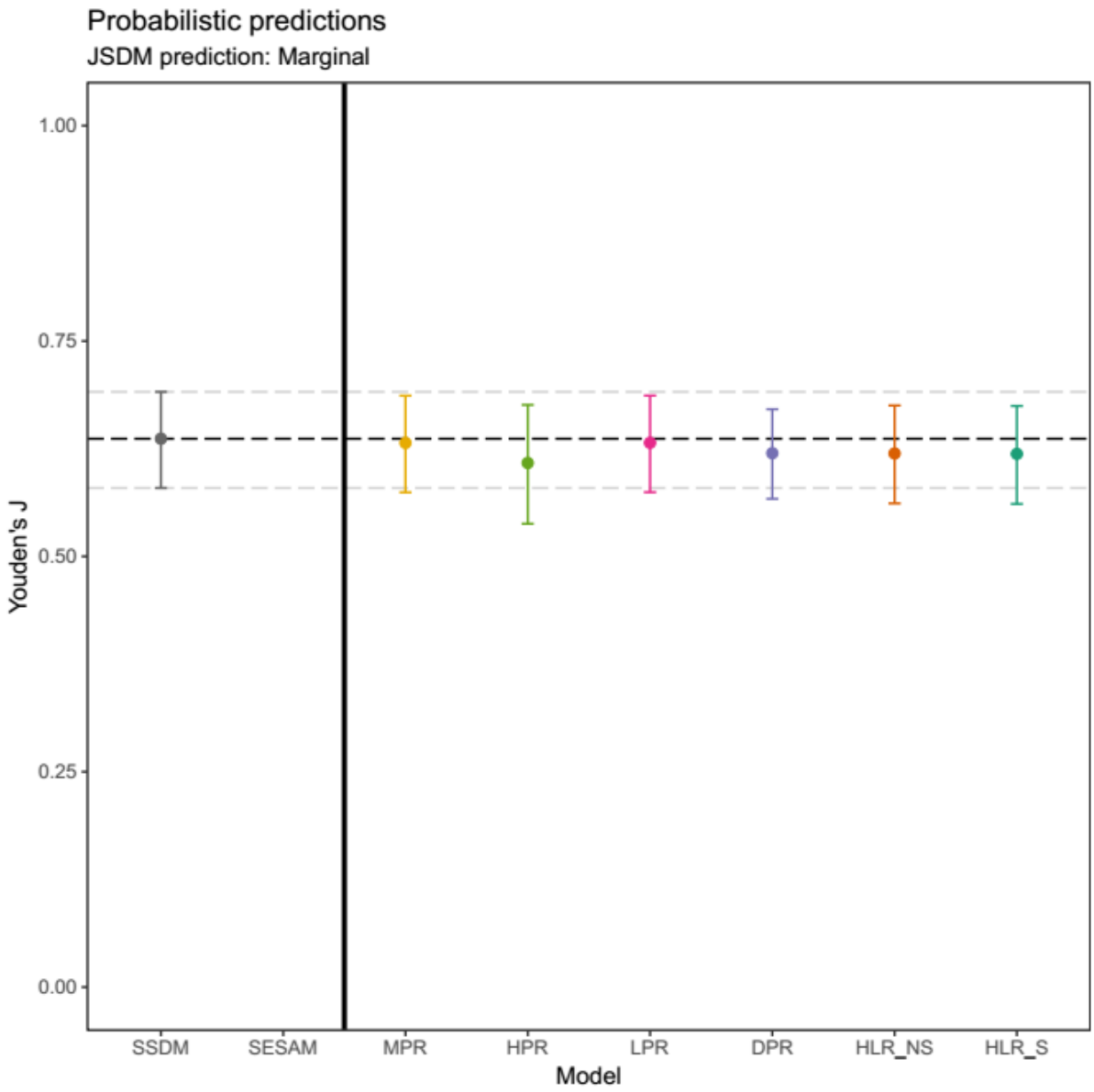


Figure S31: Model performance for the Youden's J evaluation metric for probabilistic marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

2. Marginal (binary) predictions

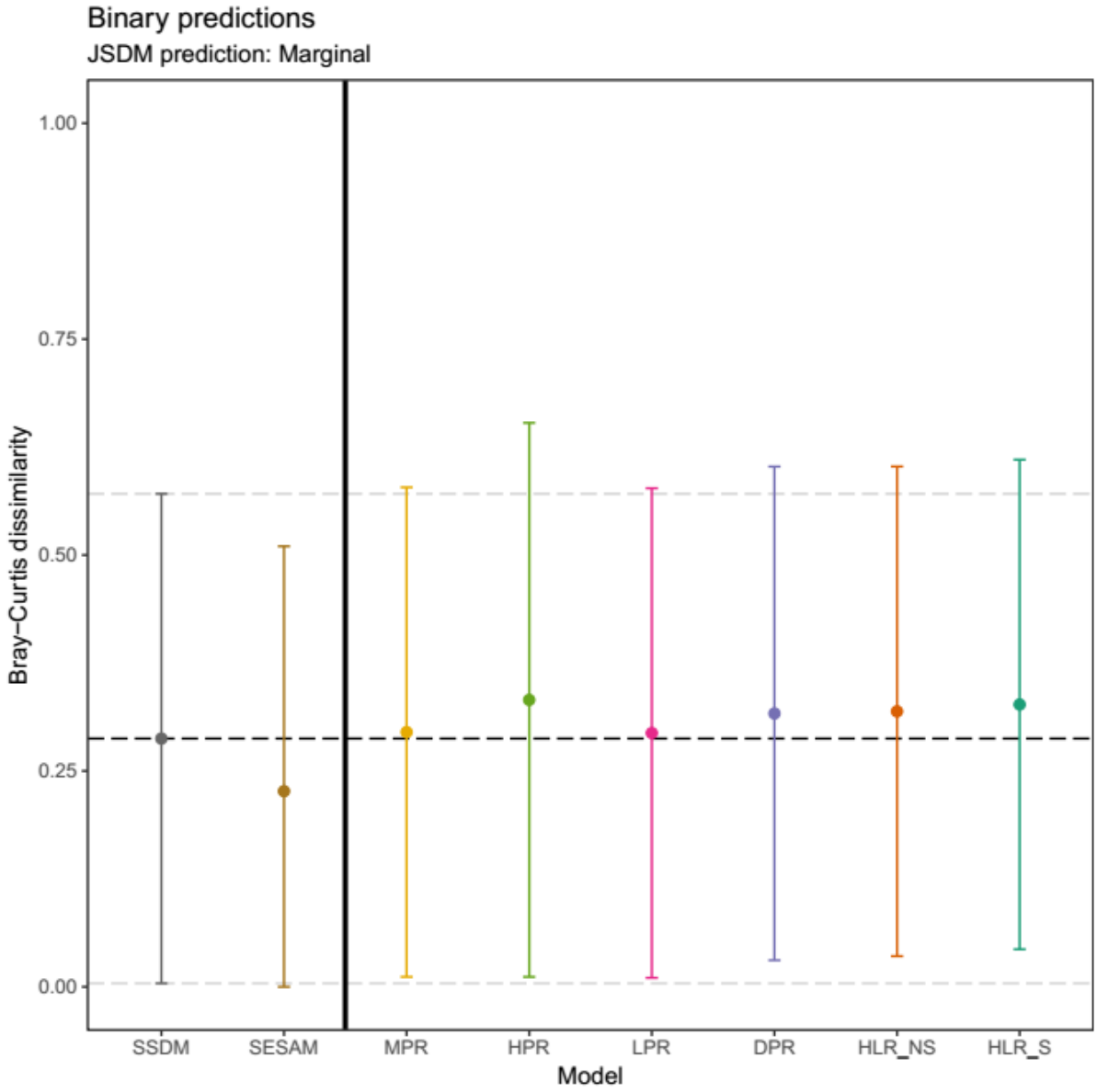


Figure S32: Model performance for the Bray-Curtis dissimilarity evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

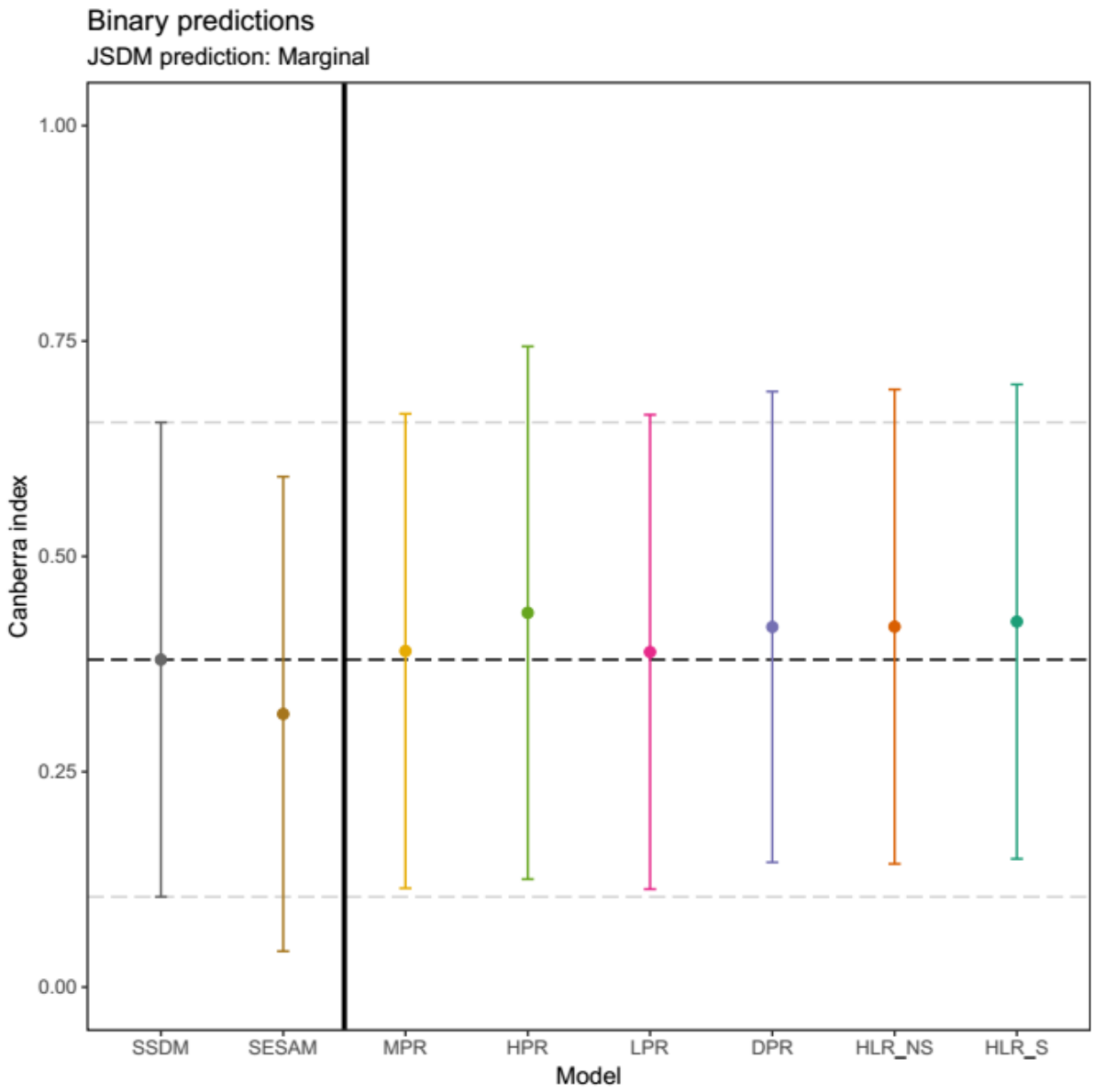


Figure S33: Model performance for the Canberra index dissimilarity evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

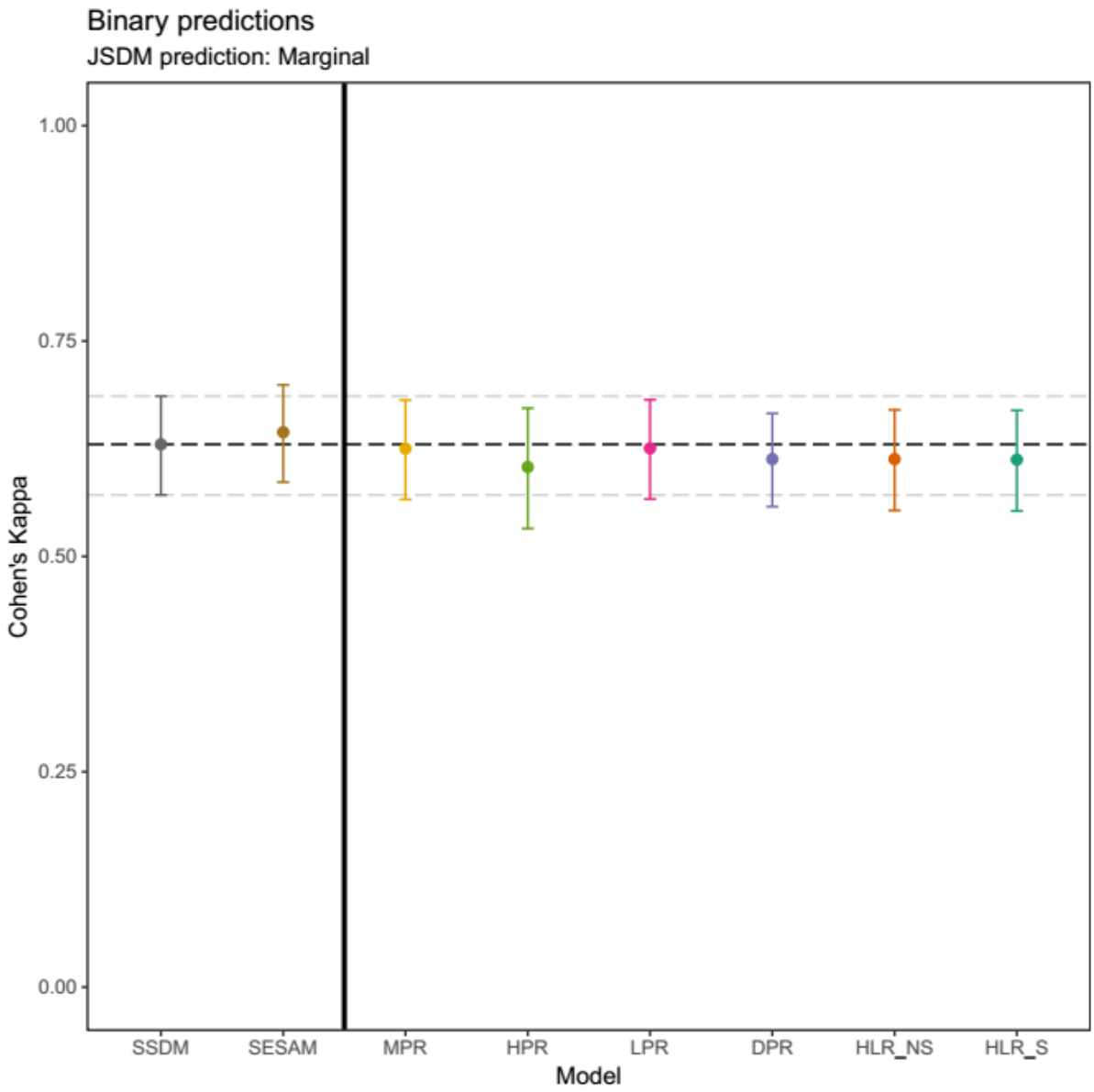


Figure S34: Model performance for Cohen's Kappa evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

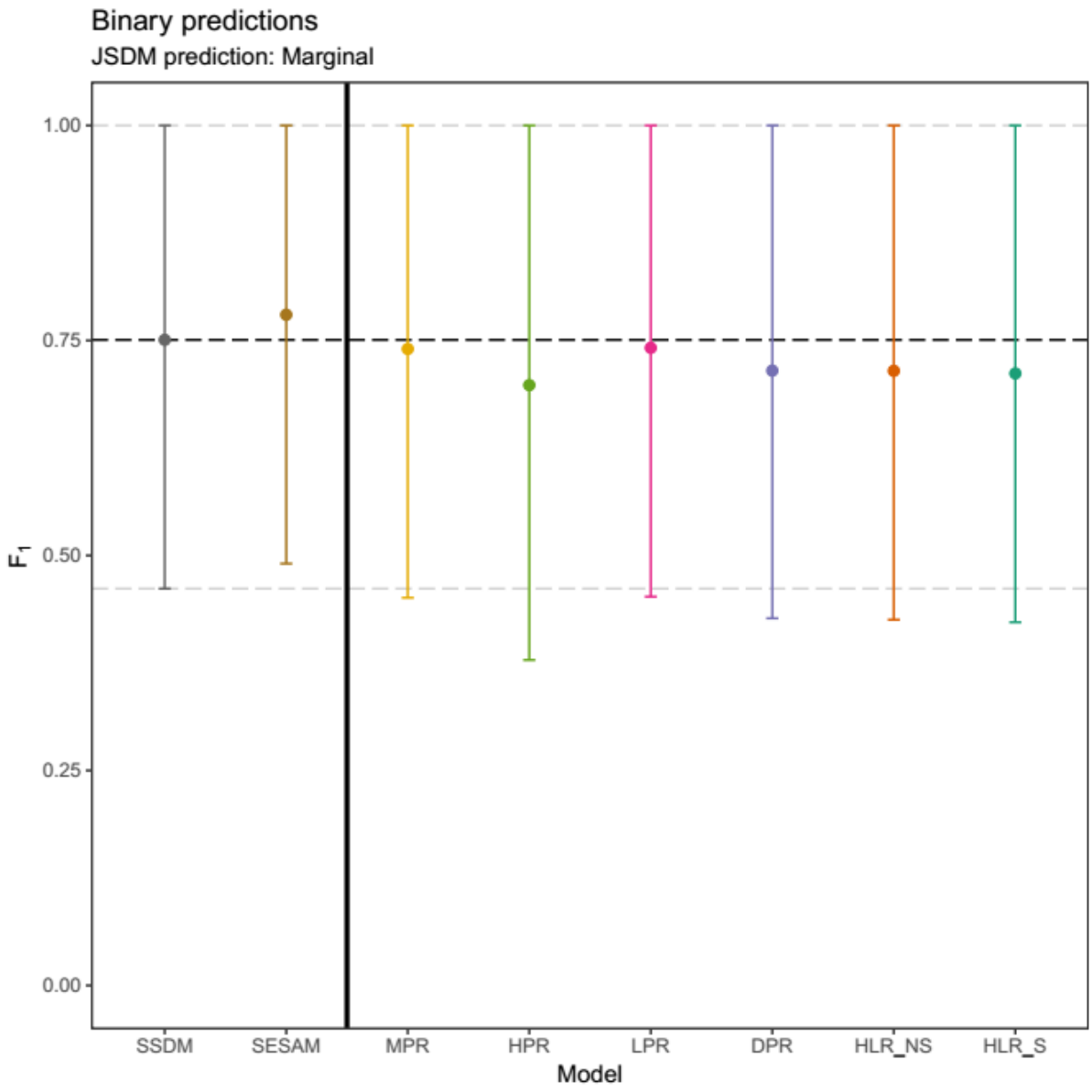


Figure S35: Model performance for the F_1 evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

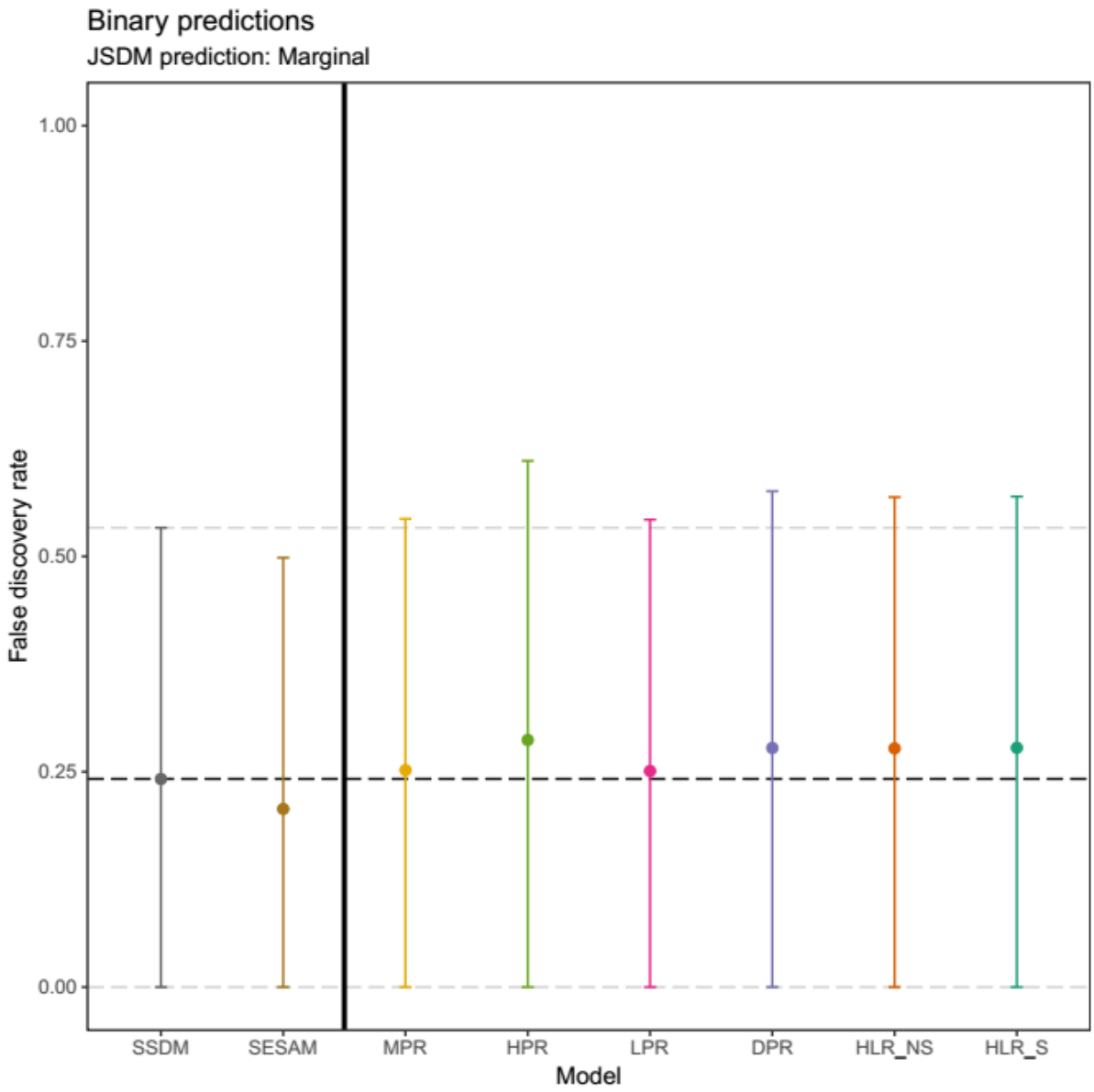


Figure S36: Model performance for the false discovery rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

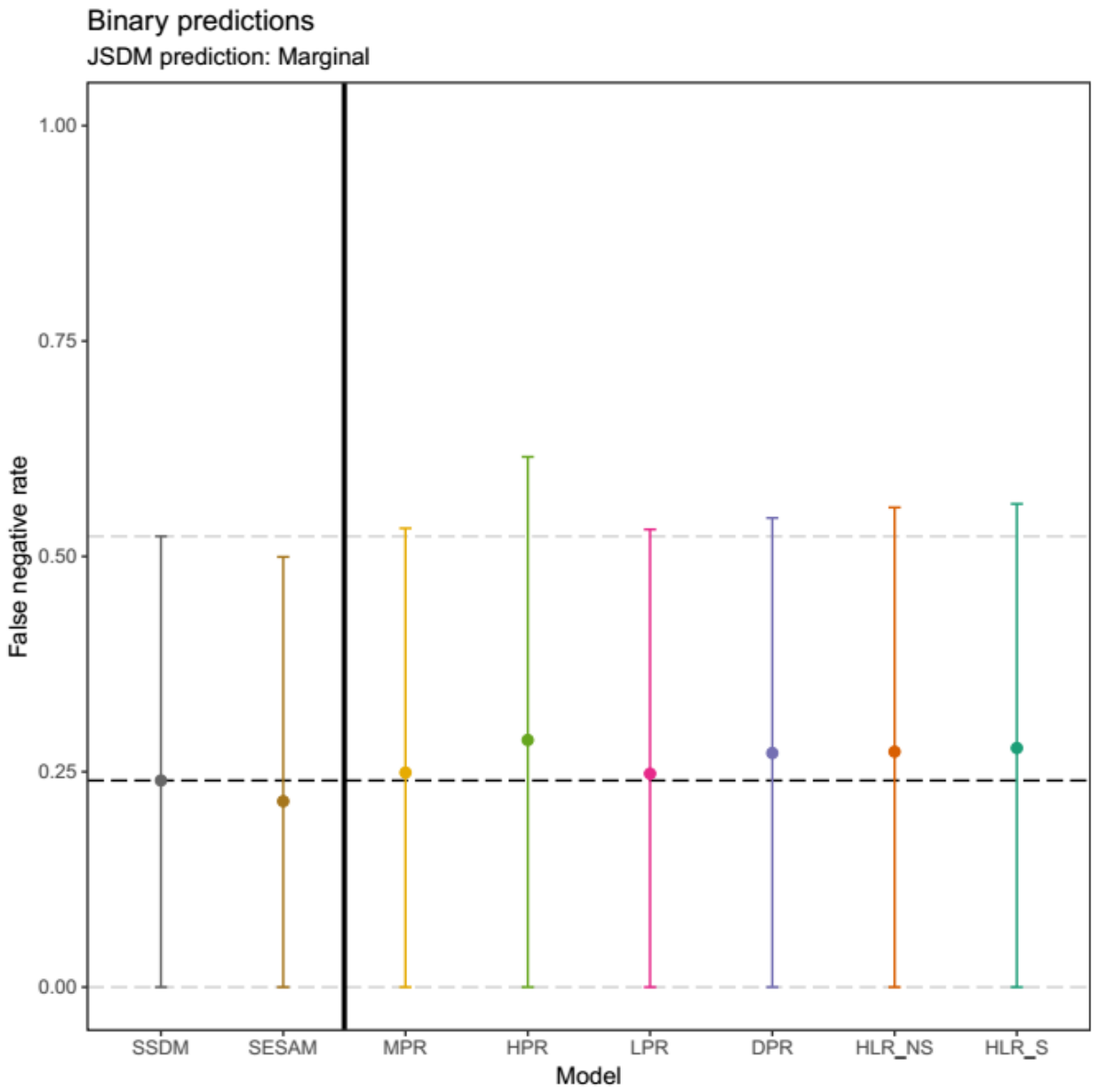


Figure S37: Model performance for the false negative rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

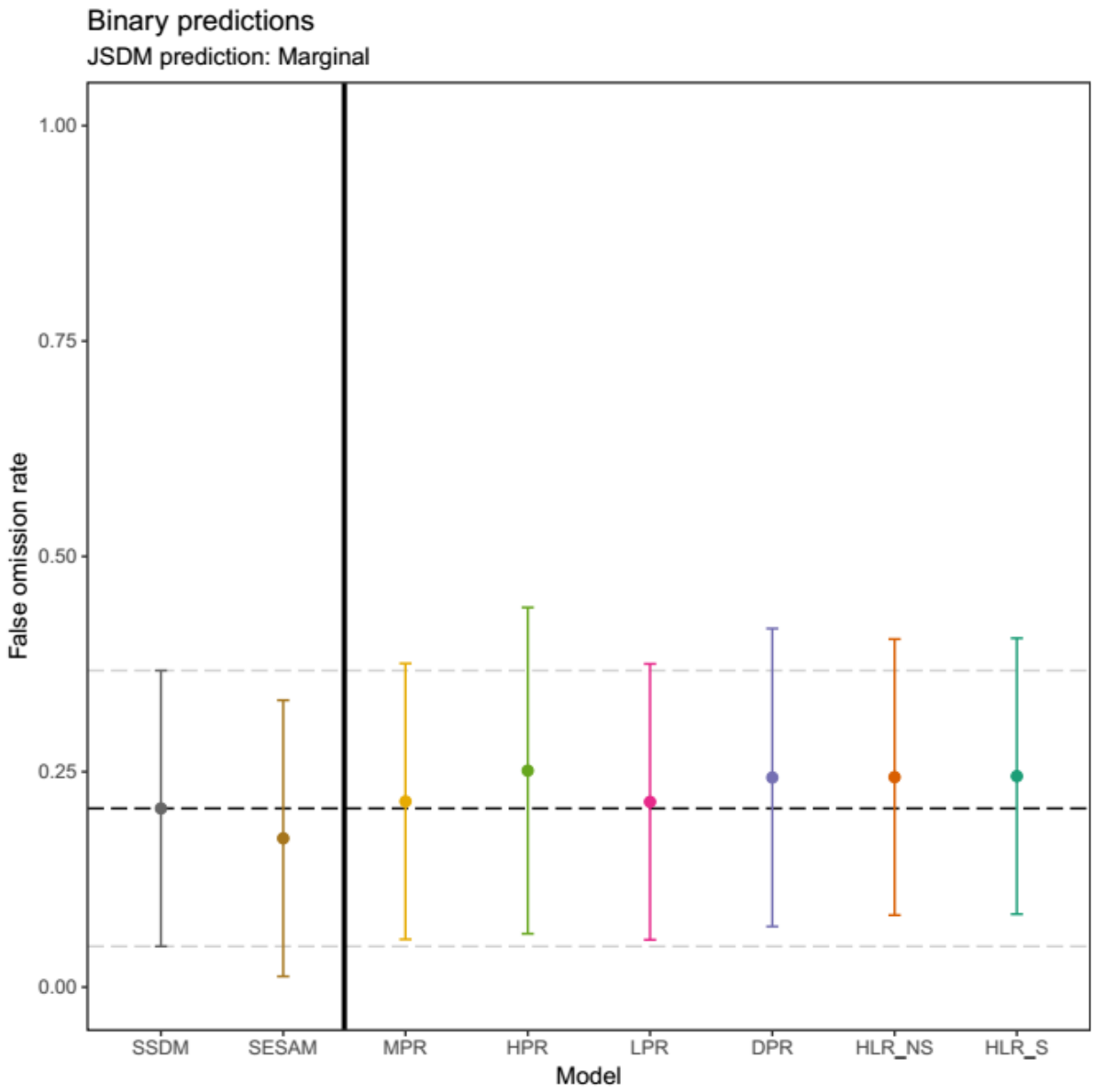


Figure S38: Model performance for the false omission rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

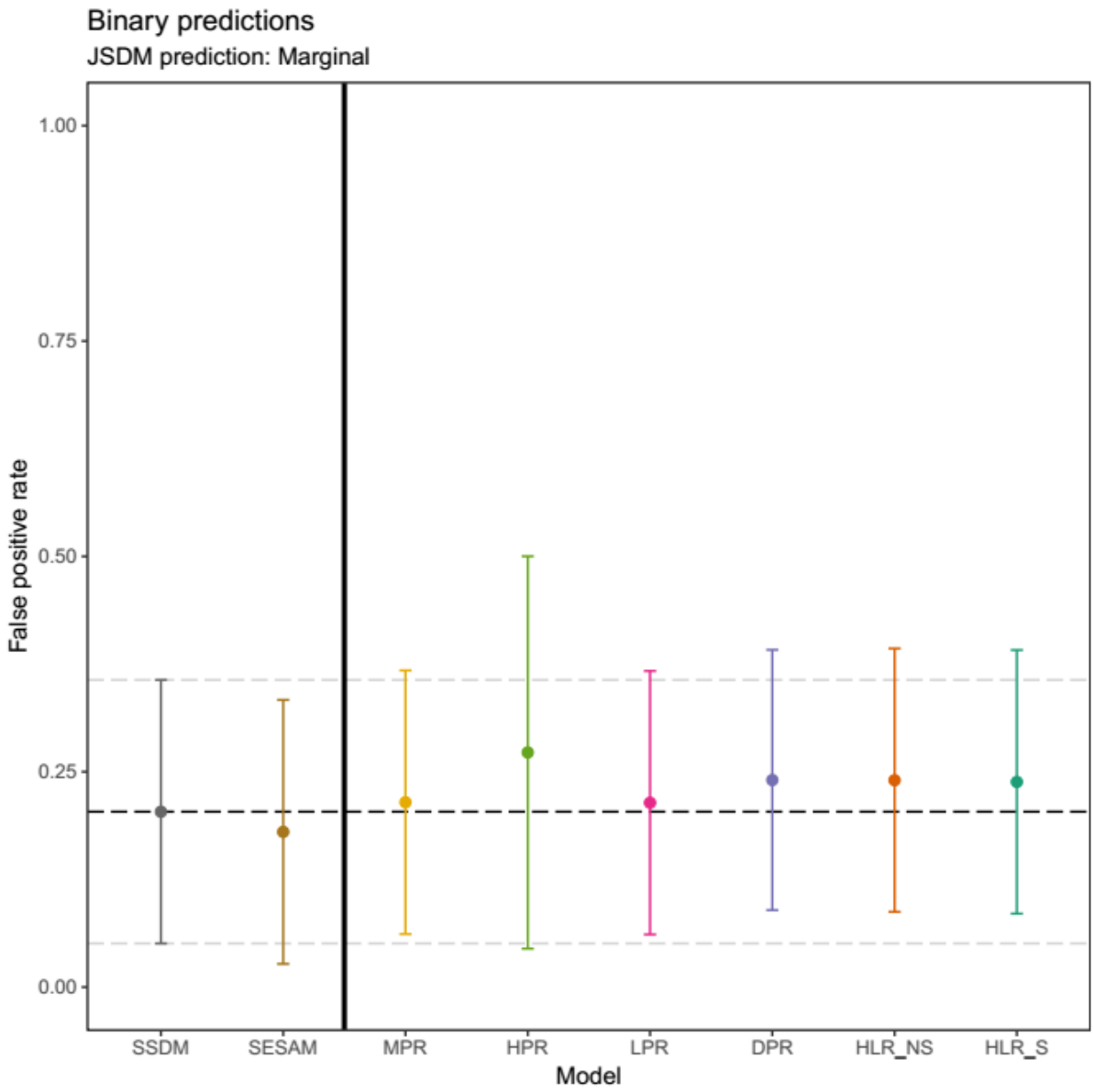


Figure S39: Model performance for the false positive rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

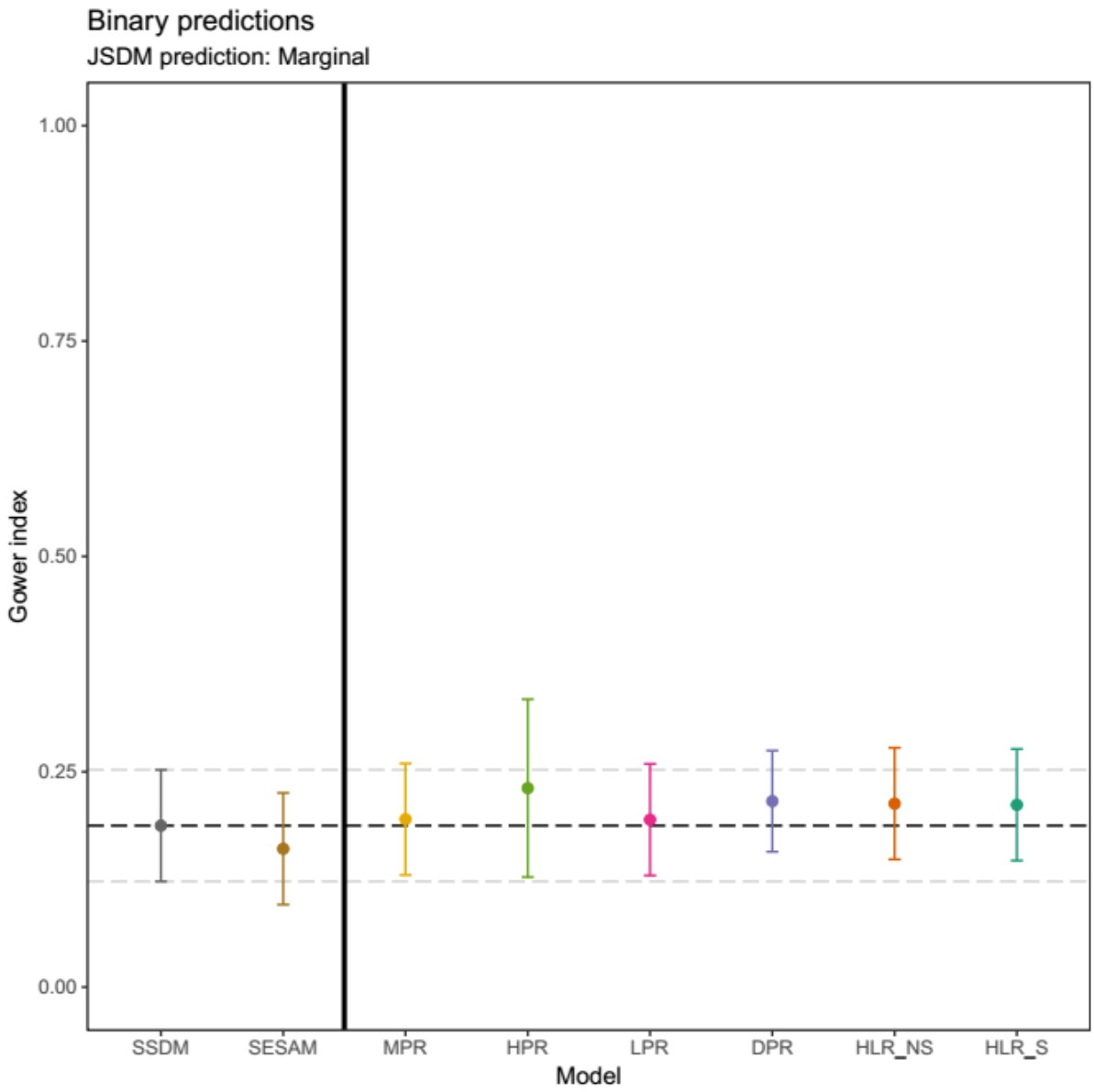


Figure S40: Model performance for the Gower index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

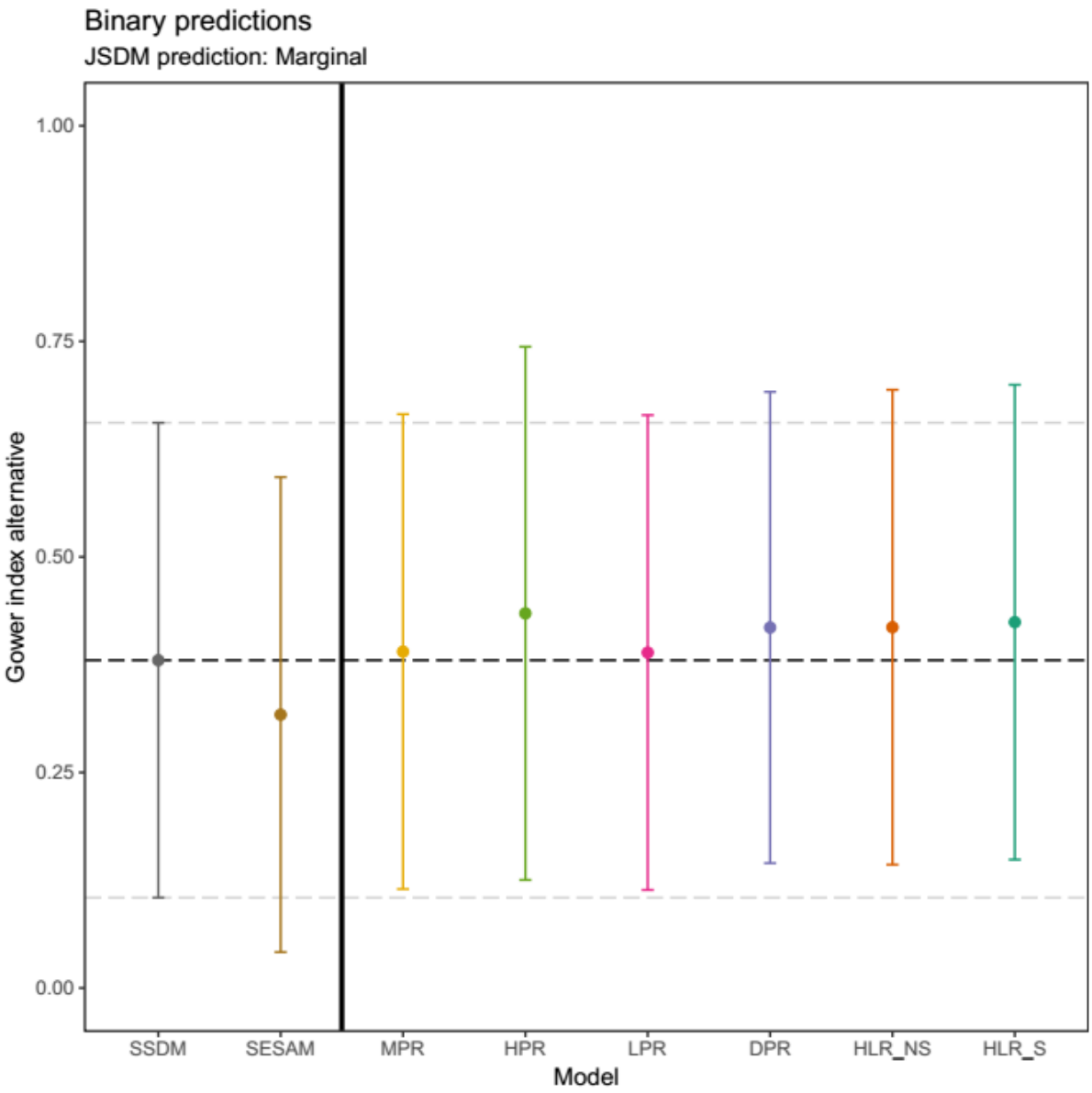


Figure S41: Model performance for the alternative Gower index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

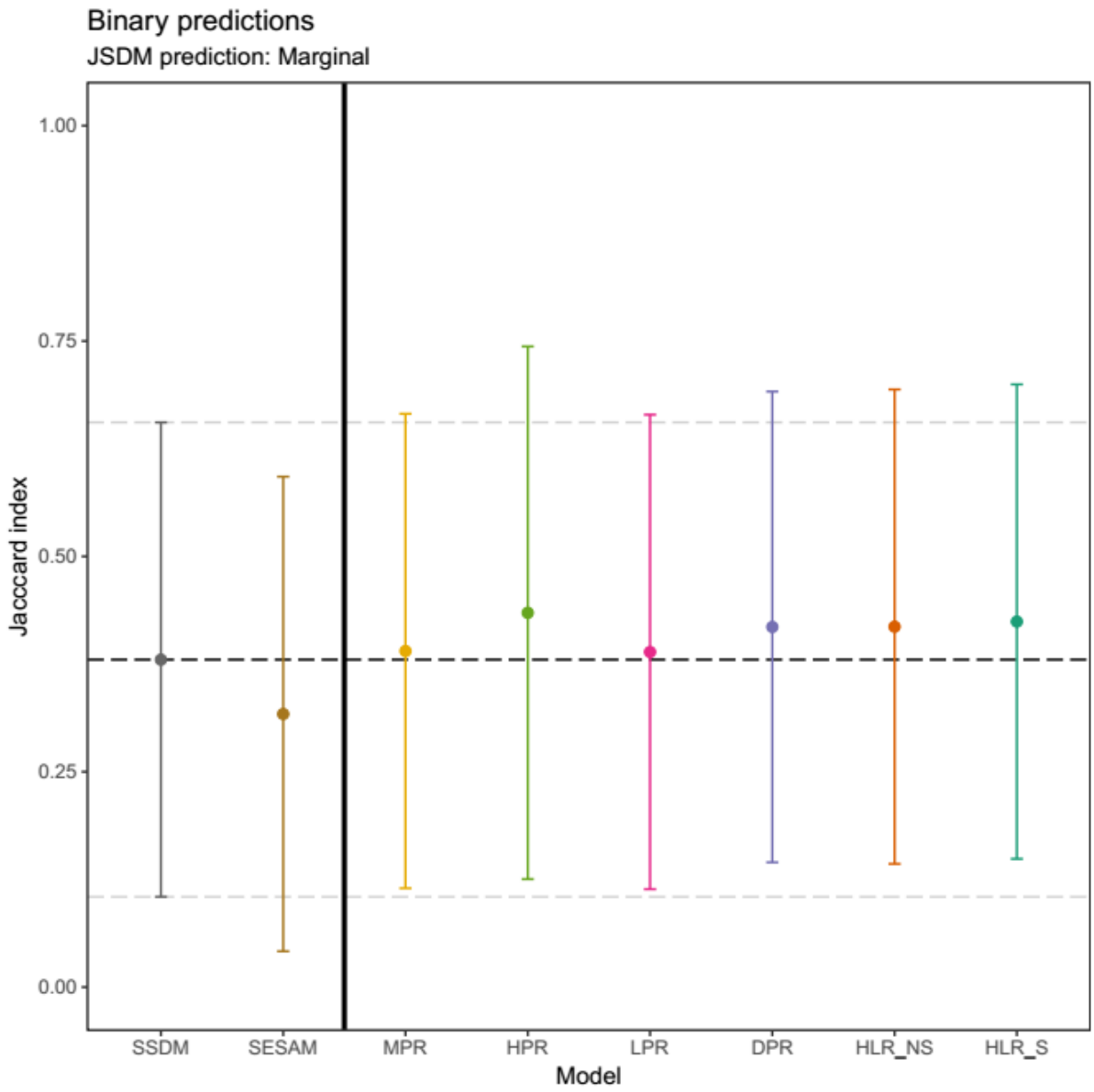


Figure S42: Model performance for the Jaccard index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

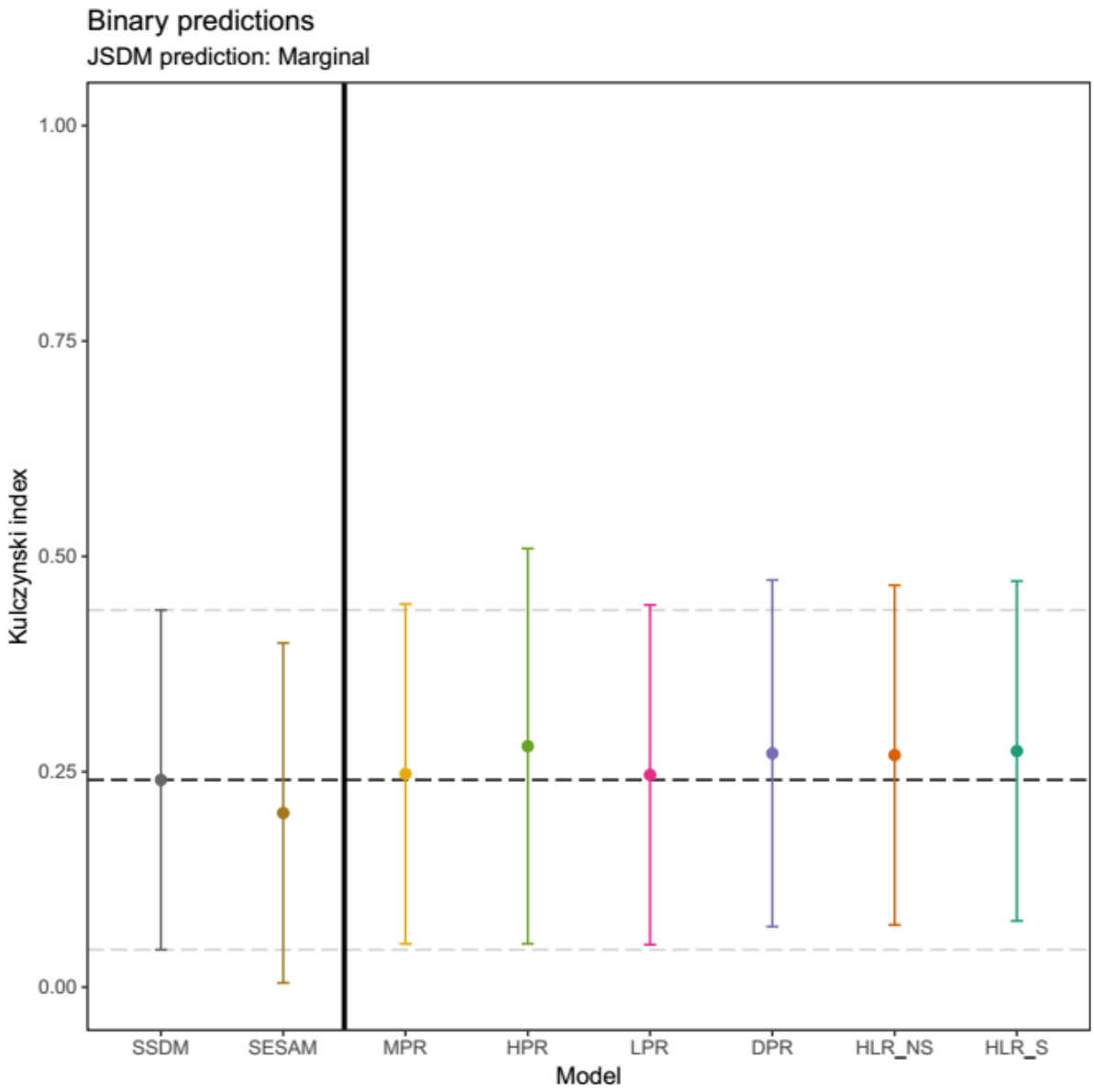


Figure S43: Model performance for the Kulczynski index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

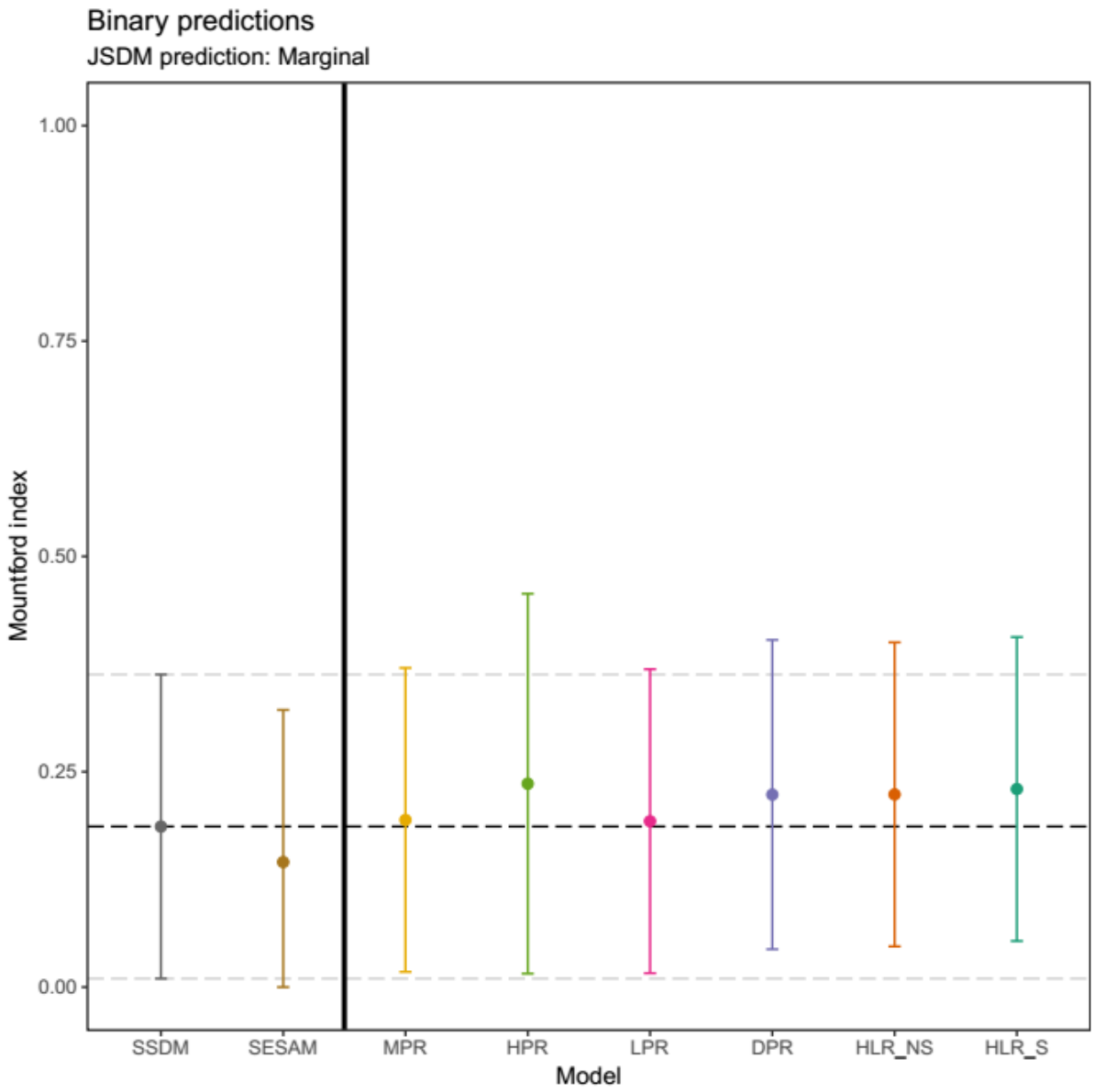


Figure S44: Model performance for the Mountford index evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

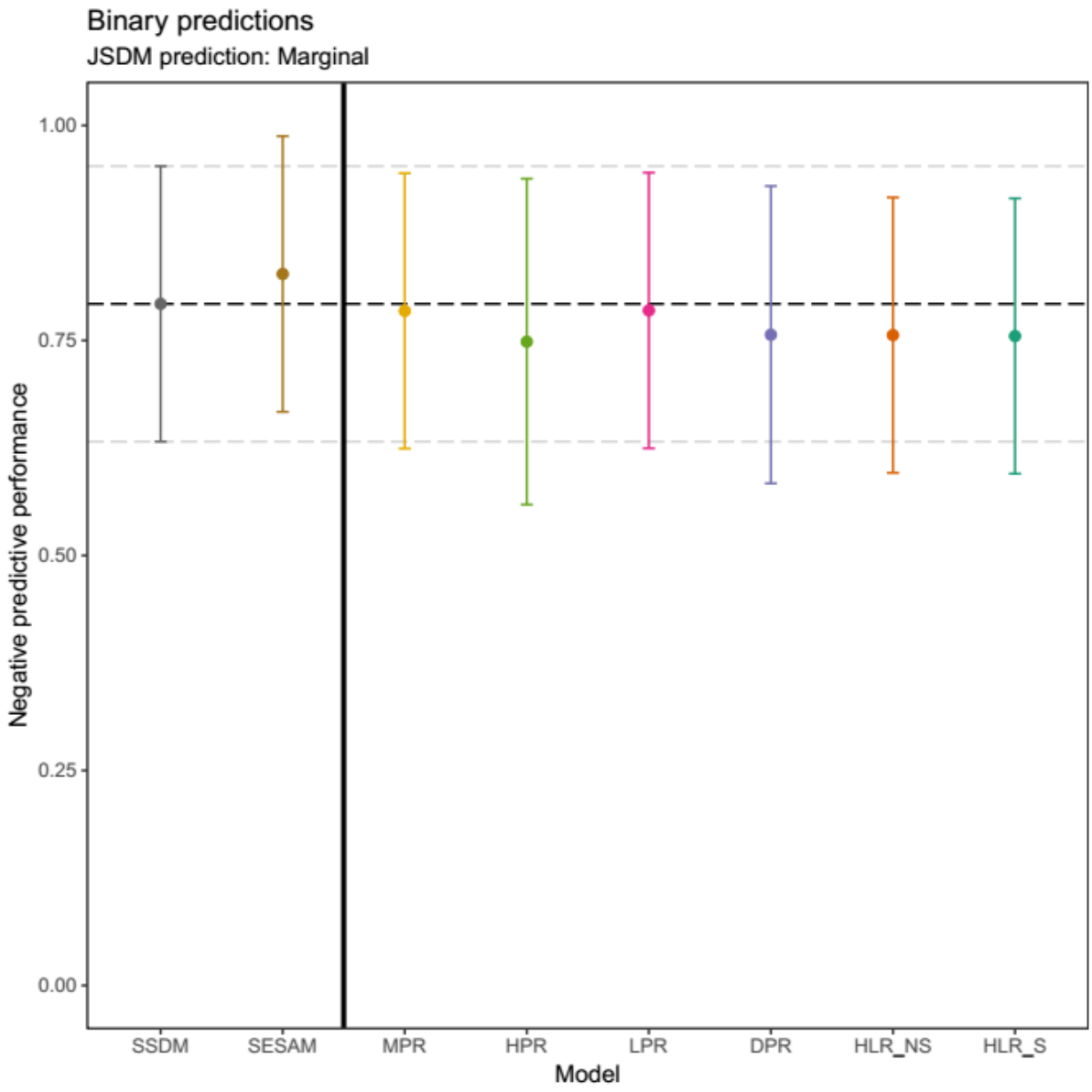


Figure S45: Model performance for the negative predictive performance evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

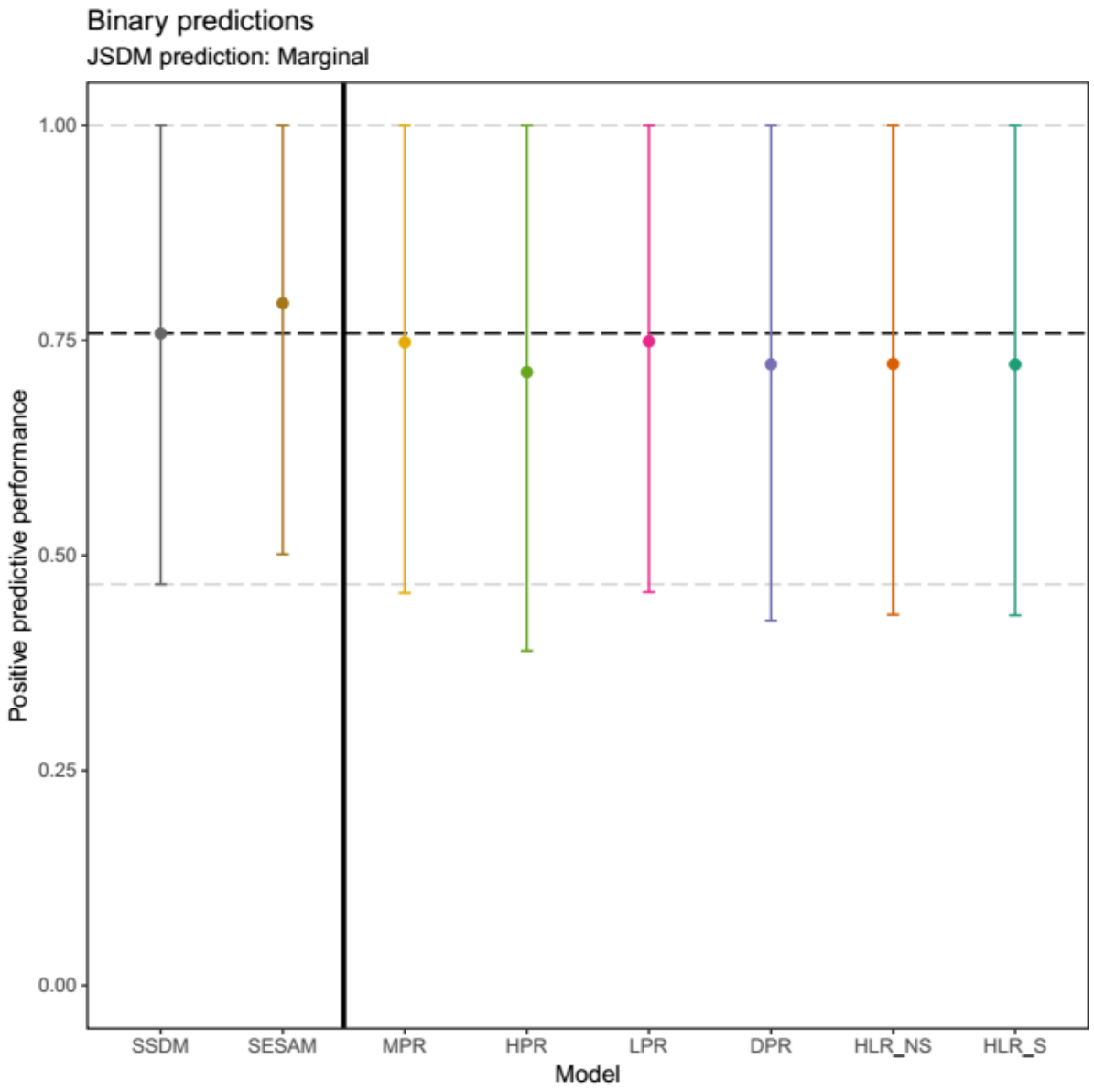


Figure S46: Model performance for the positive predictive performance evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

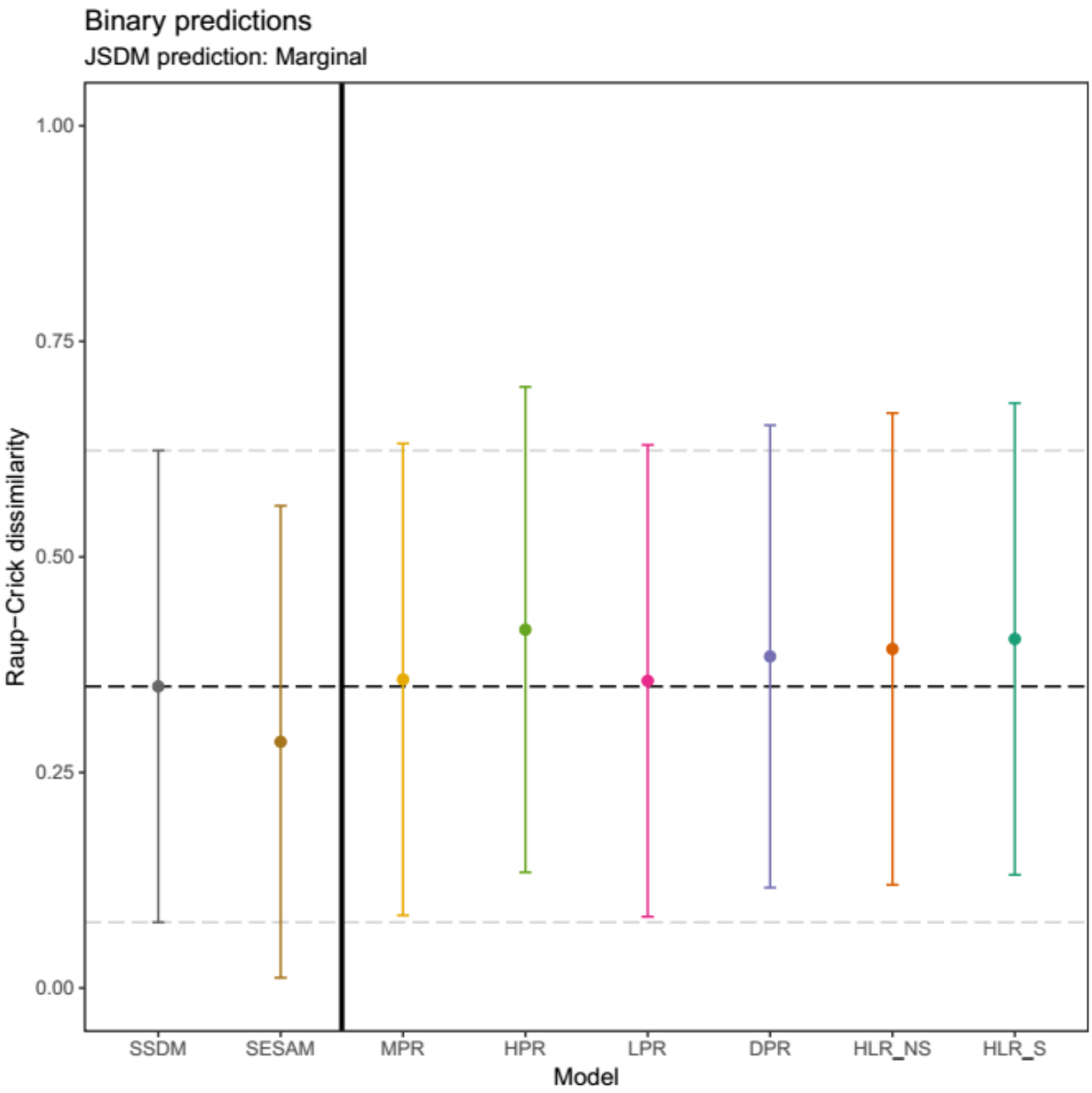


Figure S47: Model performance for the Raup-Crick dissimilarity evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

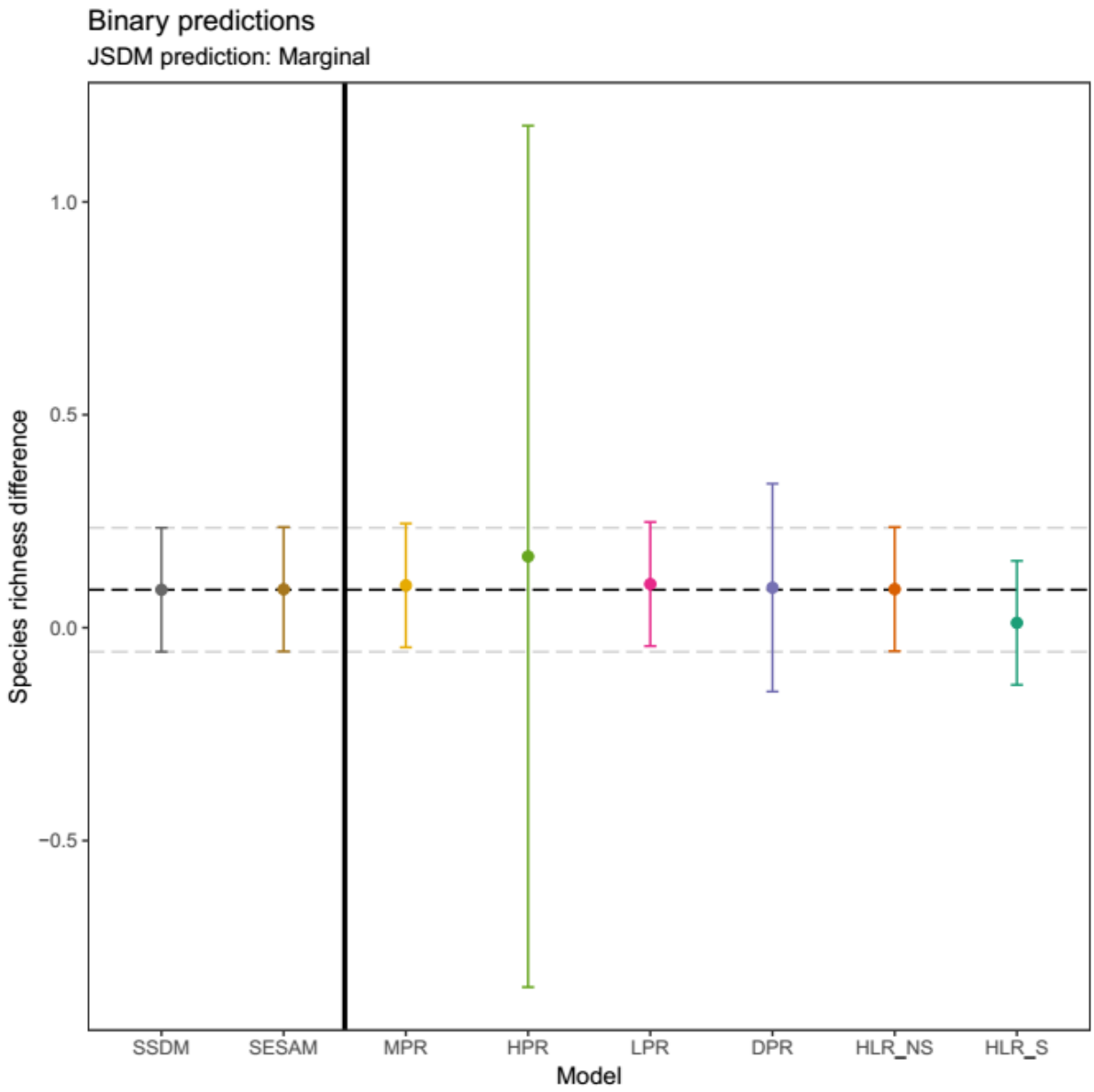


Figure S48: Model performance for the species richness difference evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

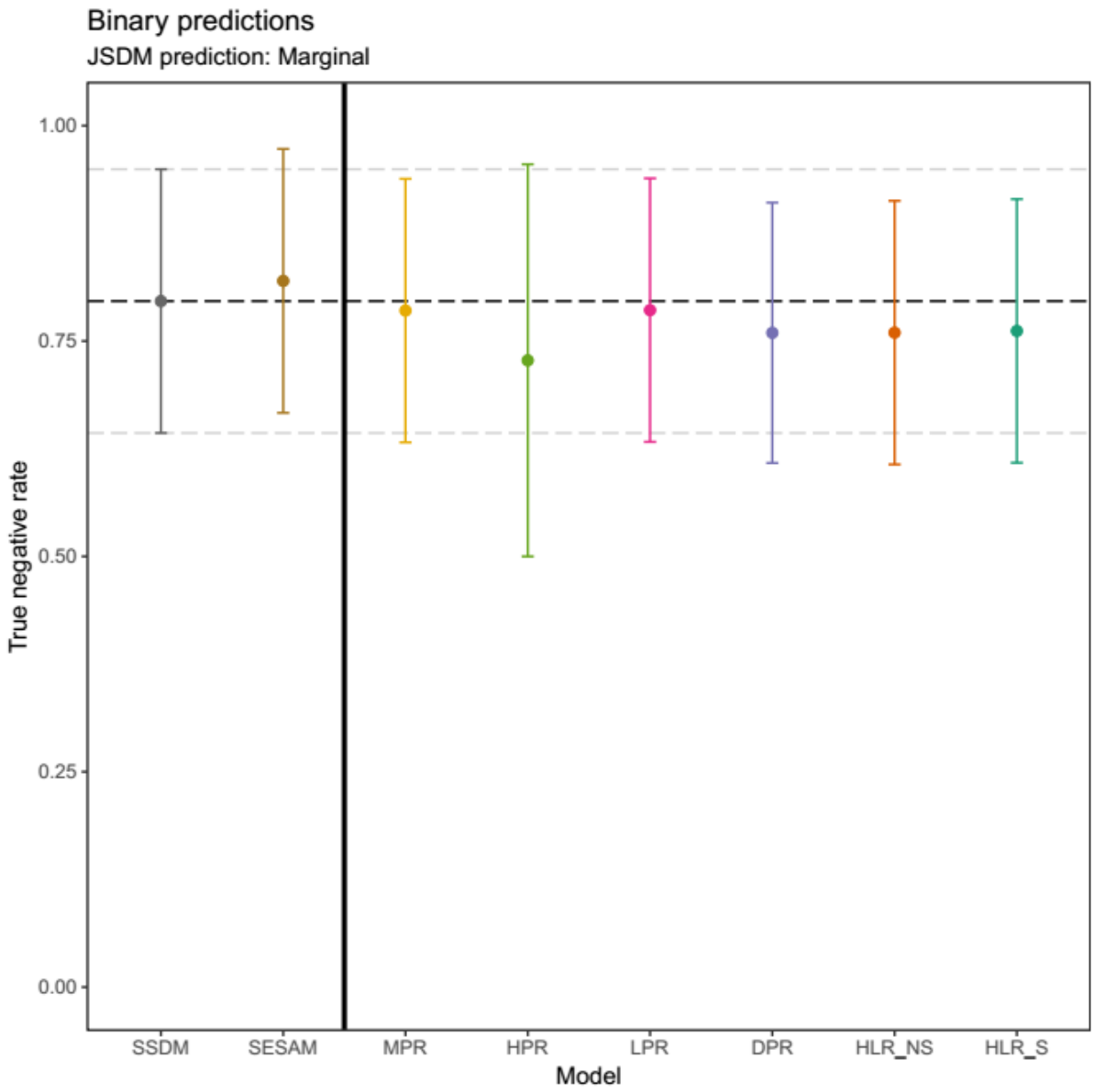


Figure S49: Model performance for the true negative rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

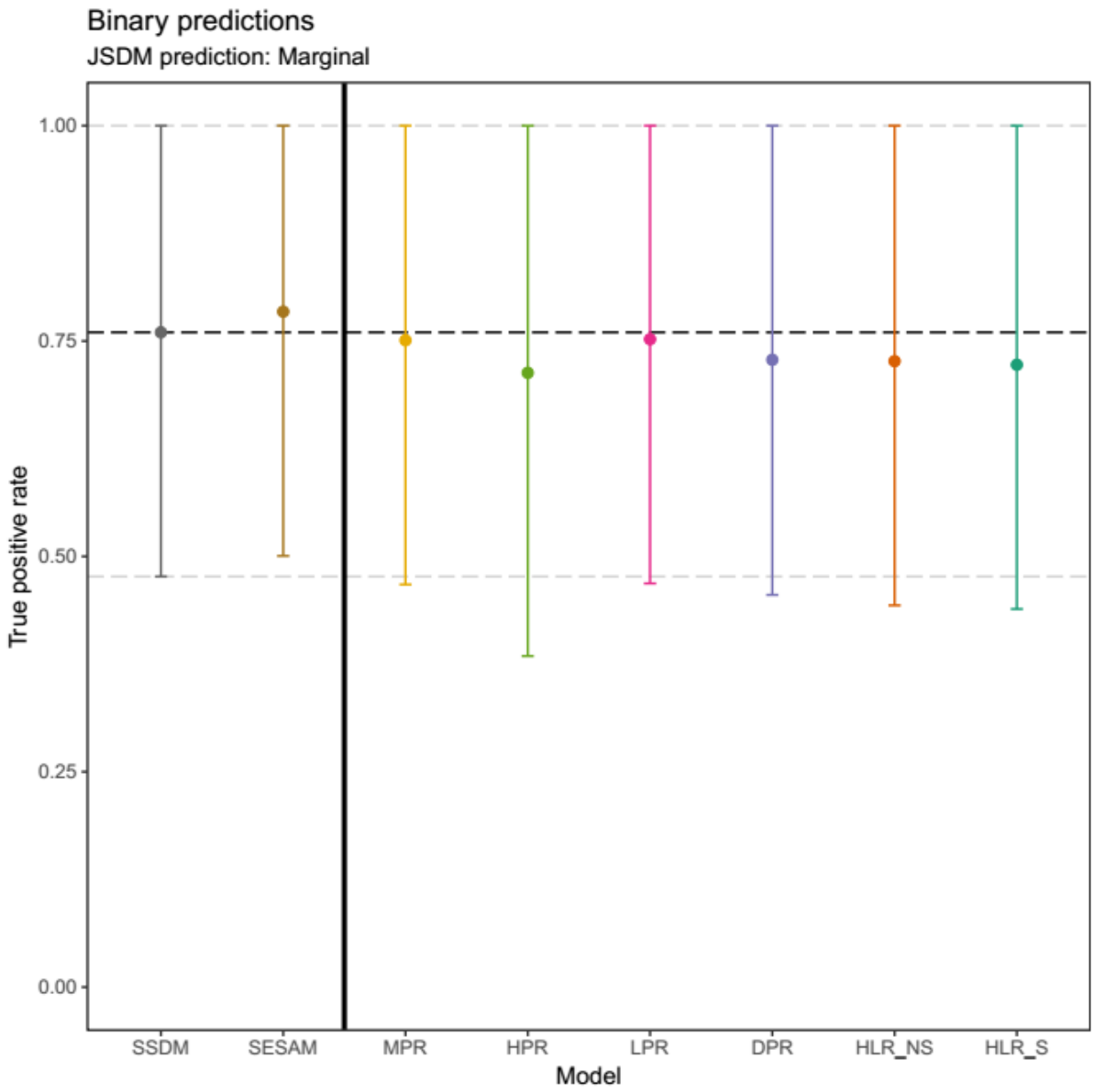


Figure S50: Model performance for the true positive rate evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

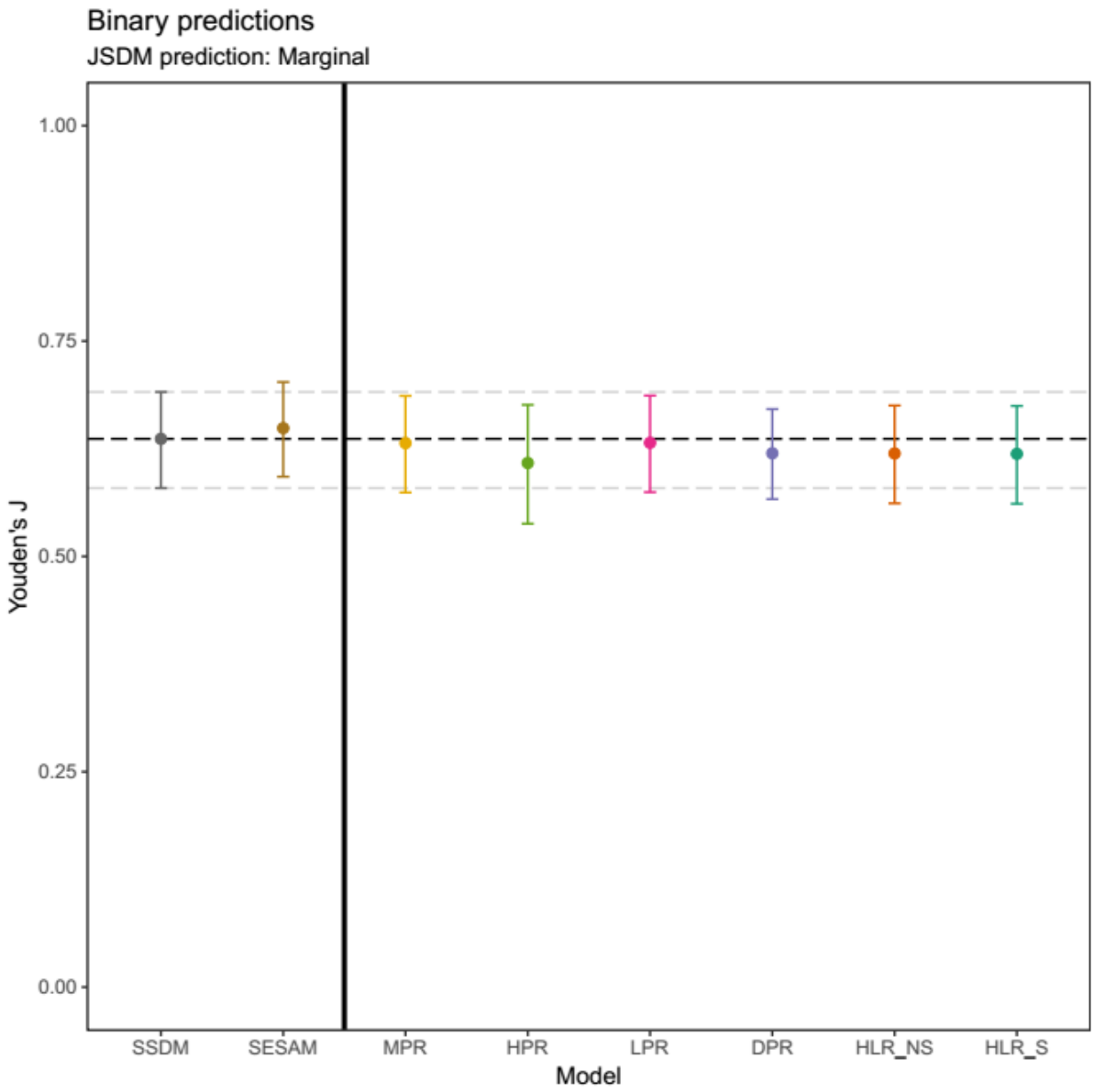


Figure S51: Model performance for the Youden's J evaluation metric for binary marginal predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

3. Joint predictions

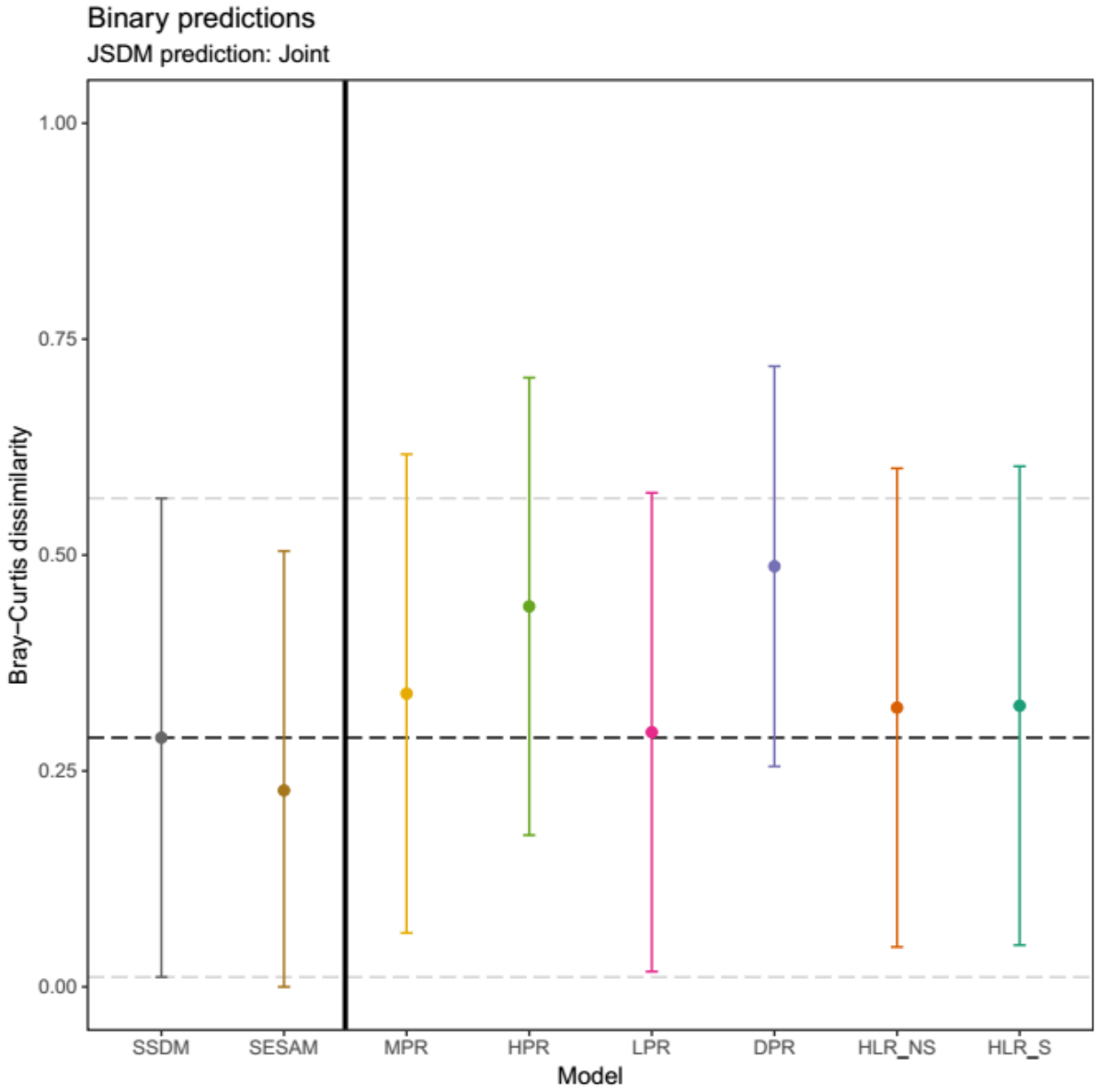


Figure S52: Model performance for the Bray-Curtis dissimilarity evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

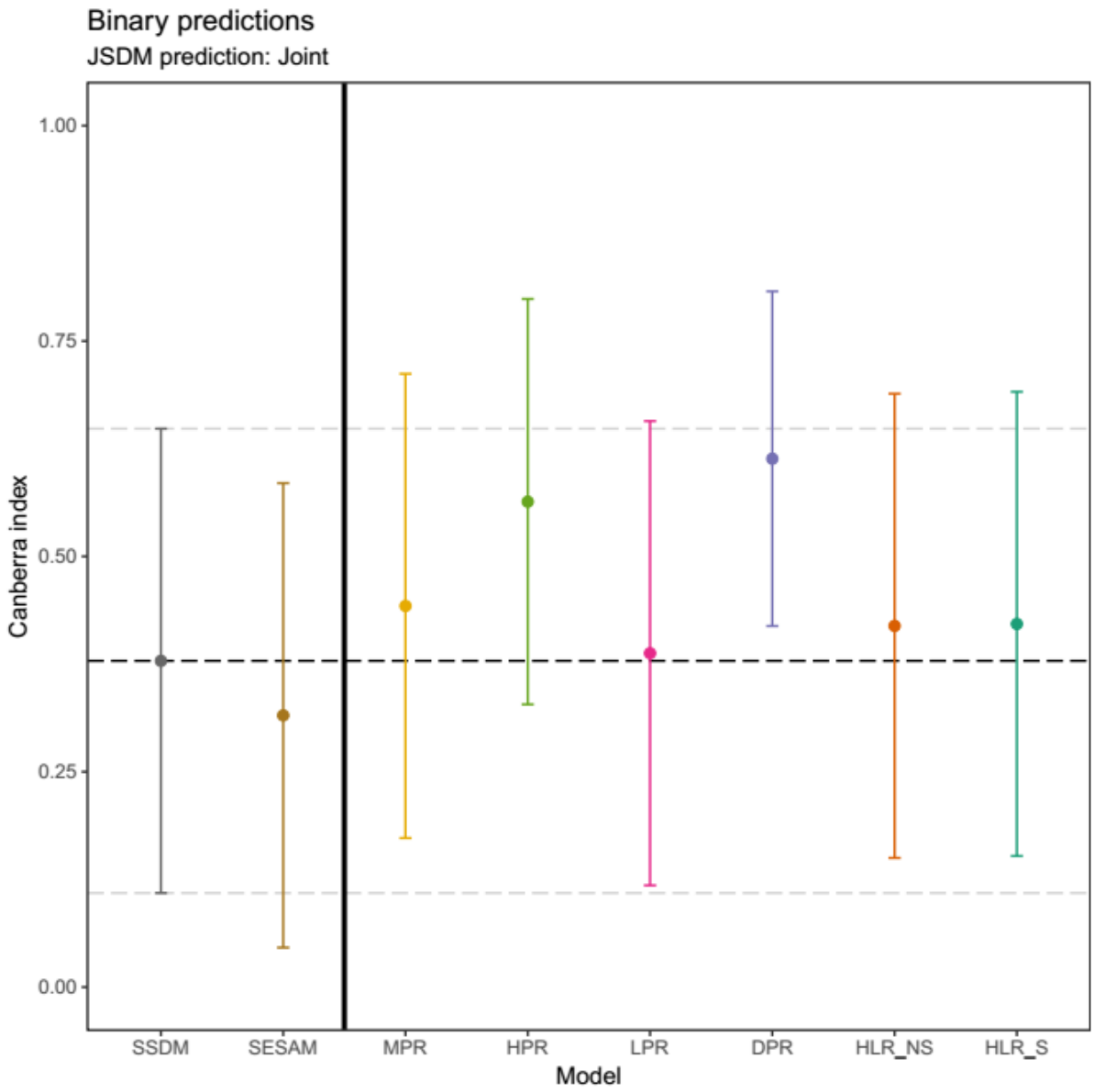


Figure S53: Model performance for the Canberra index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

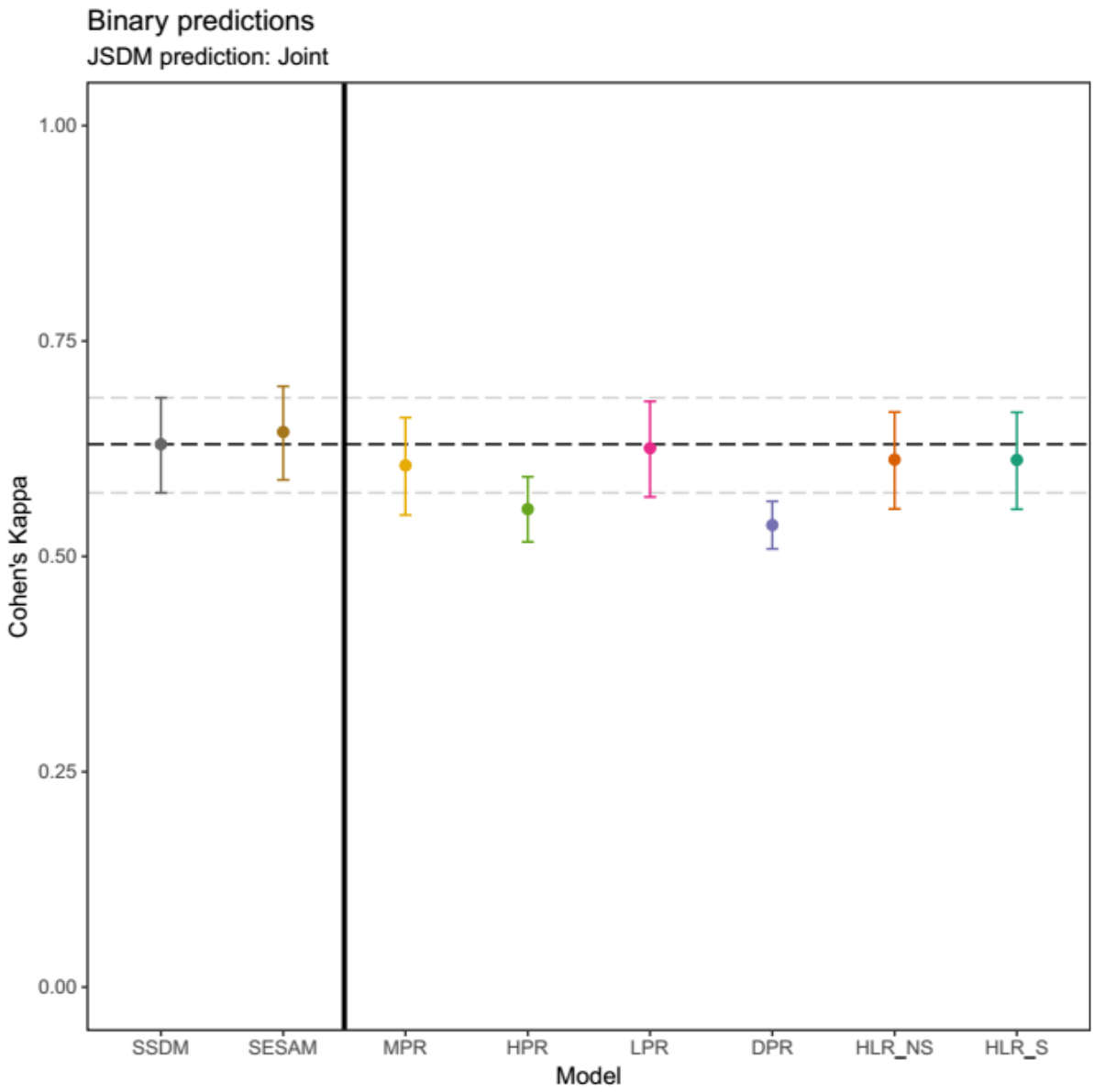


Figure S54: Model performance for the Cohen's Kappa evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

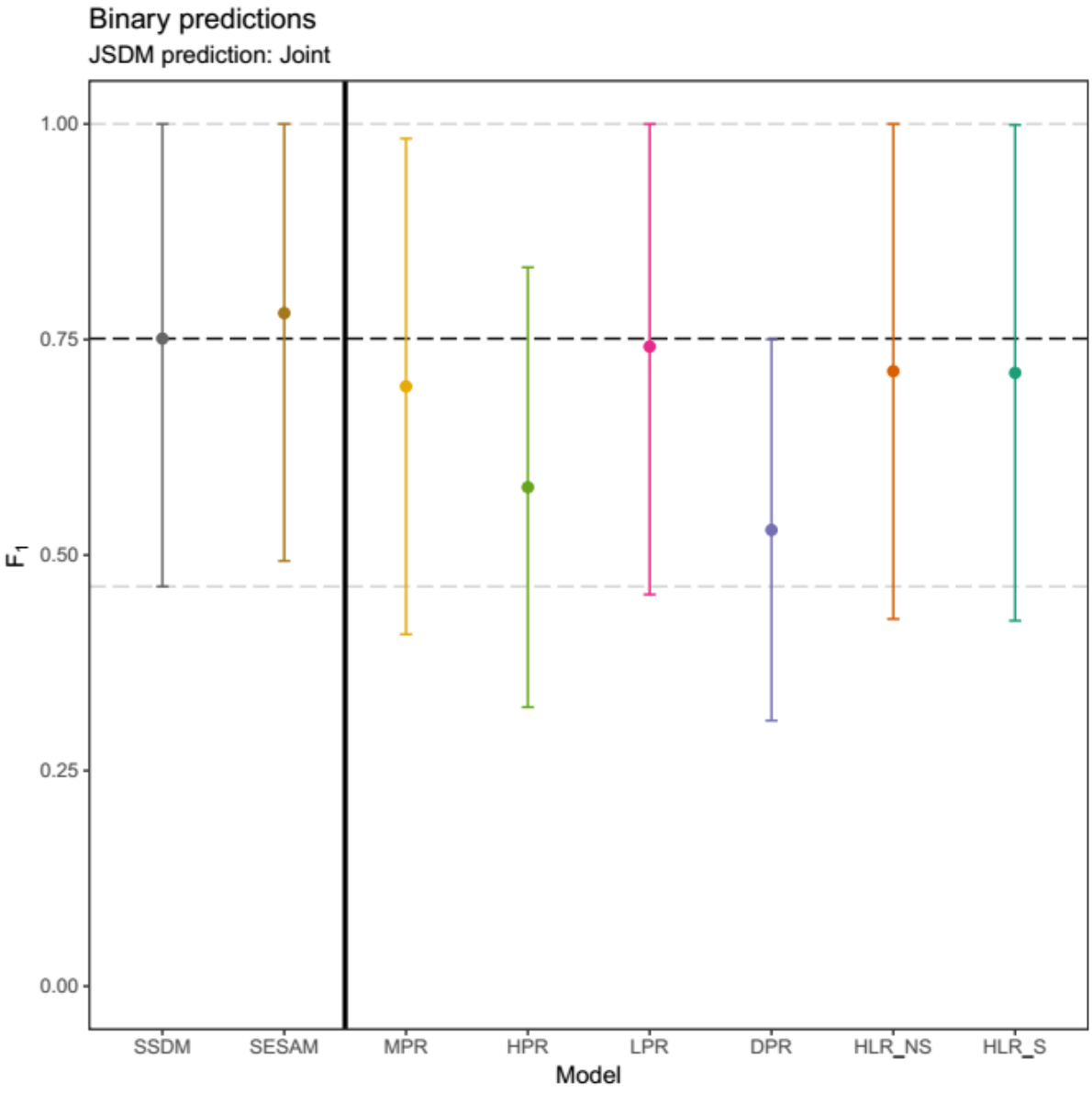


Figure S55: Model performance for the F_1 evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

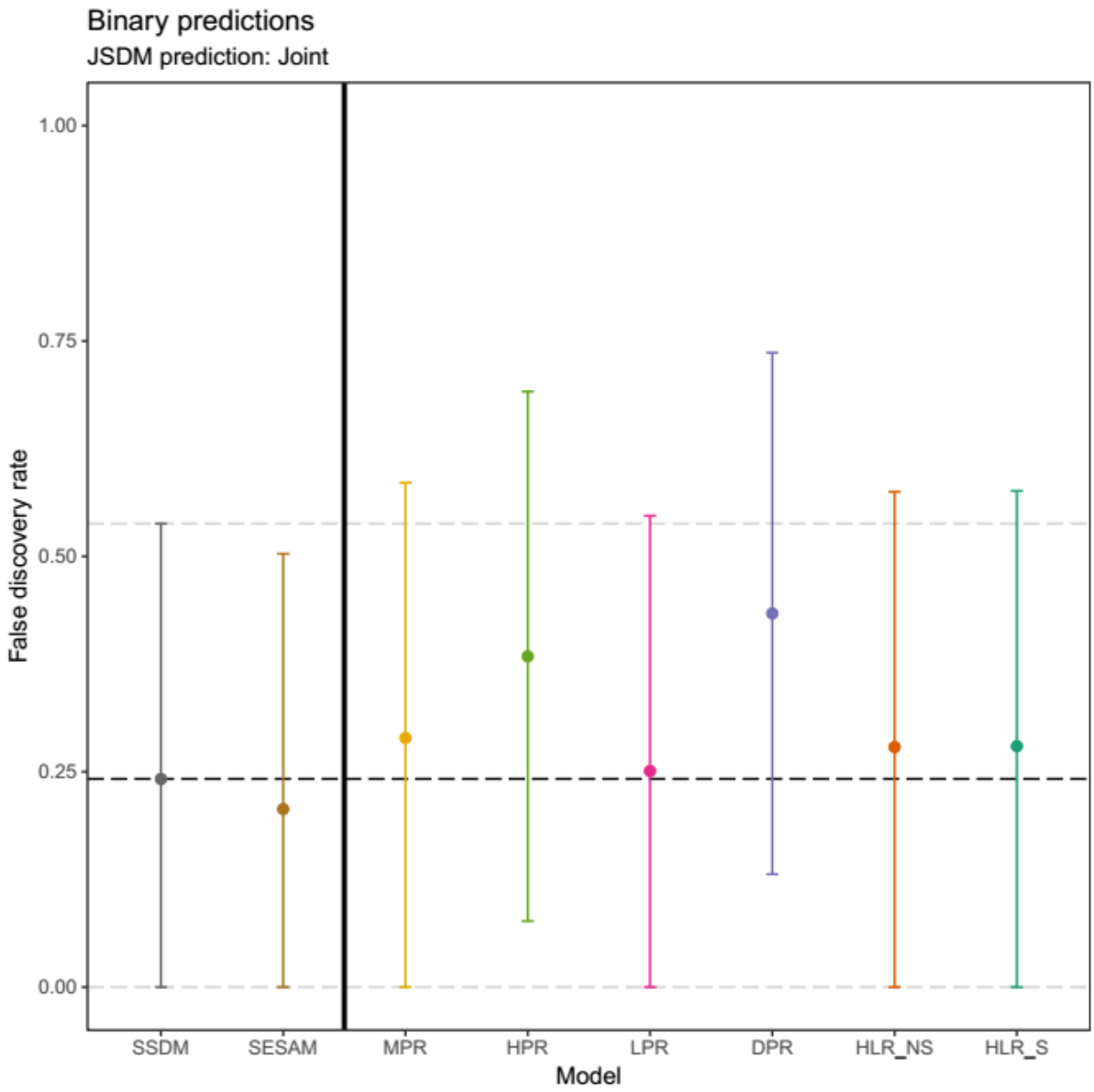


Figure S56: Model performance for the false discovery rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

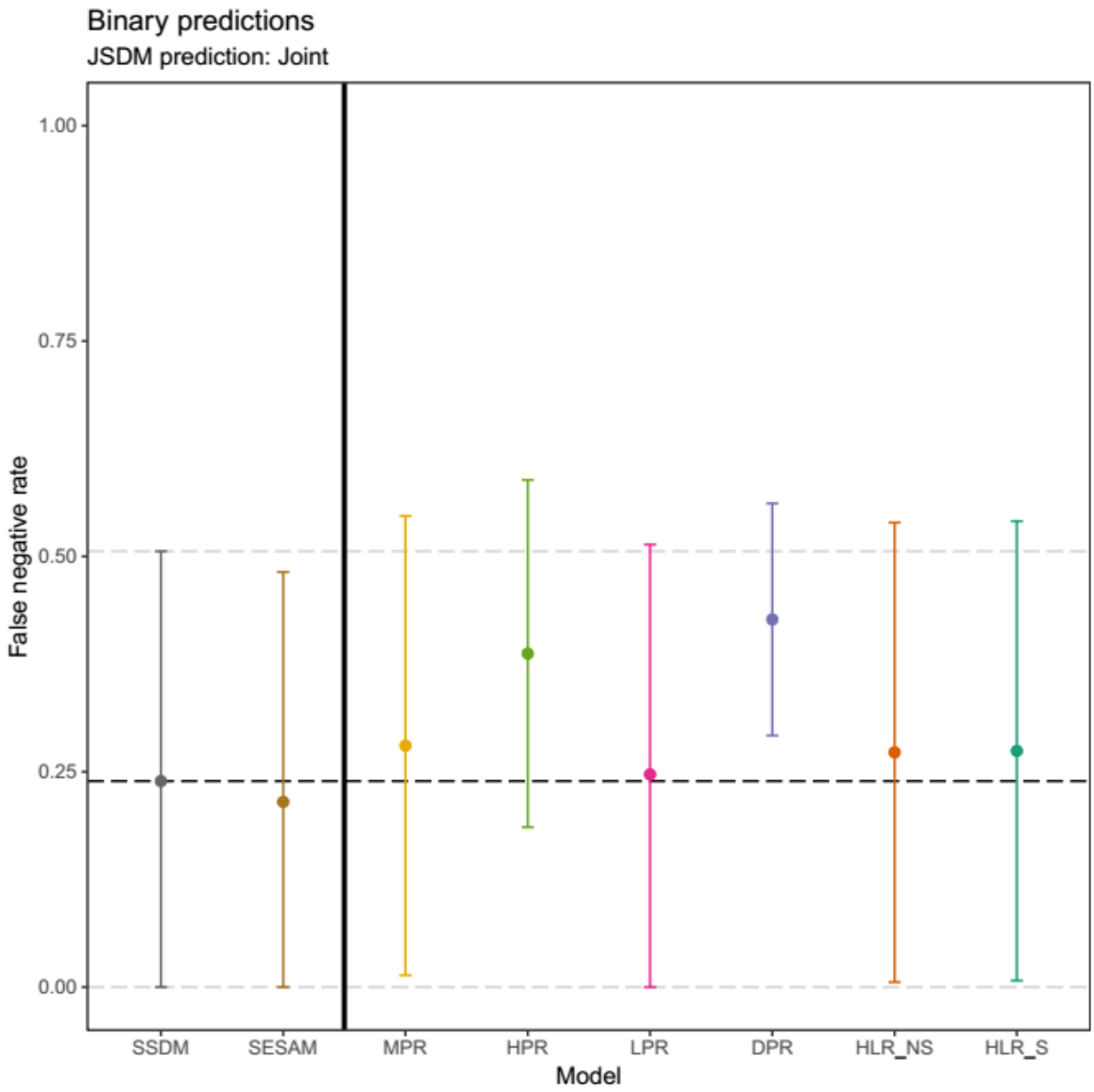


Figure S57: Model performance for the false negative rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

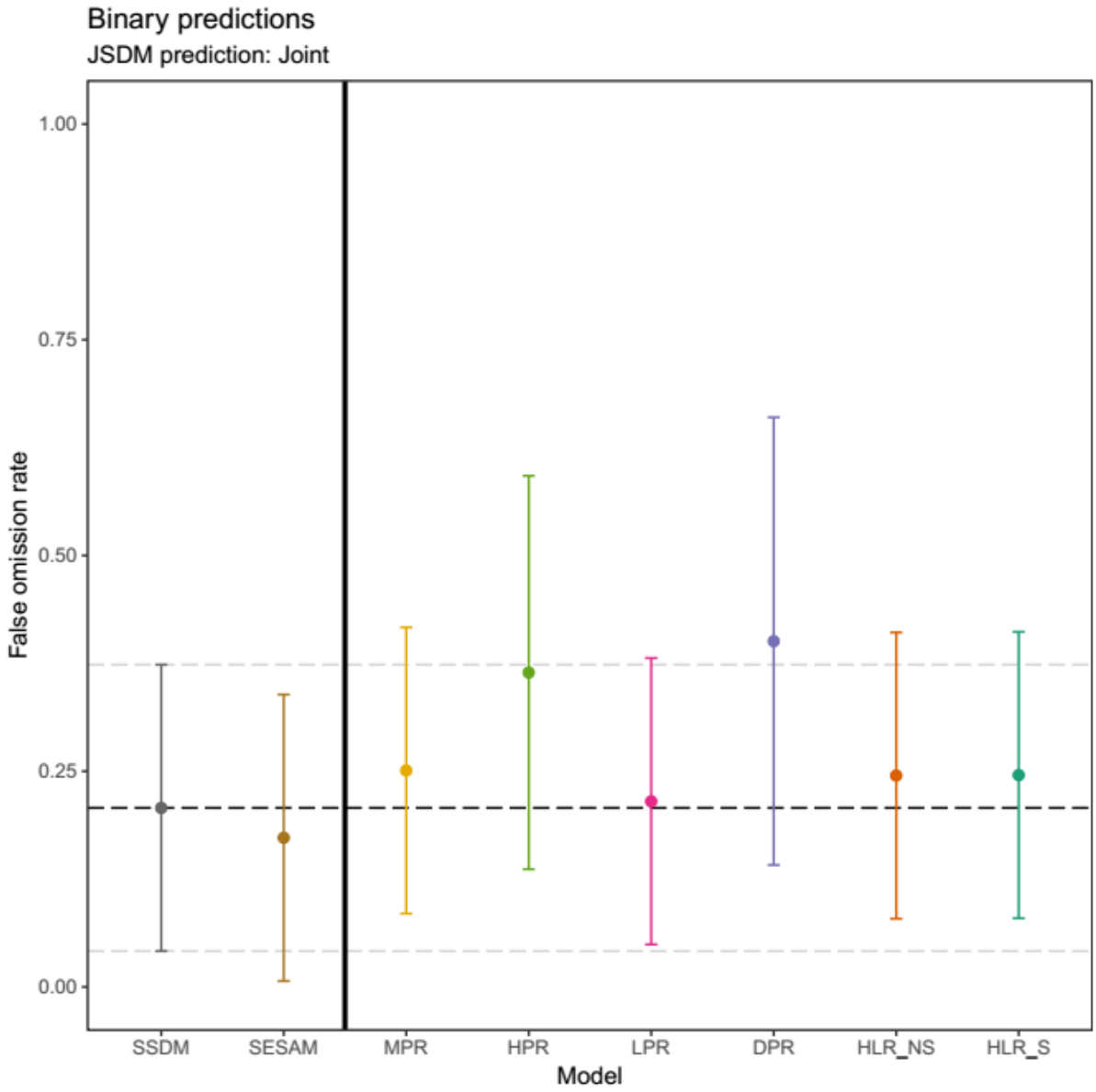


Figure S58: Model performance for the false omission rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

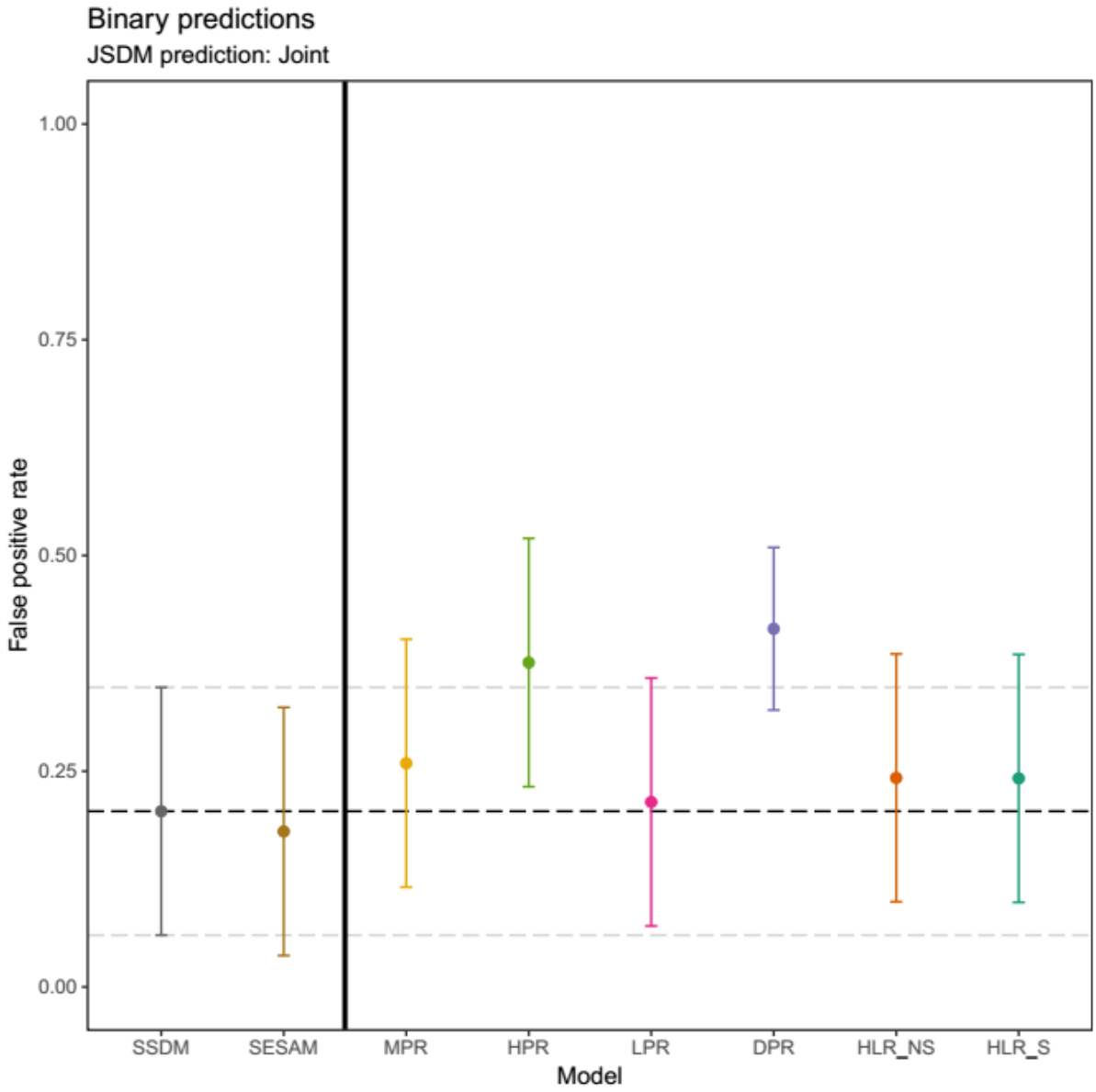


Figure S59: Model performance for the false positive rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

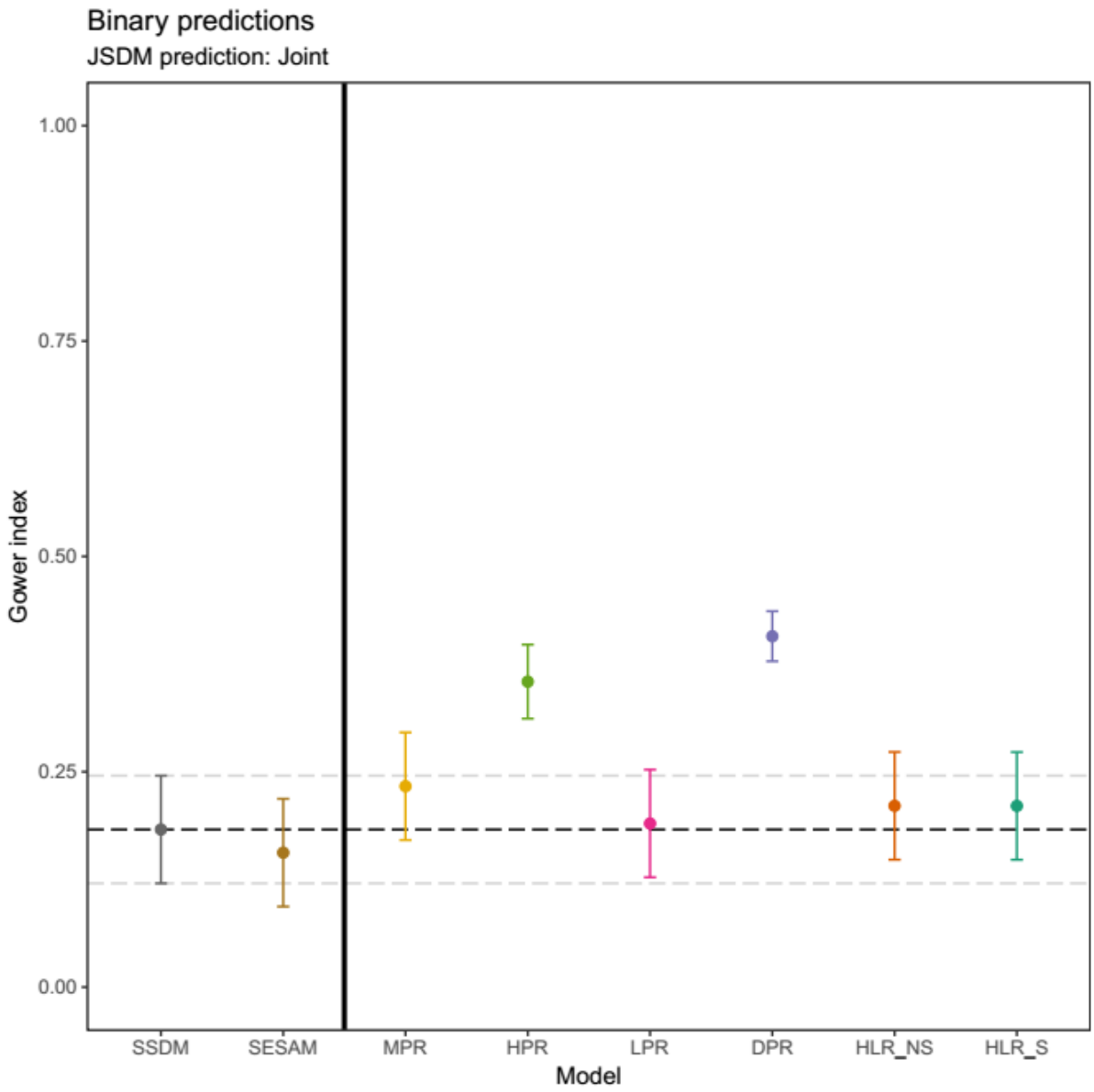


Figure S60: Model performance for the Gower index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

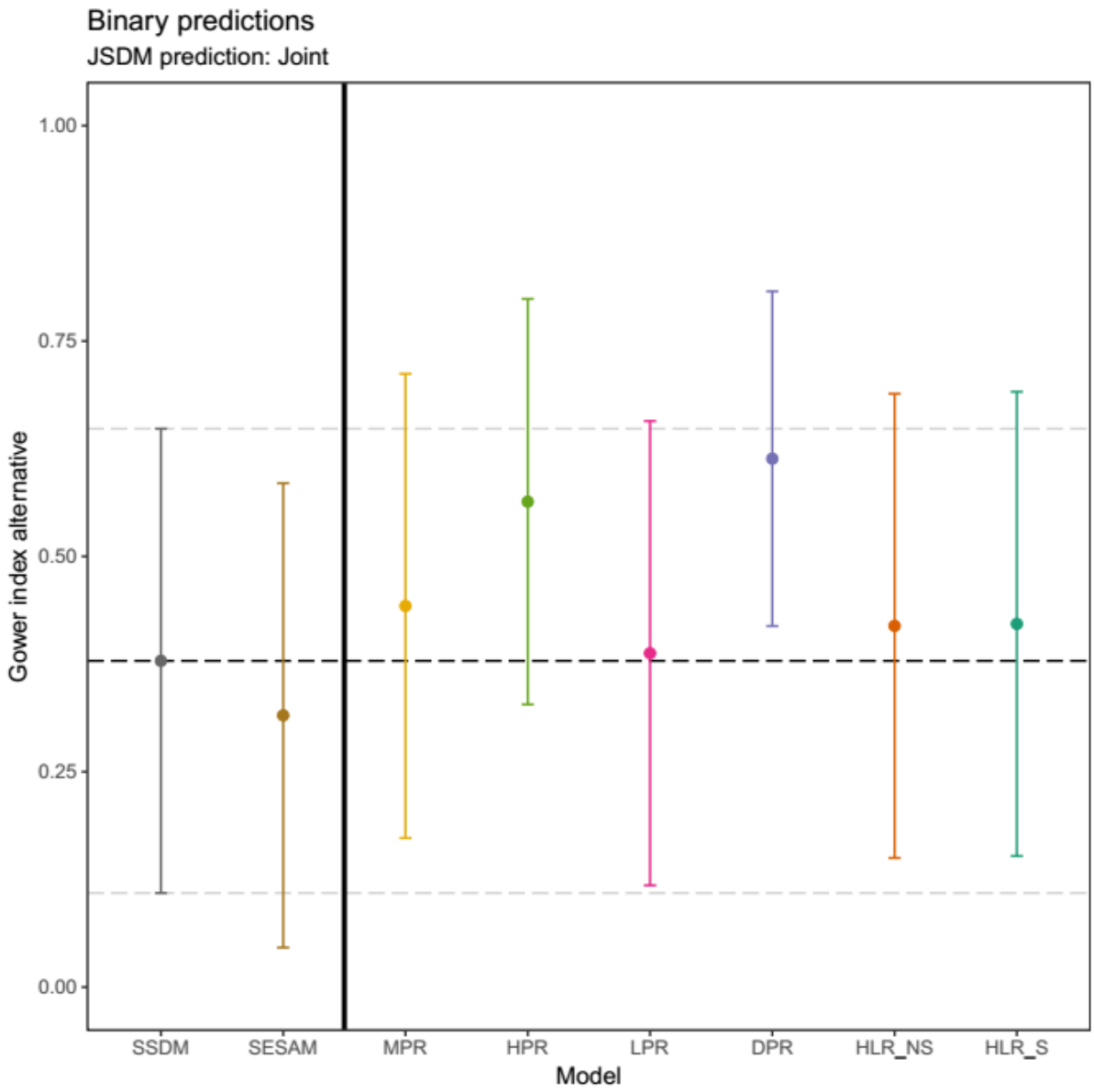


Figure S61: Model performance for the alternative Gower index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

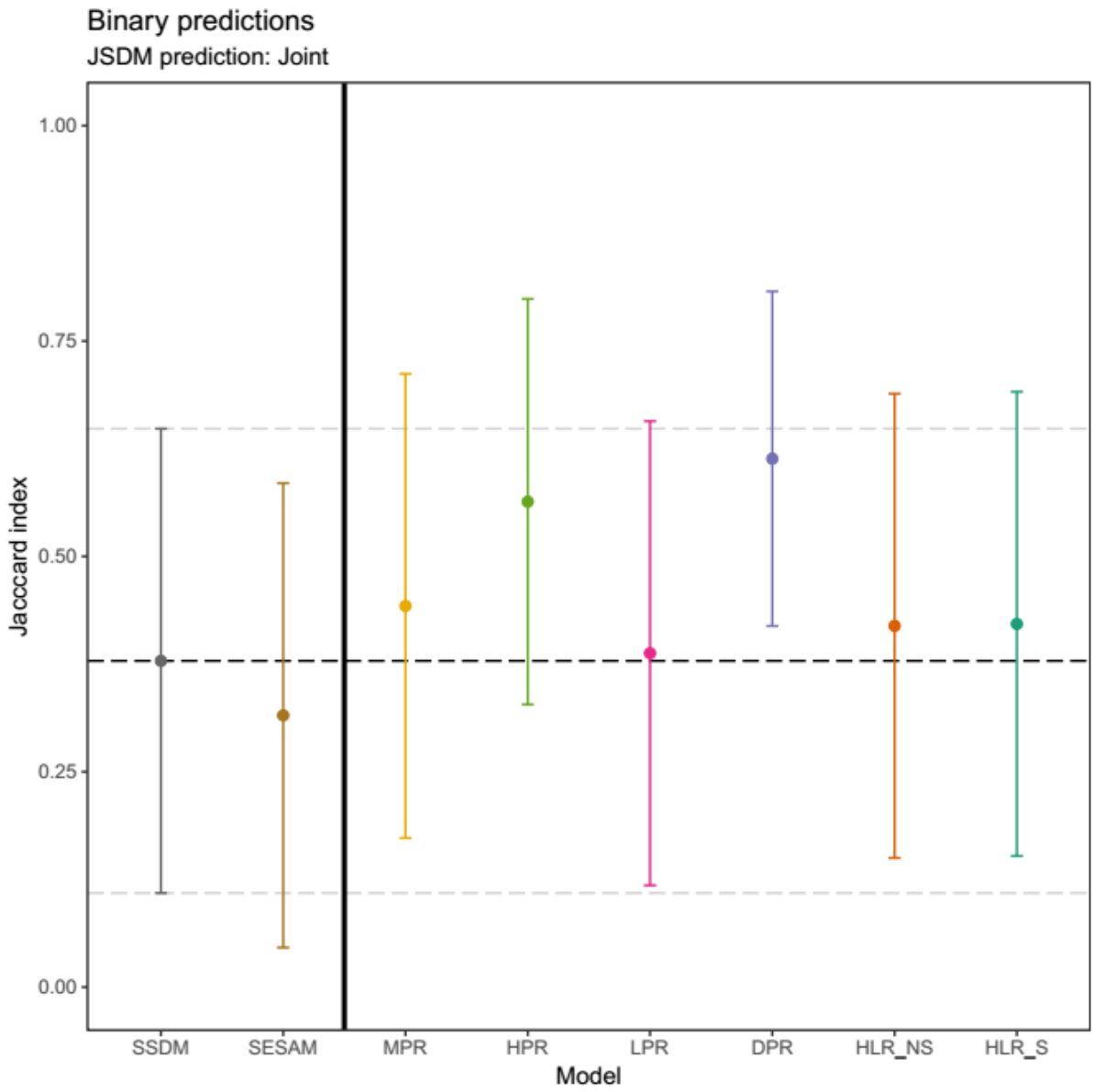


Figure S62: Model performance for the Jaccard index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

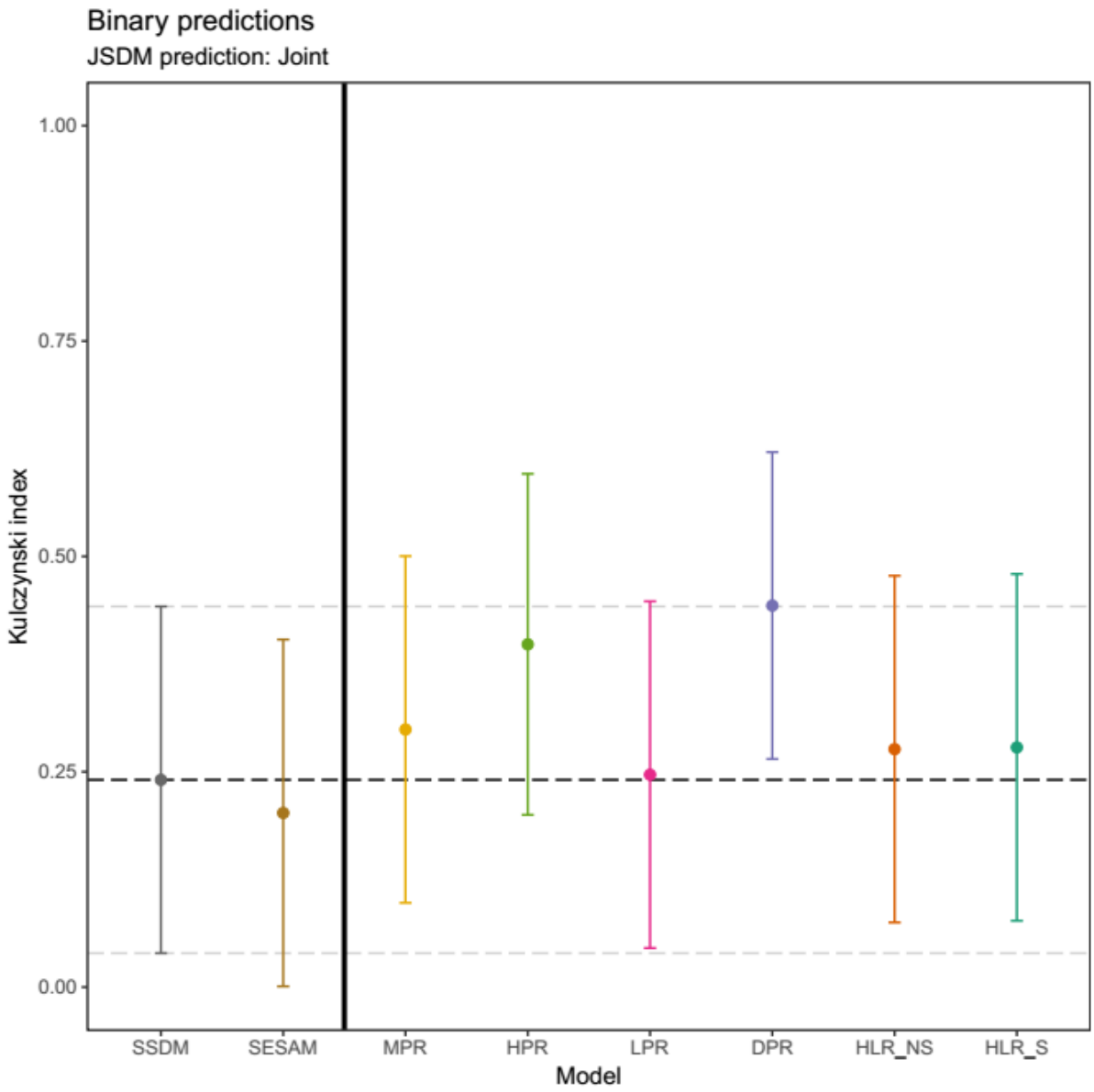


Figure S63: Model performance for the Kulczynski index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

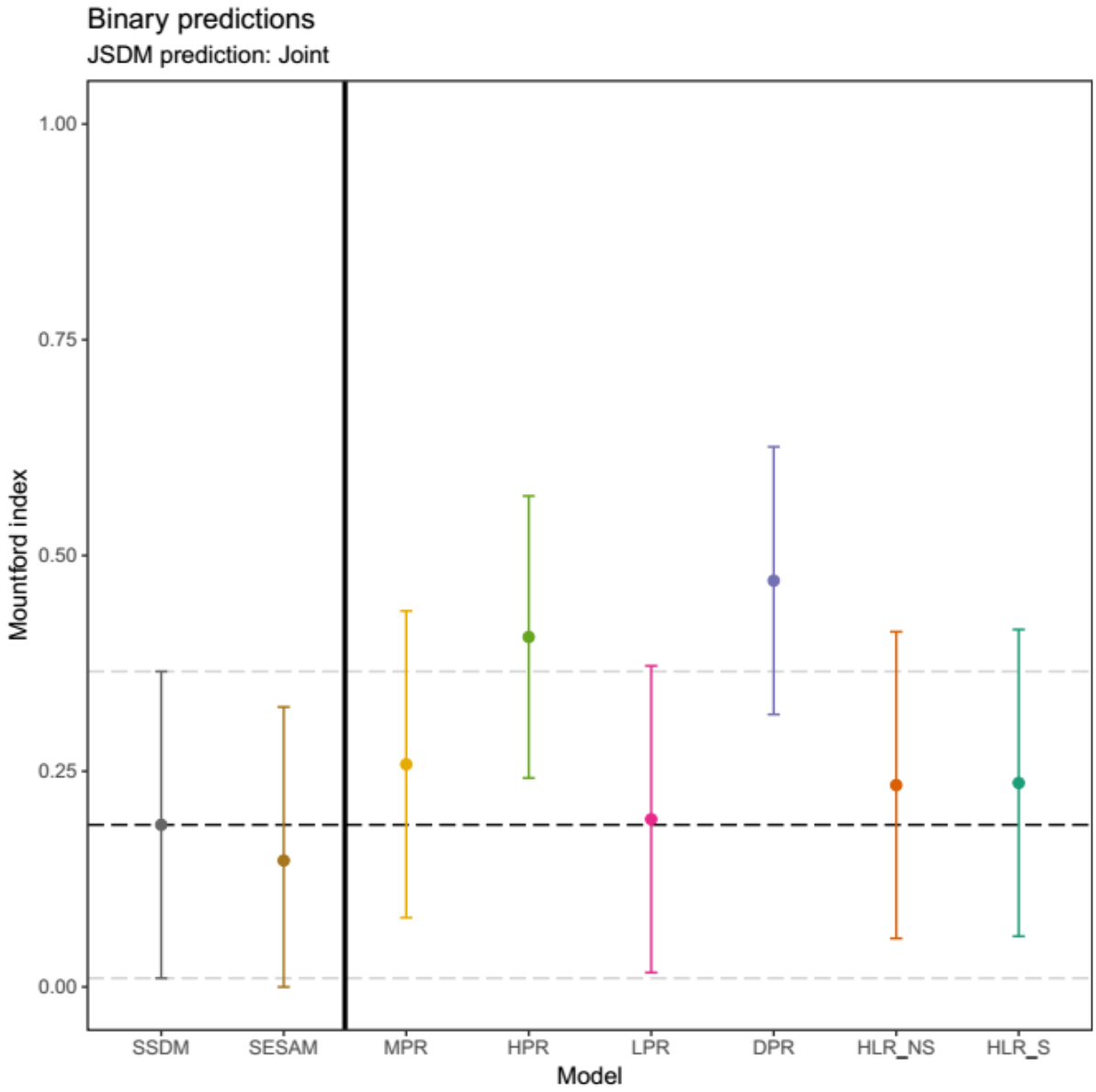


Figure S64: Model performance for the Mountford index evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

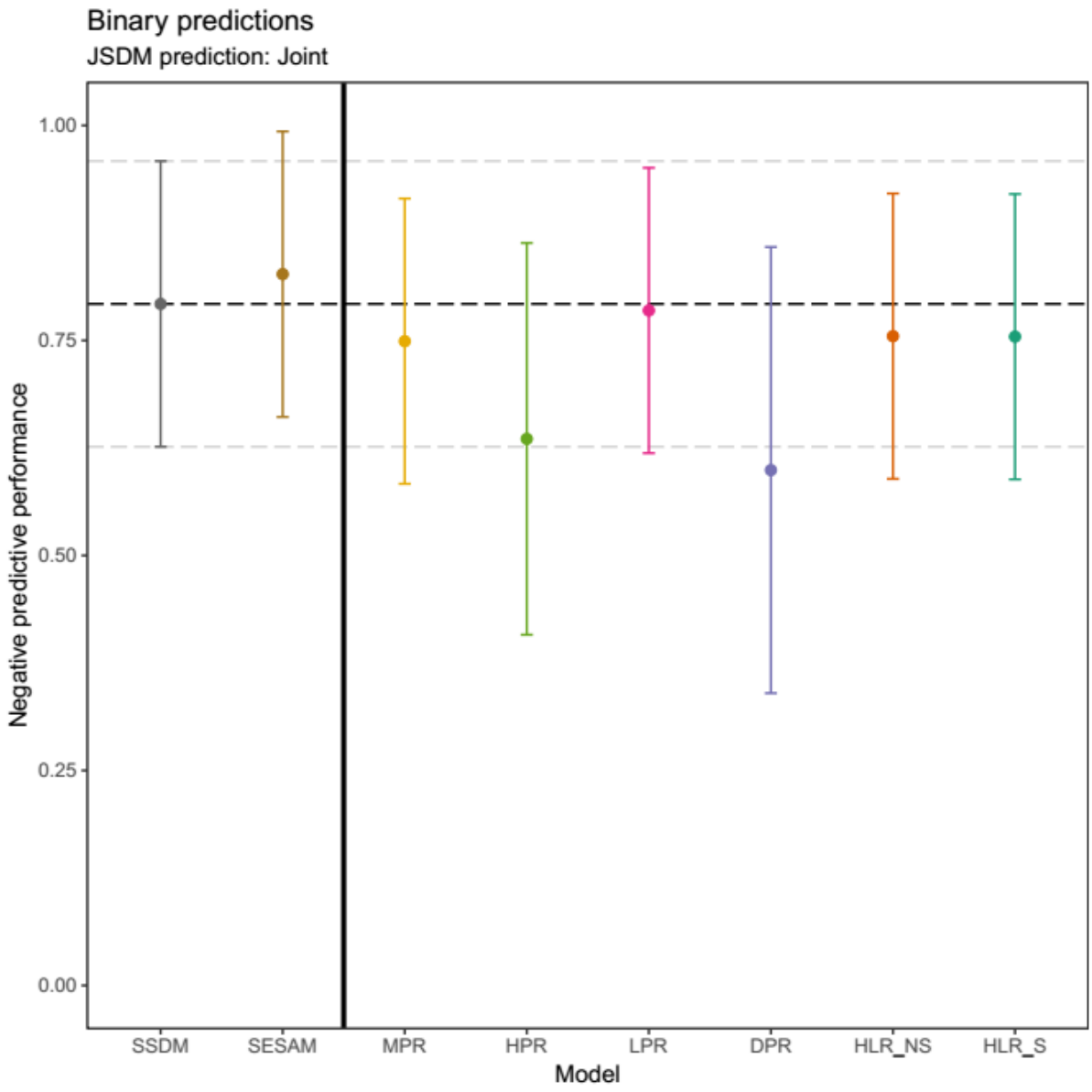


Figure S65: Model performance for the negative predictive performance evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

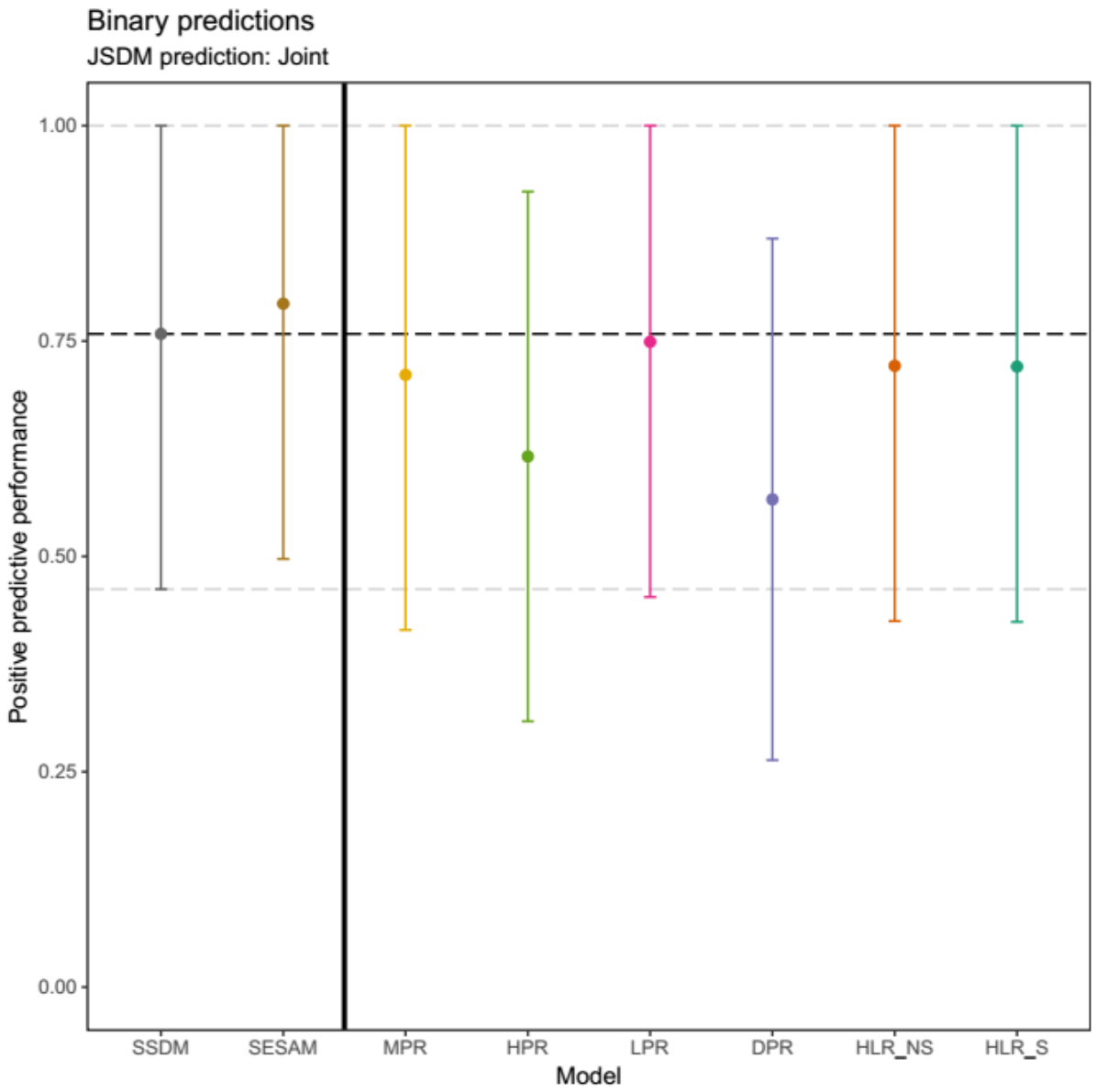


Figure S66: Model performance for the positive predictive performance evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

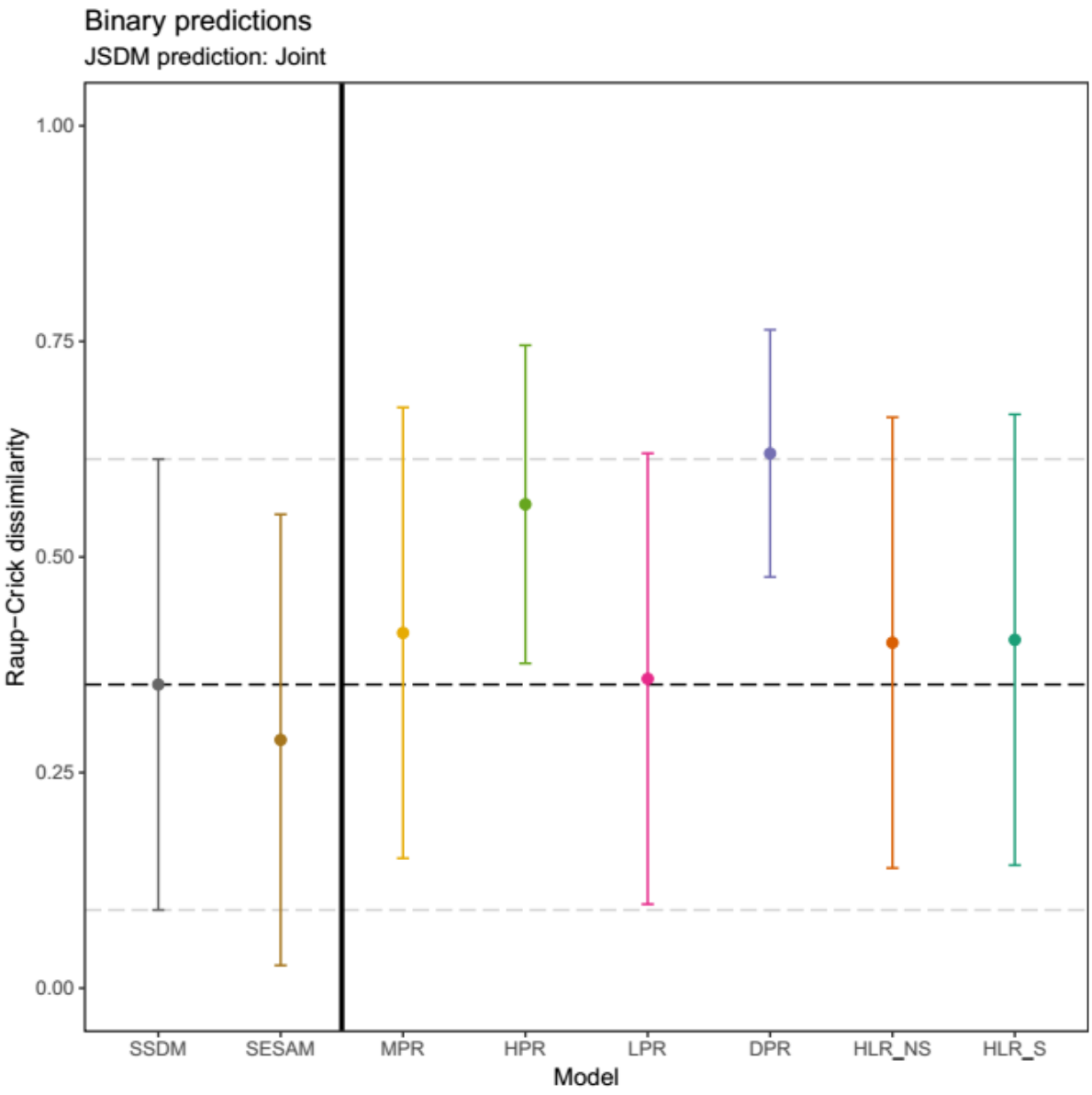


Figure S67: Model performance for the Raup-Crick dissimilarity evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

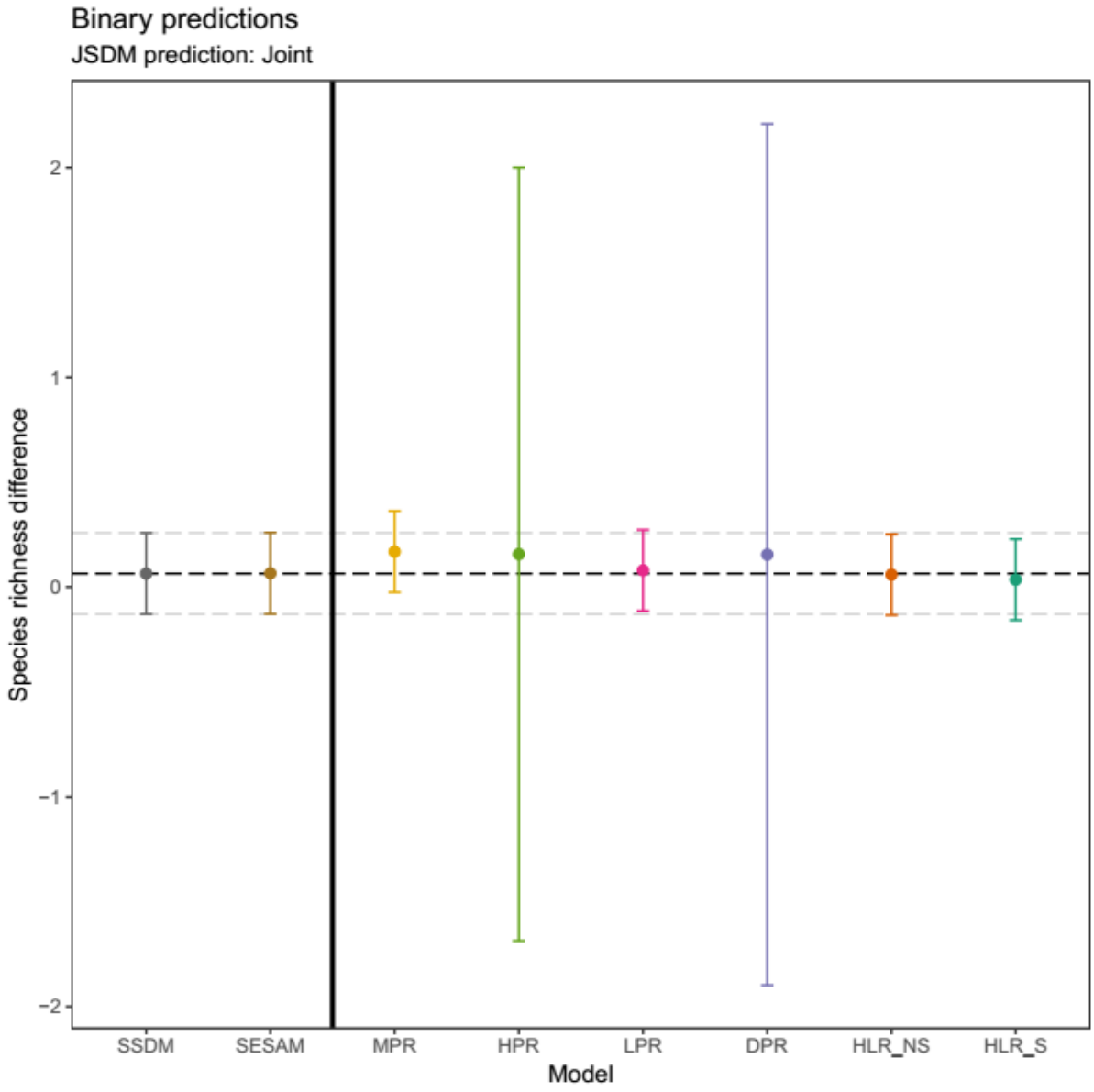


Figure S68: Model performance for the species richness difference evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

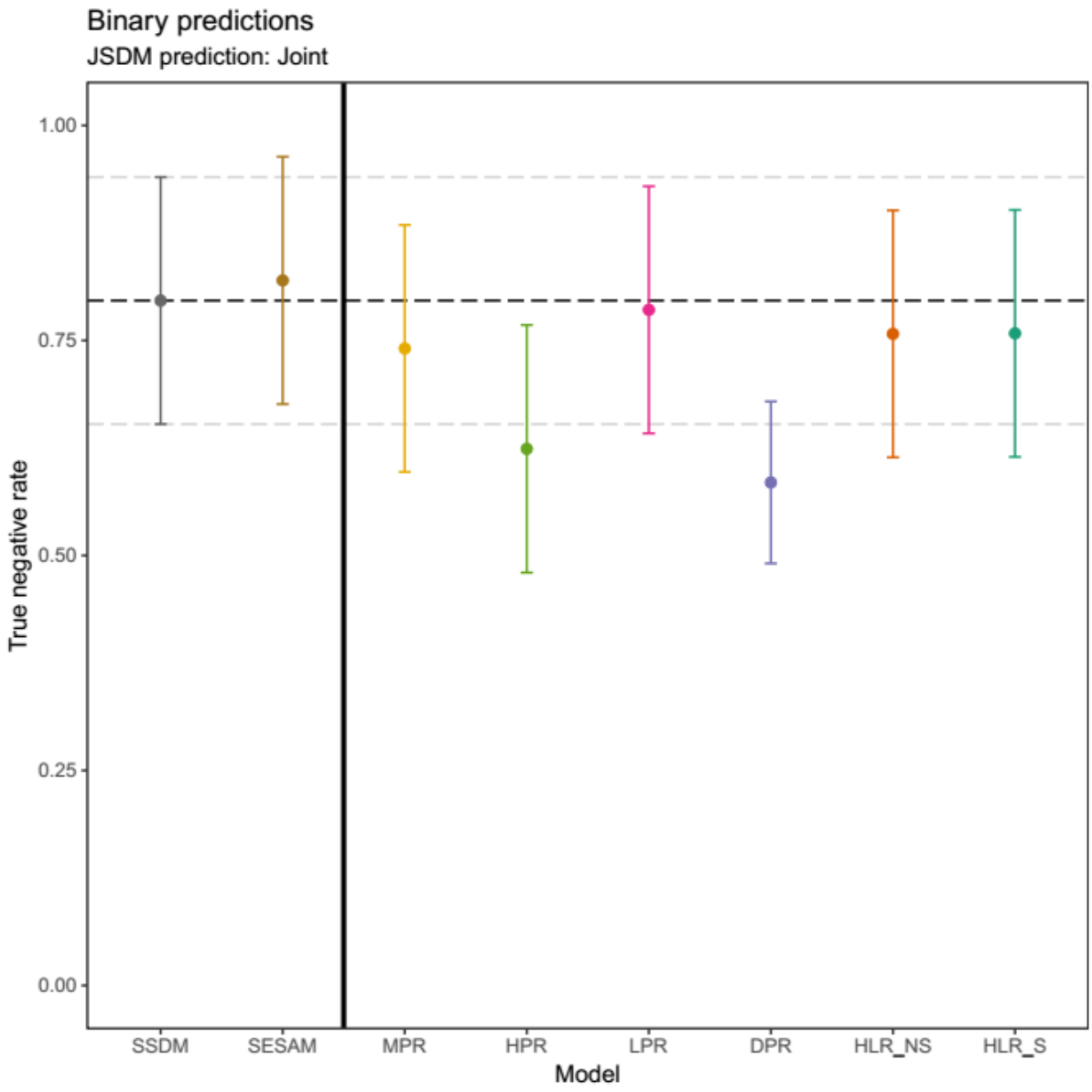


Figure S69: Model performance for the true negative rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

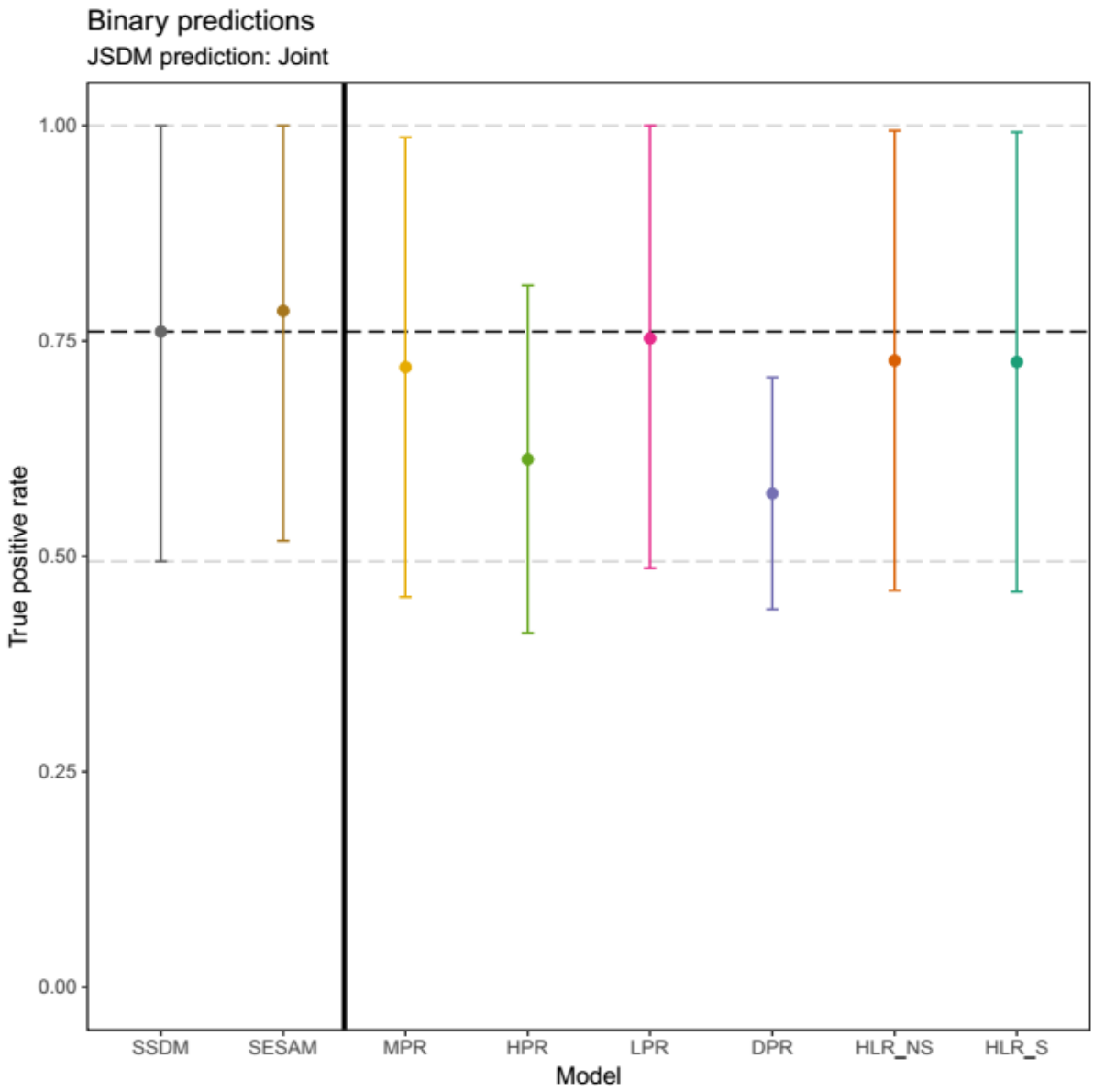


Figure S70: Model performance for the true positive rate evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

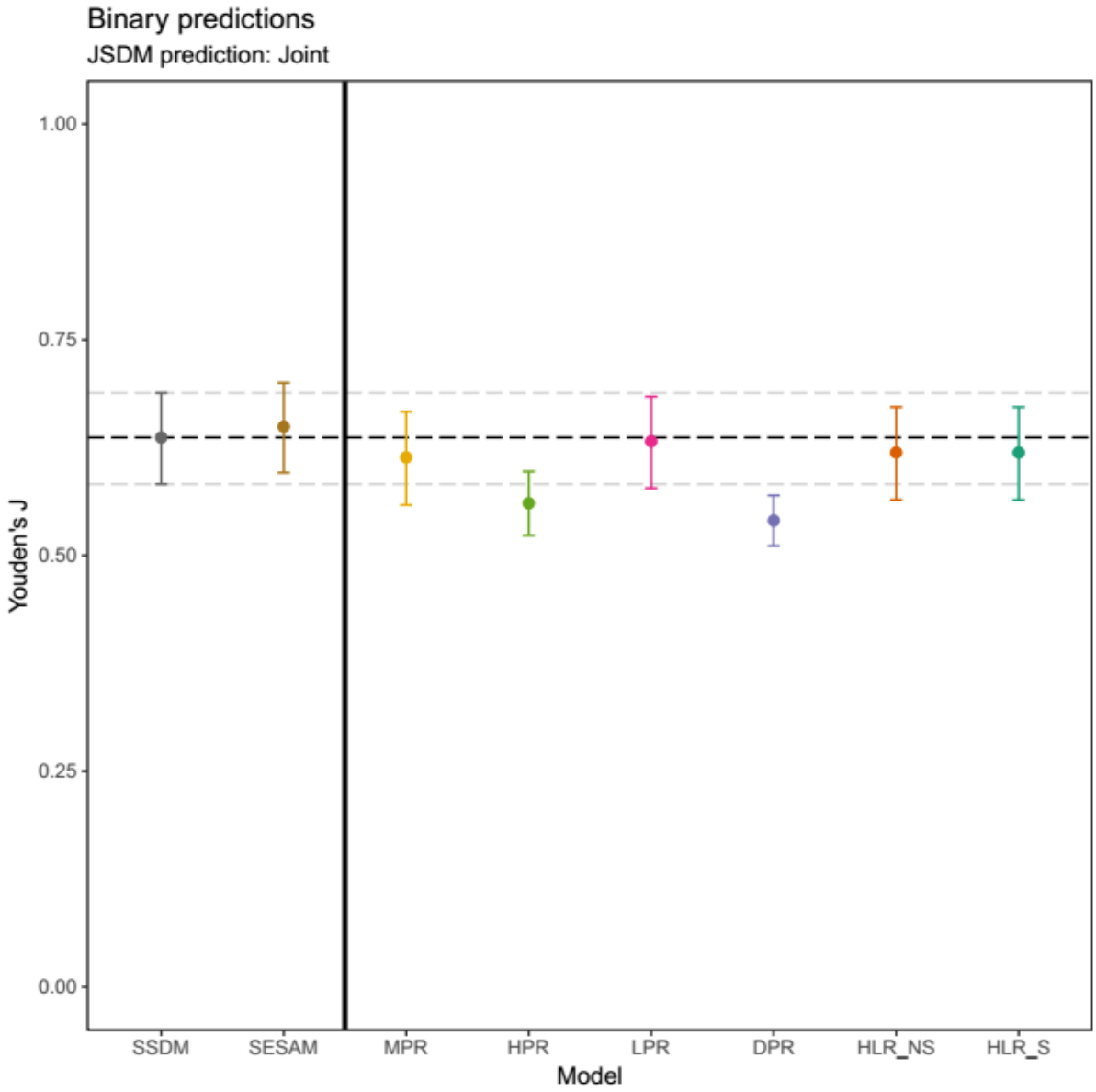


Figure S71: Model performance for the Youden's J evaluation metric for joint predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

4. Conditional (low-prevalence) predictions

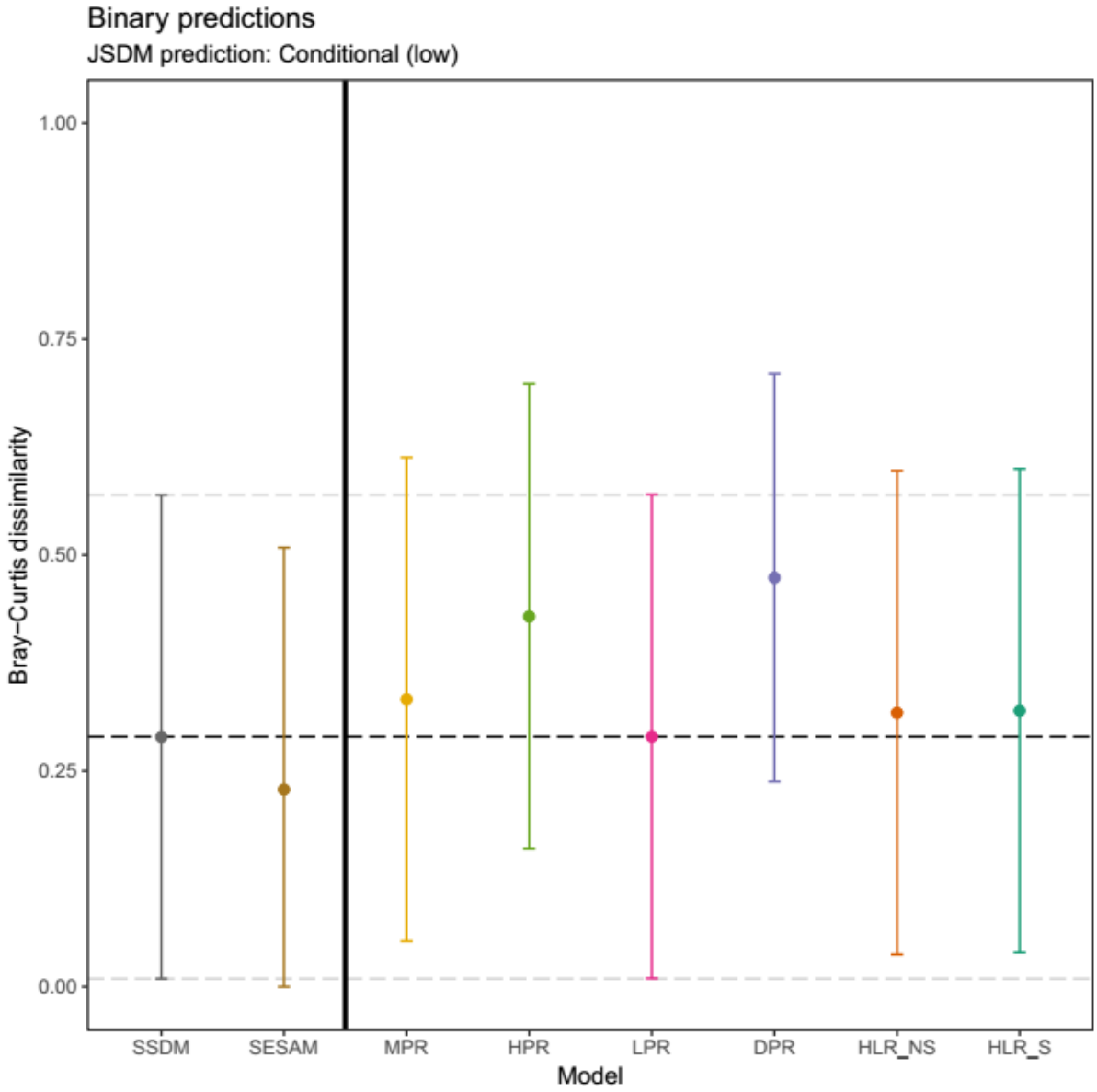


Figure S72: Model performance for the Bray-Curtis dissimilarity evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

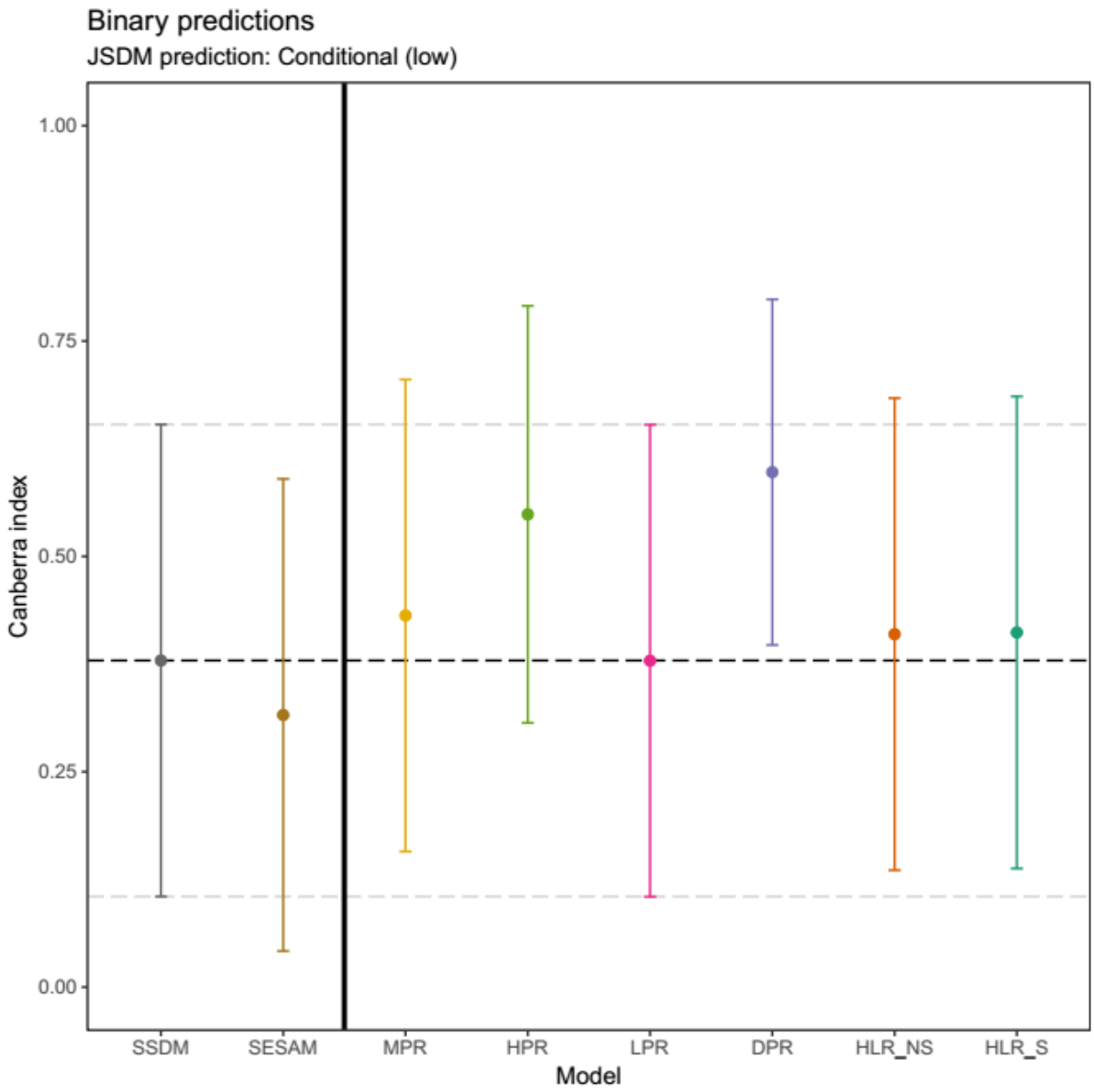


Figure S73: Model performance for the Canberra index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

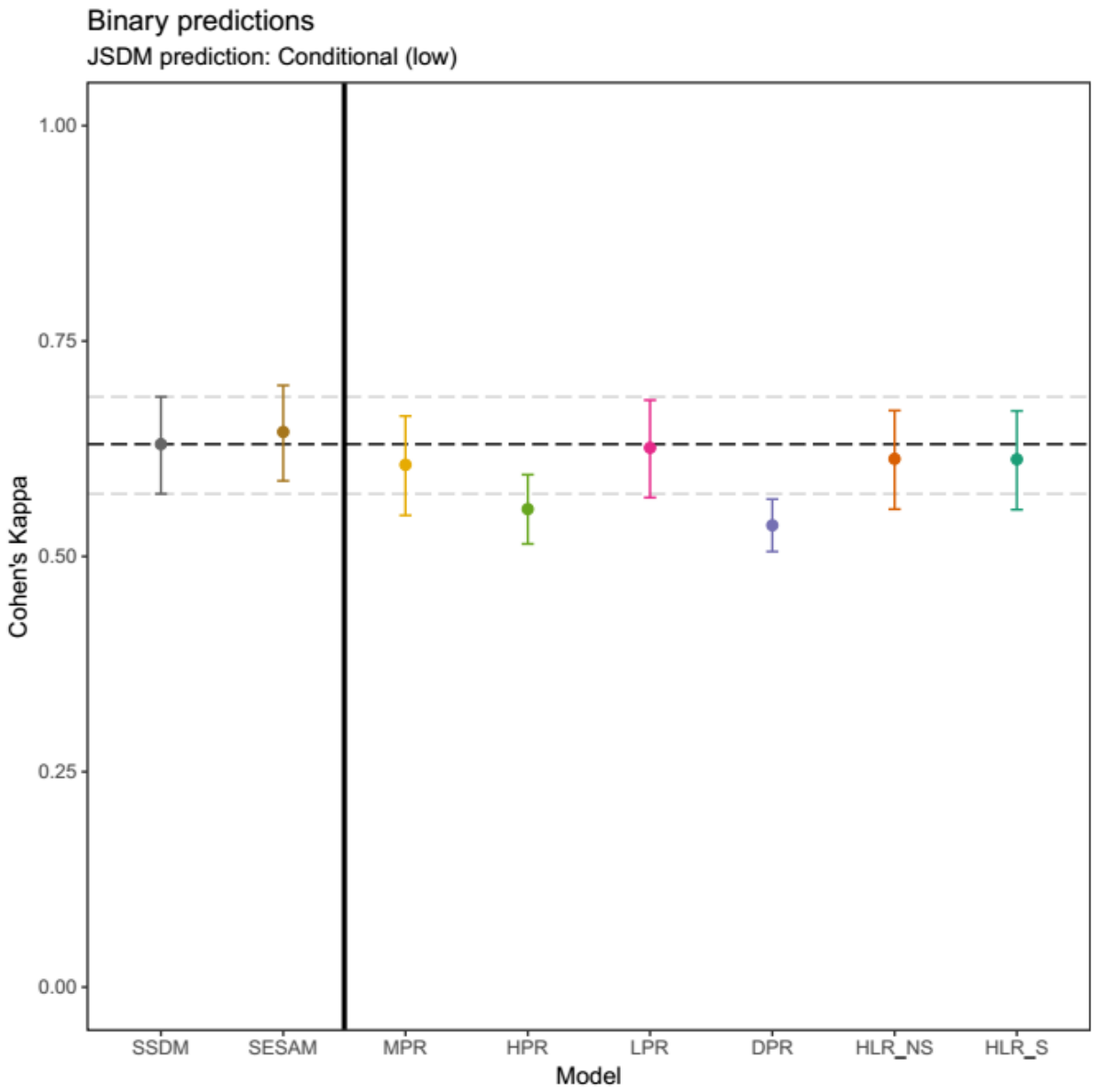


Figure S74: Model performance for the Cohen's Kappa evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

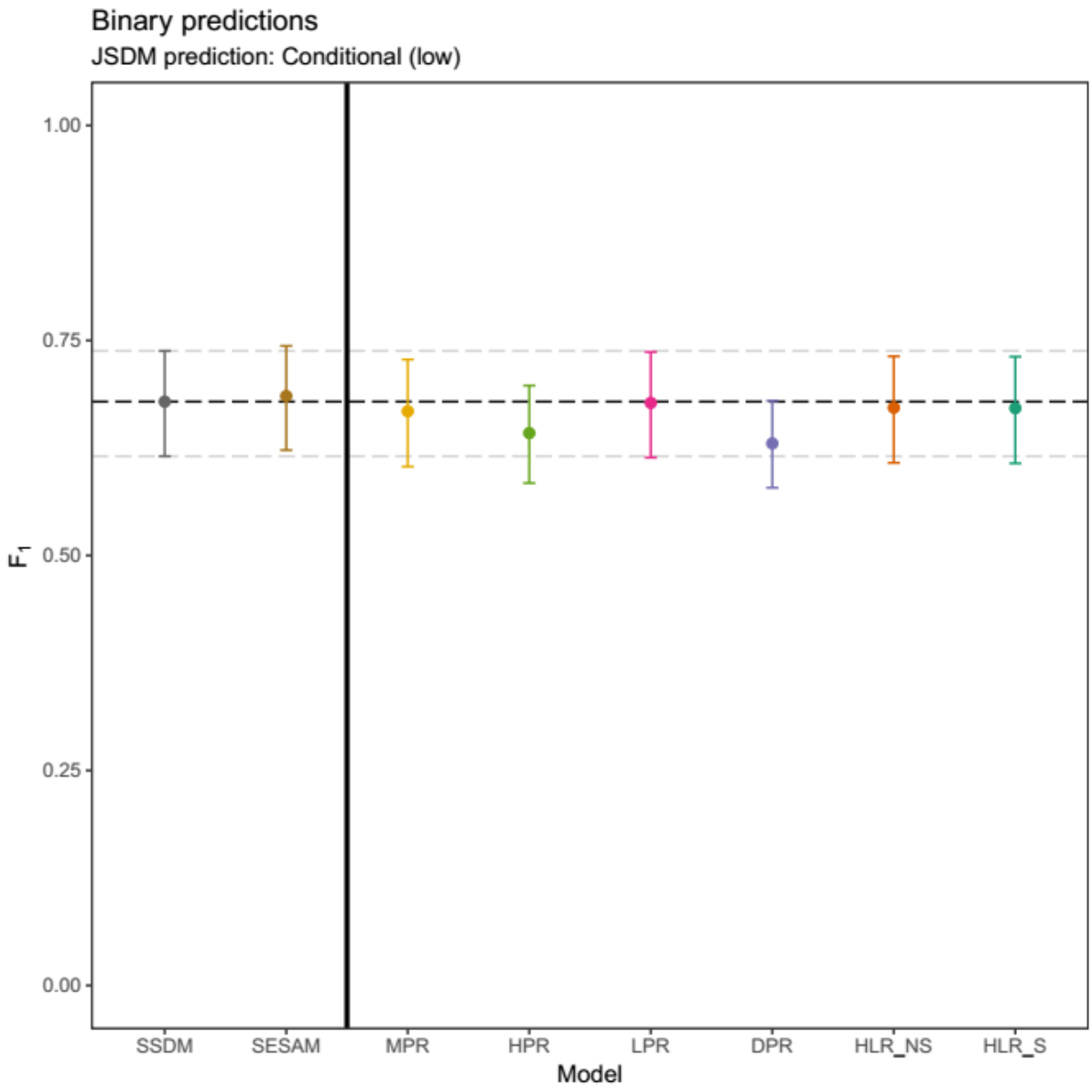


Figure S75: Model performance for the F_1 evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

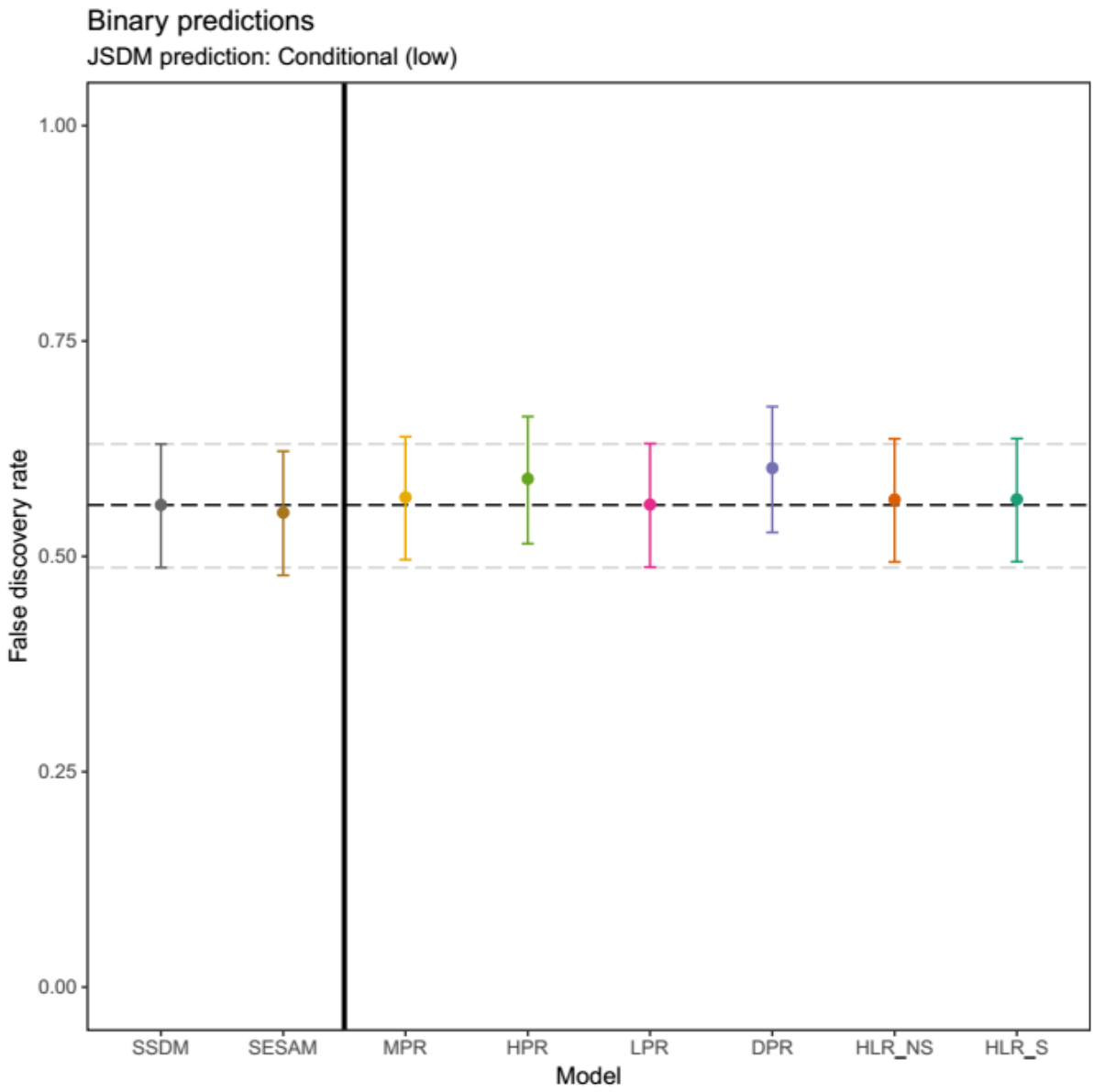


Figure S76: Model performance for the false discovery rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

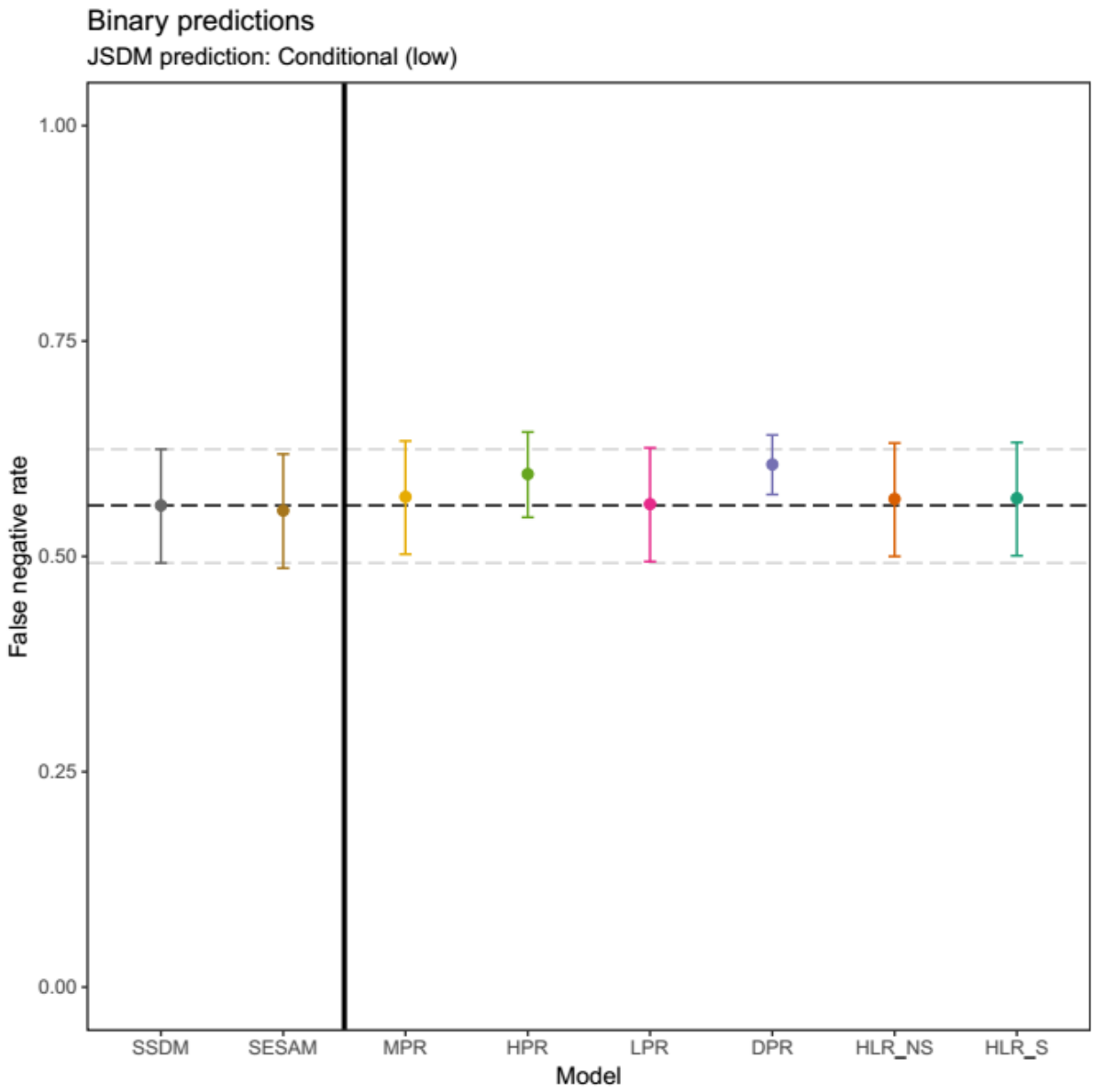


Figure S77: Model performance for the false negative rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

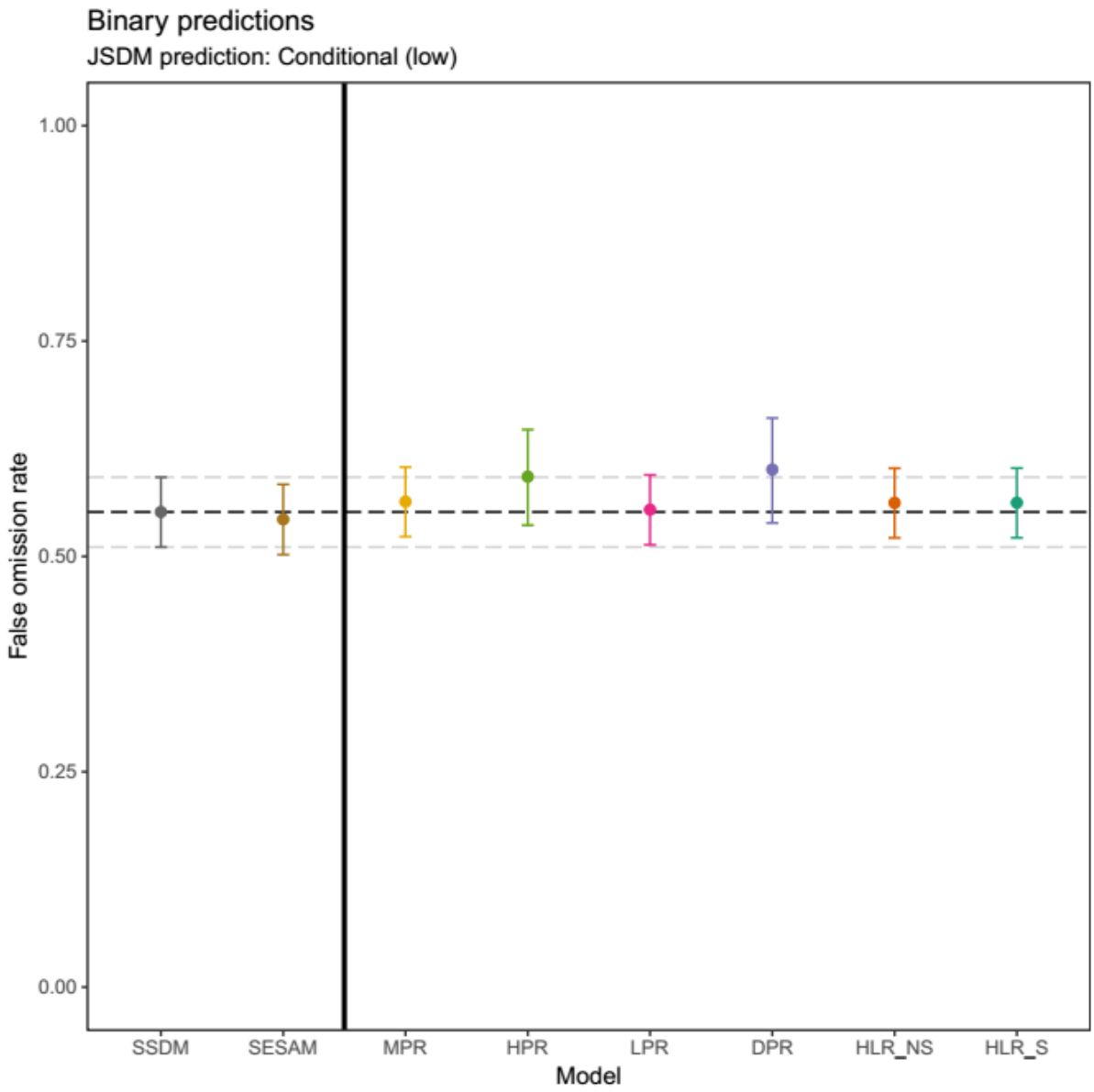


Figure S78: Model performance for the false omission rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

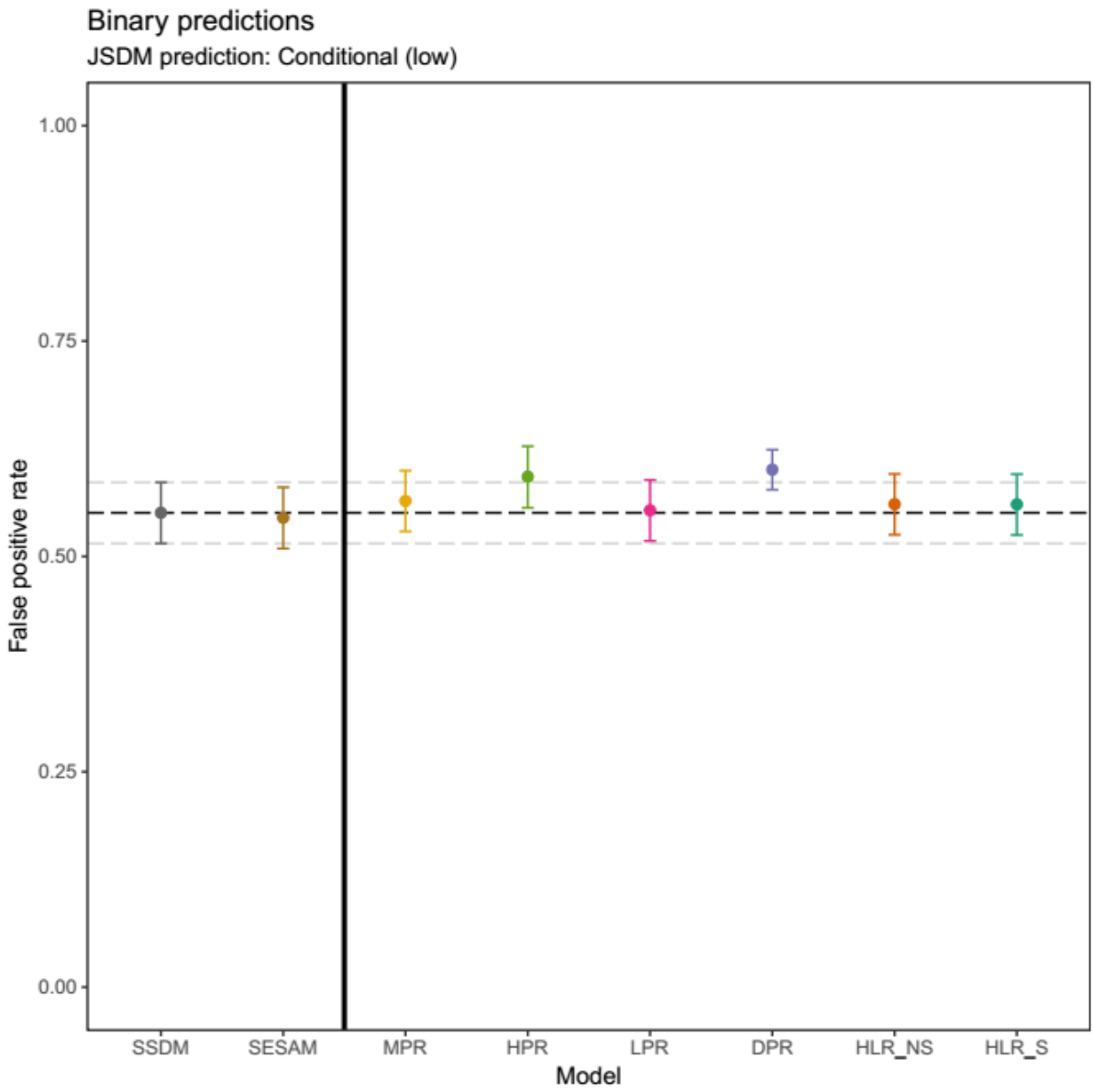


Figure S79: Model performance for the false positive rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

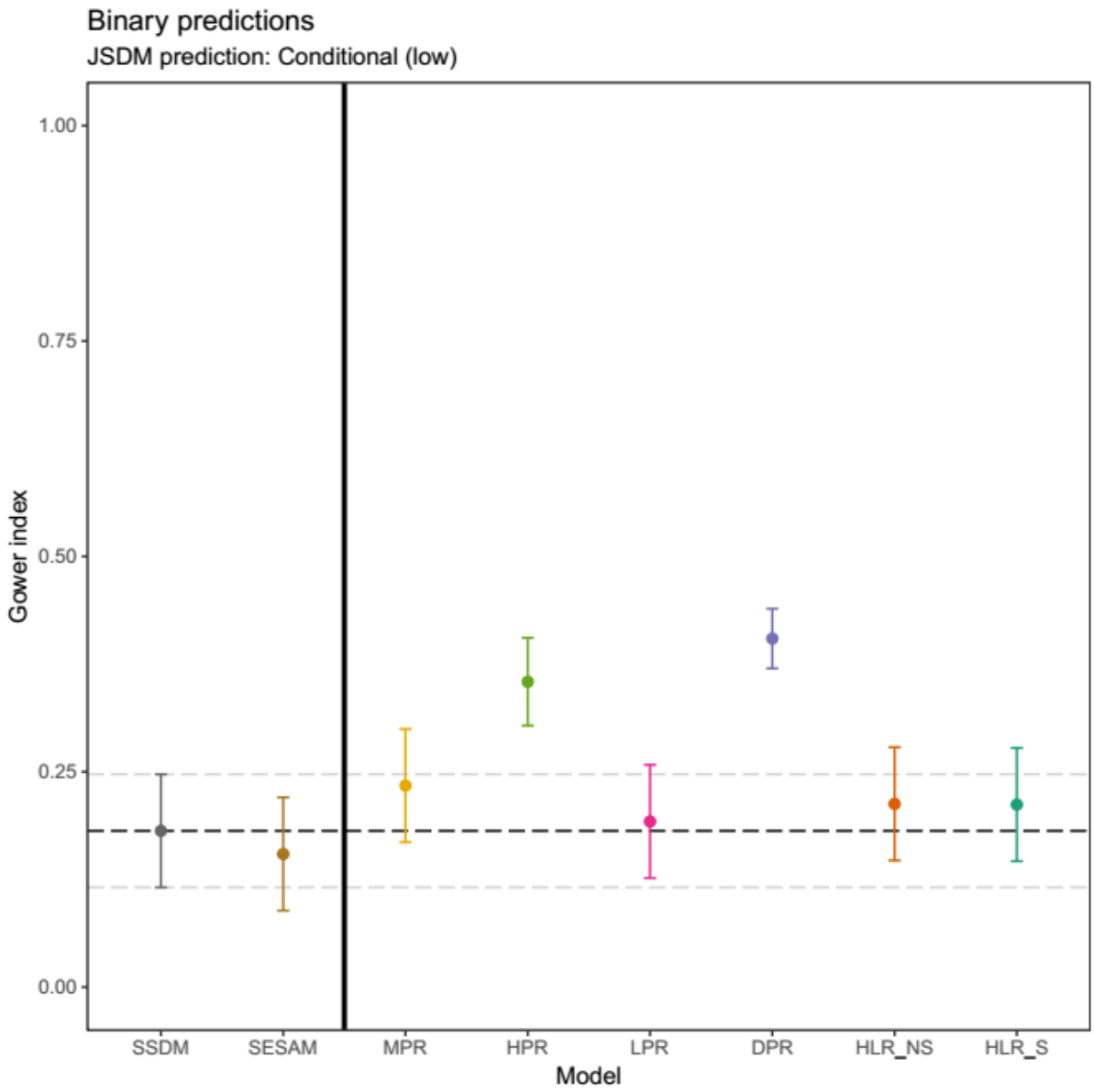


Figure S80: Model performance for the Gower index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

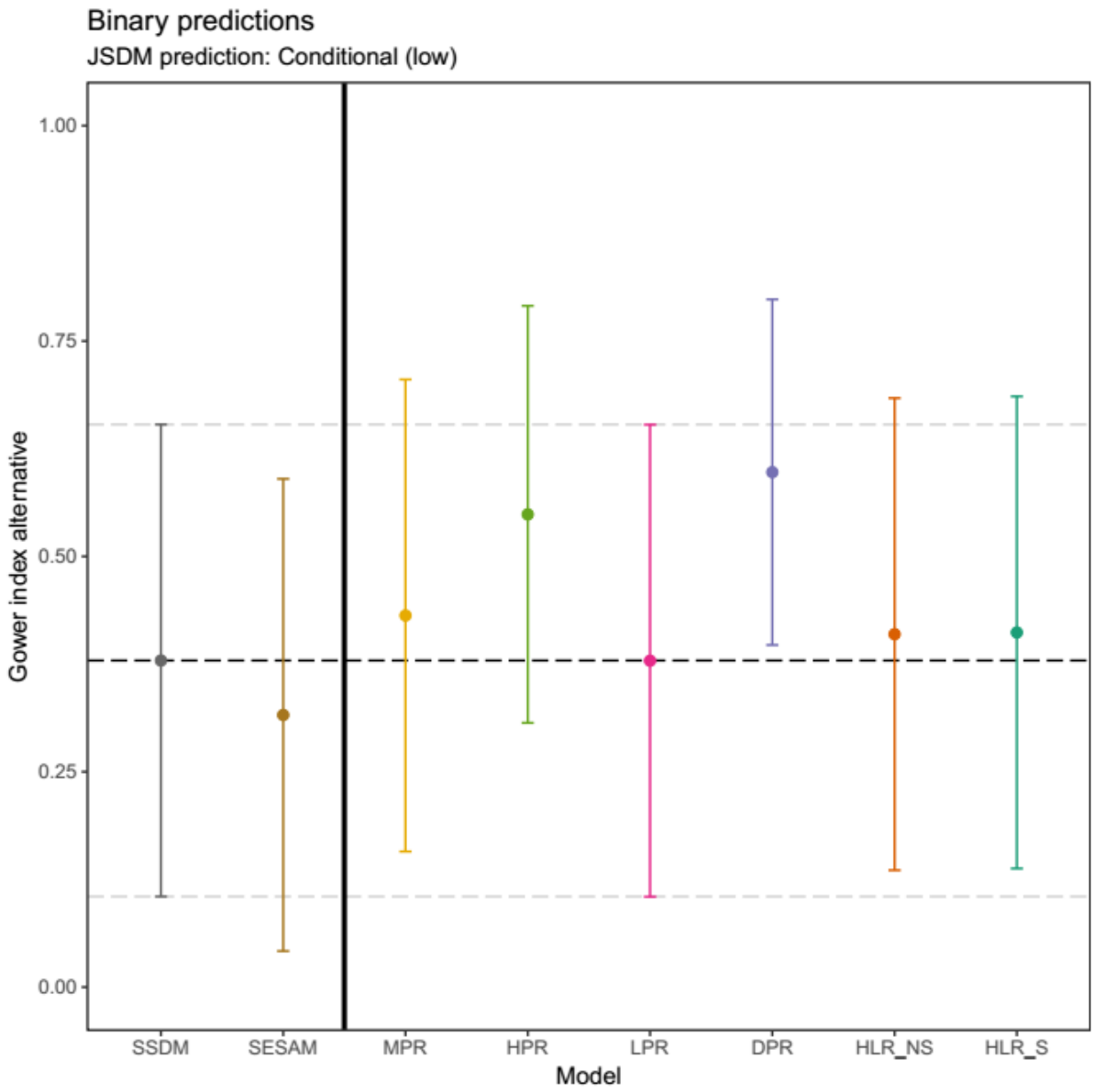


Figure S81: Model performance for the alternative Gower index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

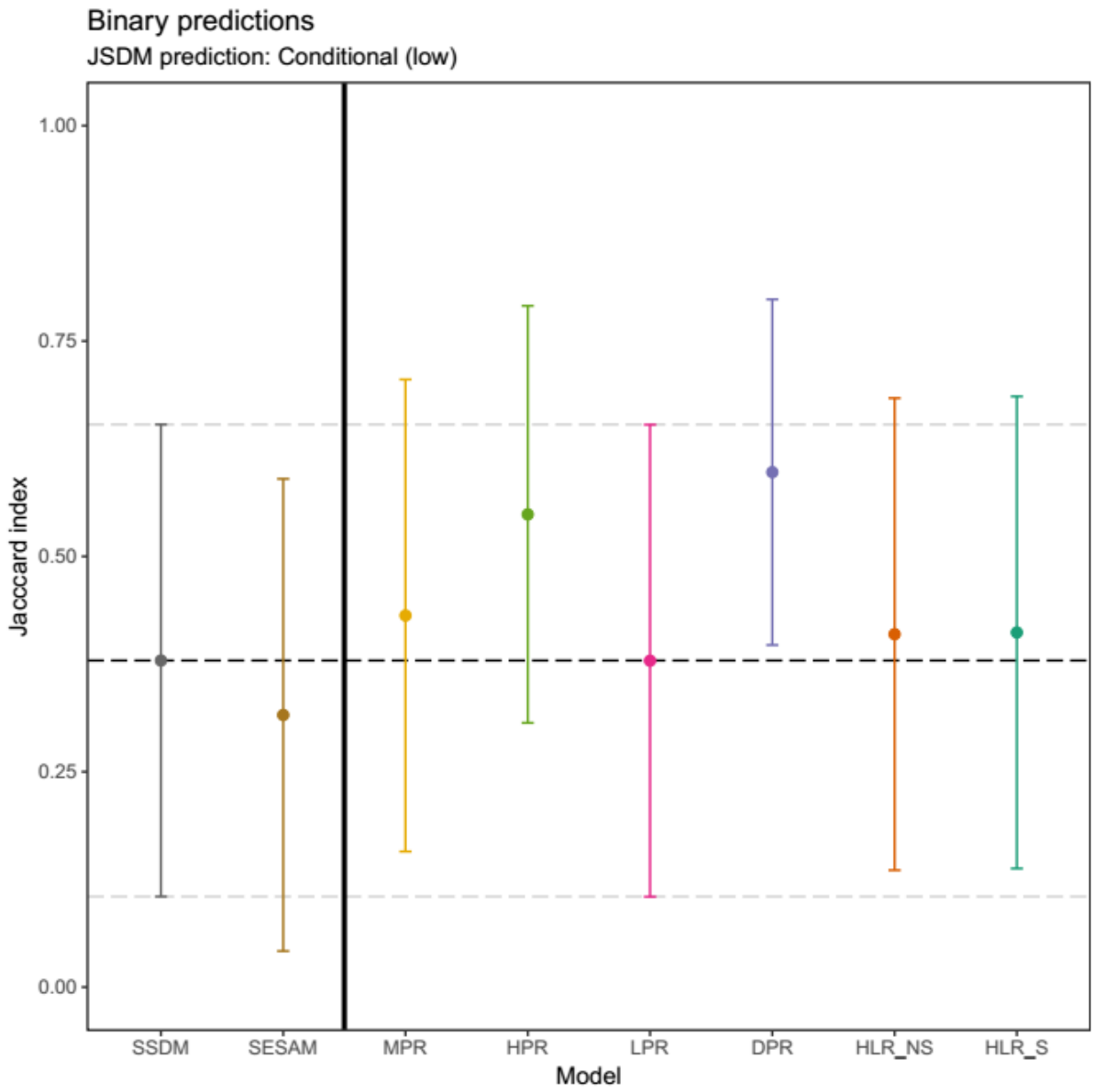


Figure S82: Model performance for the Jaccard index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

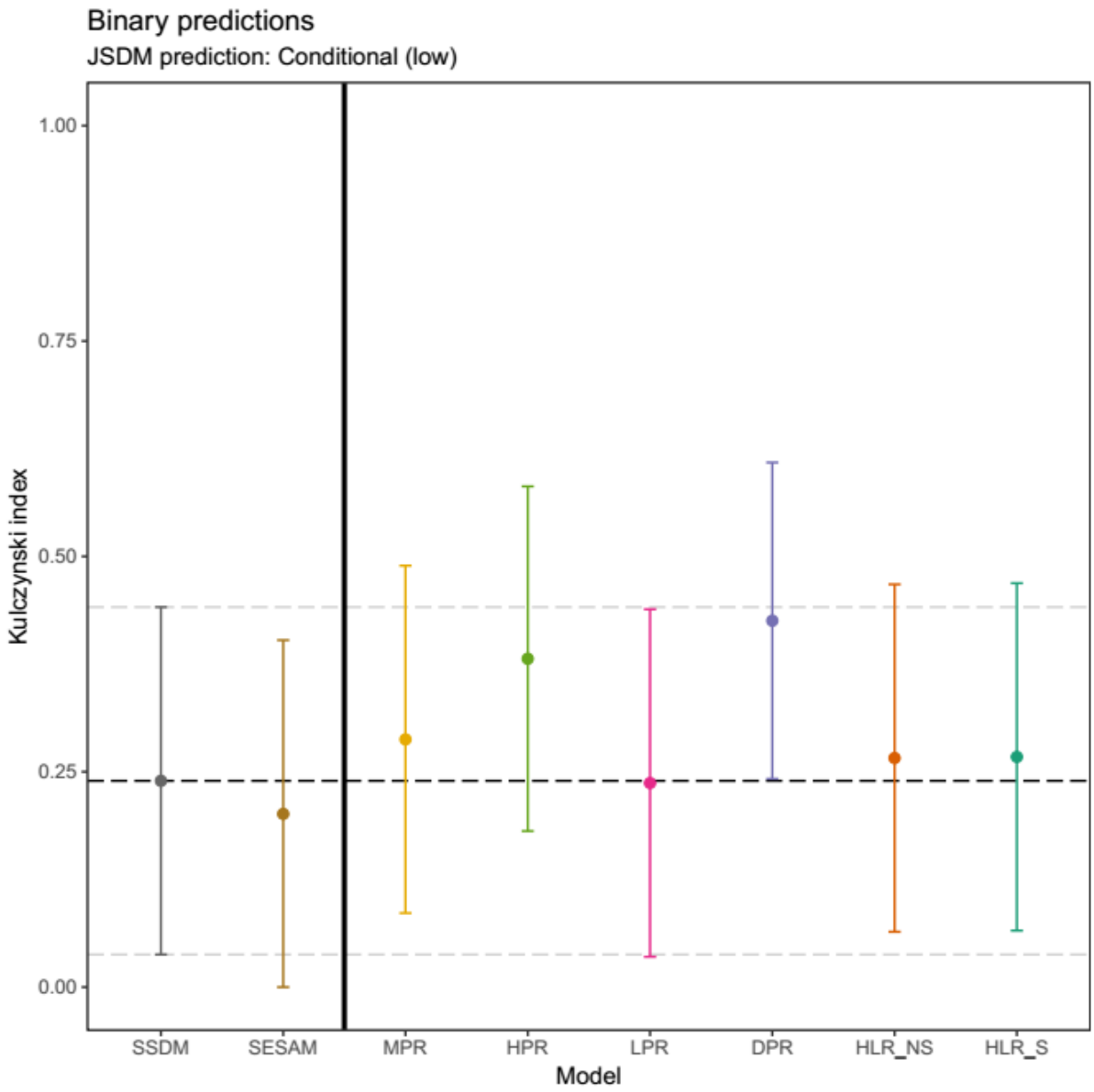


Figure S83: Model performance for the Kulczynski index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

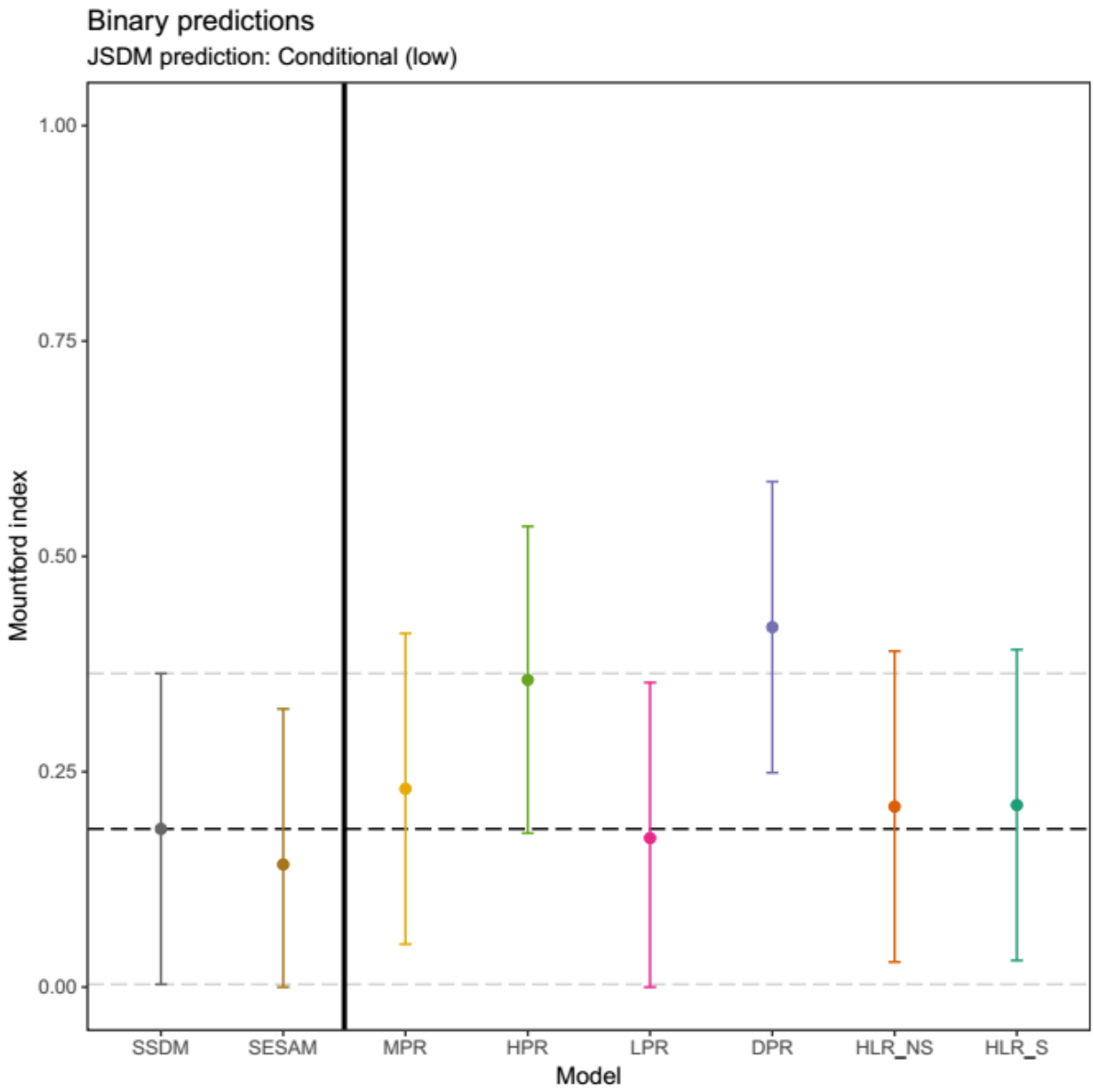


Figure S84: Model performance for the Mountford index evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

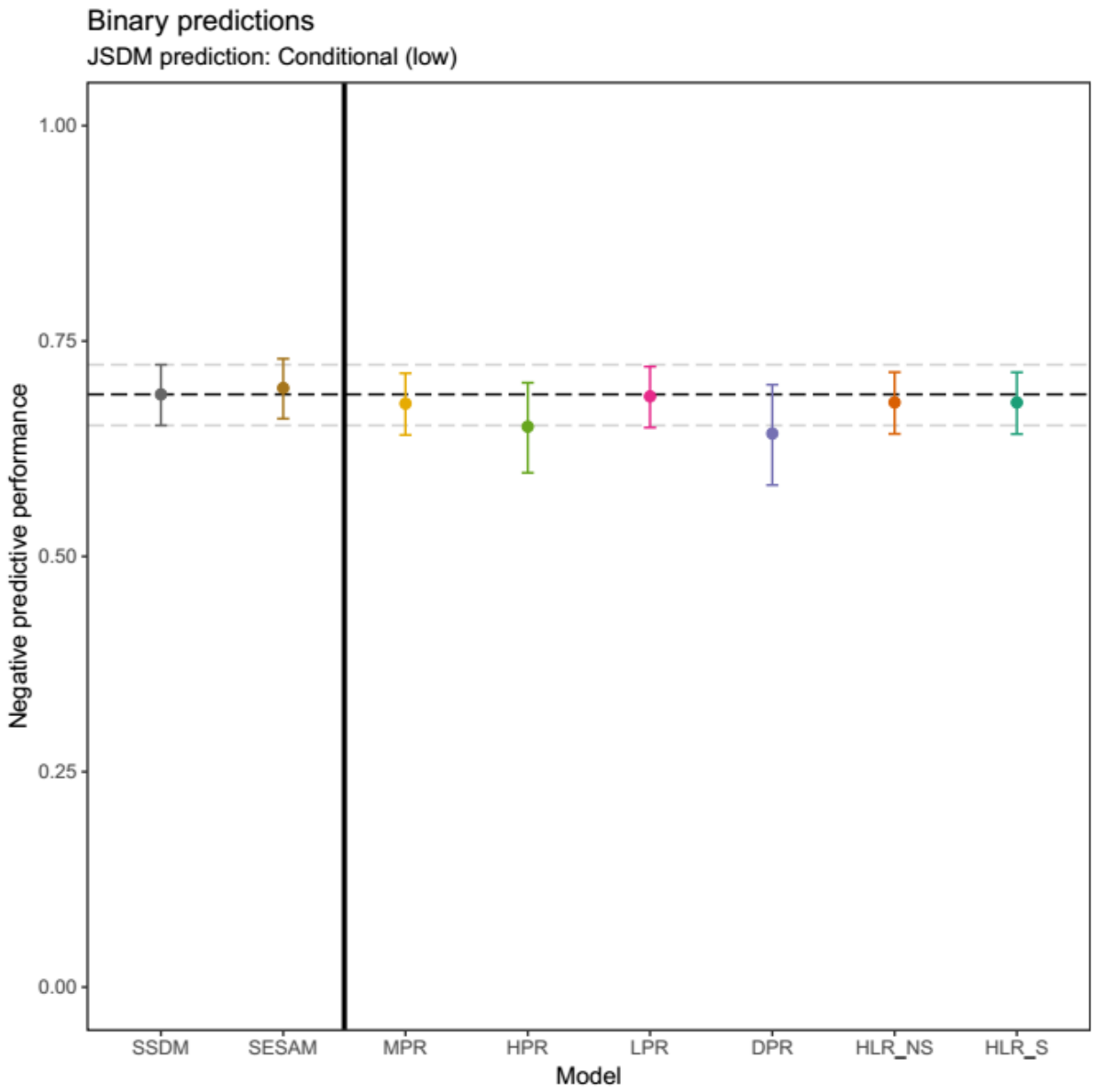


Figure S85: Model performance for the negative predictive performance evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

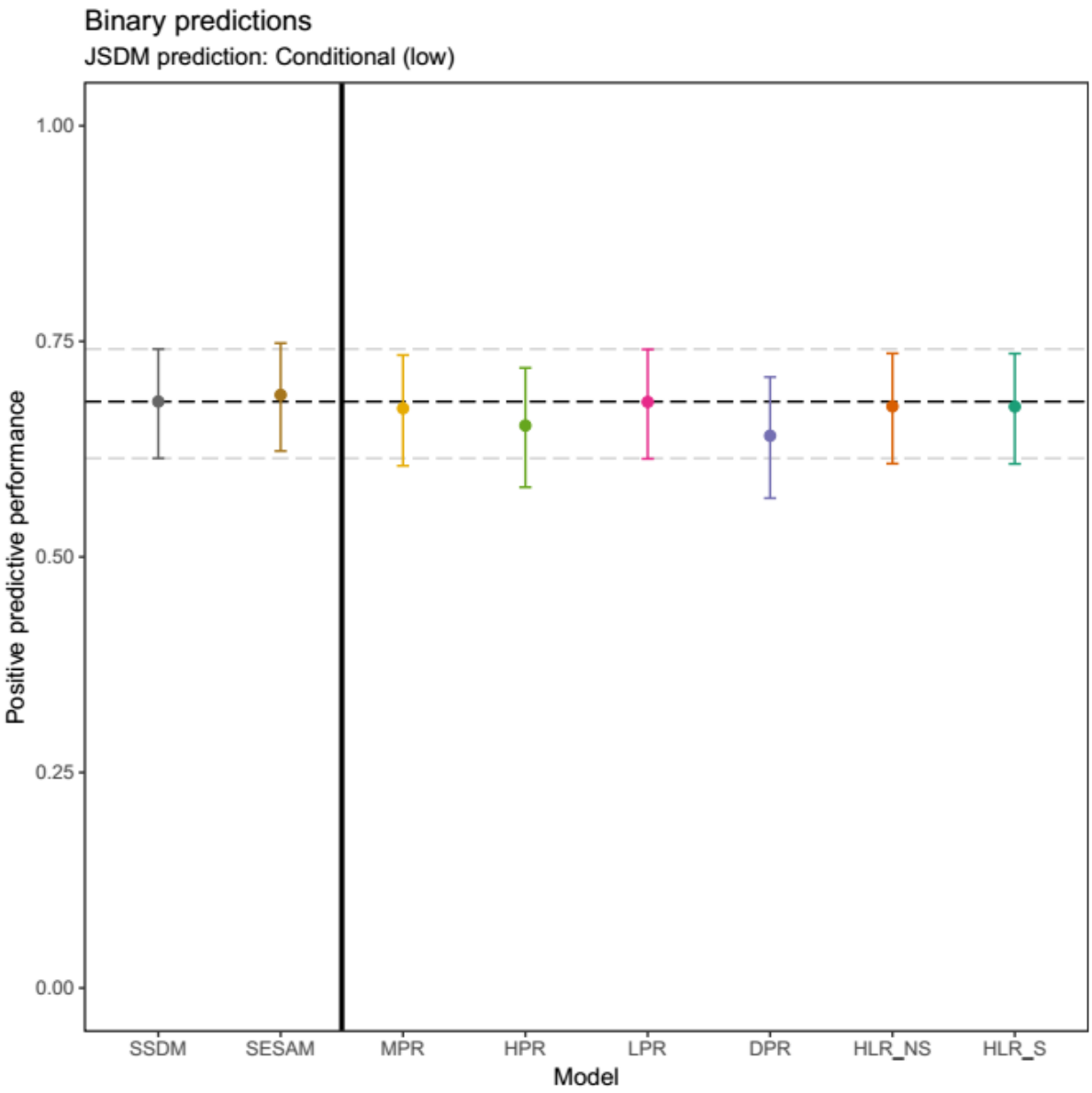


Figure S86: Model performance for the positive predictive performance evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

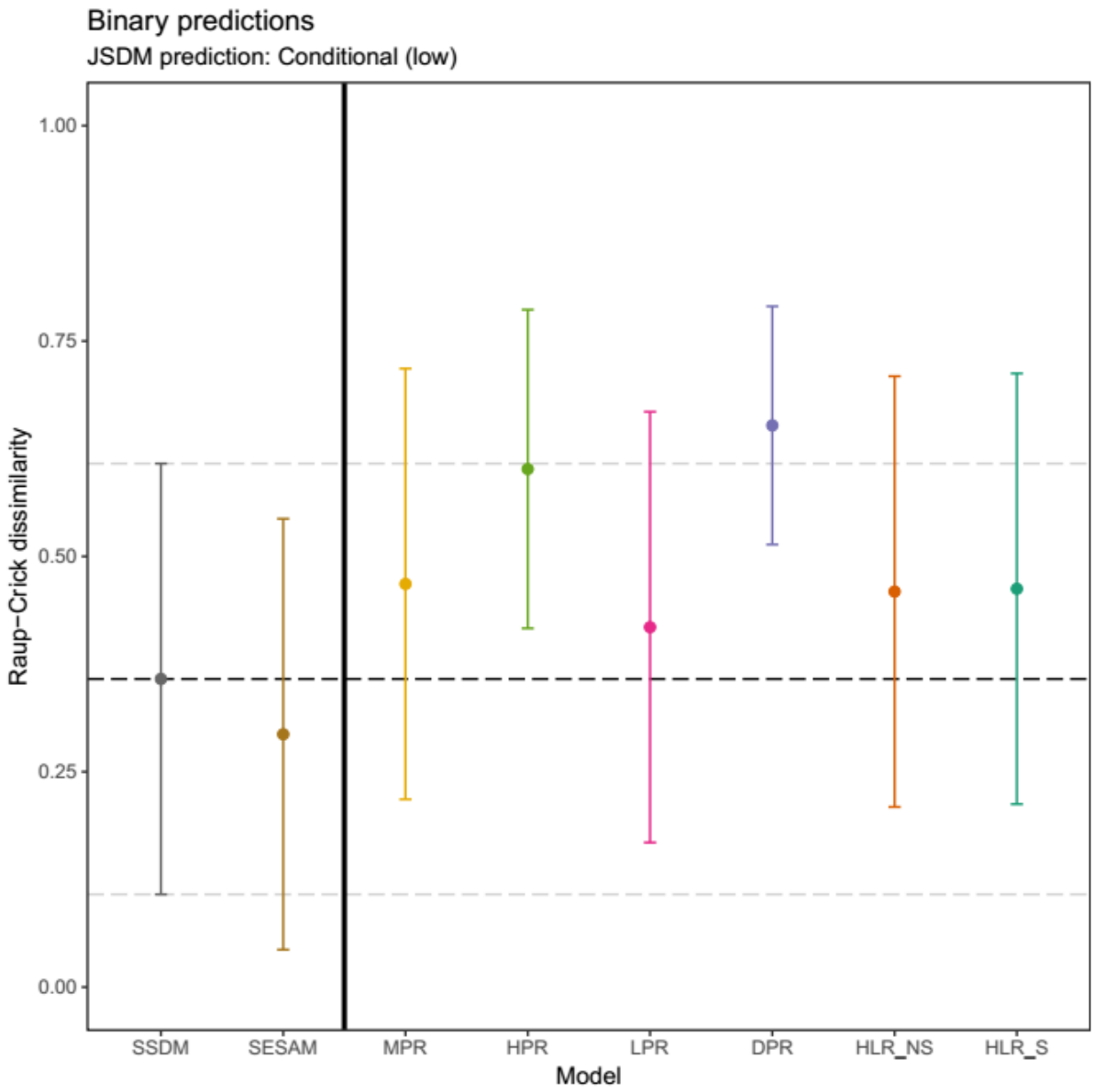


Figure S87: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

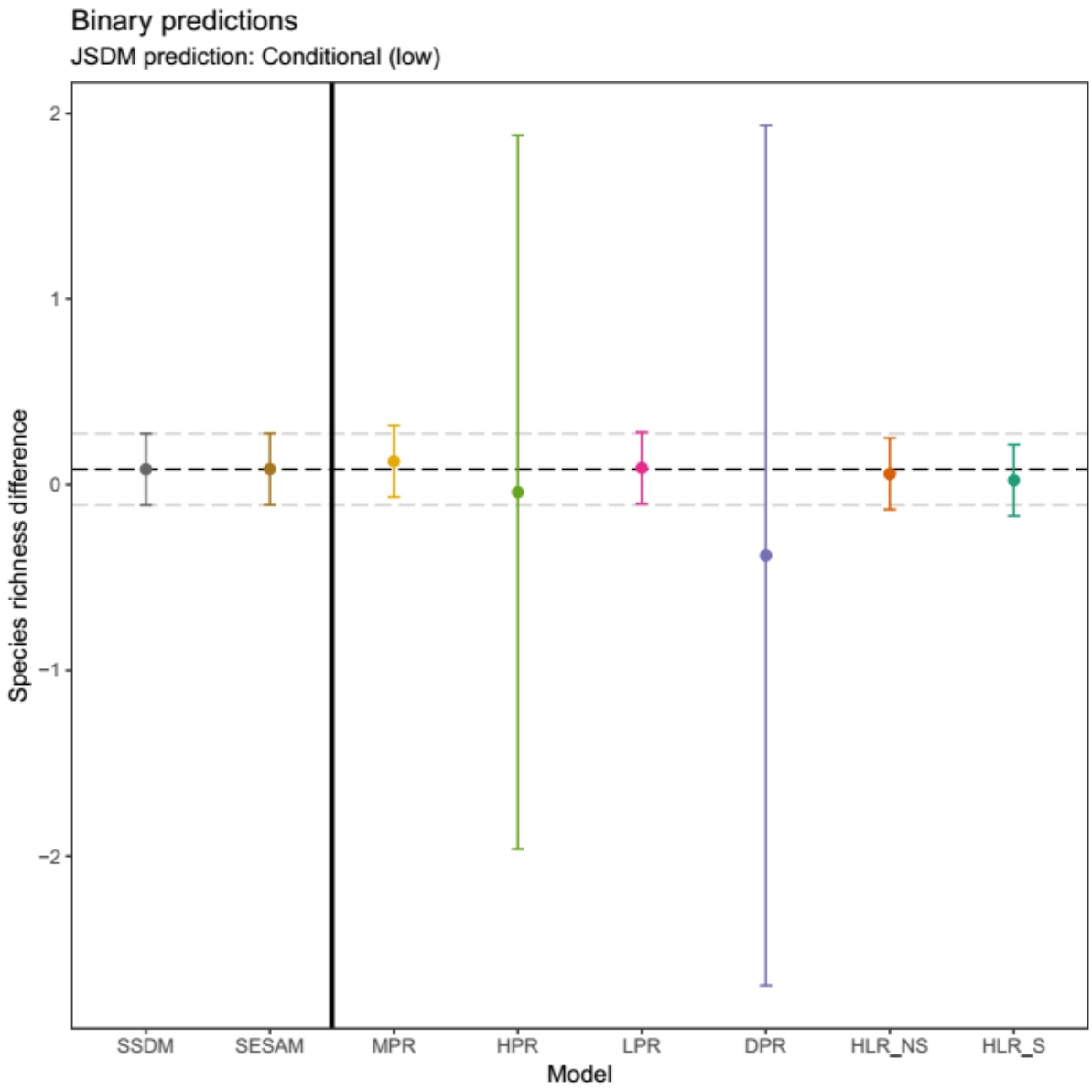


Figure S88: Model performance for the species richness difference evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

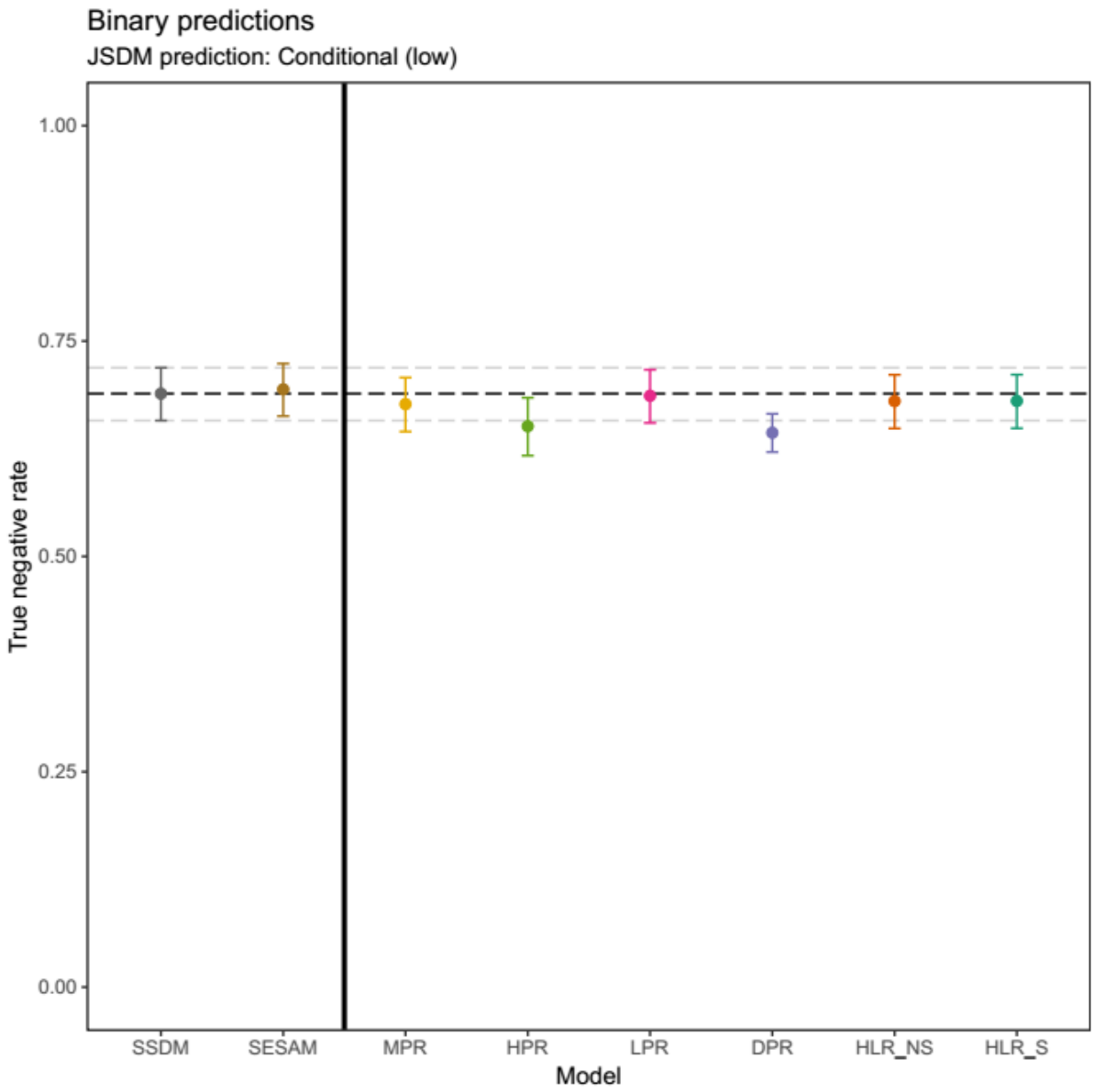


Figure S89: Model performance for the true negative rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

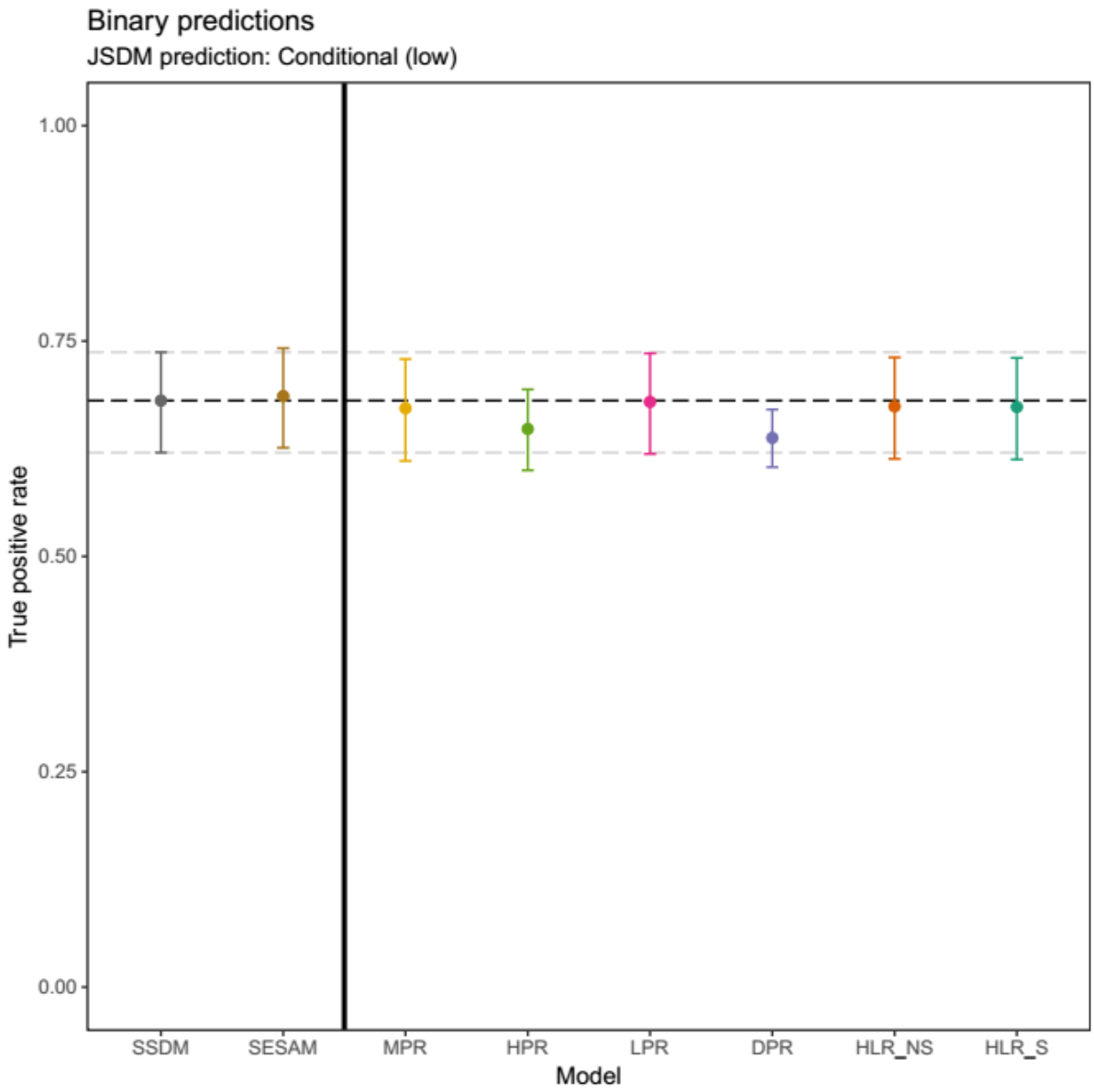


Figure S90: Model performance for the true positive rate evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

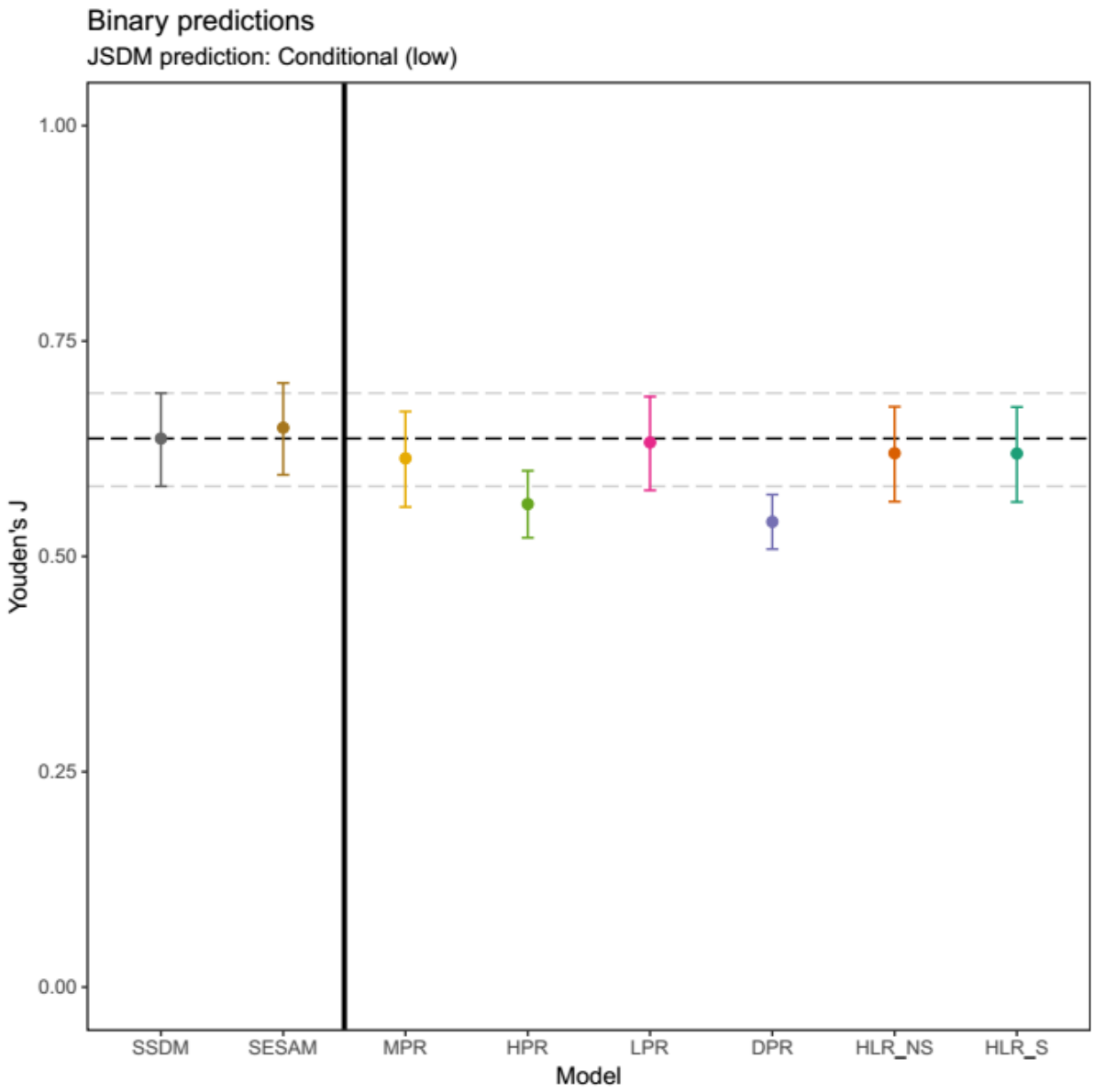


Figure S91: Model performance for the Youden's J evaluation metric for conditional (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

5. Conditional (medium-prevalence) predictions

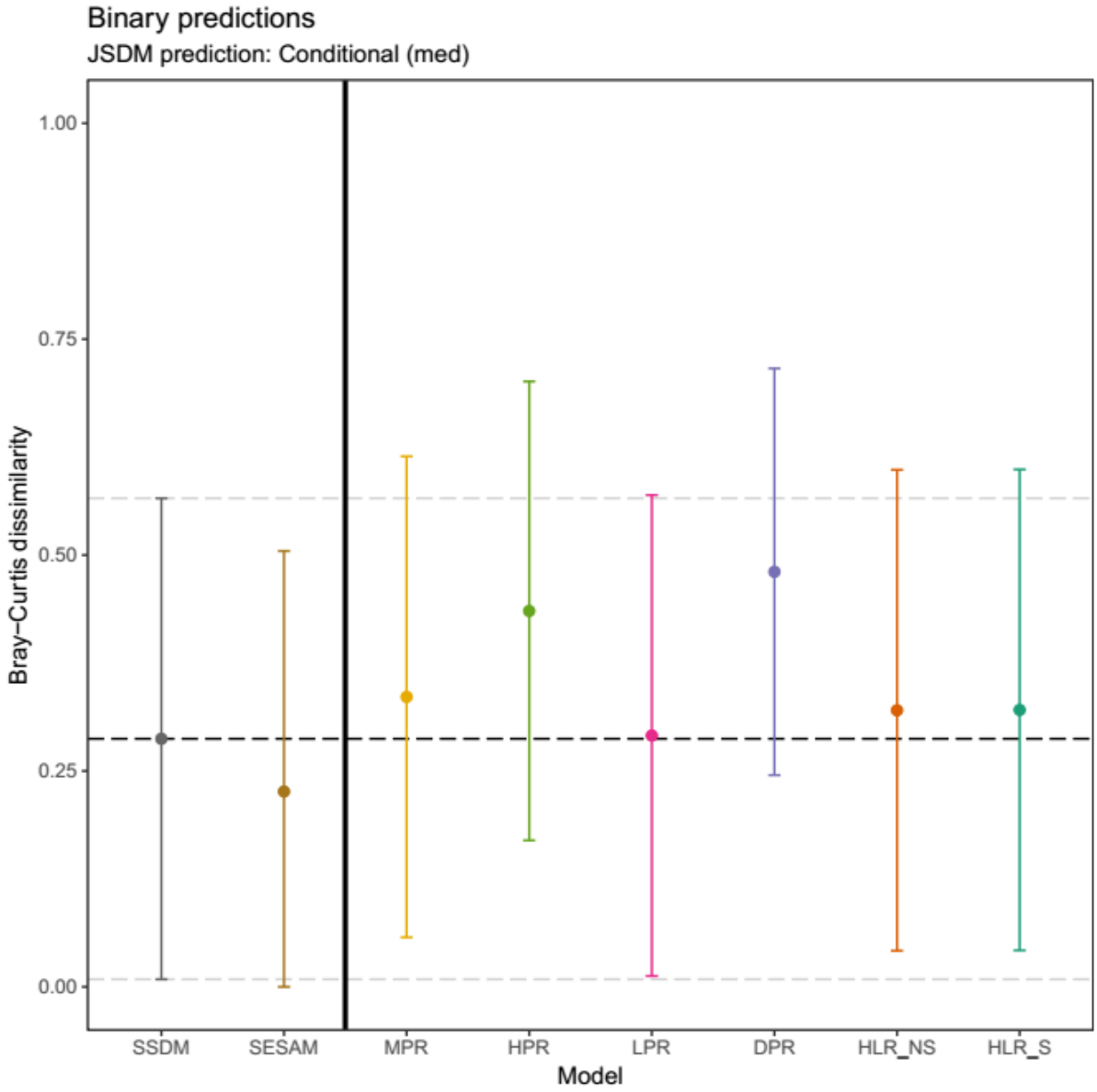


Figure S92: Model performance for the Bray-Curtis dissimilarity evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

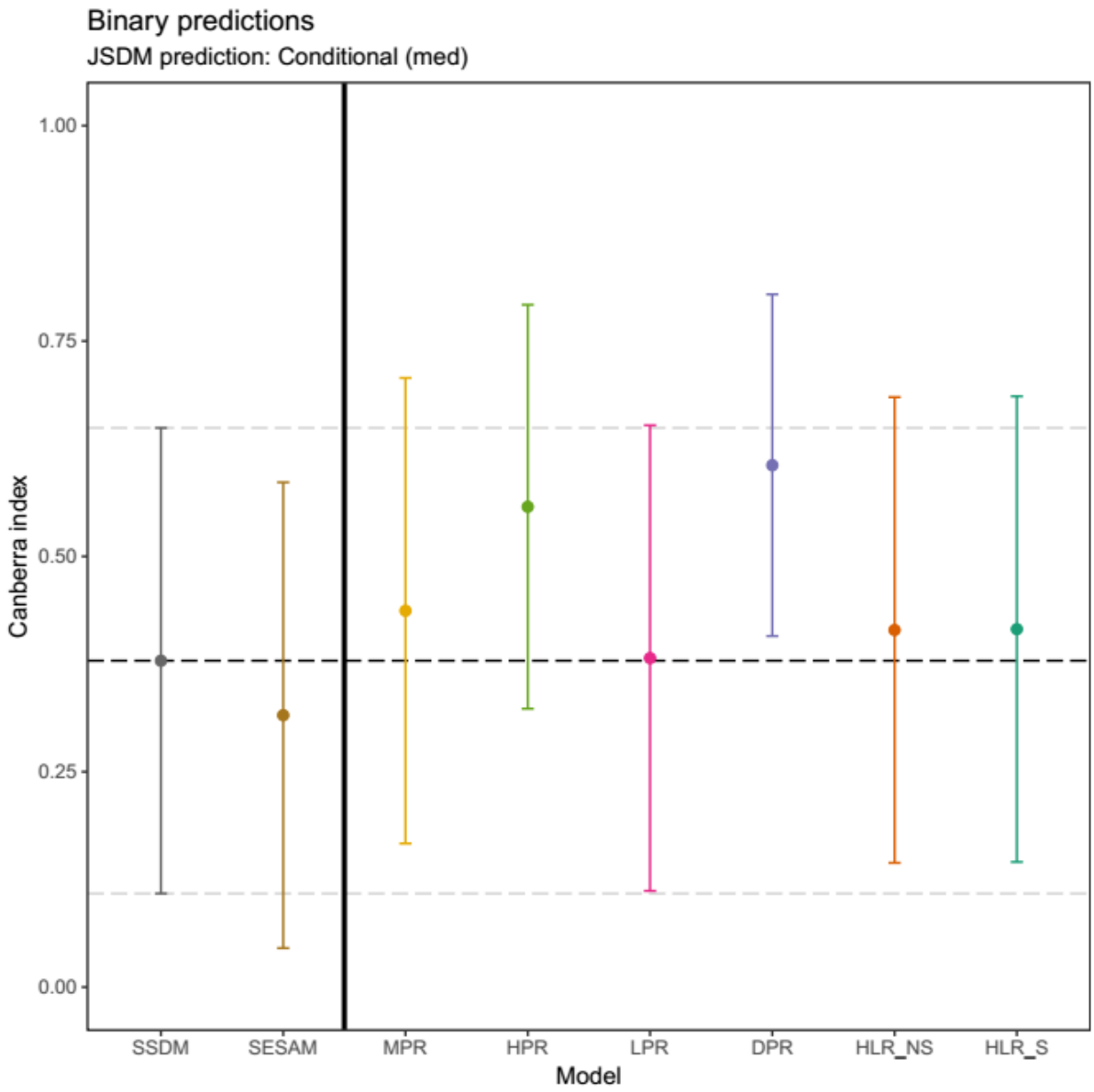


Figure S93: Model performance for the Canberra index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

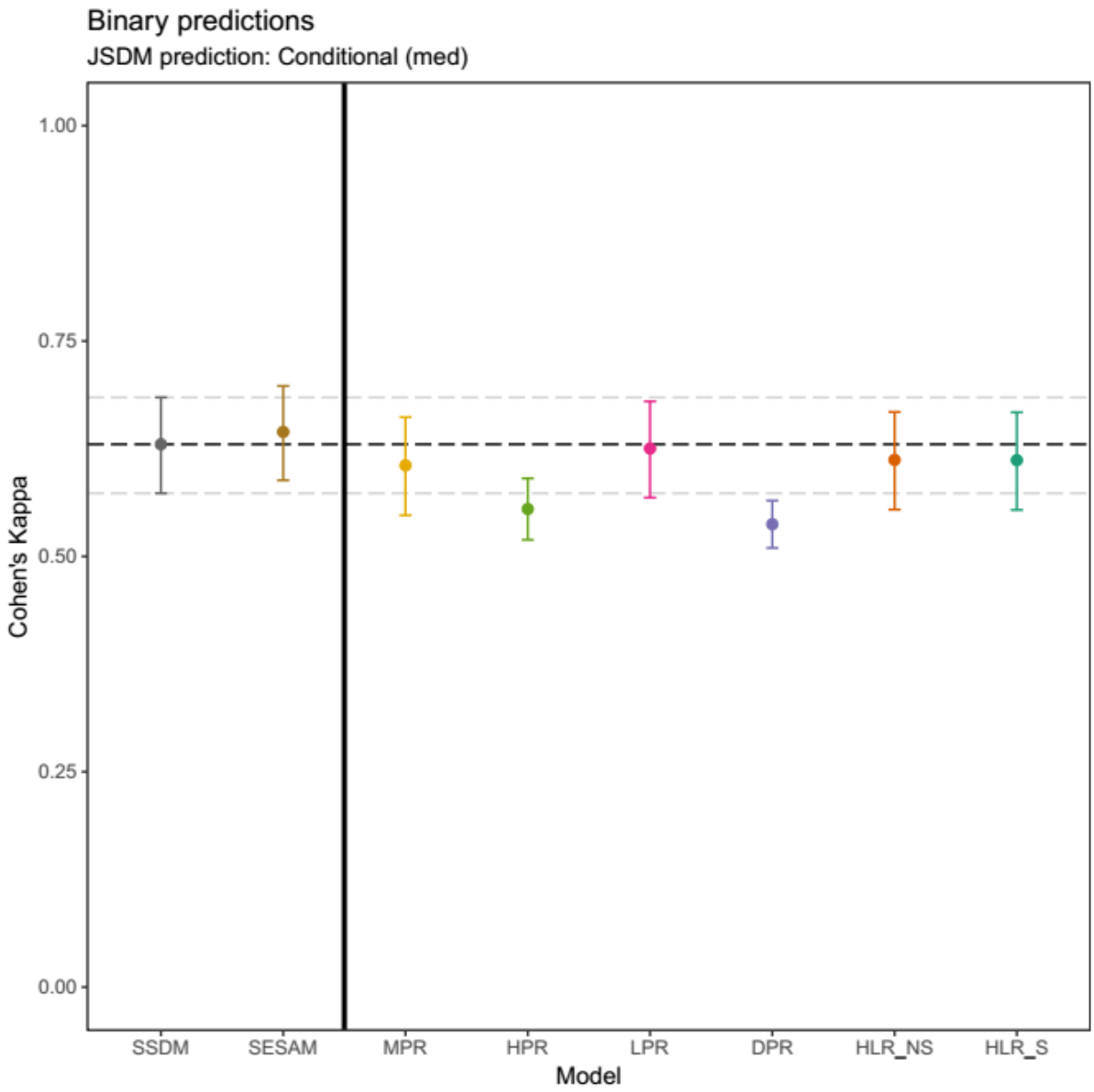


Figure S94: Model performance for the Cohen's Kappa evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

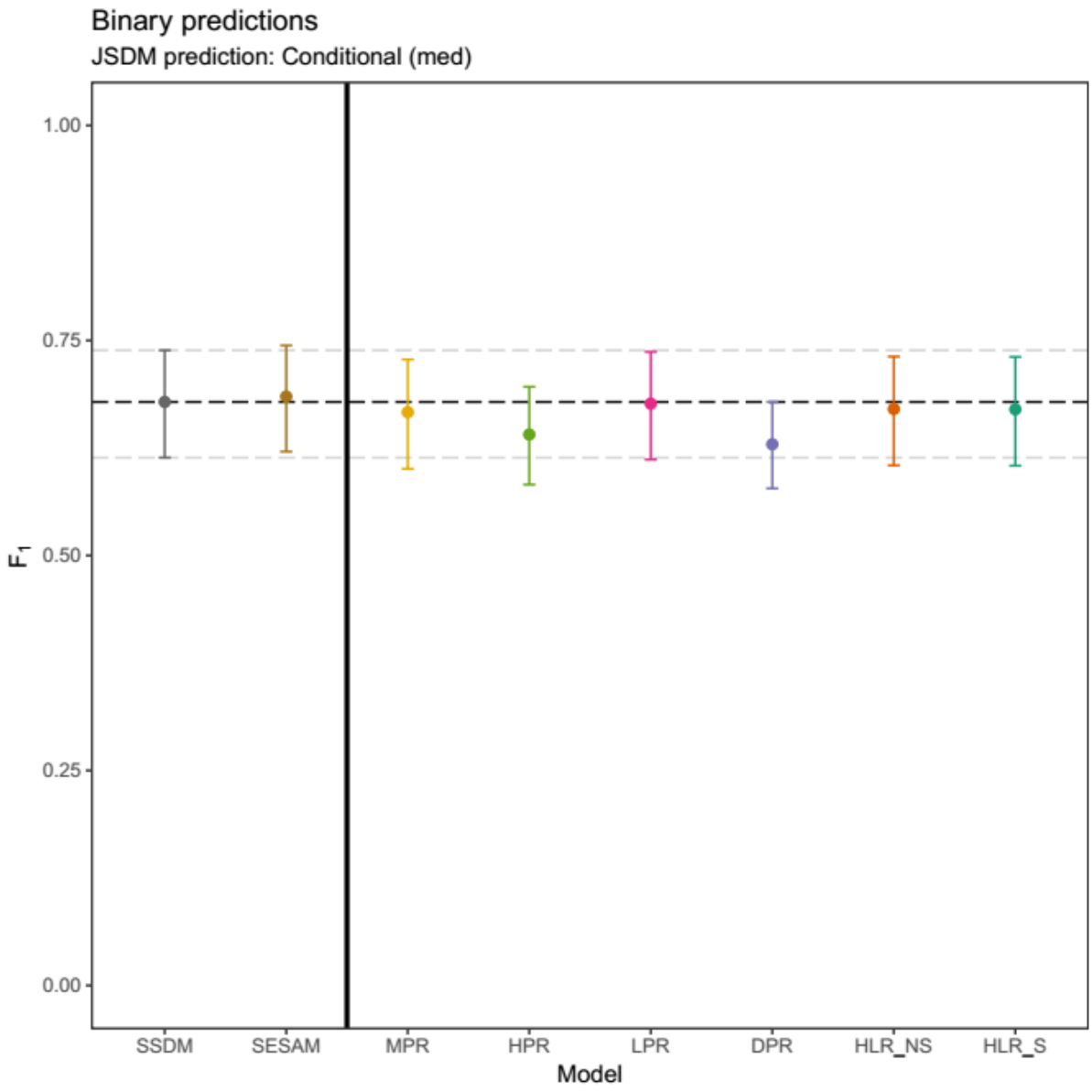


Figure S95: Model performance for the F_1 evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

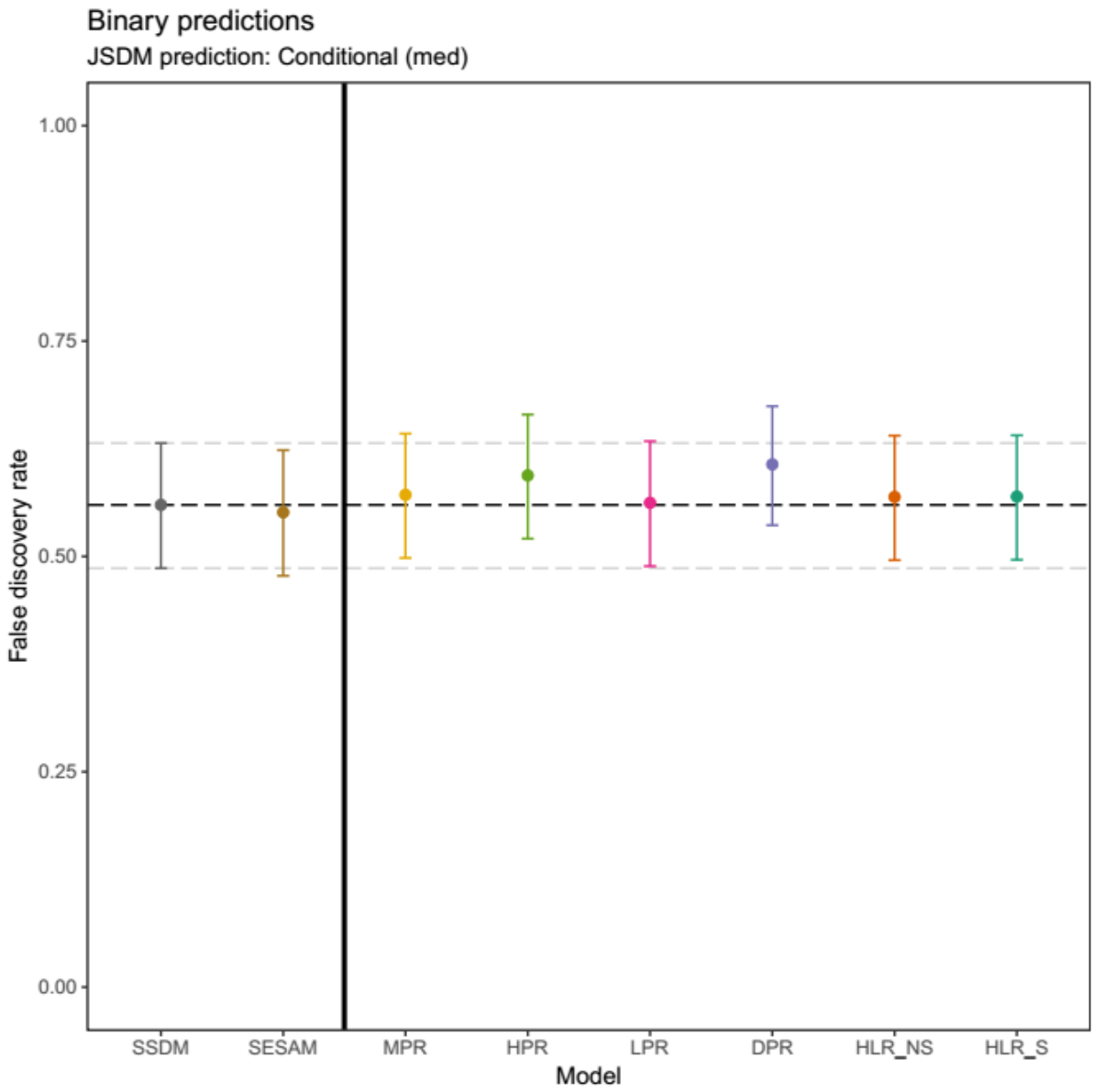


Figure S96: Model performance for the false discovery rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

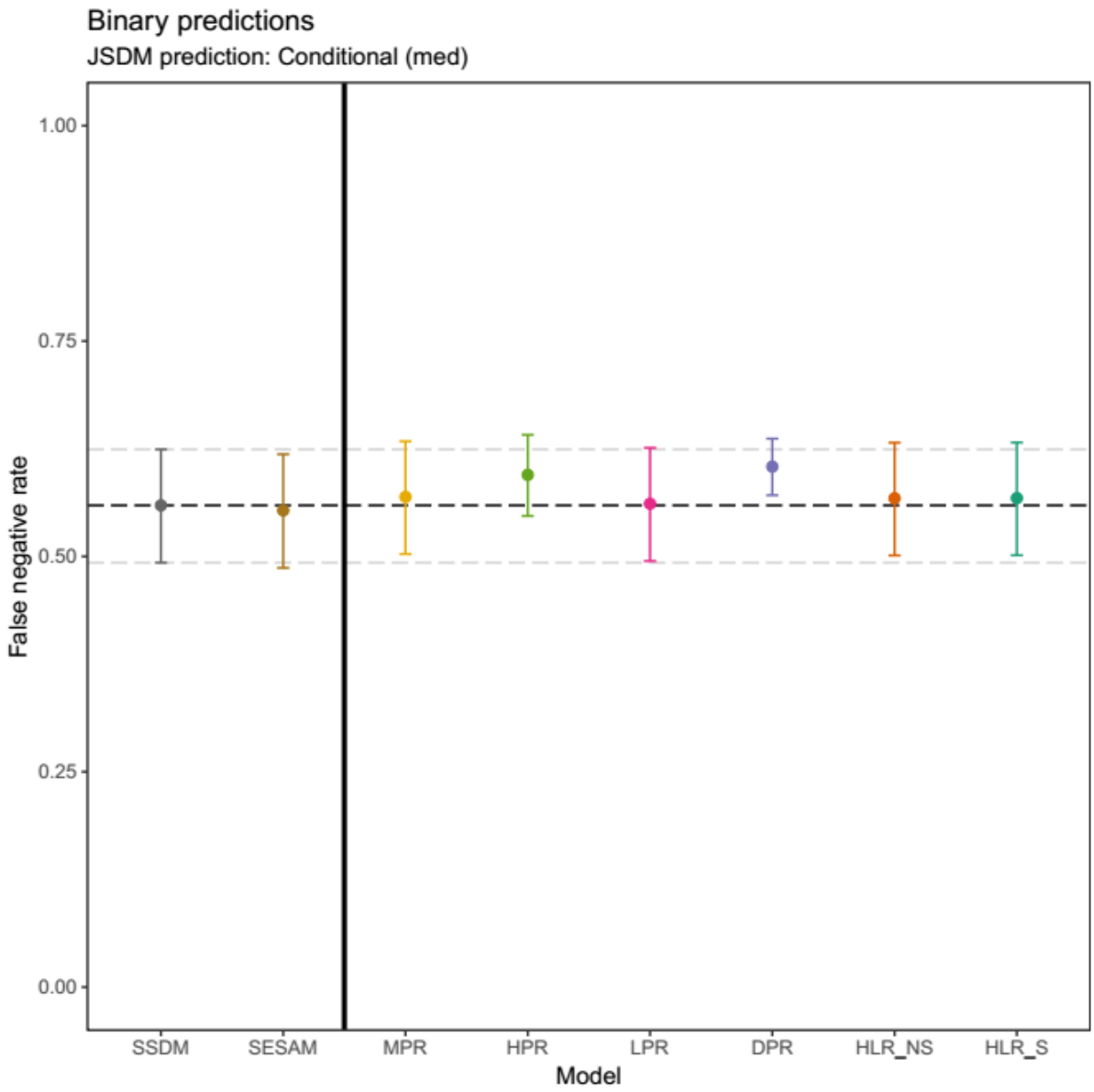


Figure S97: Model performance for the false negative rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

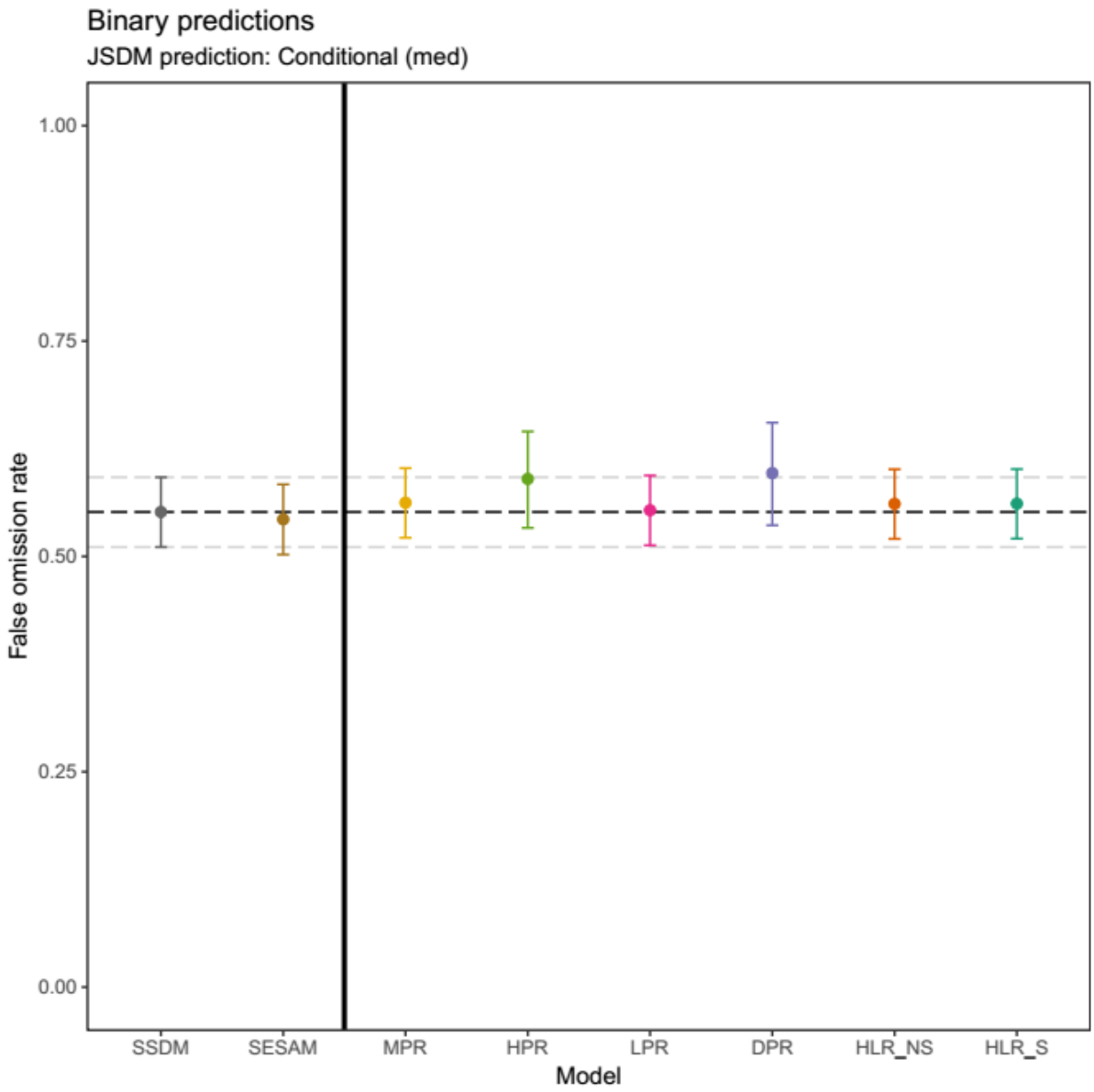


Figure S98: Model performance for the false omission rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

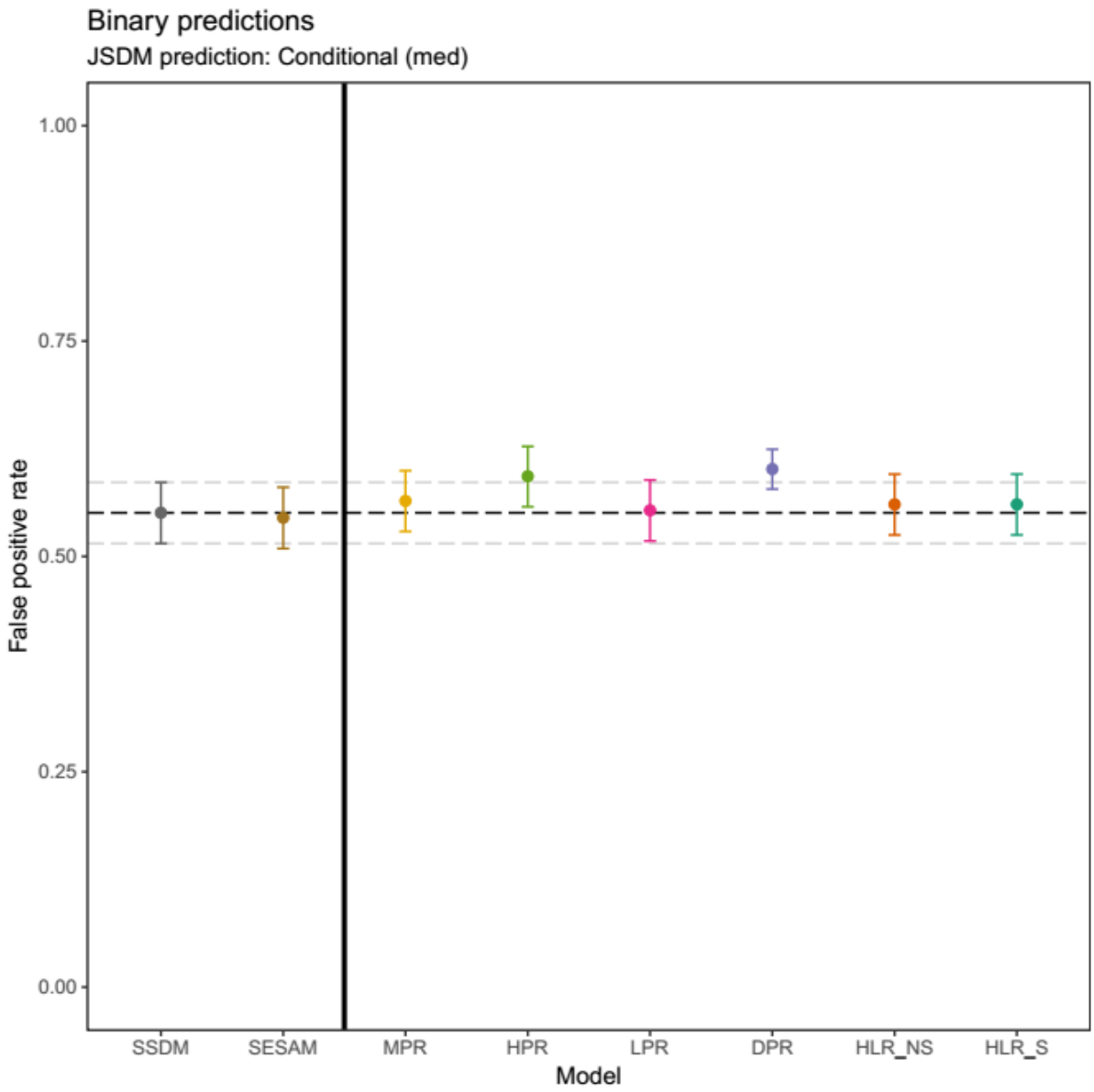


Figure S99: Model performance for the false positive rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

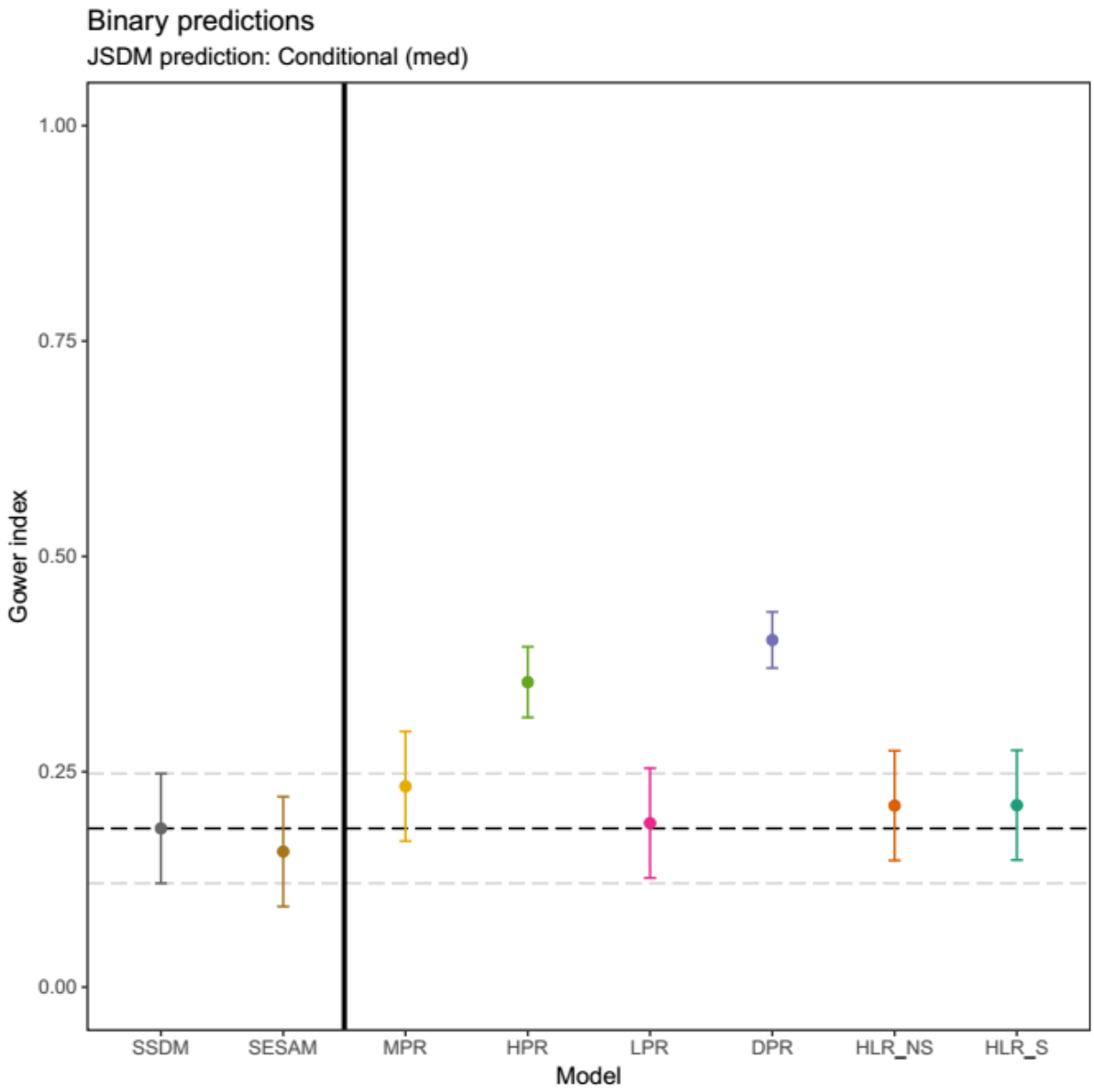


Figure S100: Model performance for the Gower index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

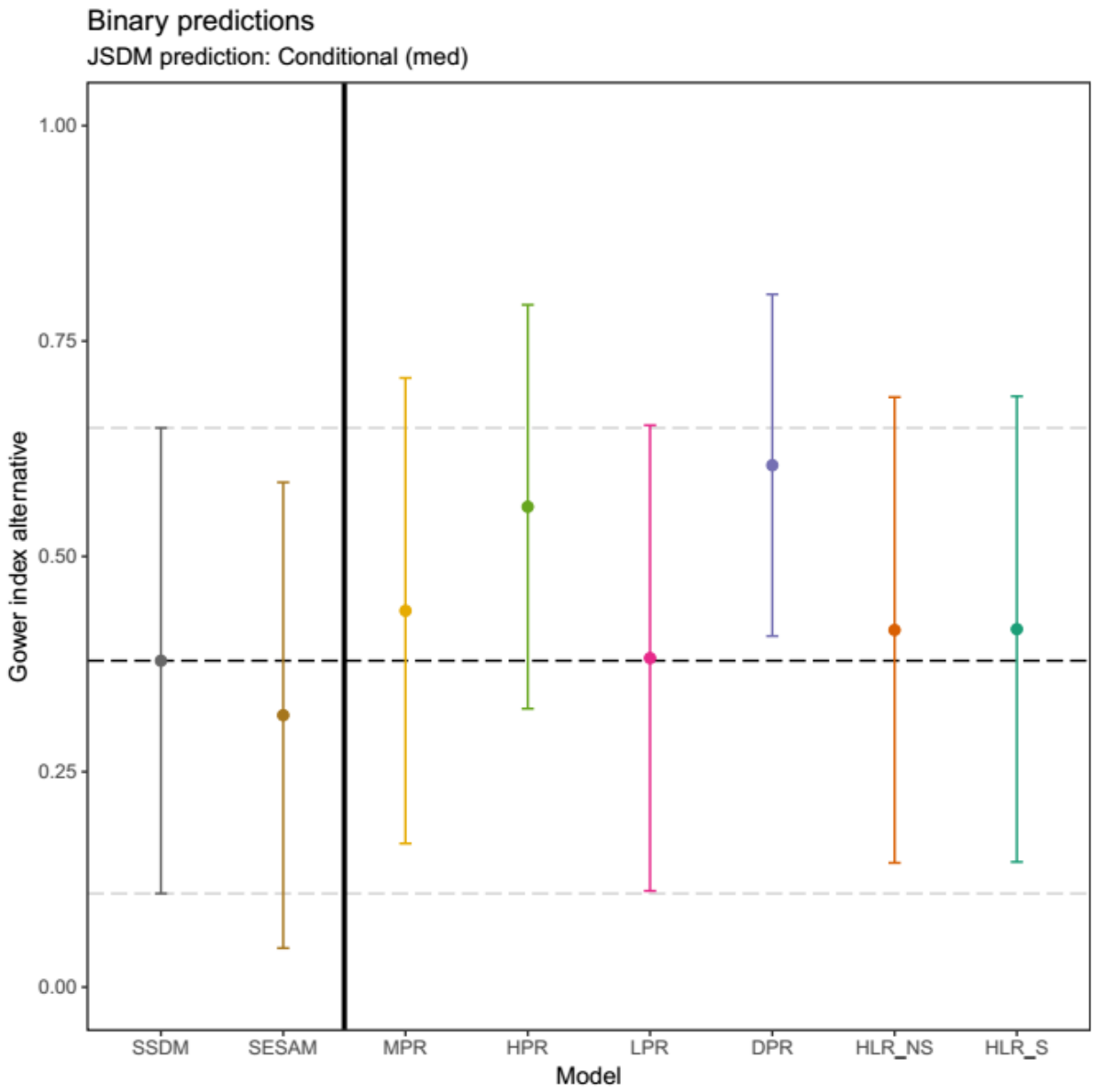


Figure S101: Model performance for the alternative Gower index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

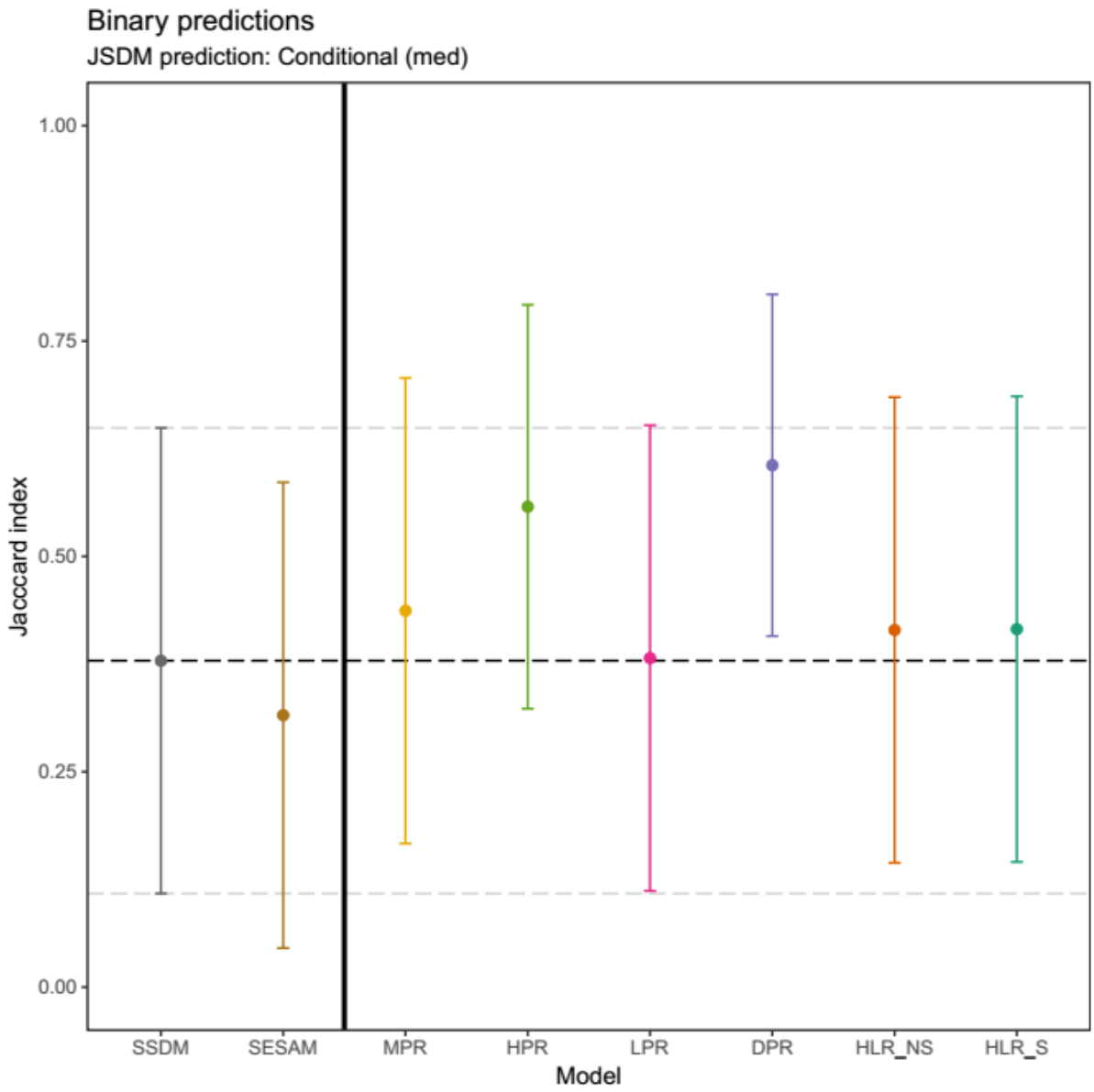


Figure S102: Model performance for the Jaccard index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

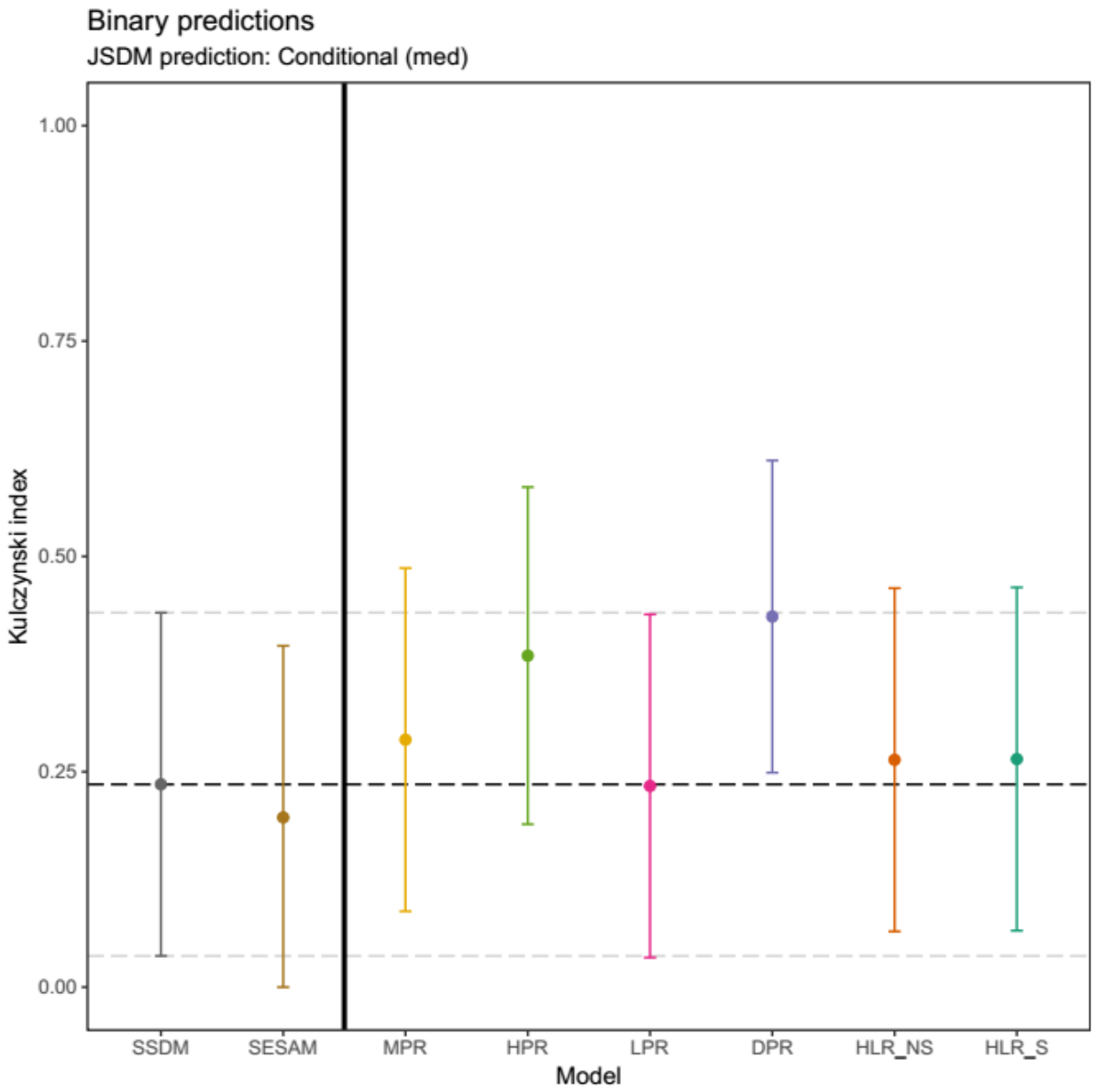


Figure S103: Model performance for the Kulczynski index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

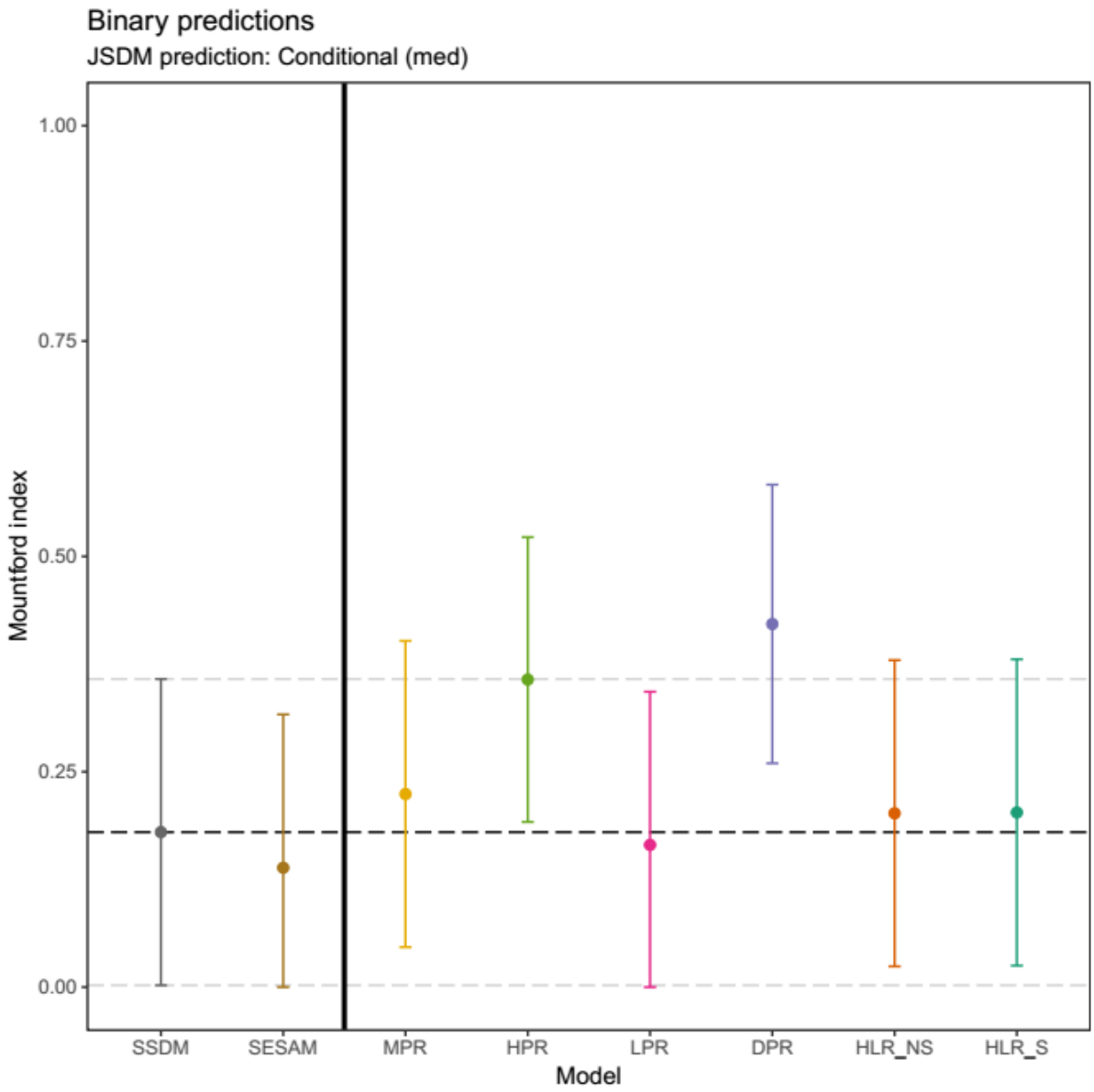


Figure S104: Model performance for the Mountford index evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

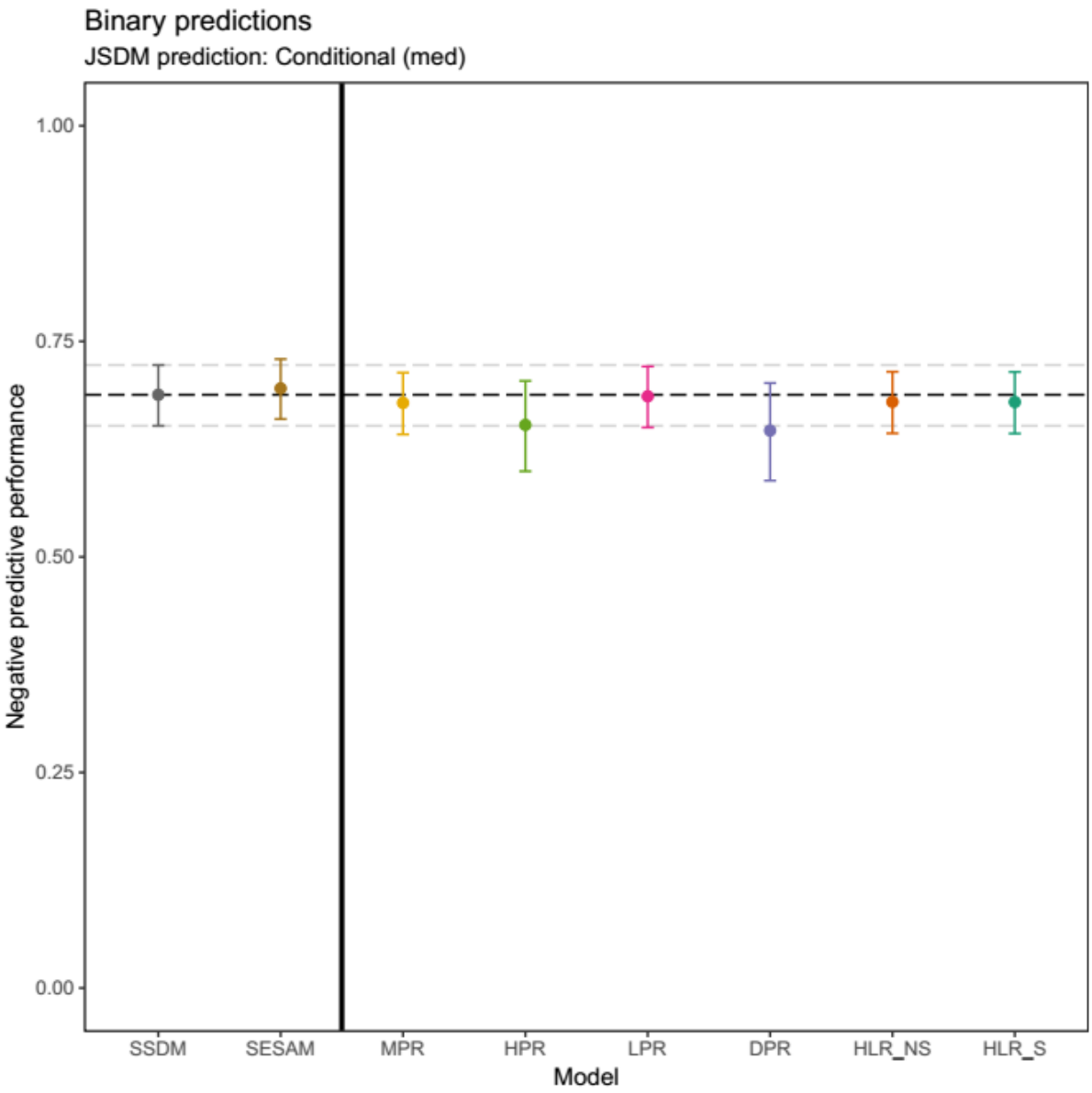


Figure S105: Model performance for the Negative predictive performance evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

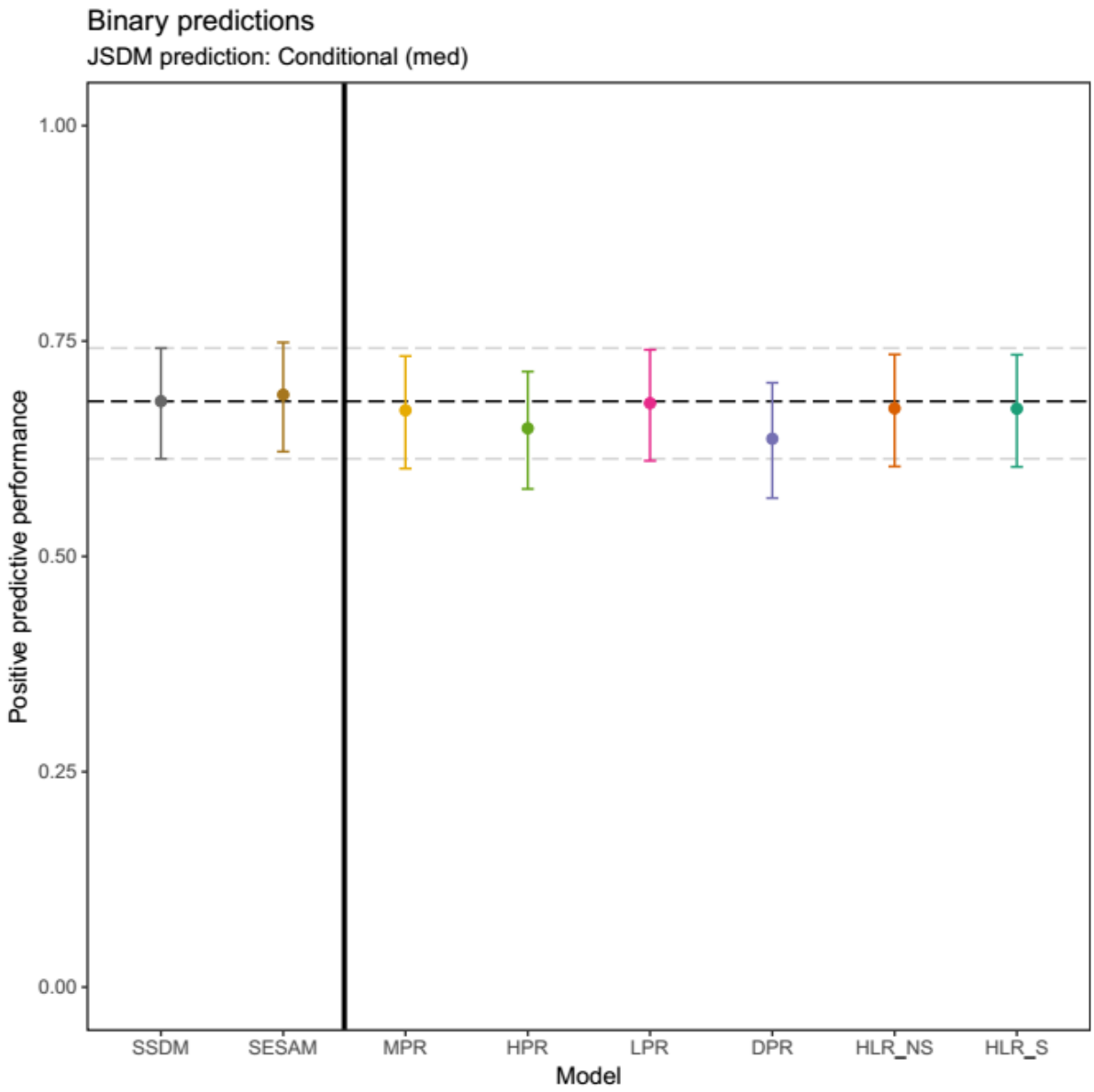


Figure S106: Model performance for the positive predictive performance evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

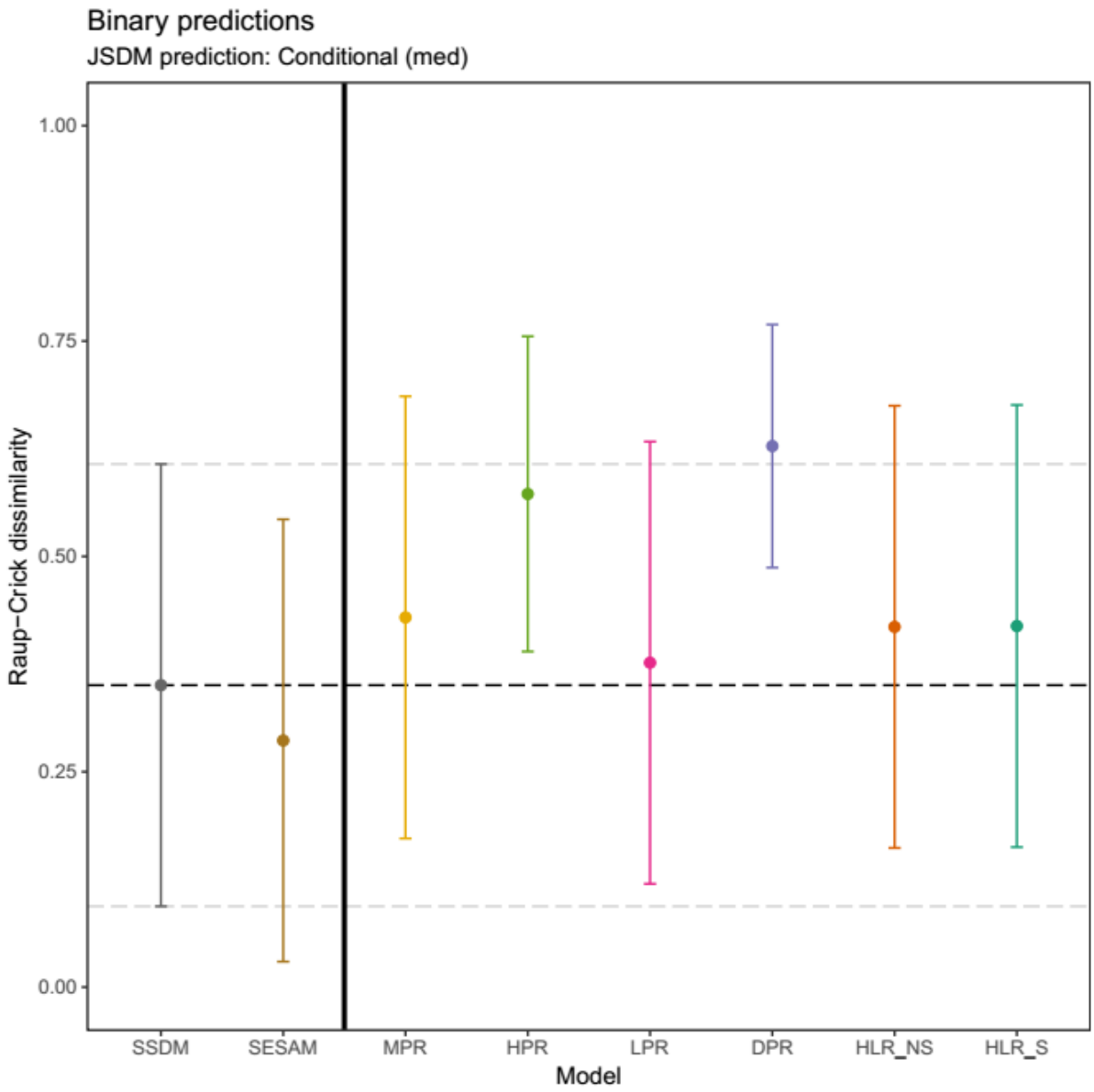


Figure S106: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

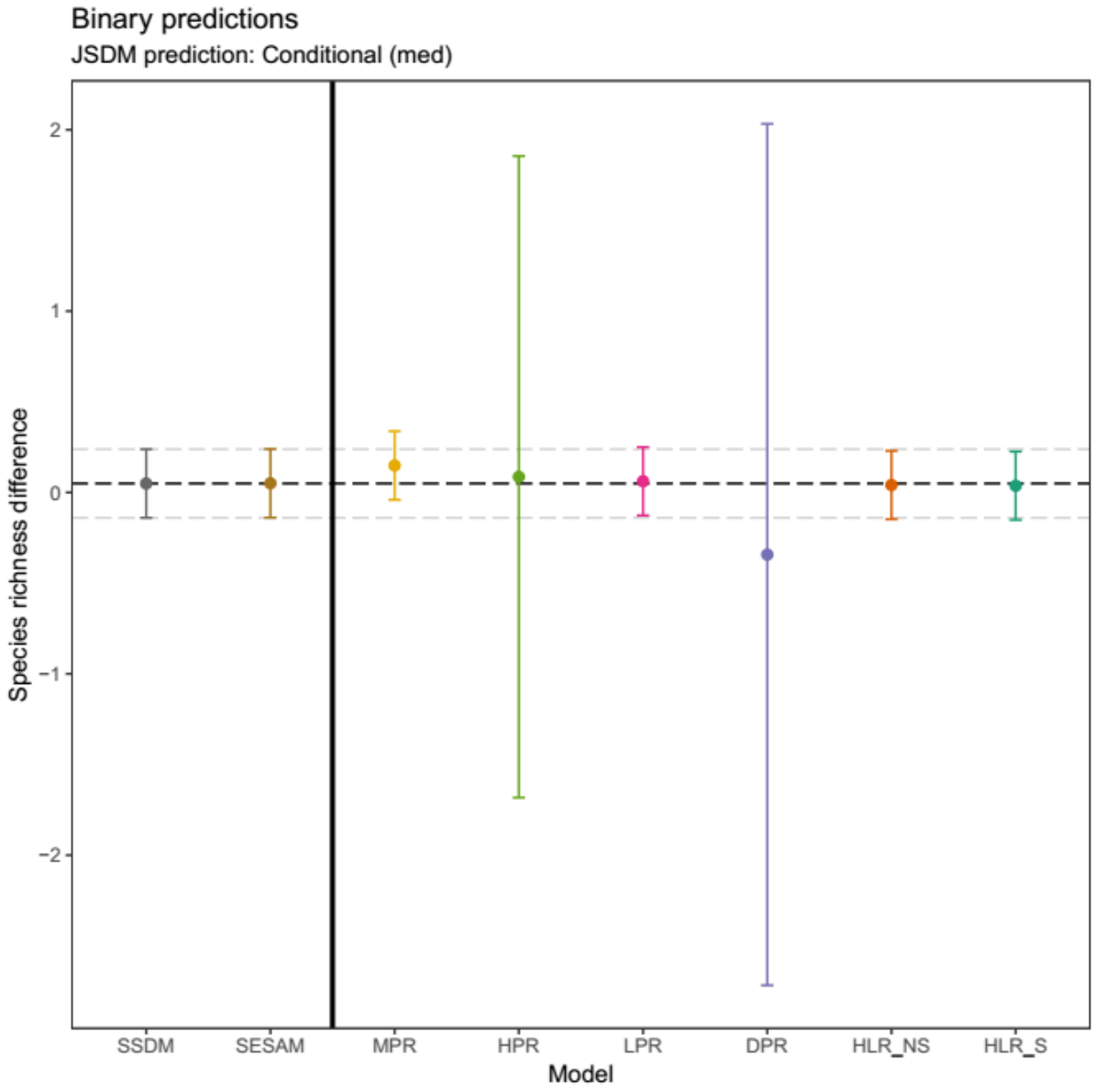


Figure S108: Model performance for the species richness difference evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

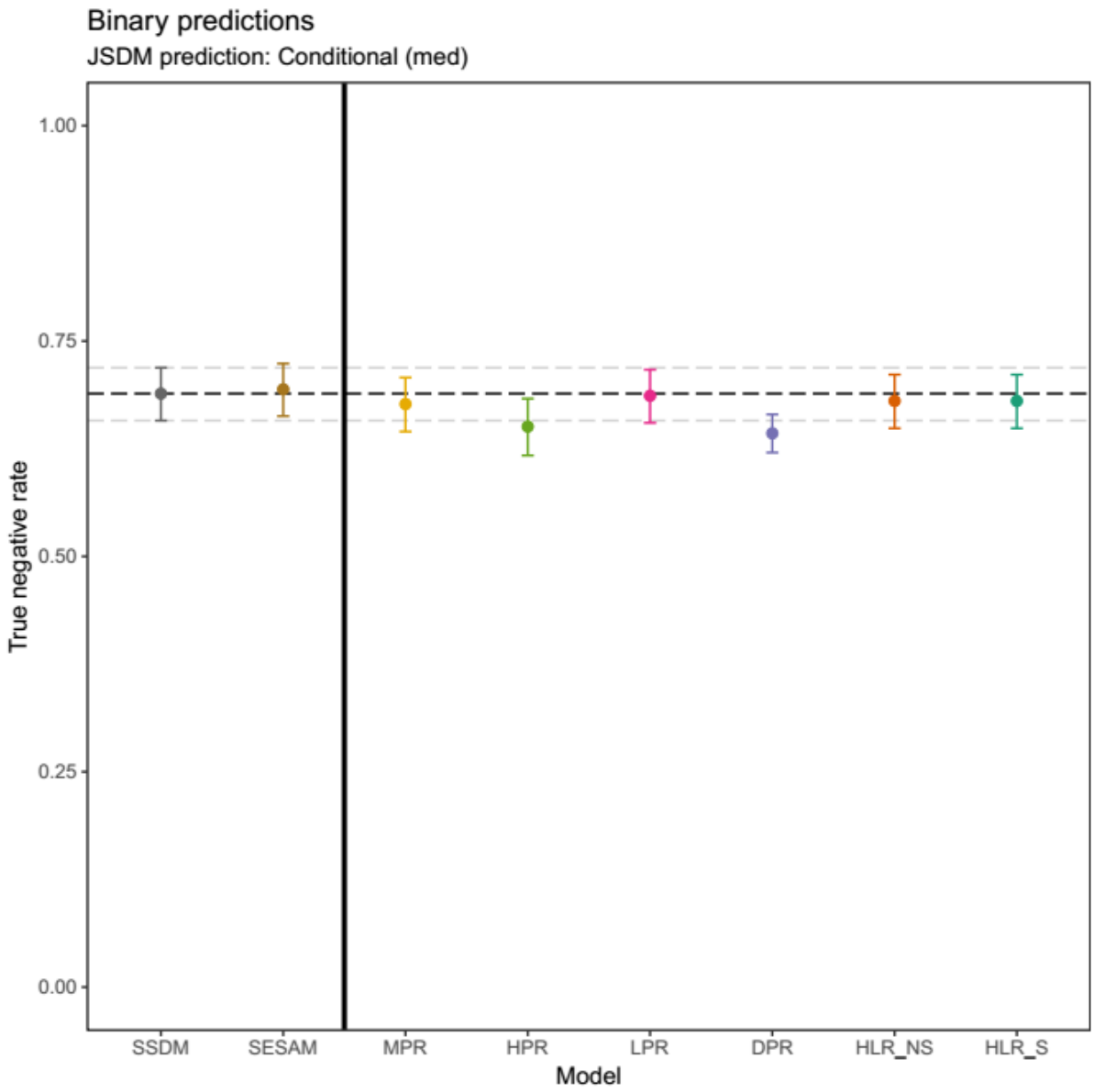


Figure S109: Model performance for the true negative rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

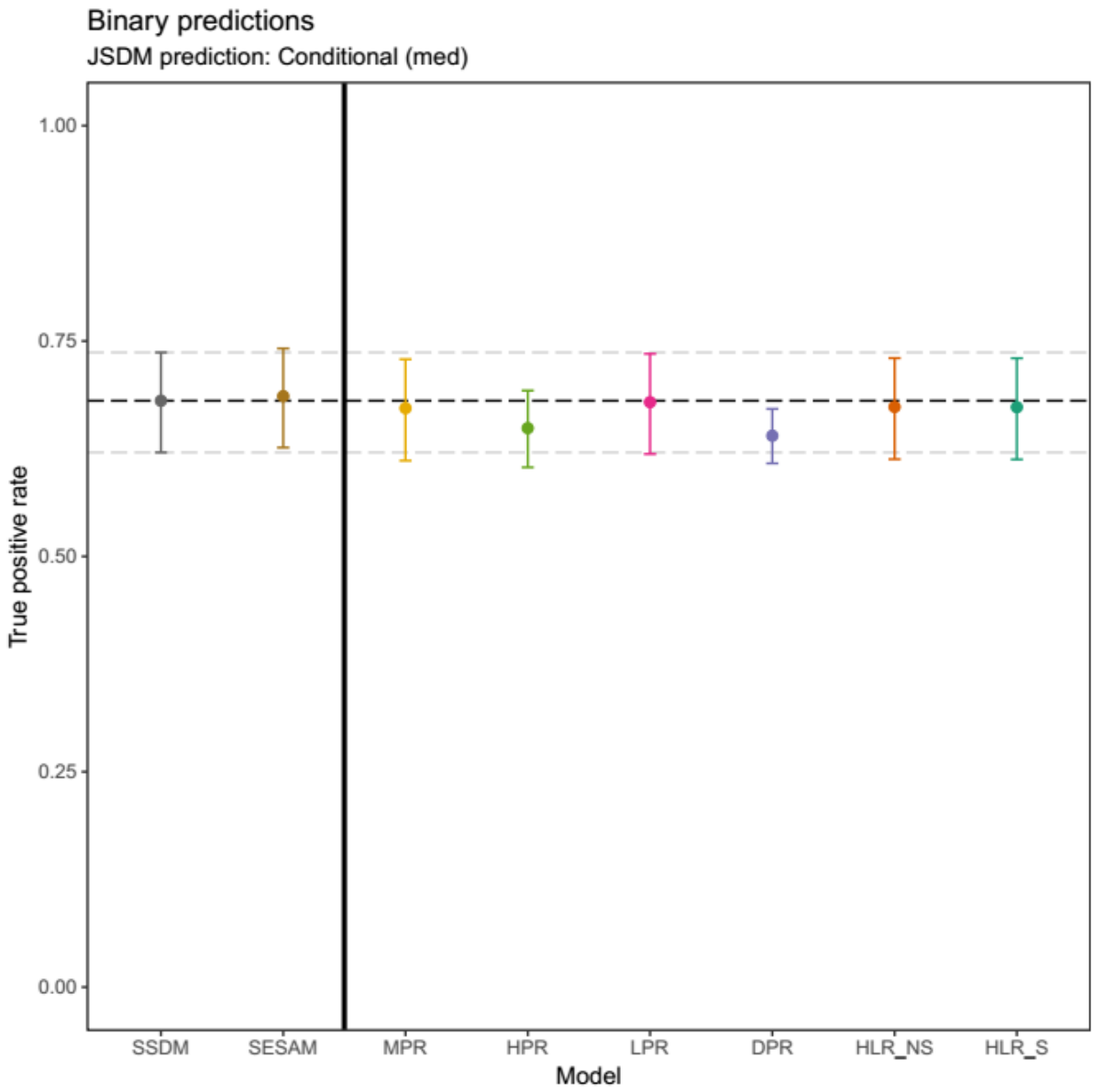


Figure S110: Model performance for the true positive rate evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

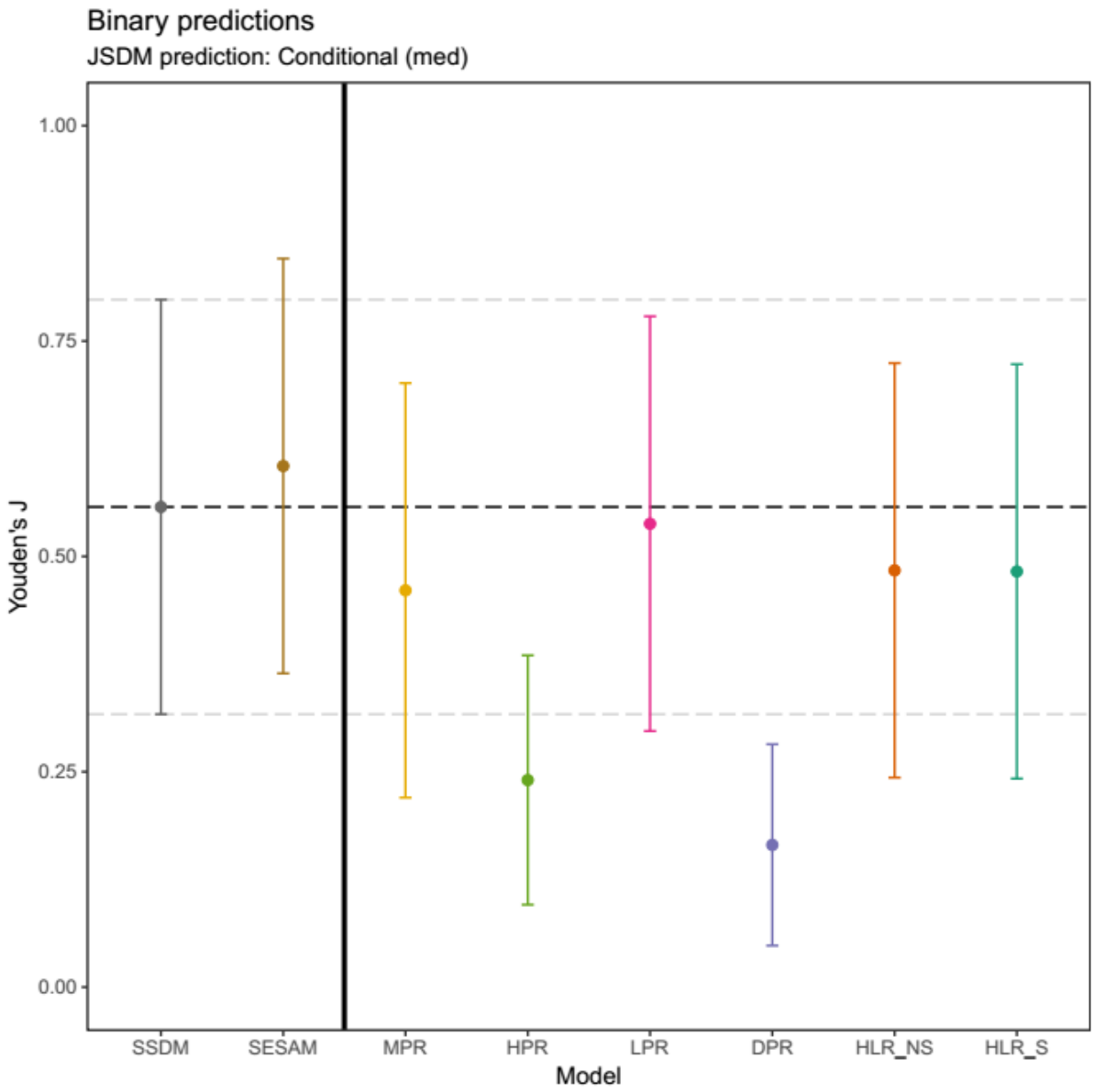


Figure S111: Model performance for the Youden's J evaluation metric for conditional (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

6. Conditional (high-prevalence) predictions

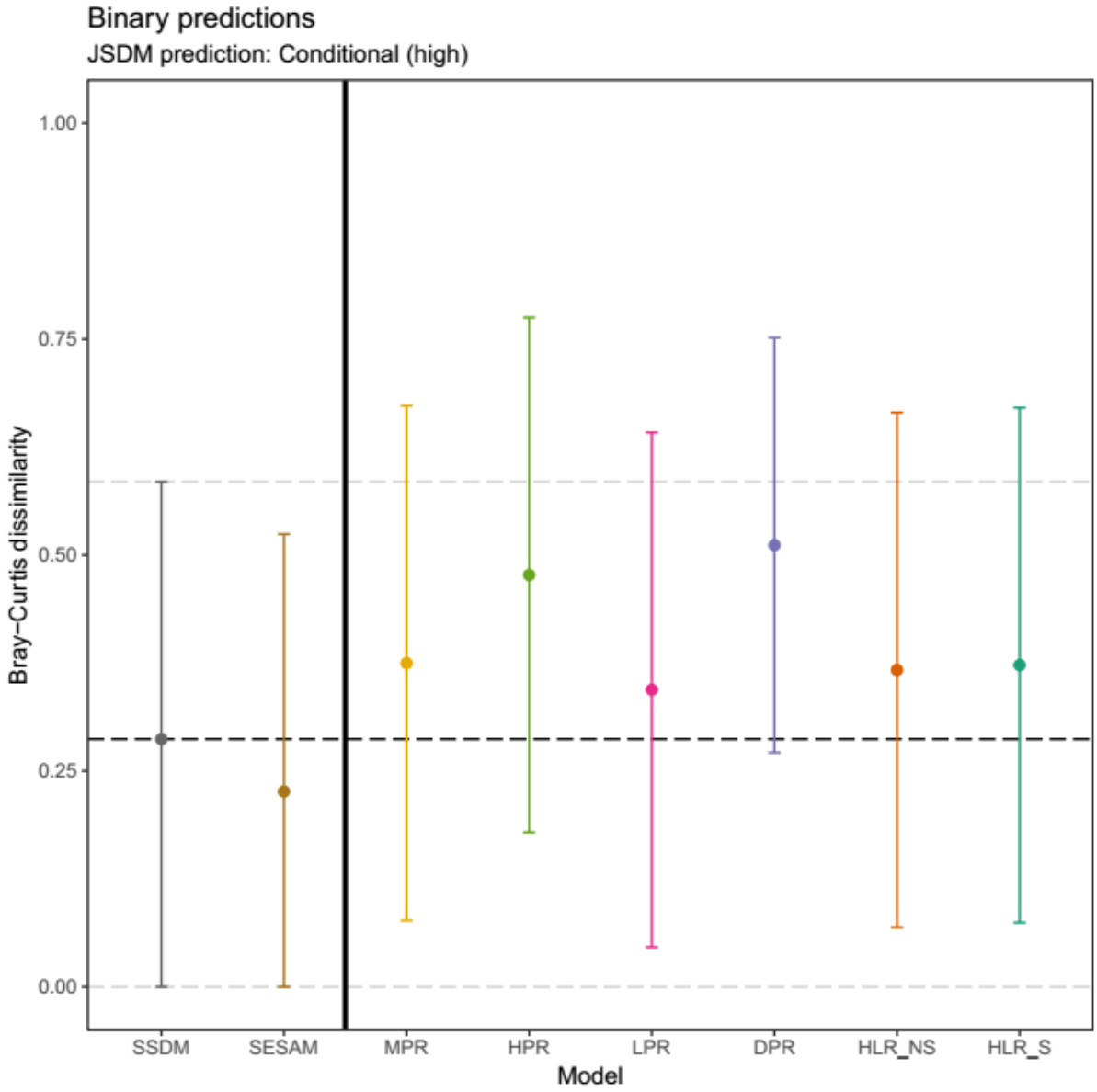


Figure 112: Model performance for the Bray-Curtis dissimilarity evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

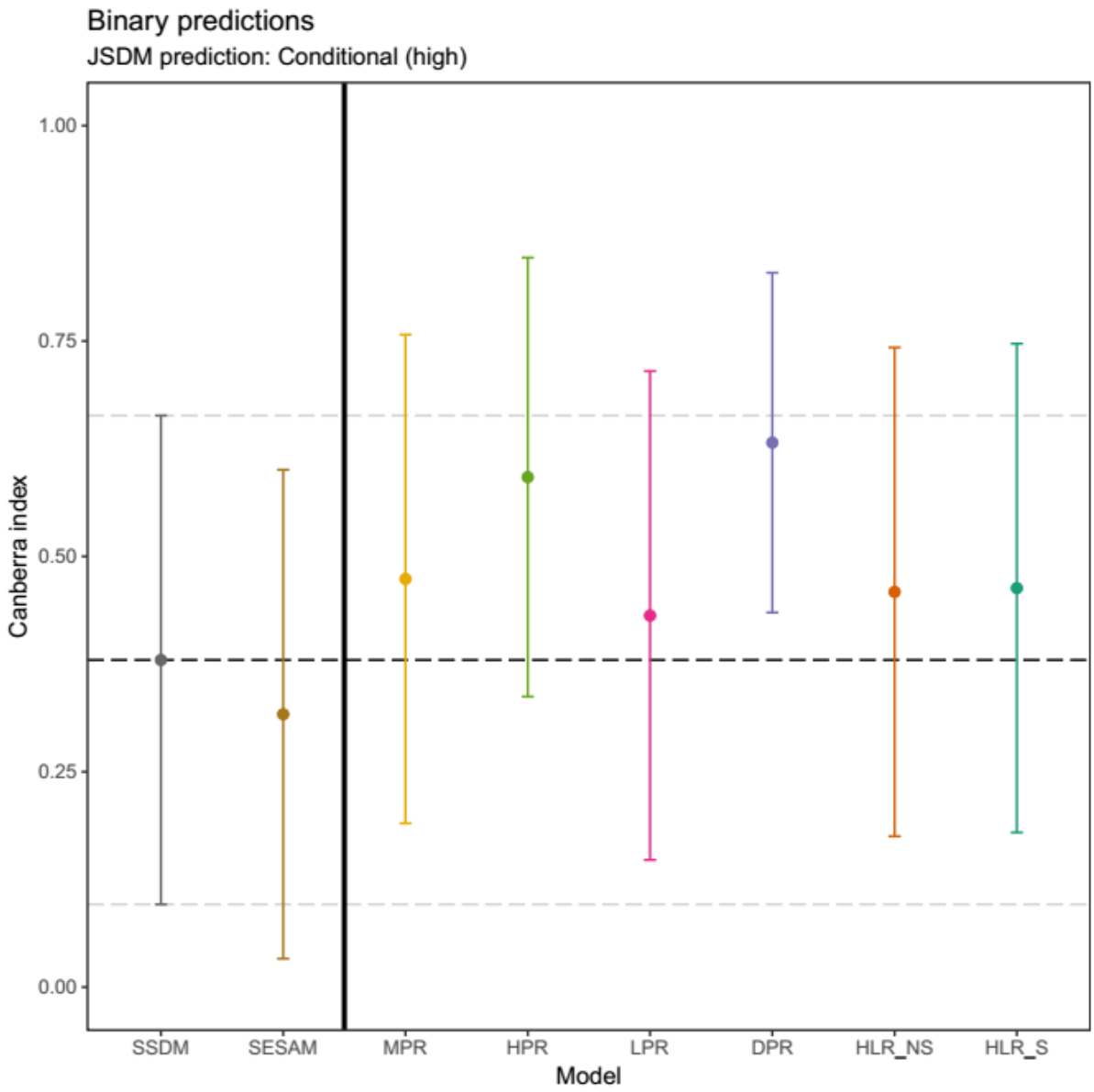


Figure 113: Model performance for the Canberra index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

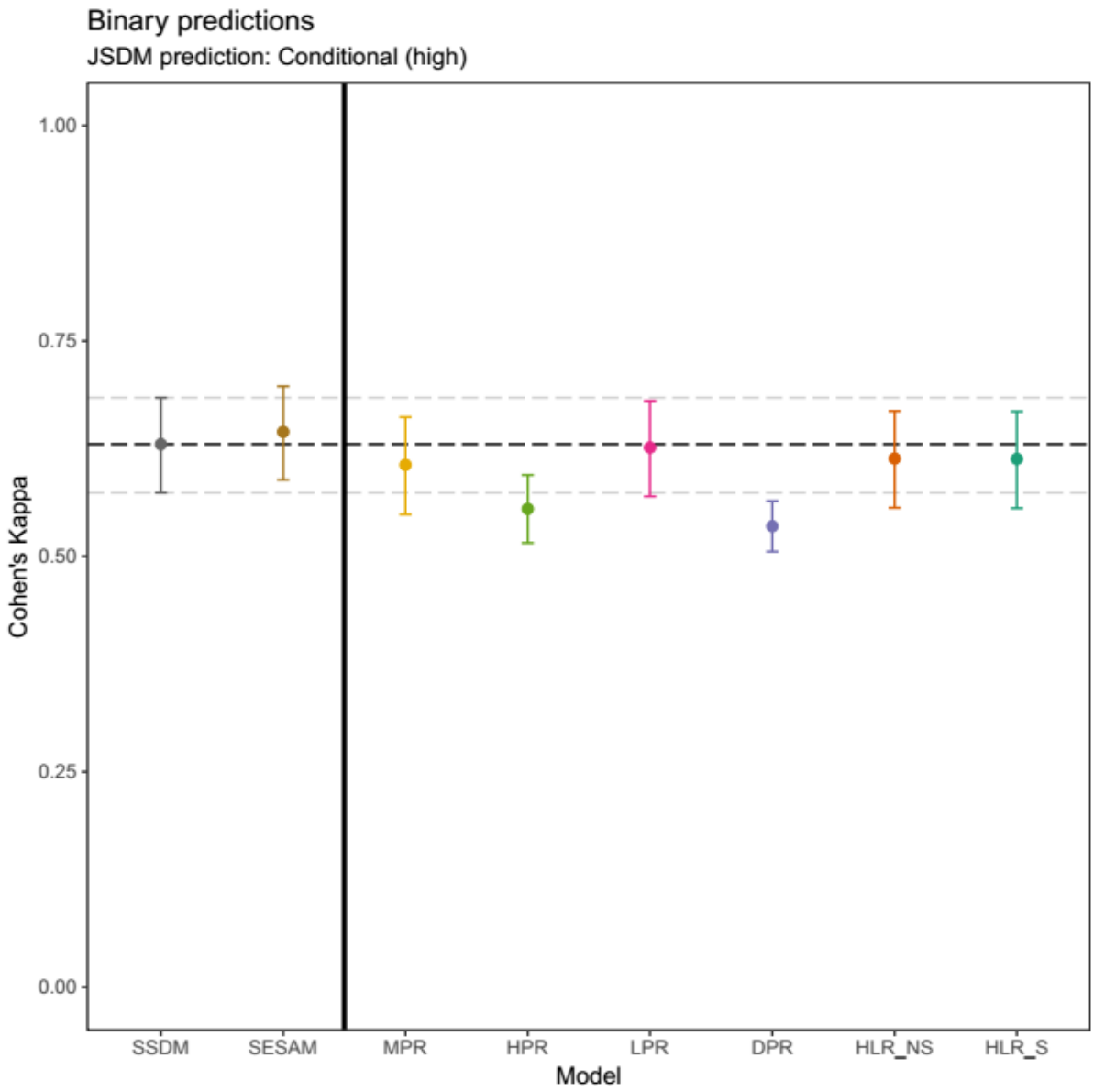


Figure 114: Model performance for the Cohen's Kappa evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

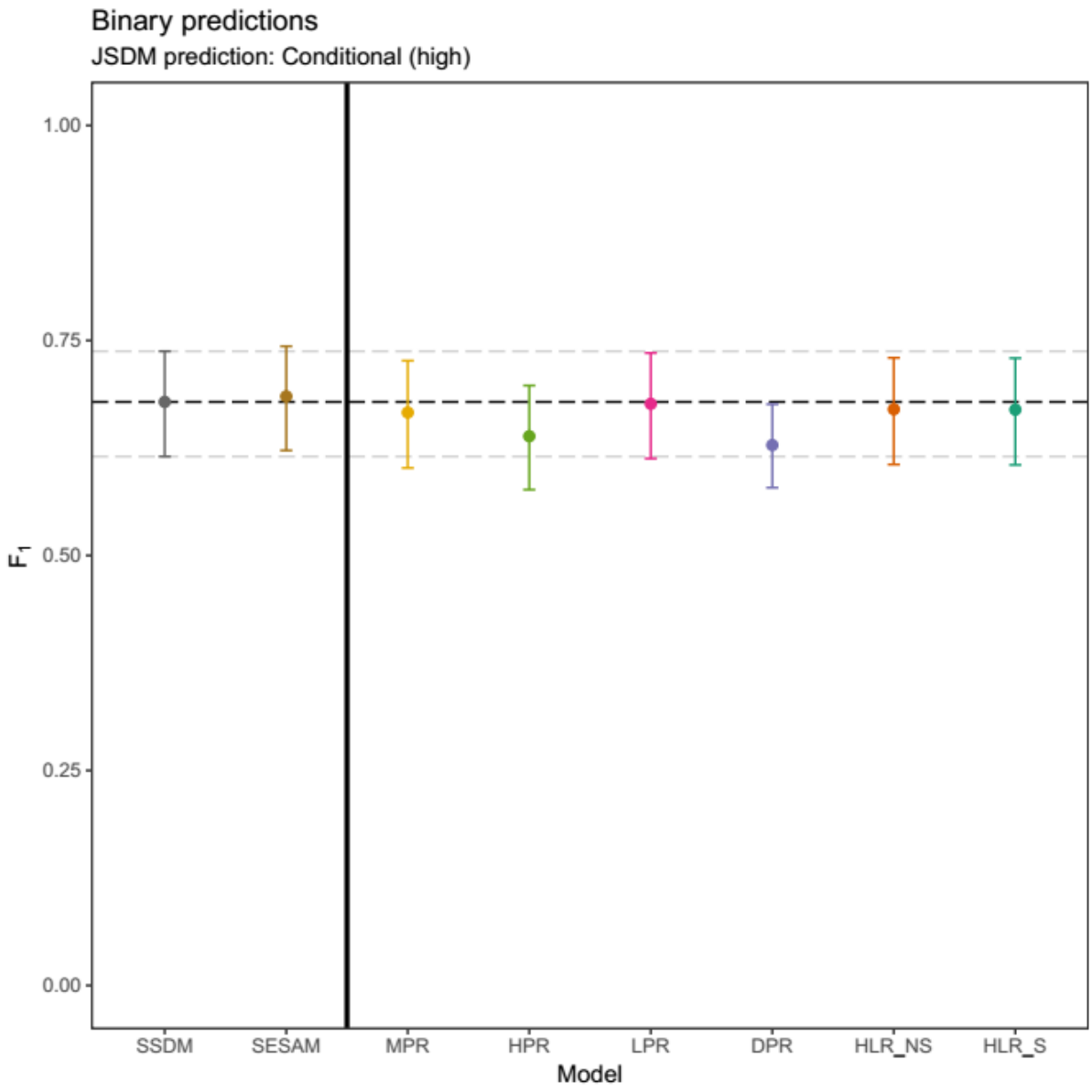


Figure 115: Model performance for the F_1 evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

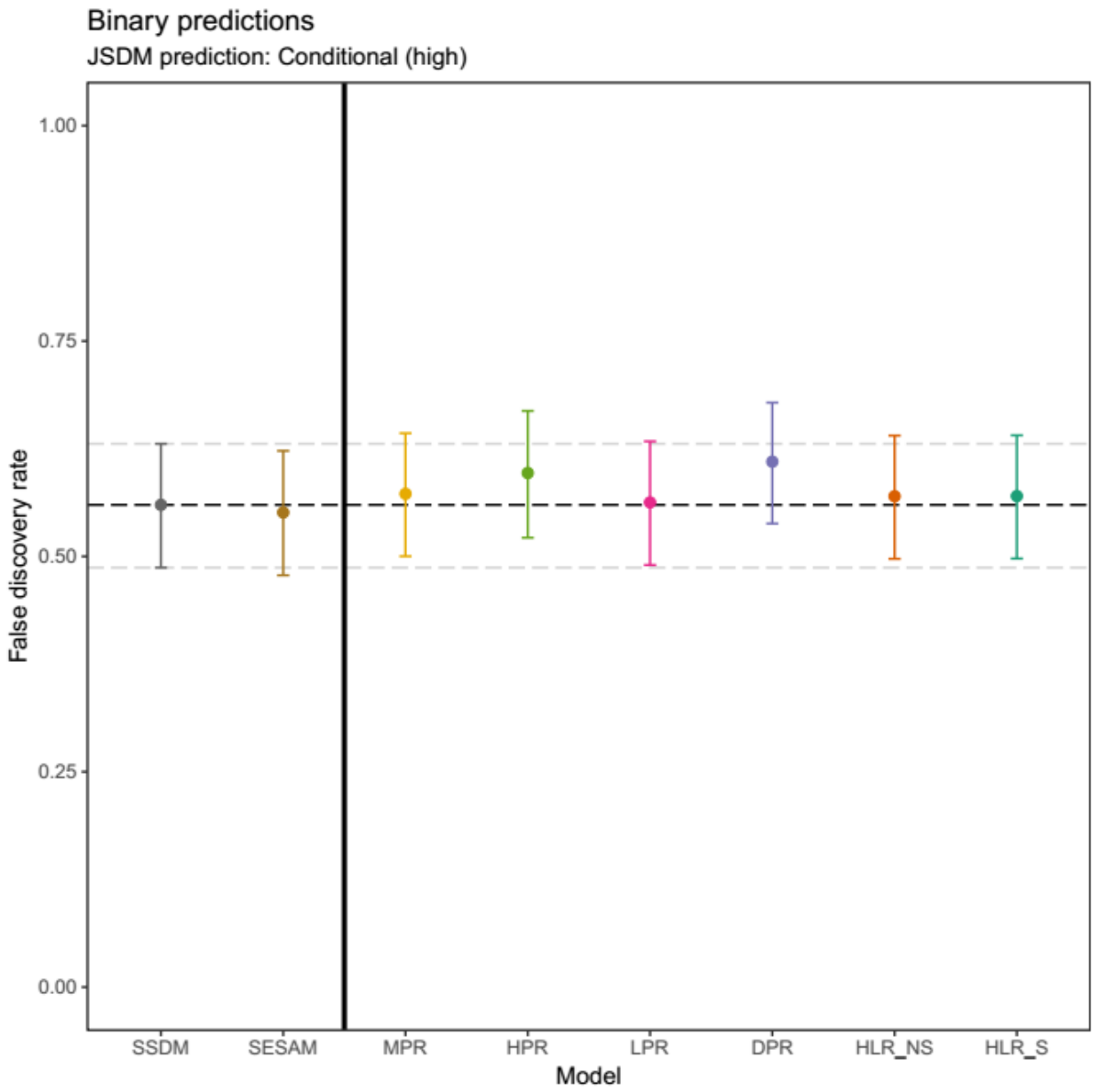


Figure 116: Model performance for the false discovery rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

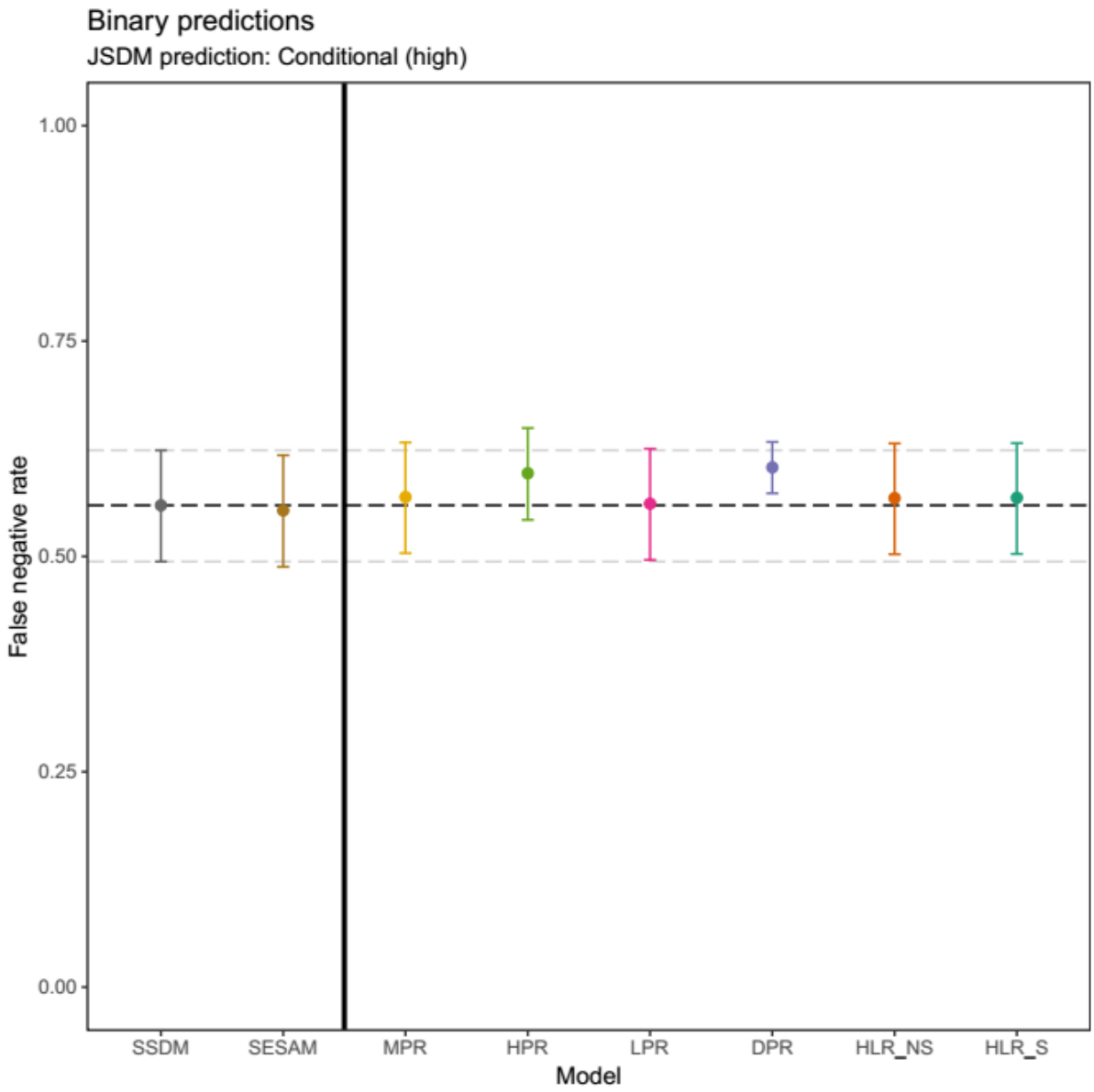


Figure 117: Model performance for the false negative rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

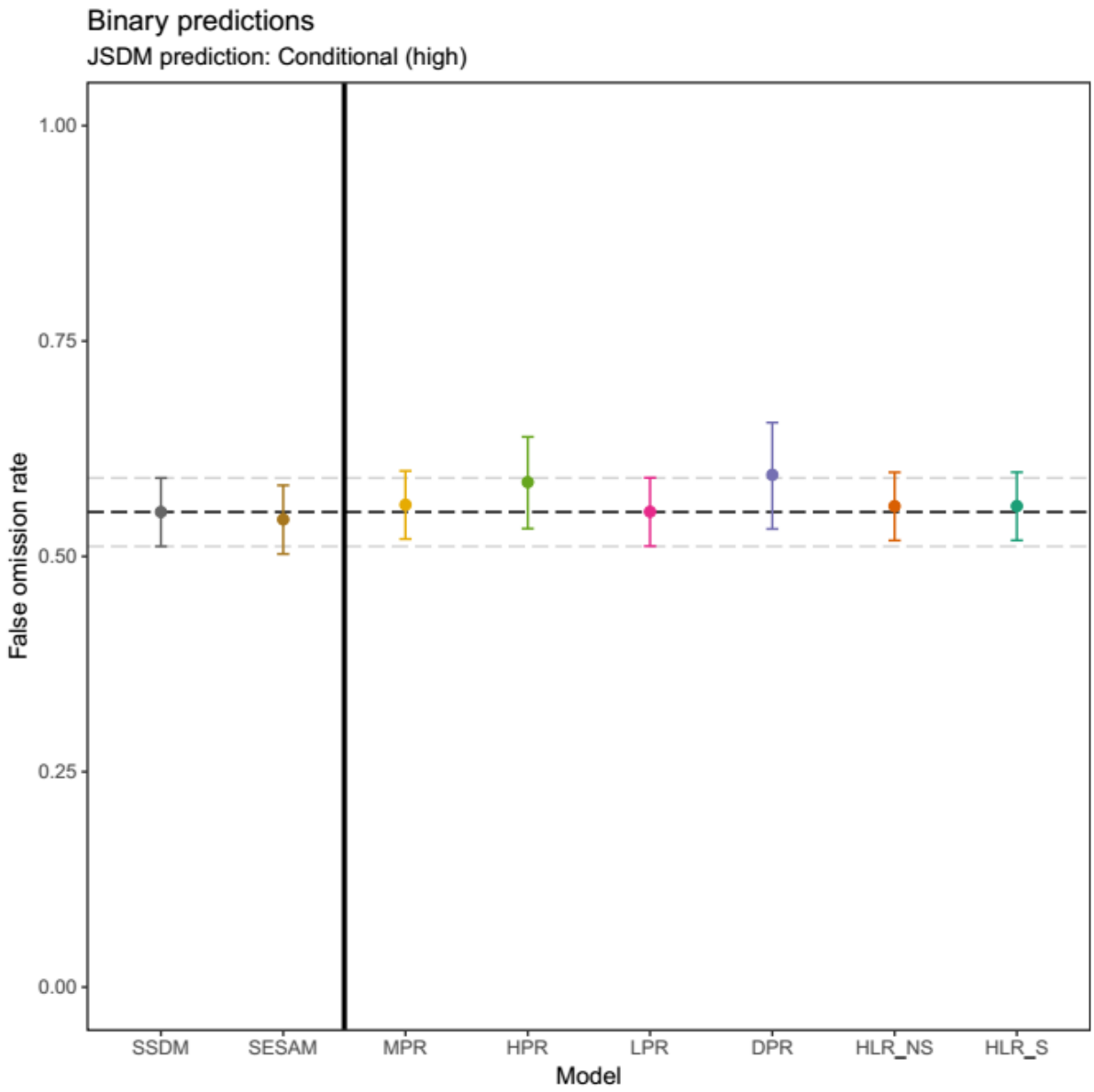


Figure 118: Model performance for the false omission rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

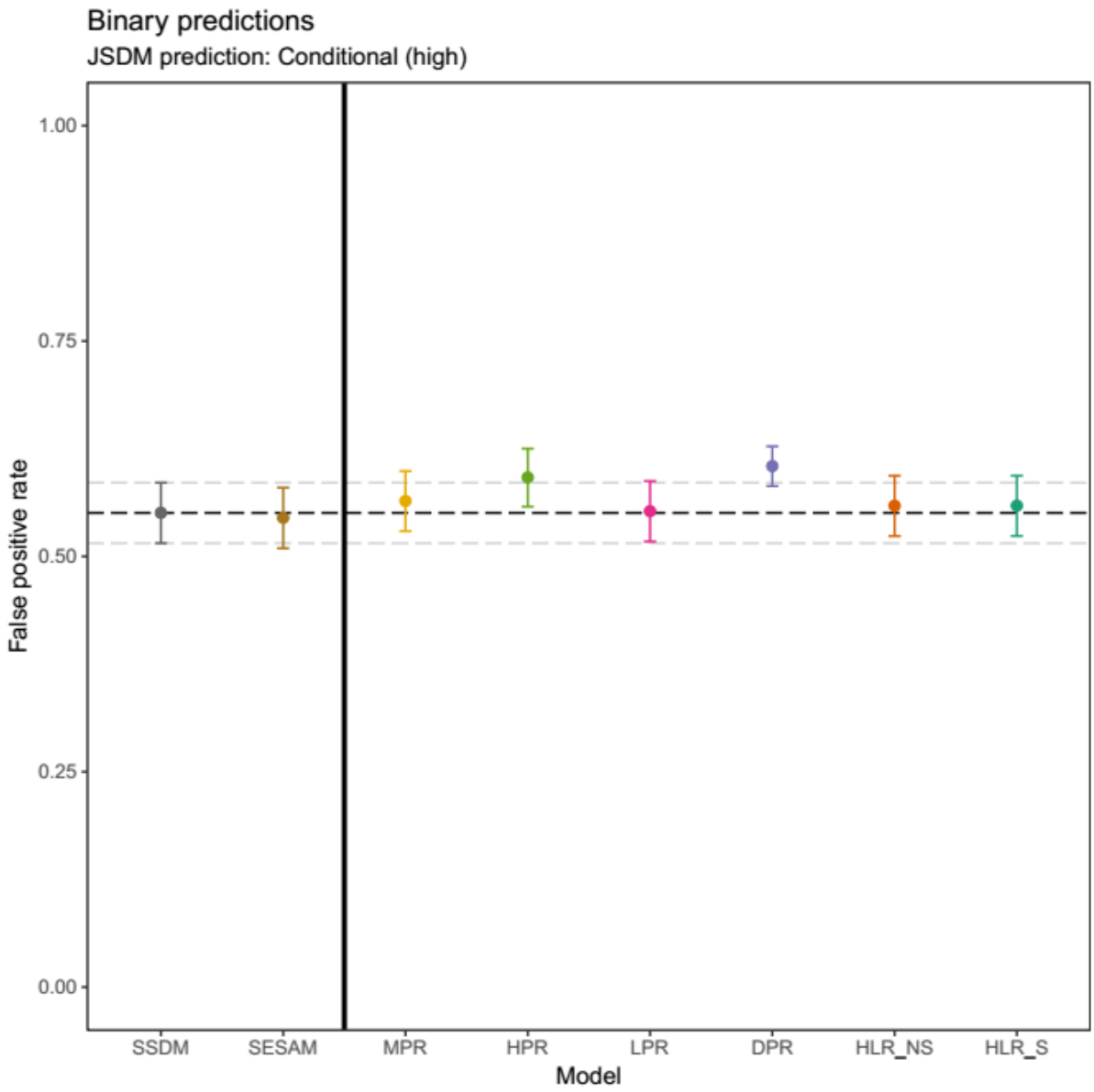


Figure 119: Model performance for the false positive rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

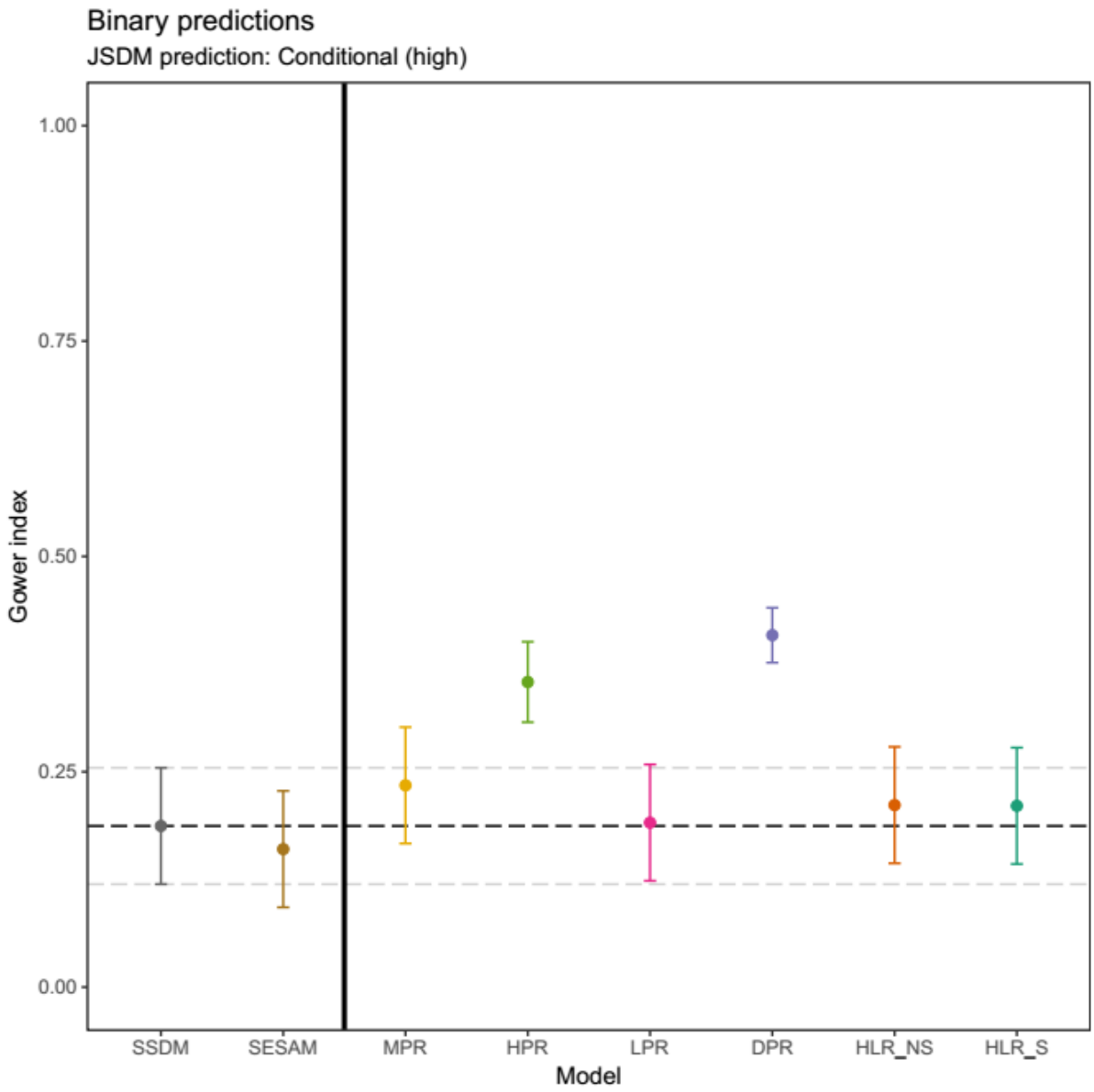


Figure 120: Model performance for the Gower index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

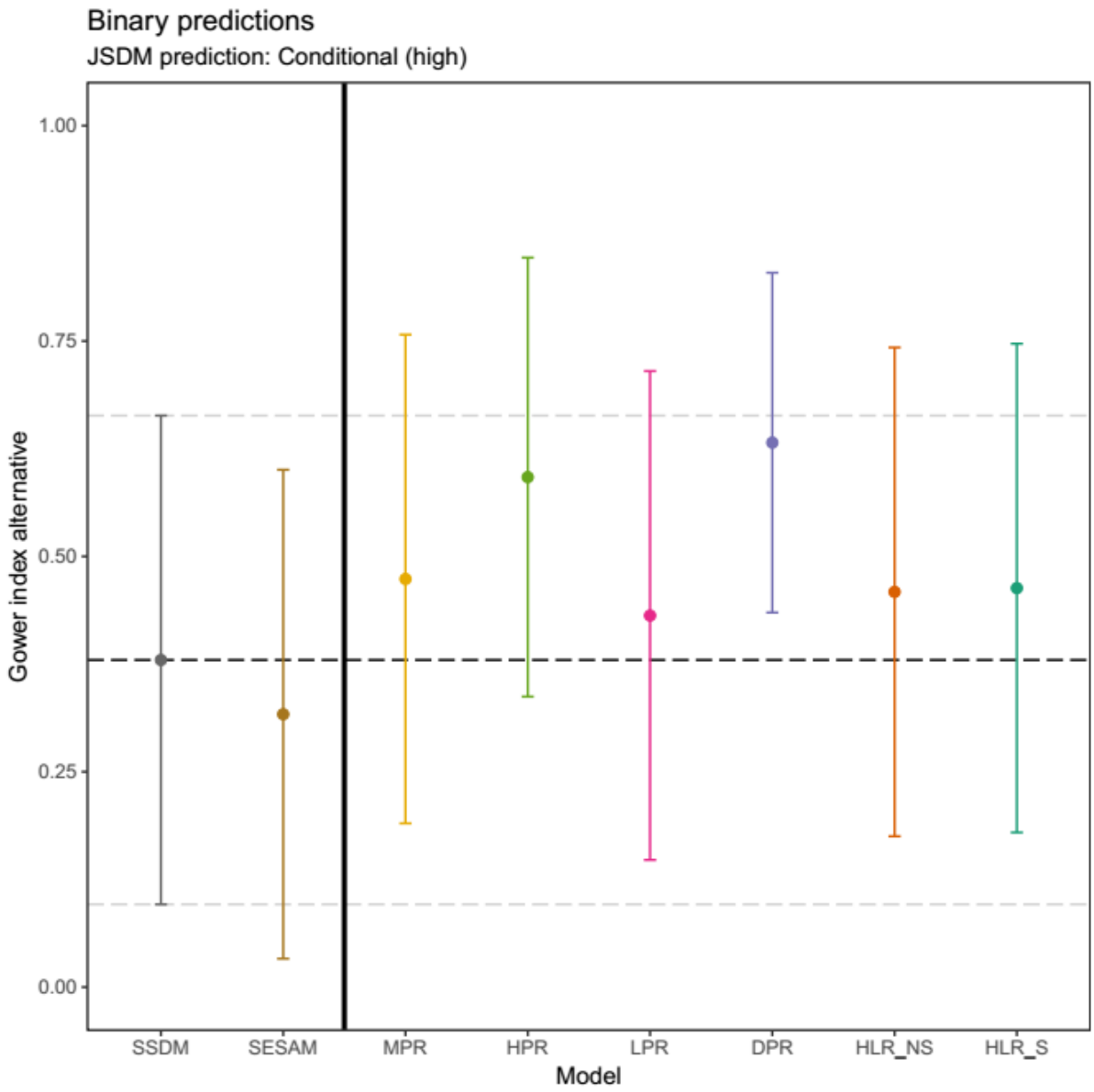


Figure 121: Model performance for the alternative Gower index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

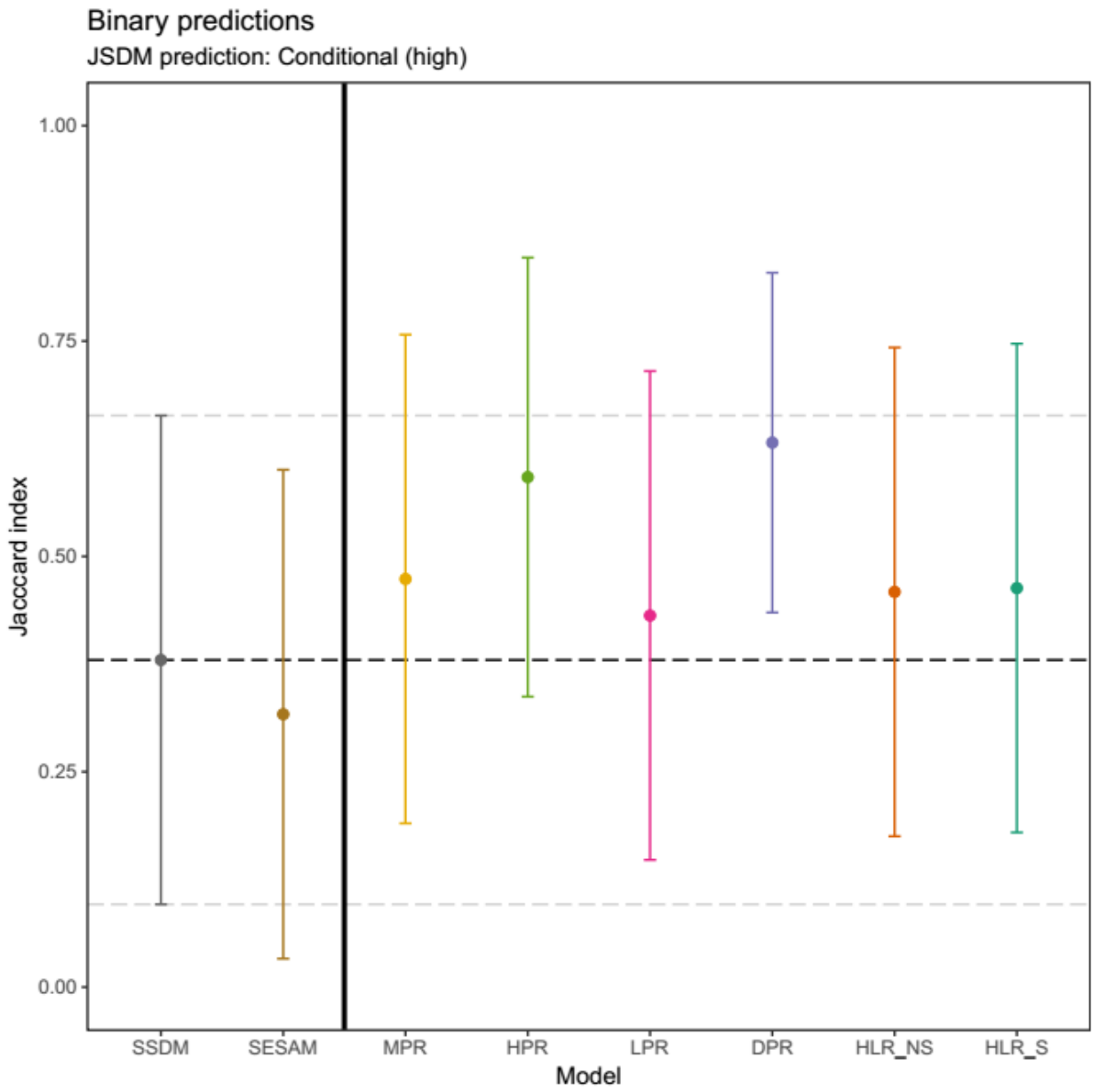


Figure 122: Model performance for the Jaccard index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

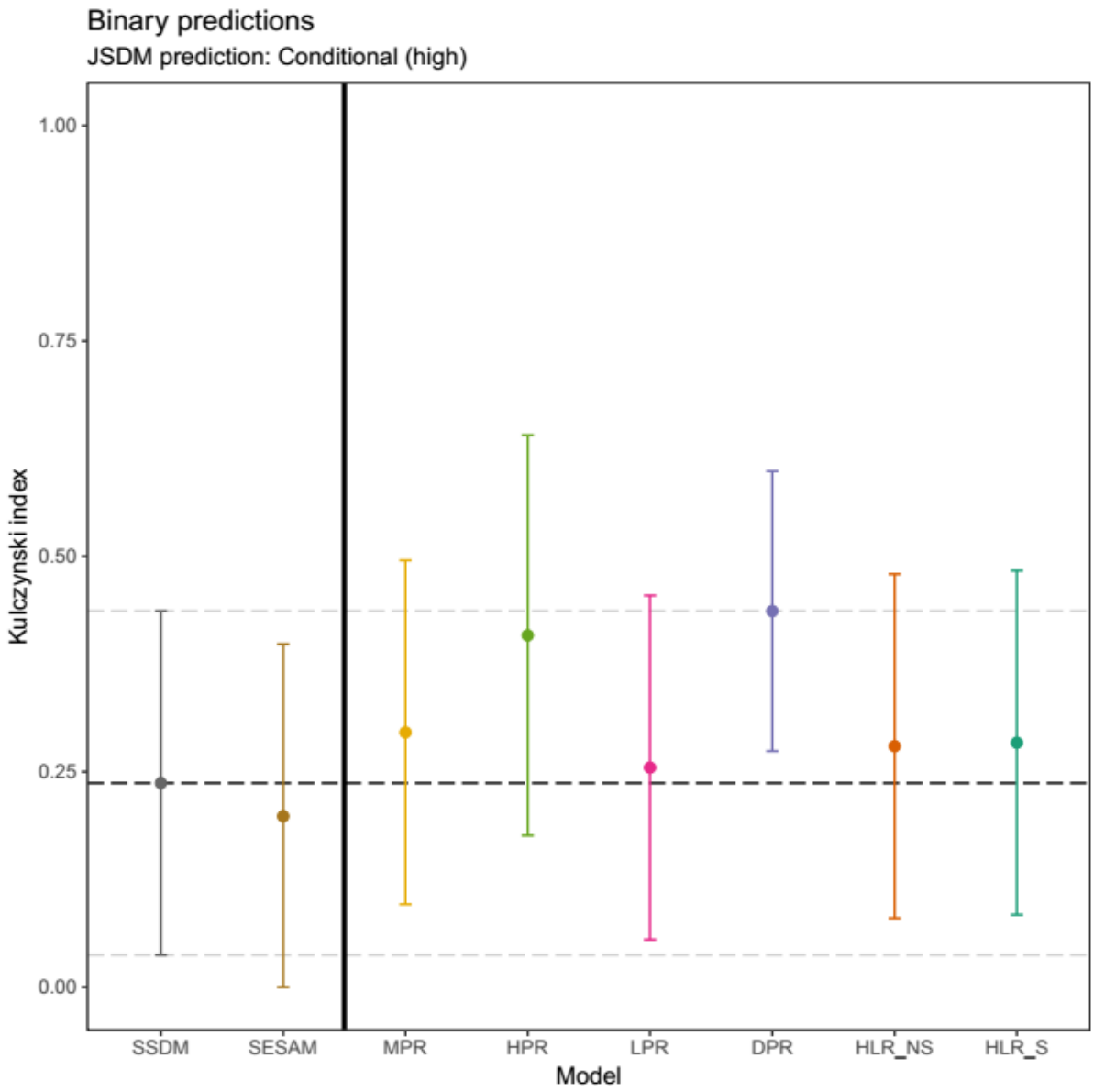


Figure 123: Model performance for the Kulczynski index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

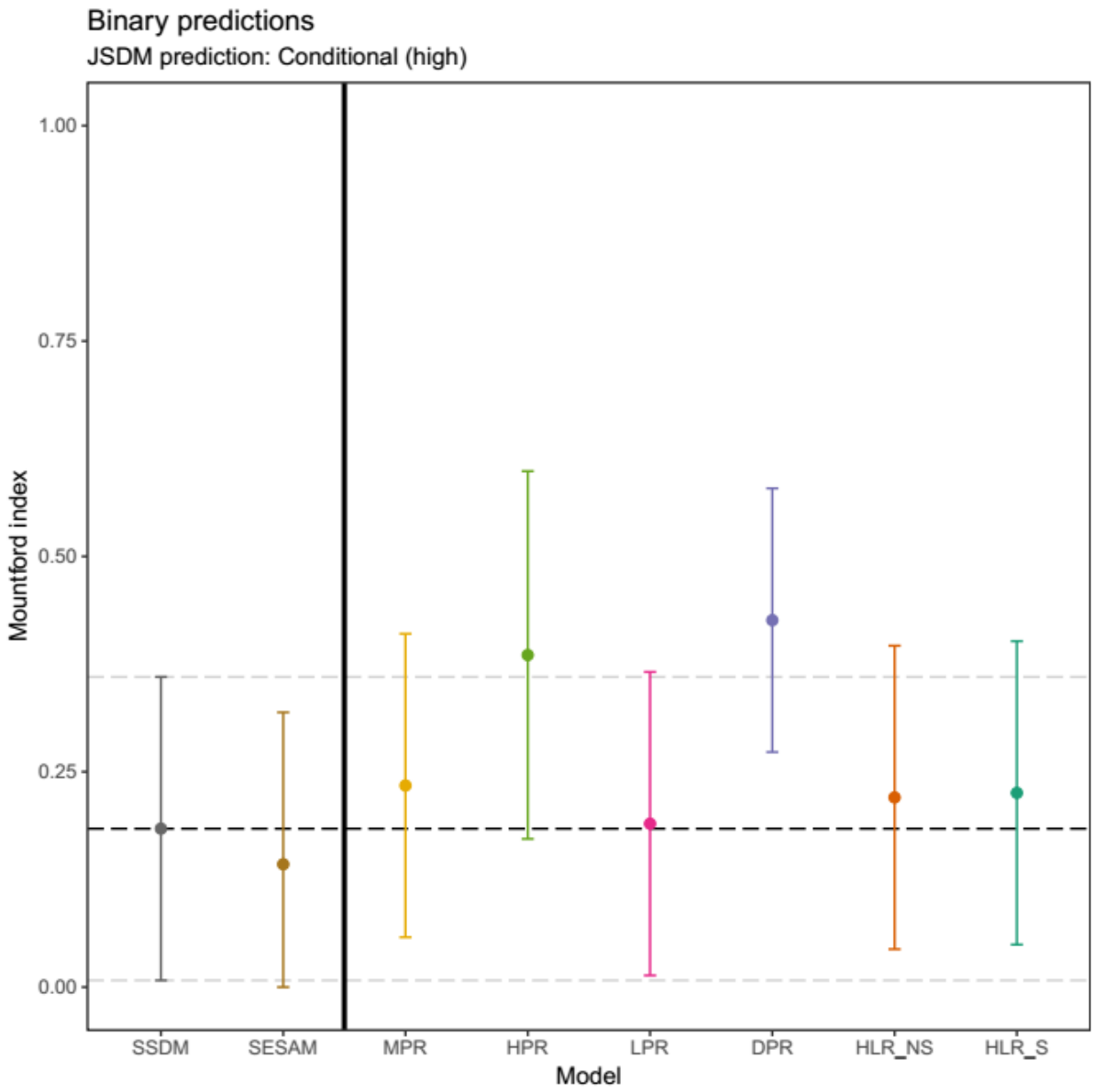


Figure 124: Model performance for the Mountford index evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

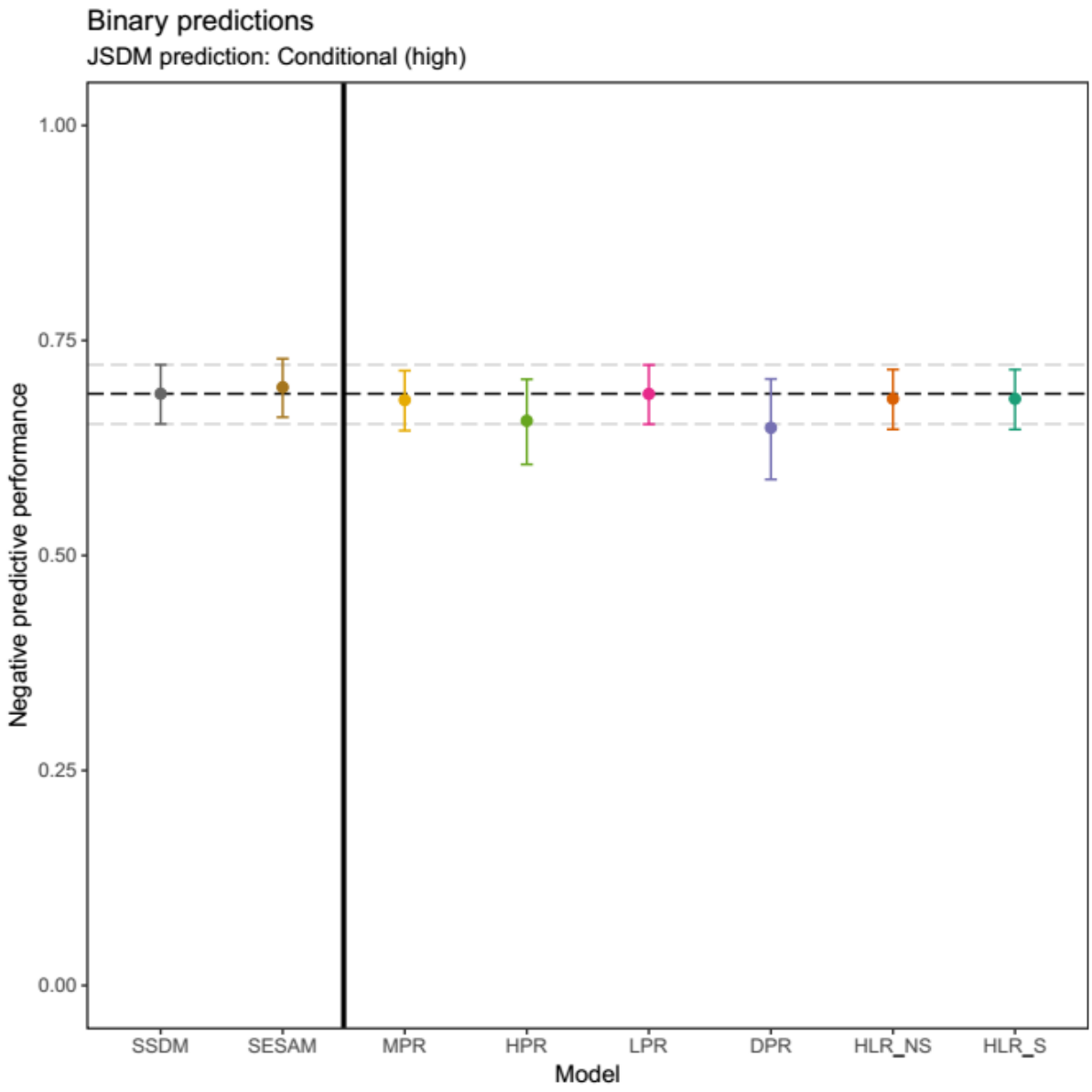


Figure 125: Model performance for the negative predictive performance evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

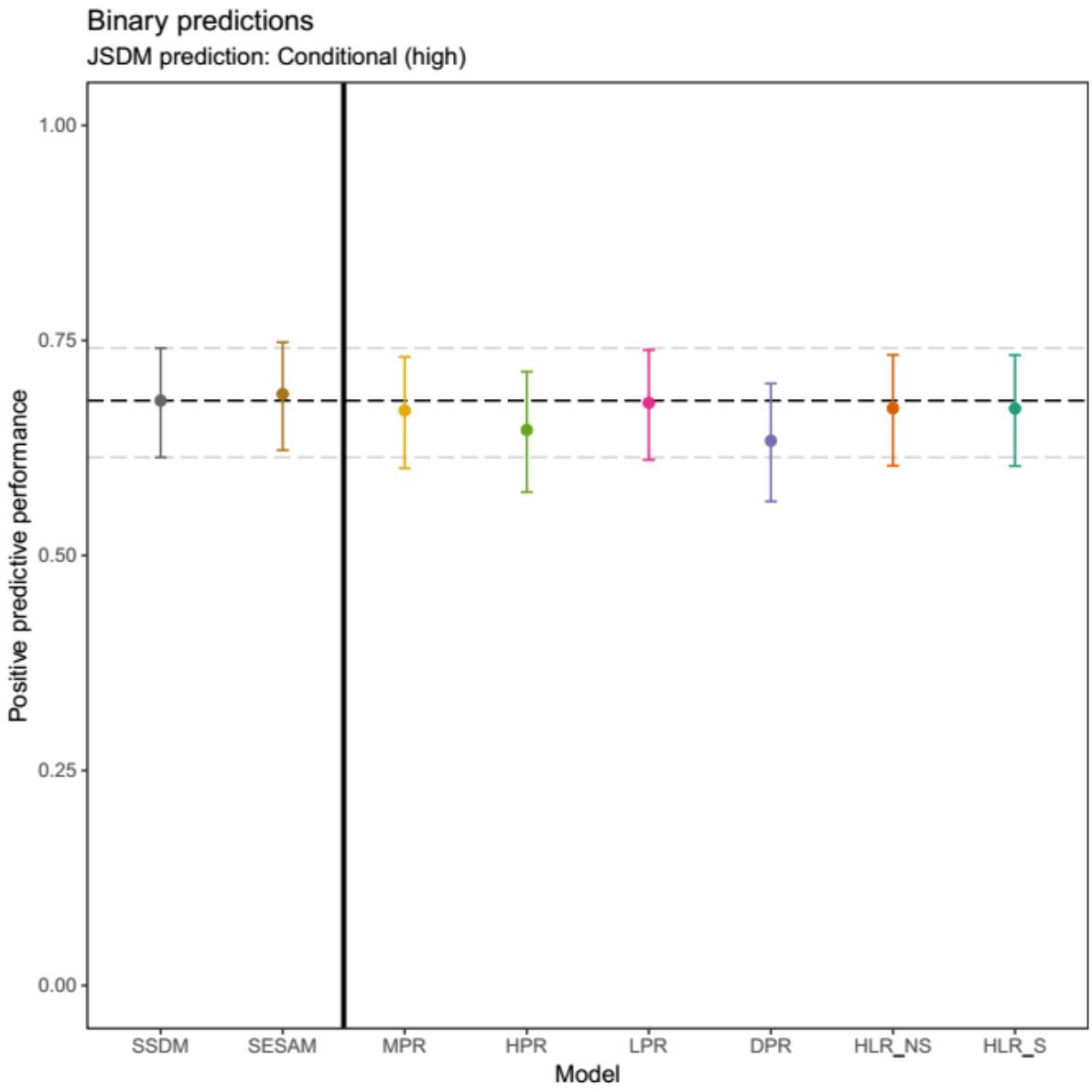


Figure 126: Model performance for the positive predictive performance evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

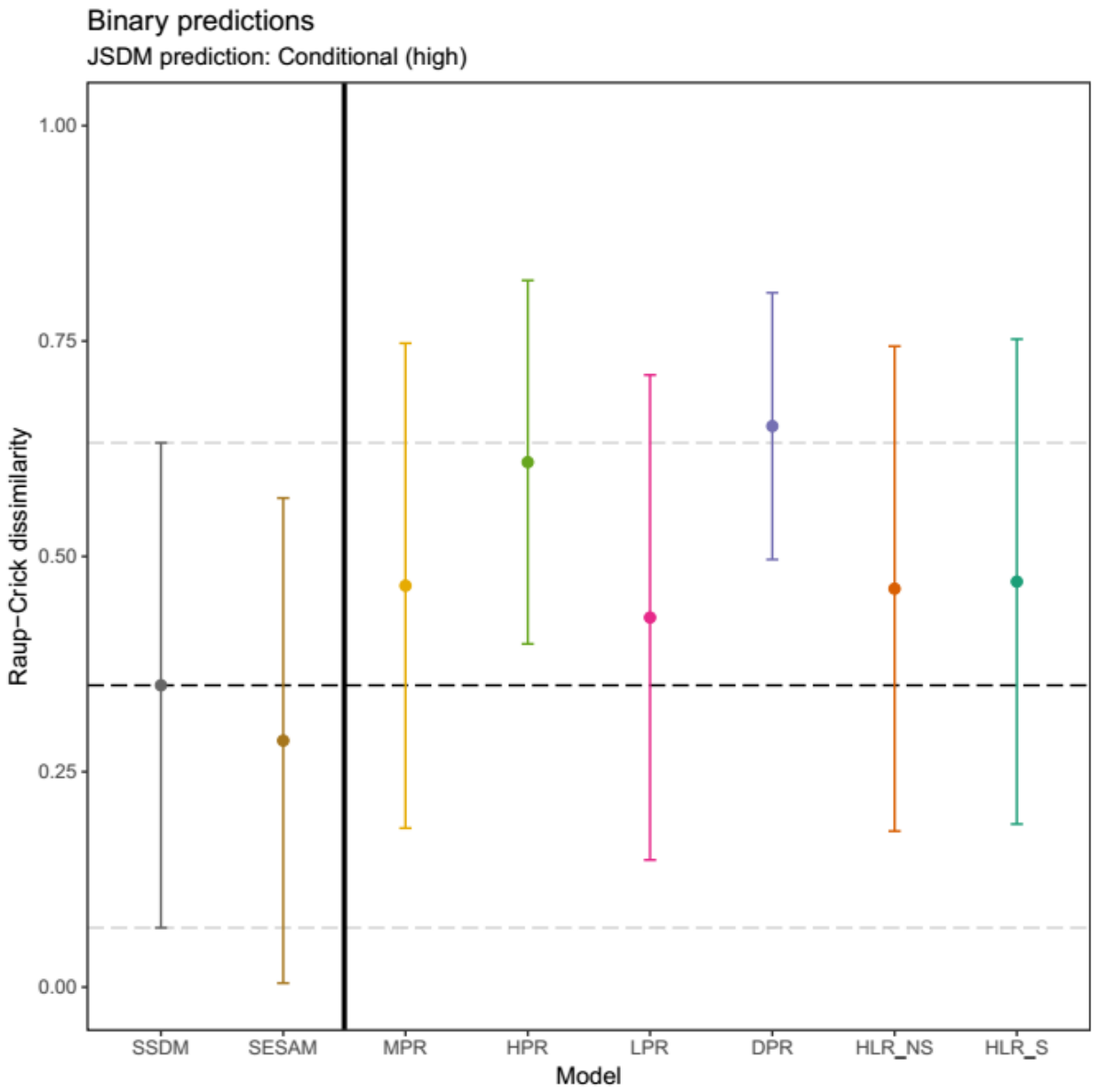


Figure 127: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

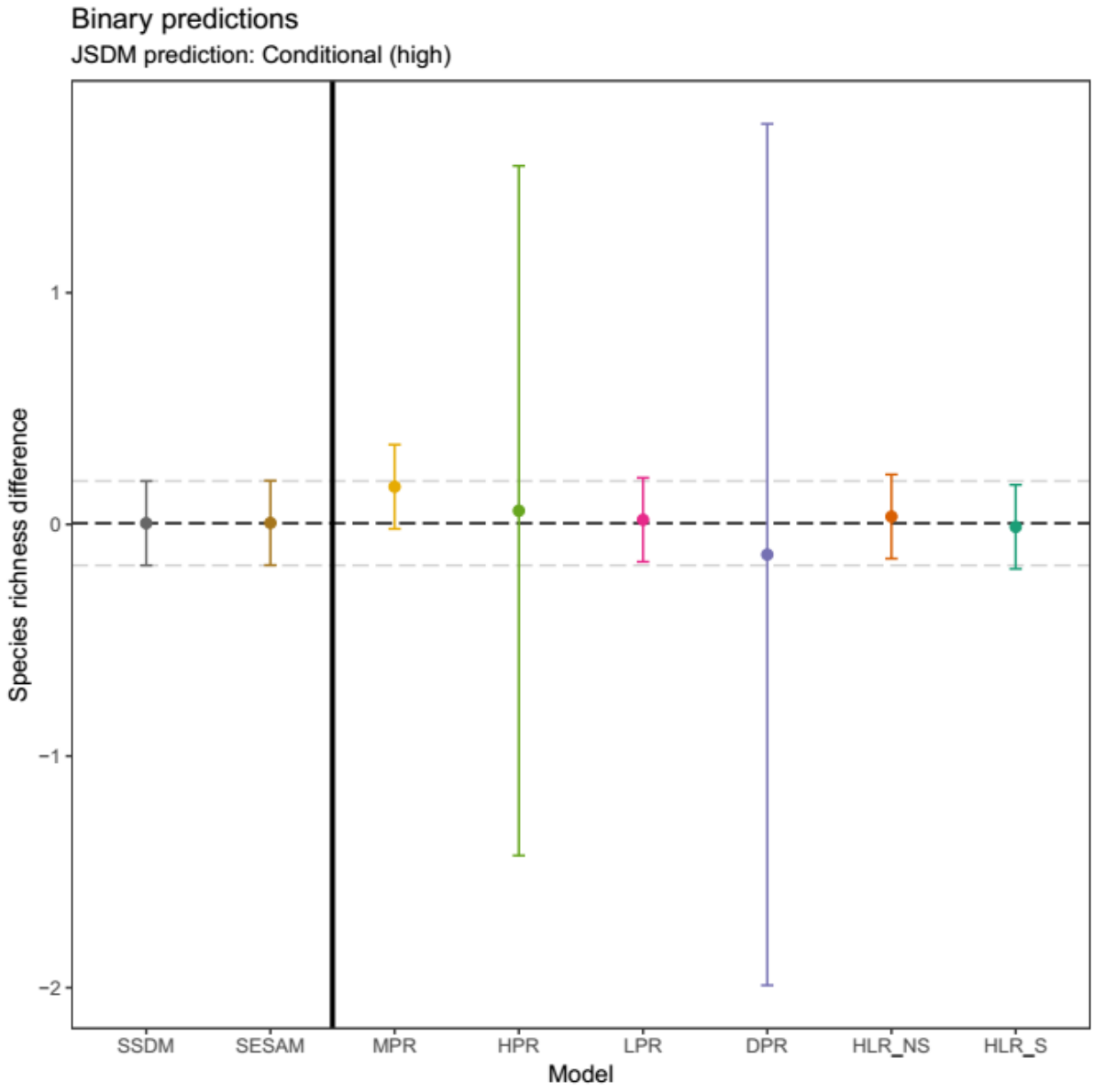


Figure 128: Model performance for the species richness difference evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

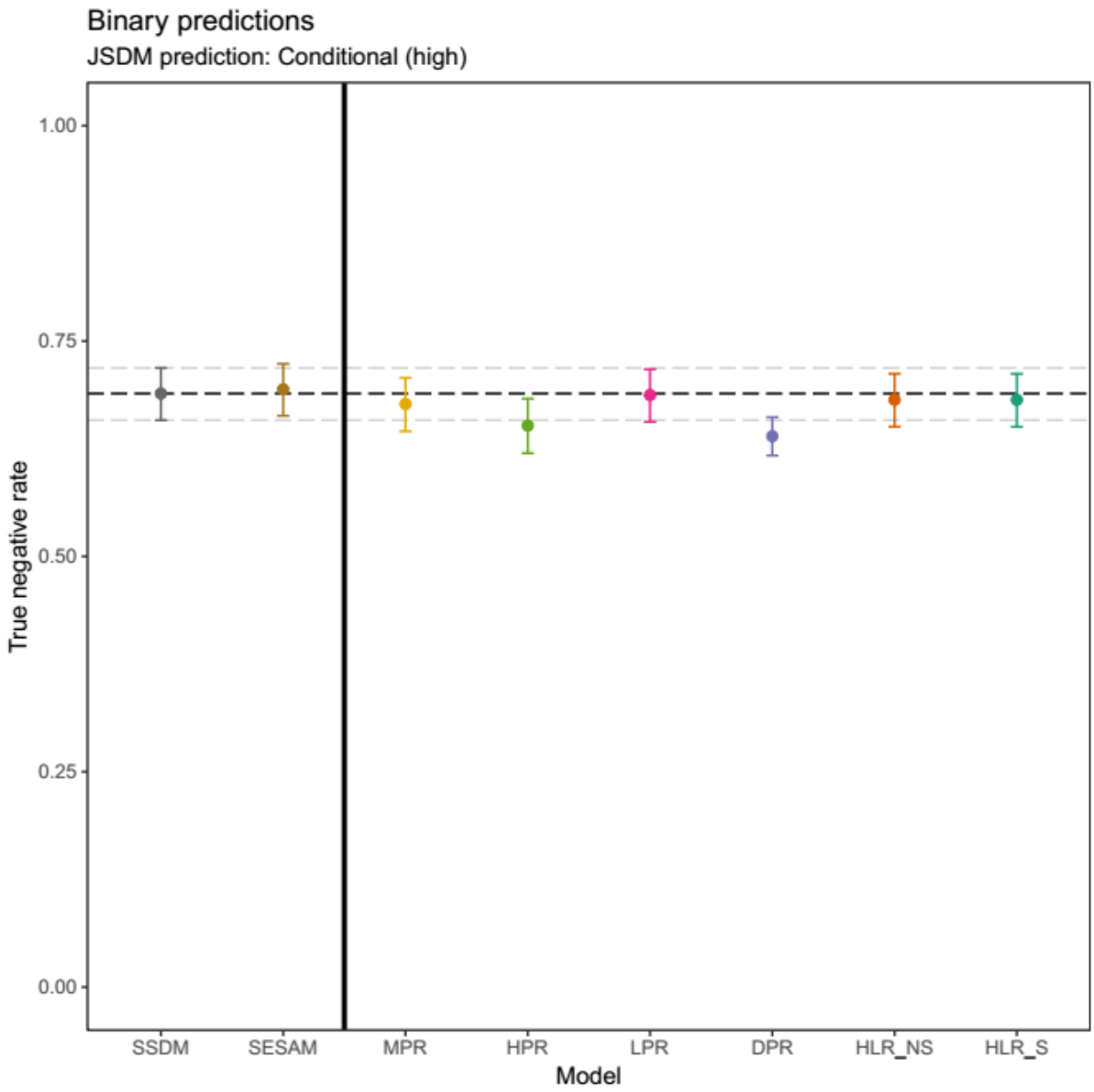


Figure 129: Model performance for the true negative rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

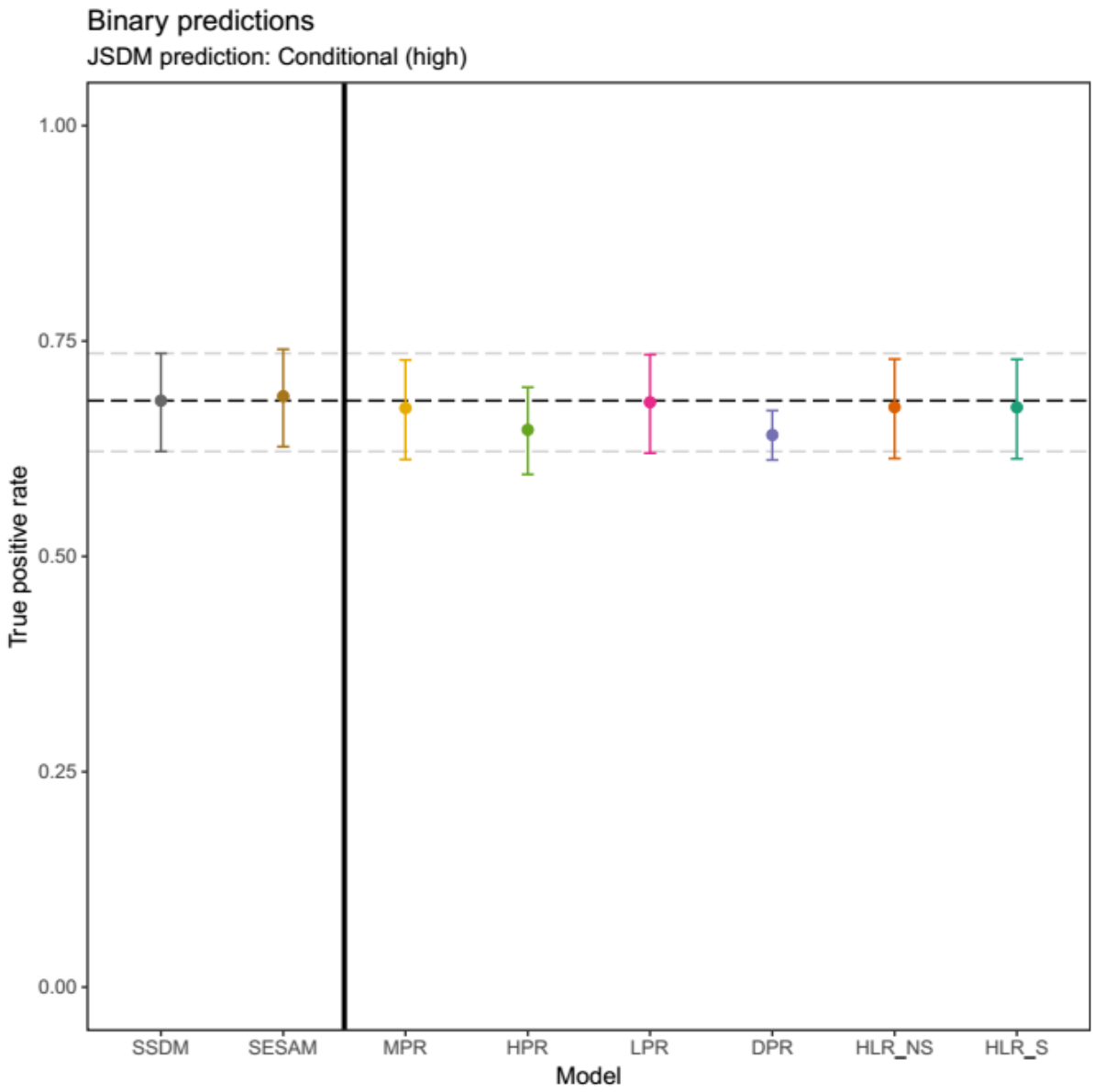


Figure 130: Model performance for the true positive rate evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

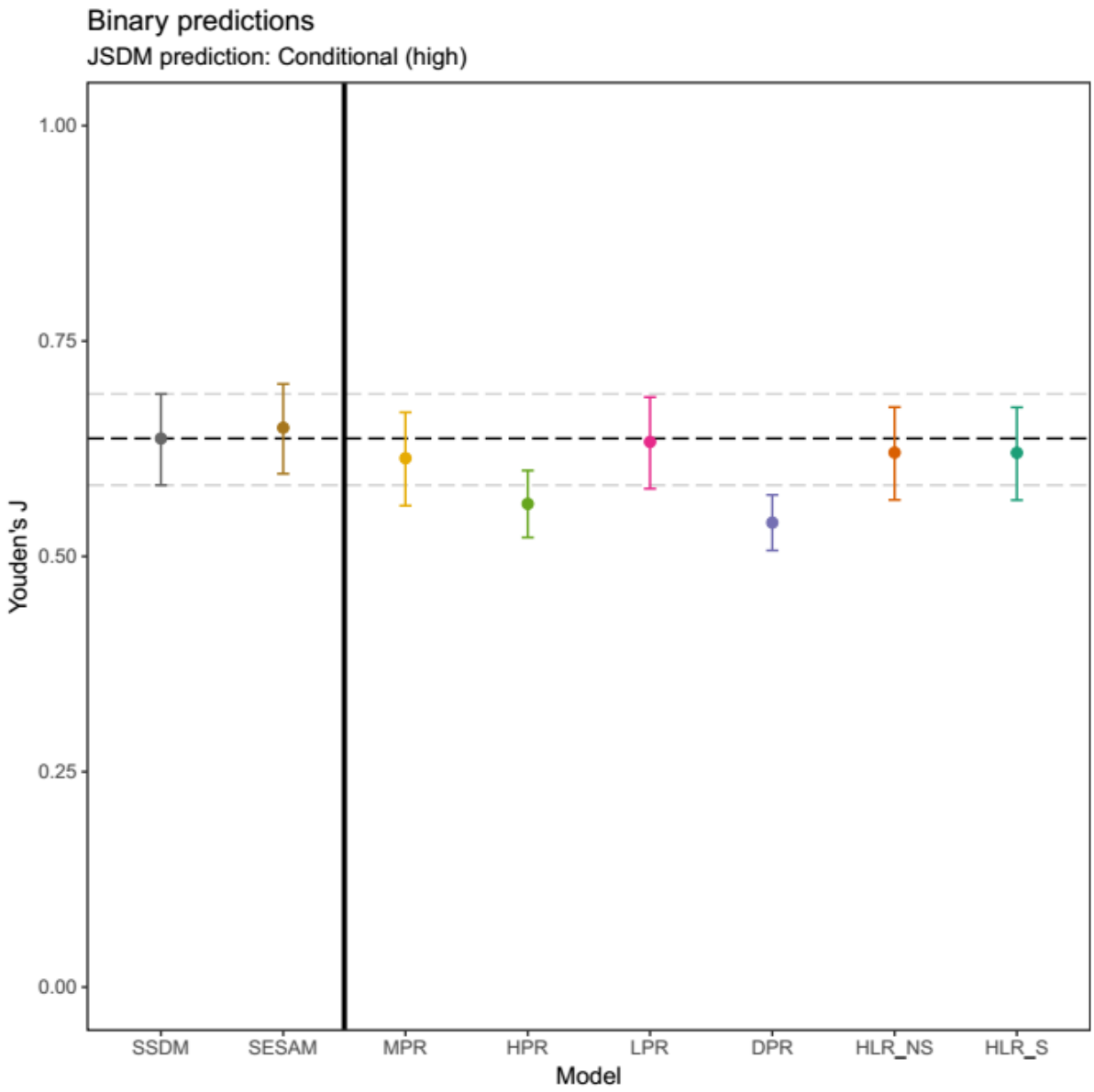


Figure 131: Model performance for the Youden's J evaluation metric for conditional (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

7. Conditional marginal (low-prevalence) predictions

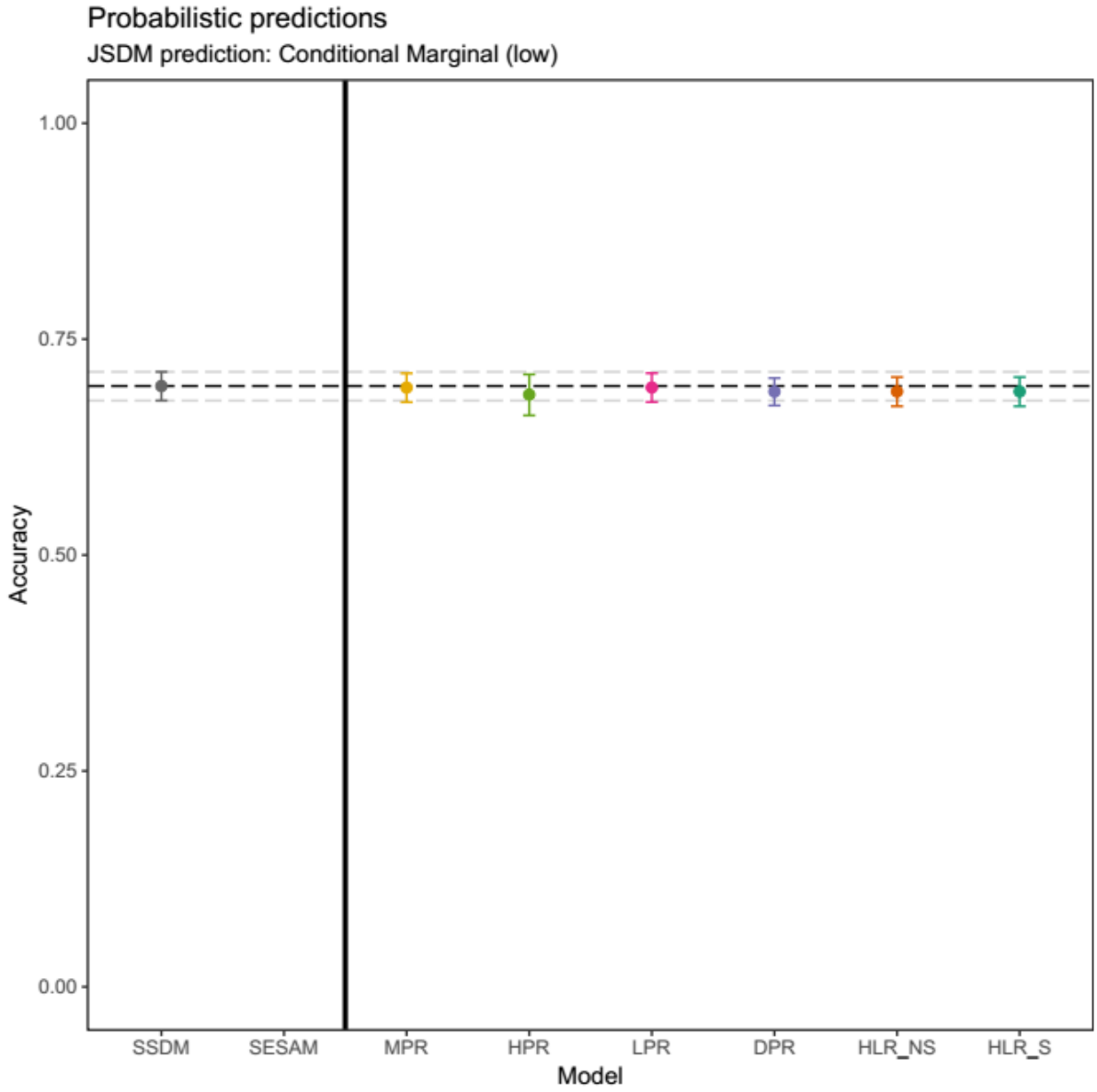


Figure 132: Model performance for the accuracy evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

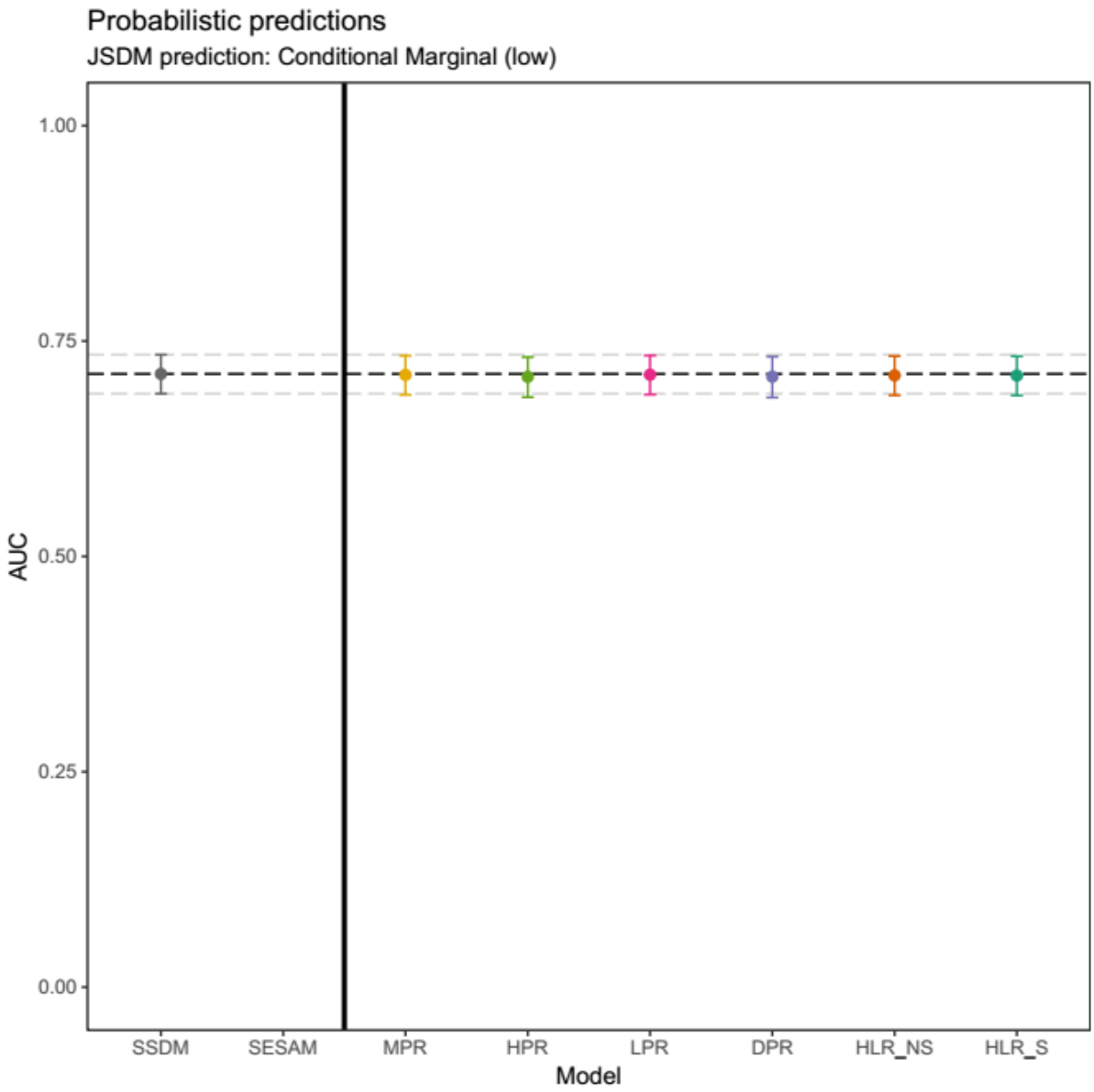


Figure 133: Model performance for the AUC evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

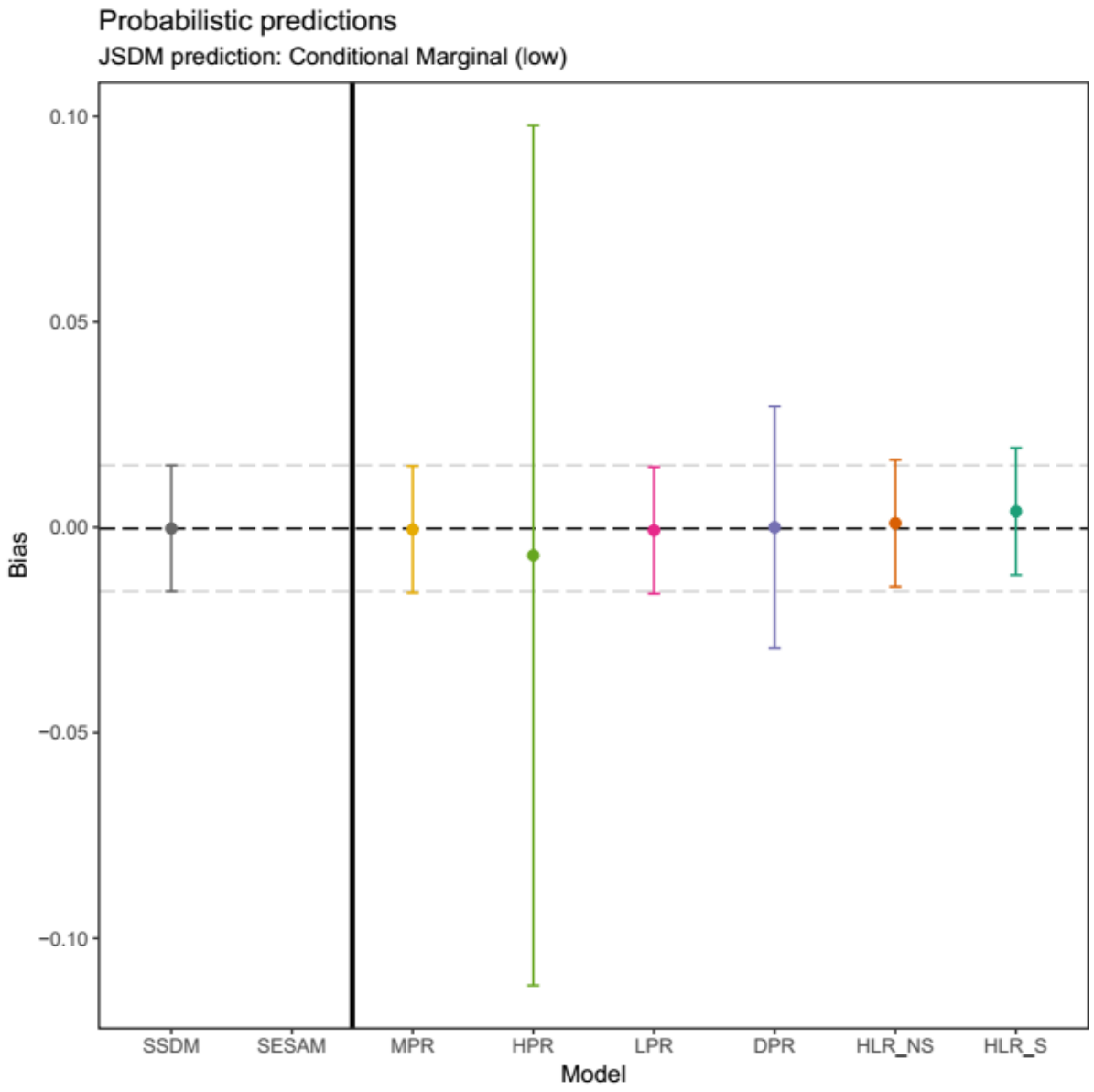


Figure 134: Model performance for the bias (mean error) evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

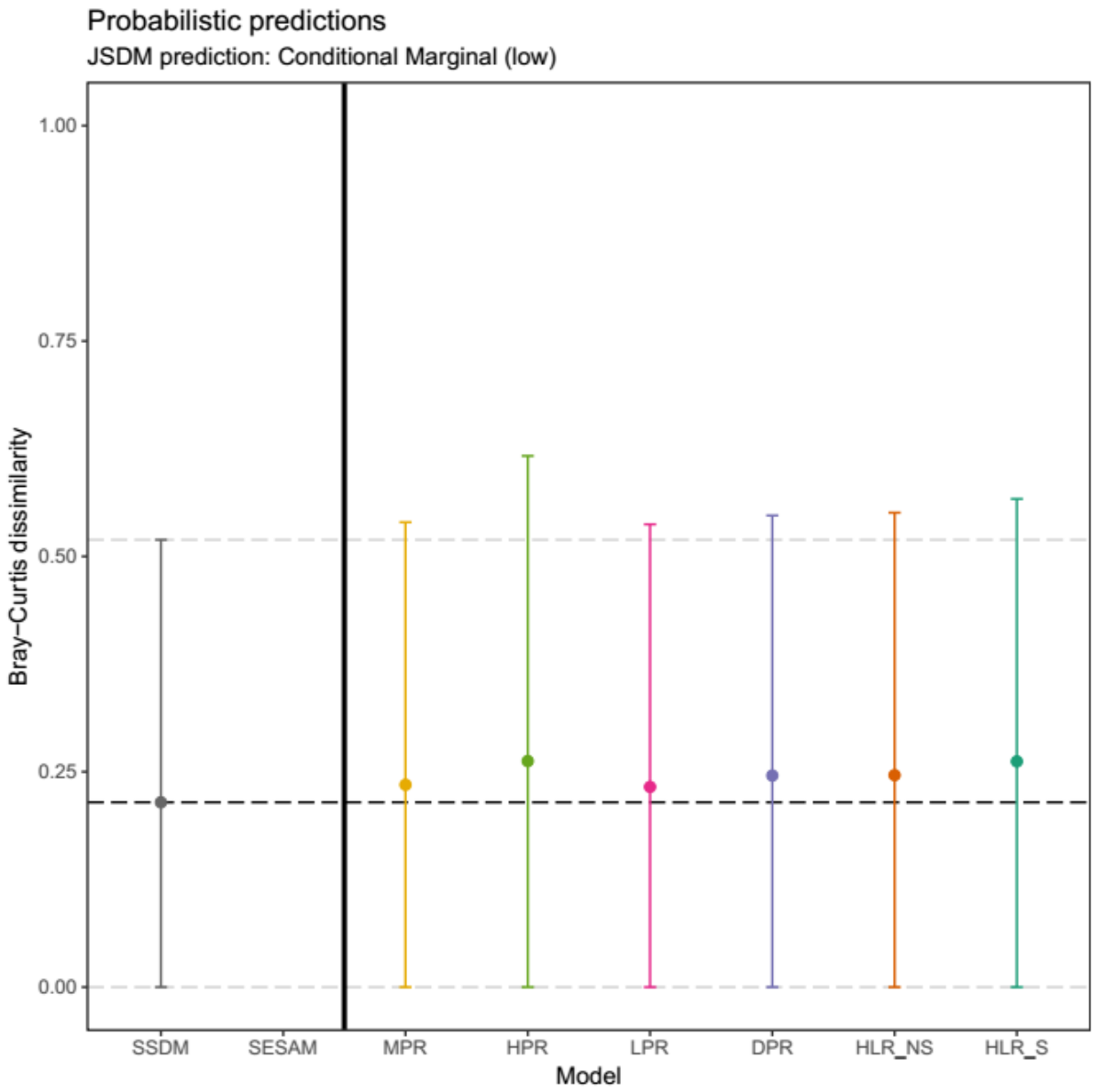


Figure 135: Model performance for the Bray-Curtis dissimilarity evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

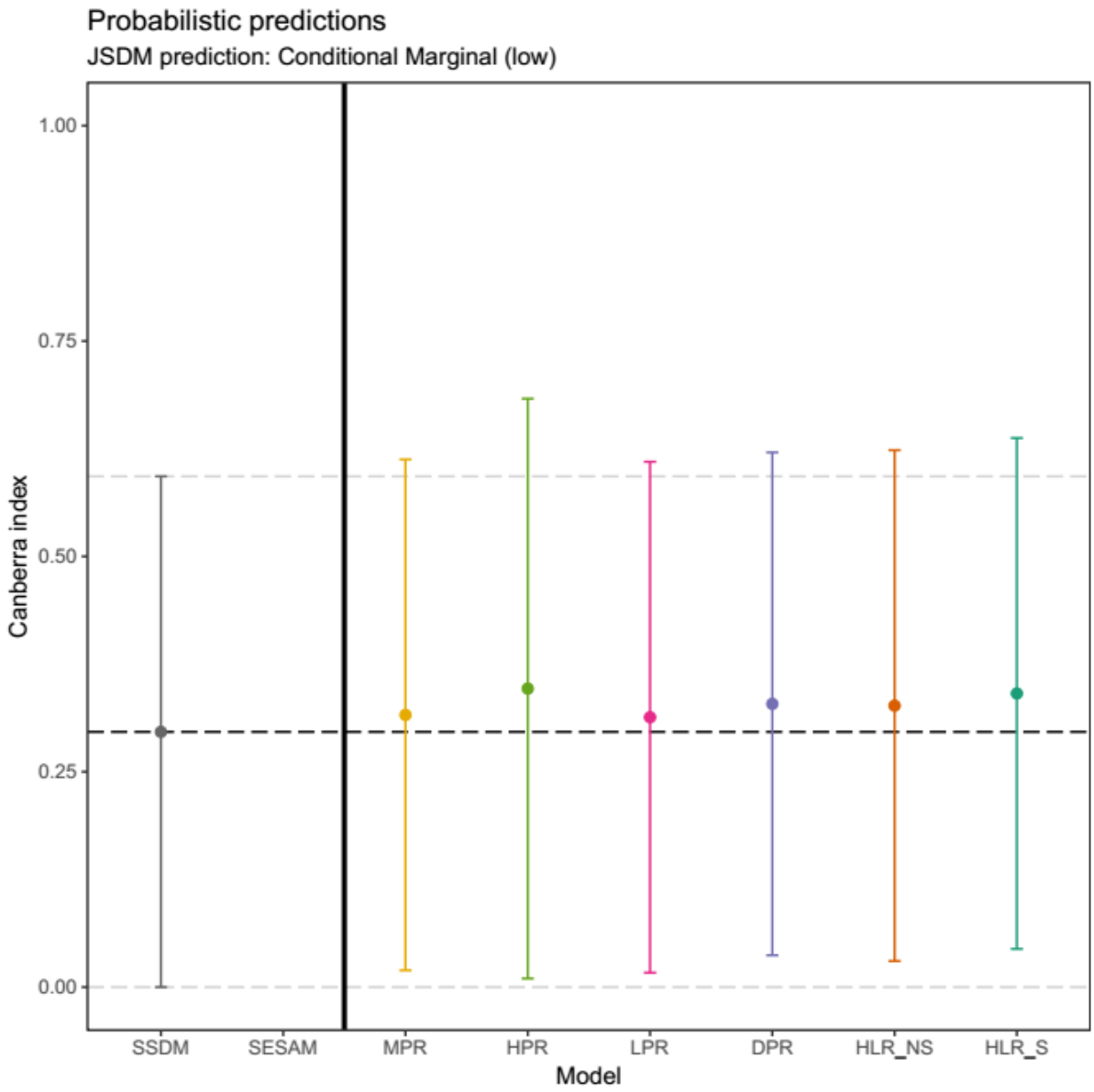


Figure 136: Model performance for the Canberra index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

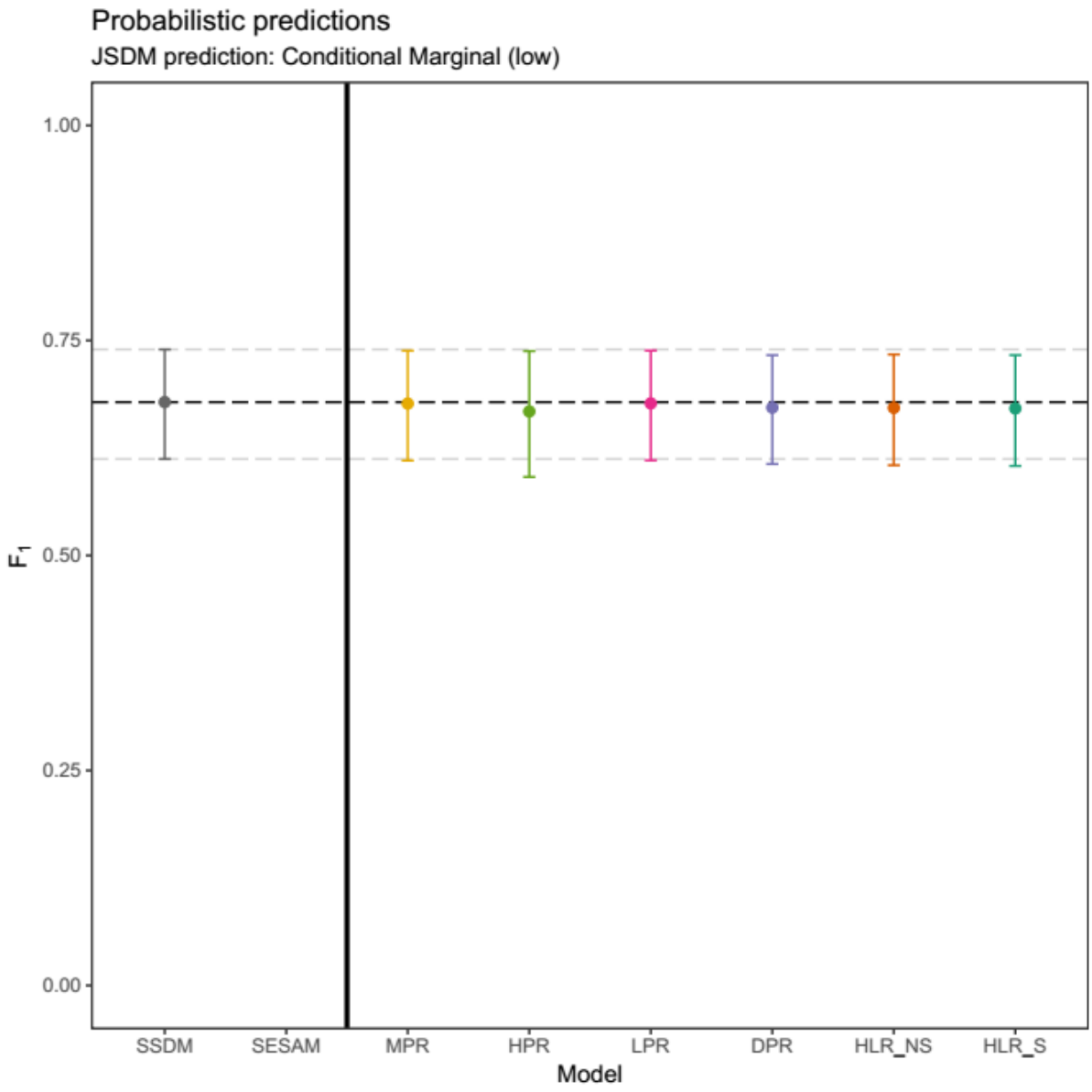


Figure 137: Model performance for the F_1 evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

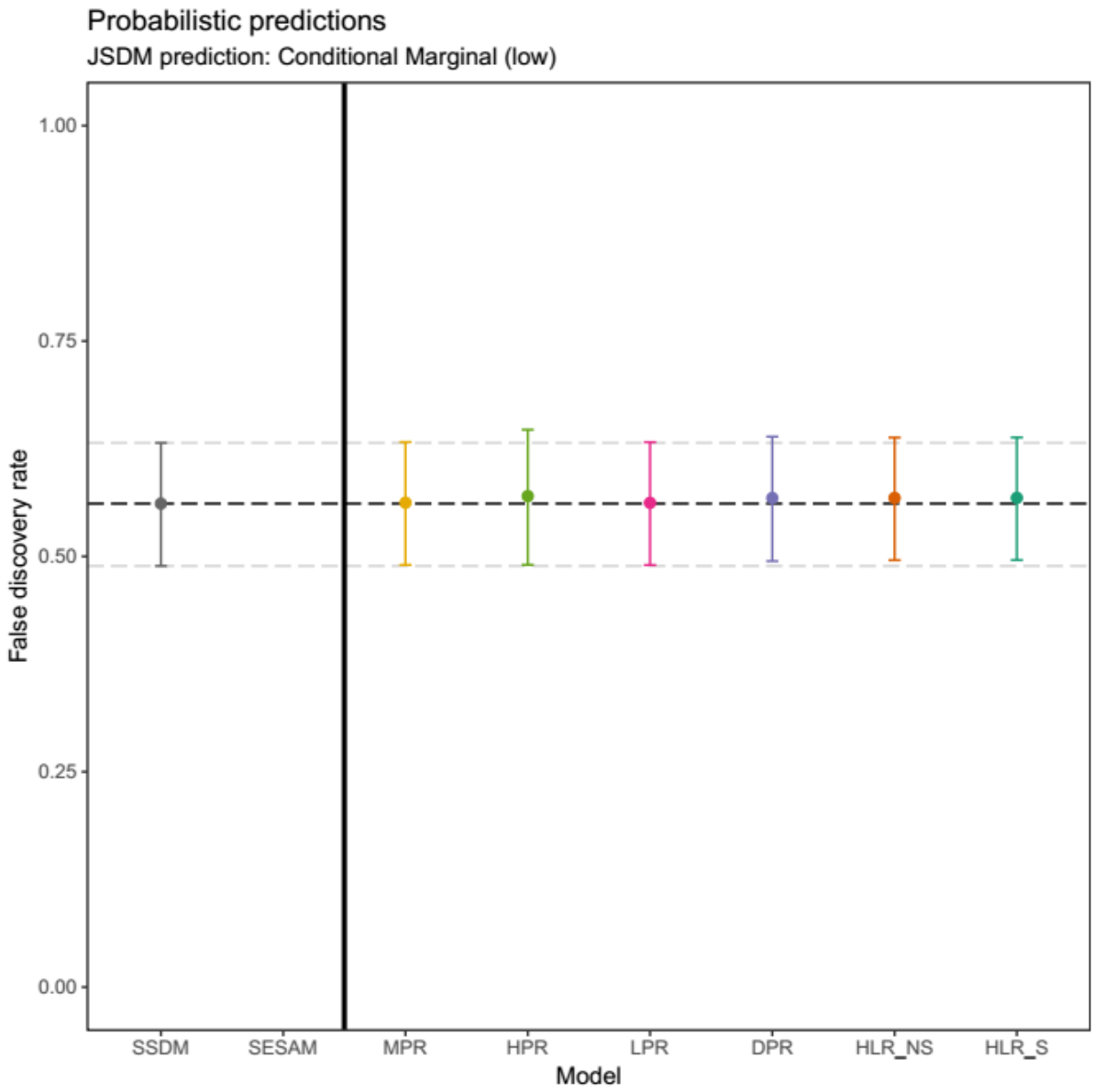


Figure 138: Model performance for the false discovery rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

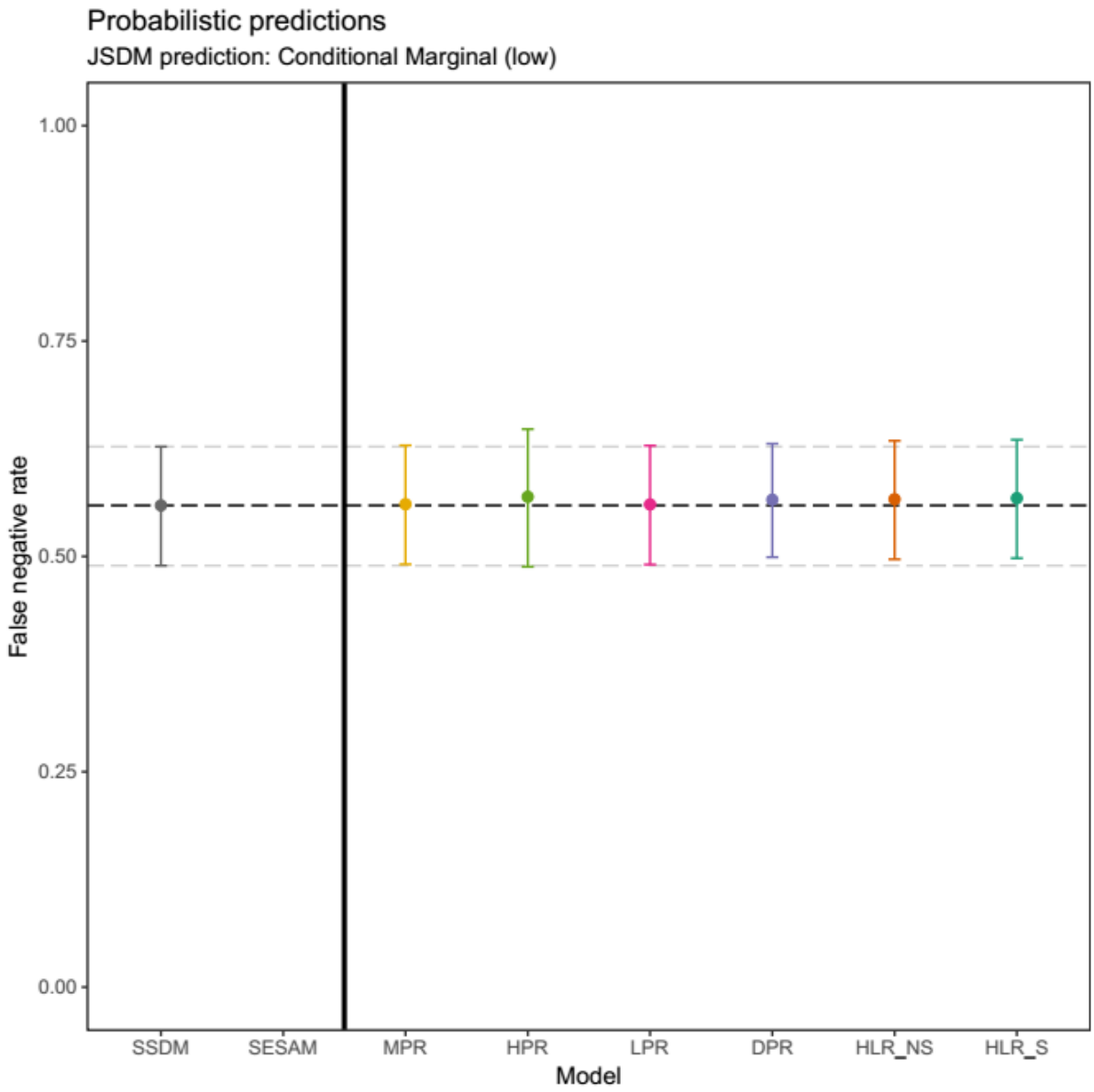


Figure 139: Model performance for the false negative rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

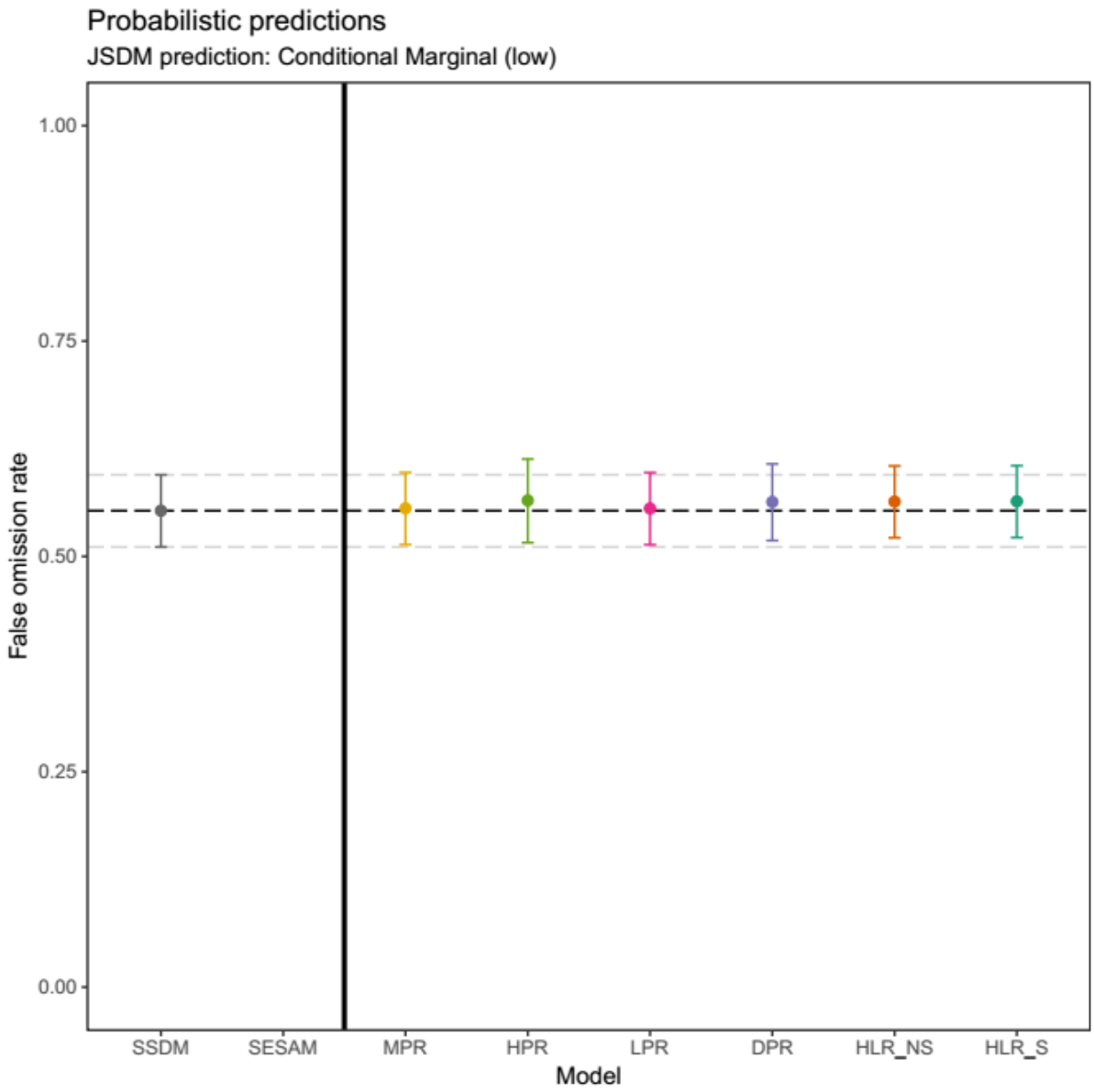


Figure 140: Model performance for the false omission rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

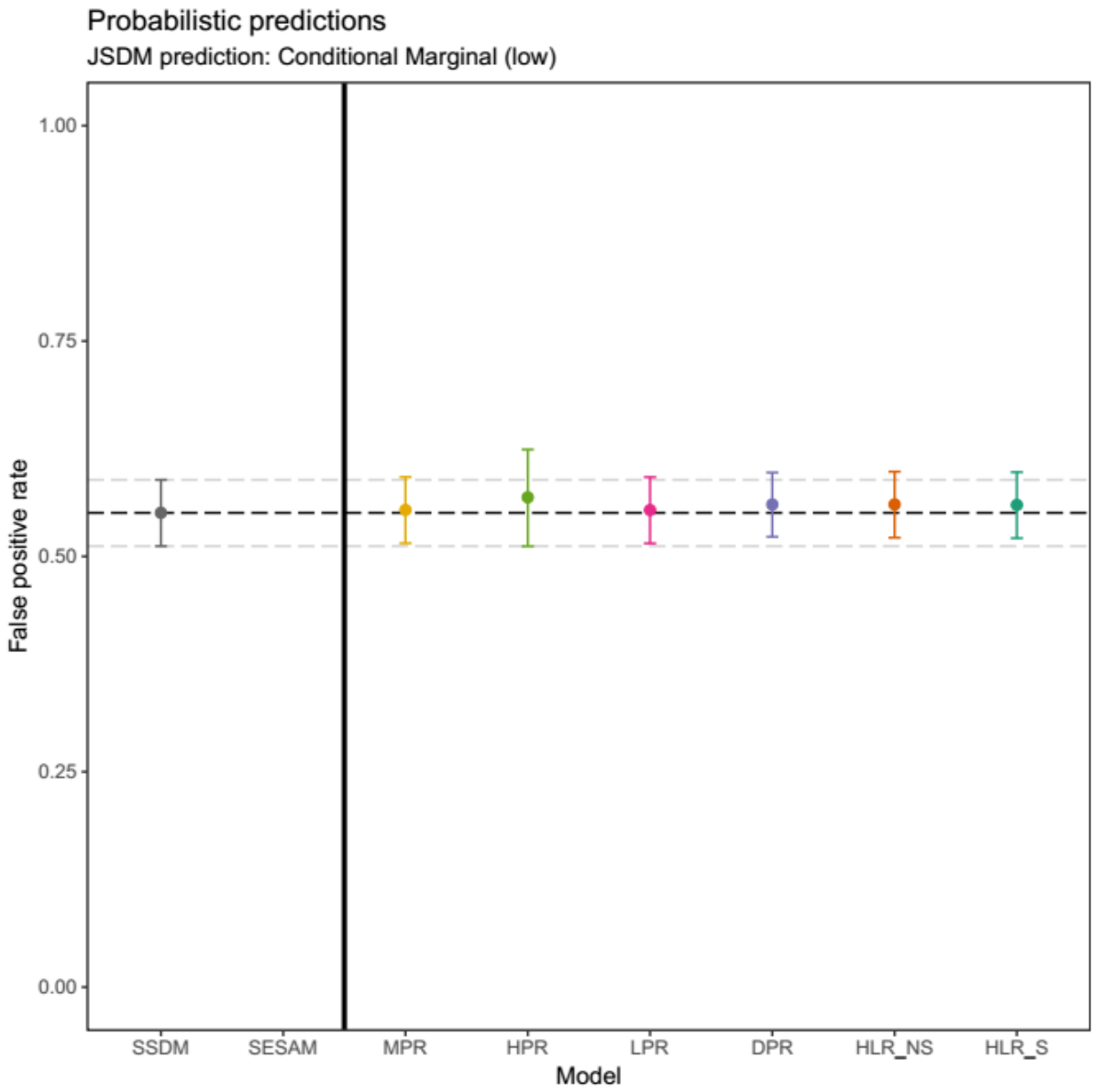


Figure 141: Model performance for the false positive rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

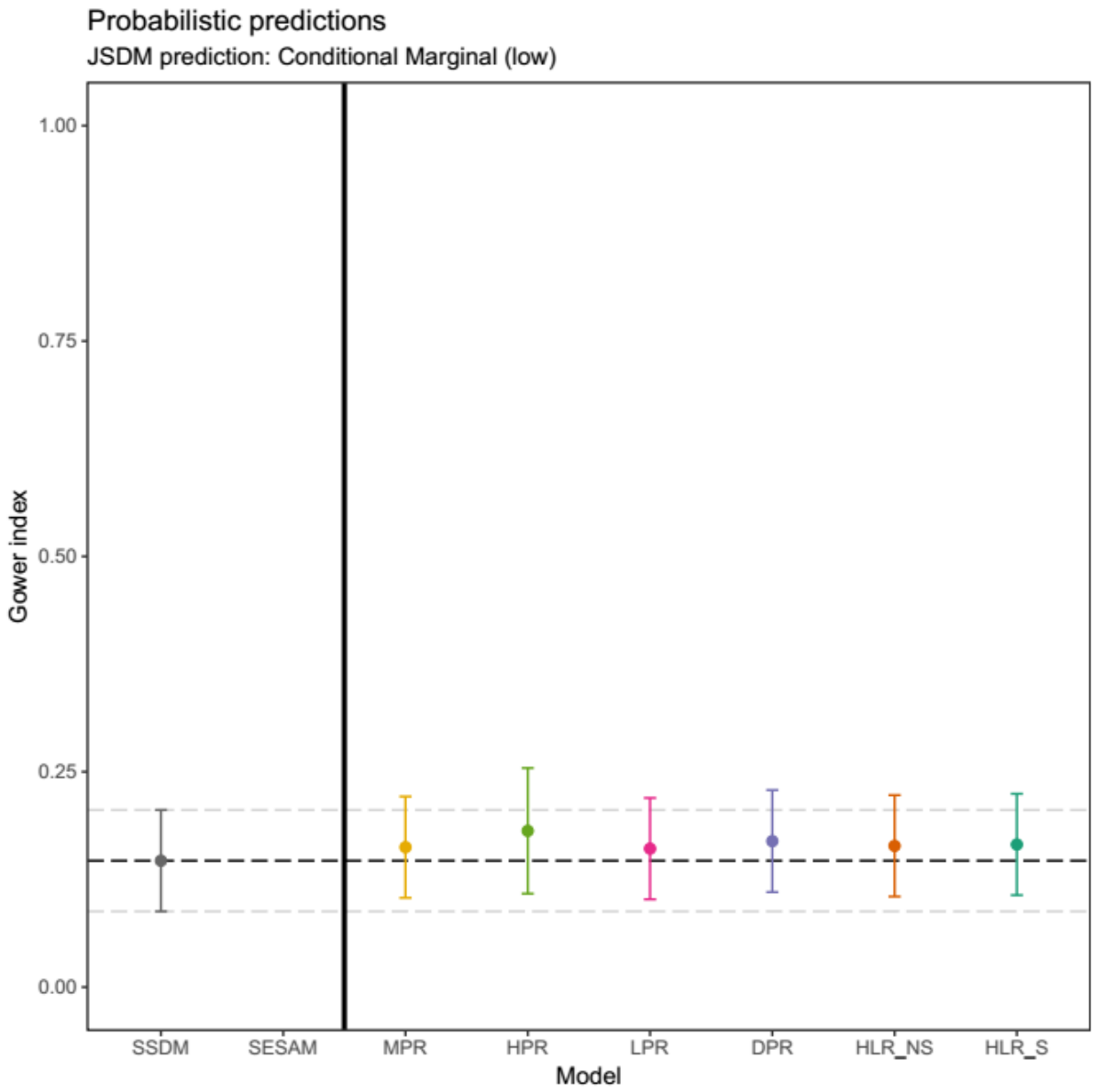


Figure 142: Model performance for the Gower index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

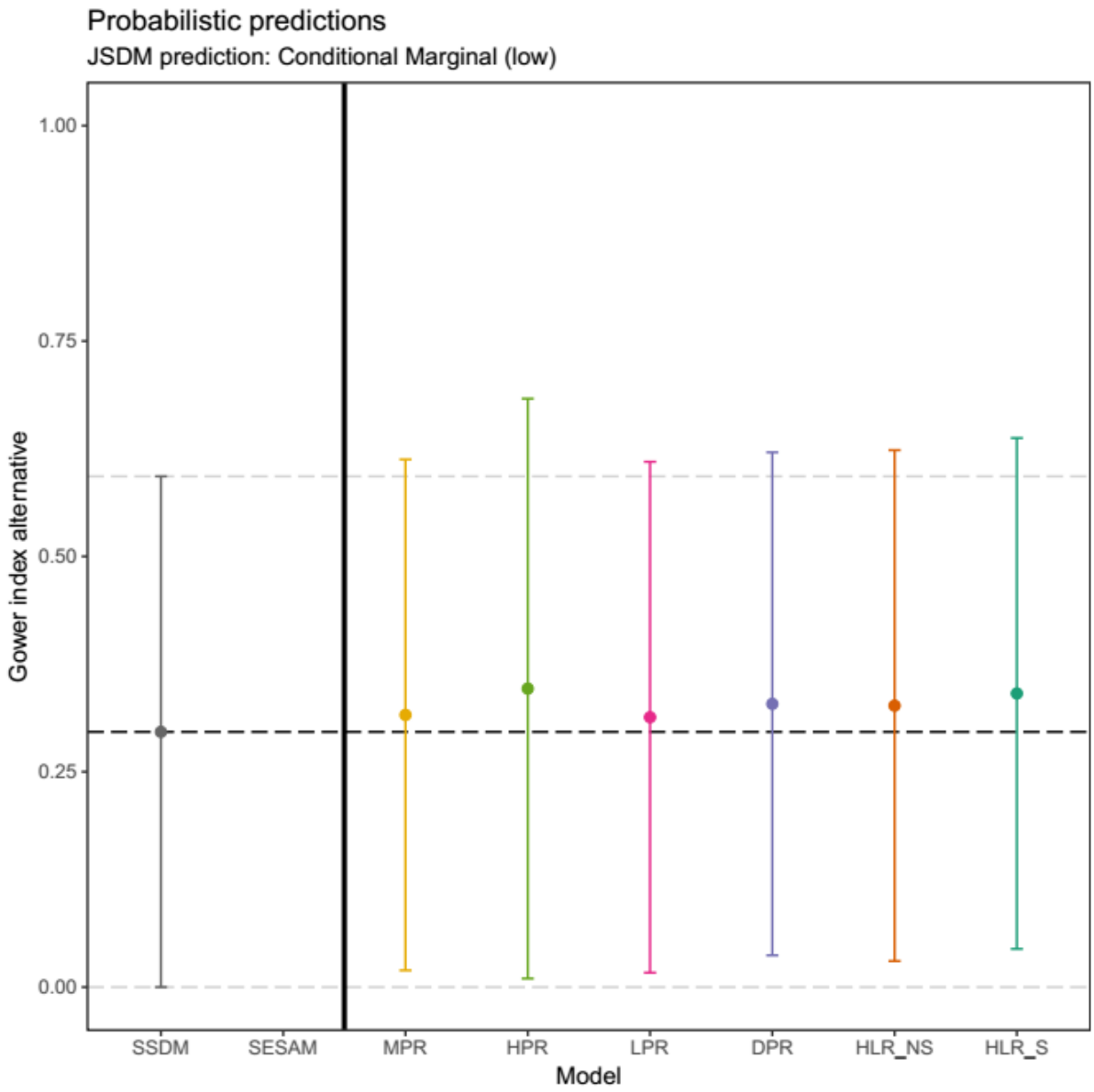


Figure 143: Model performance for the alternative Gower index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

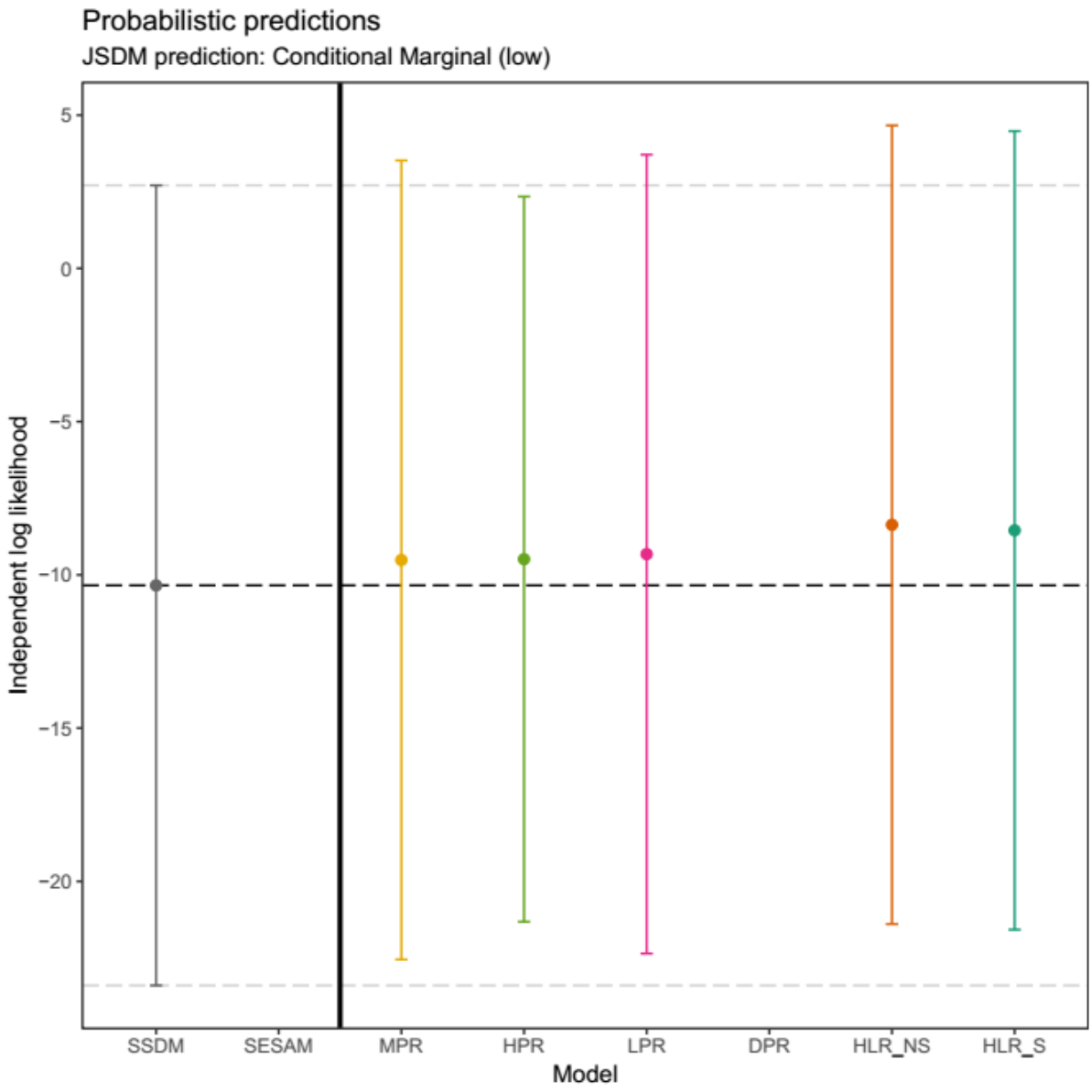


Figure 144: Model performance for the independent log-likelihood evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

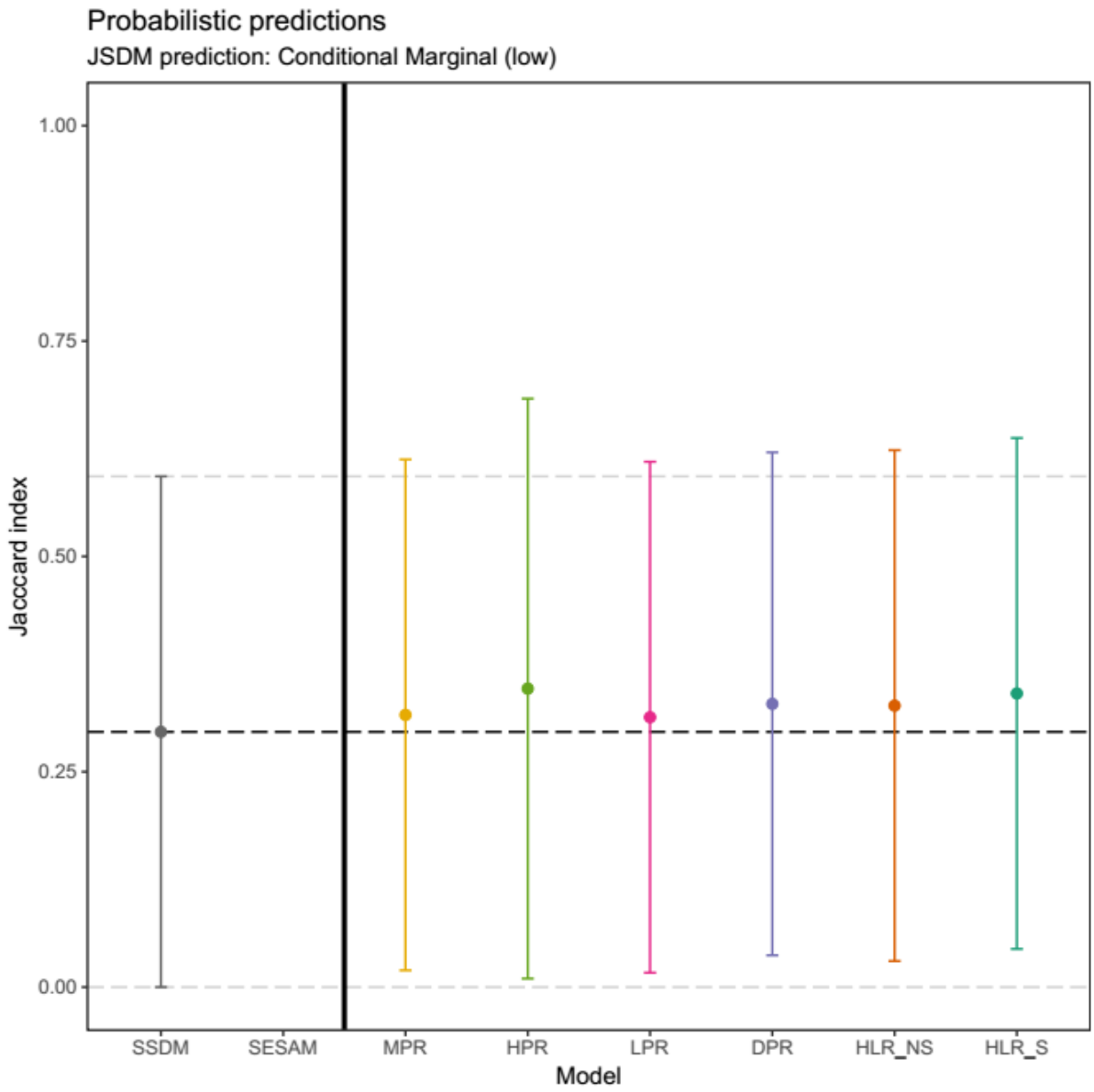


Figure 145: Model performance for the Jaccard index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

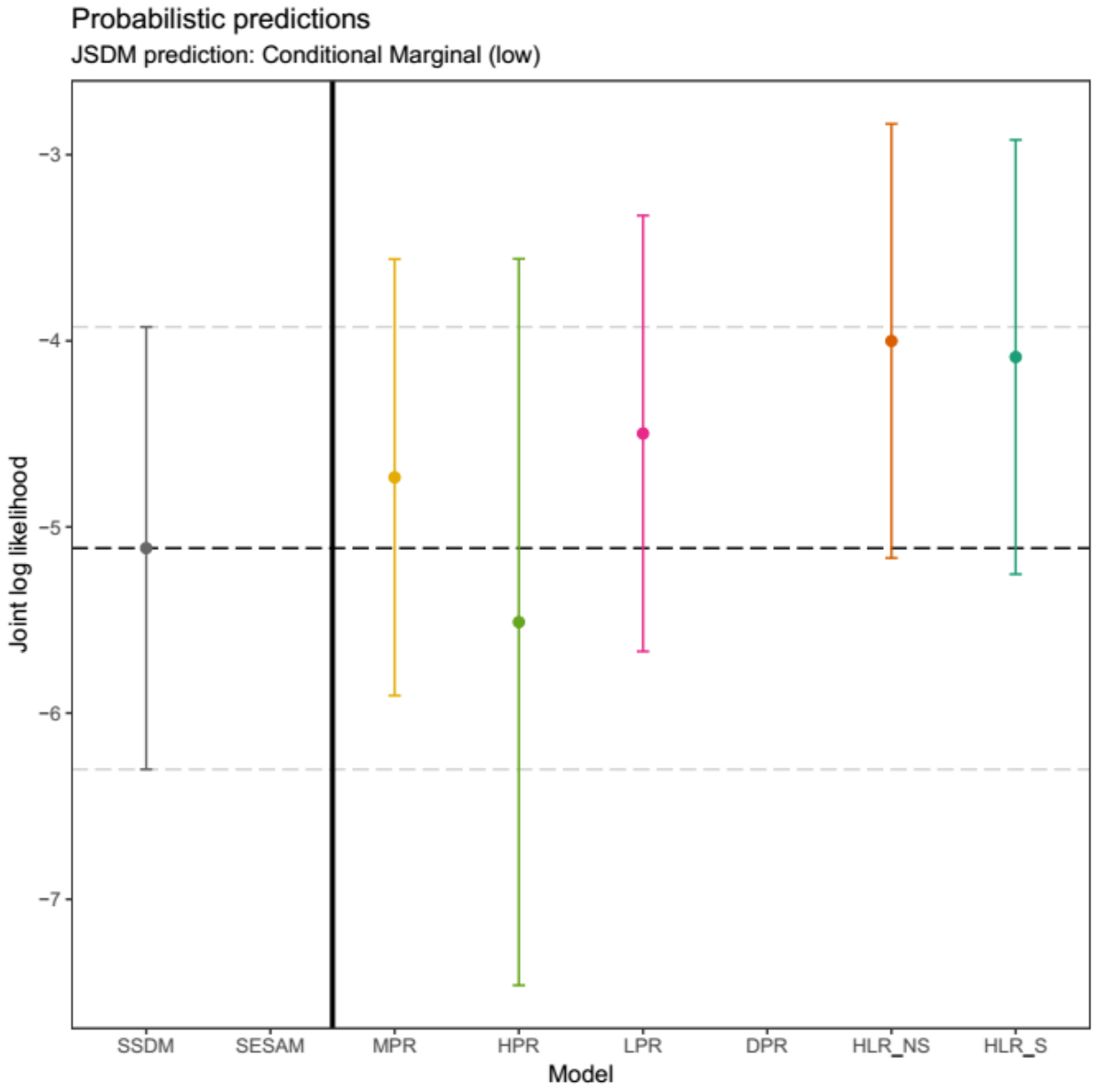


Figure 146: Model performance for the joint log-likelihood evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

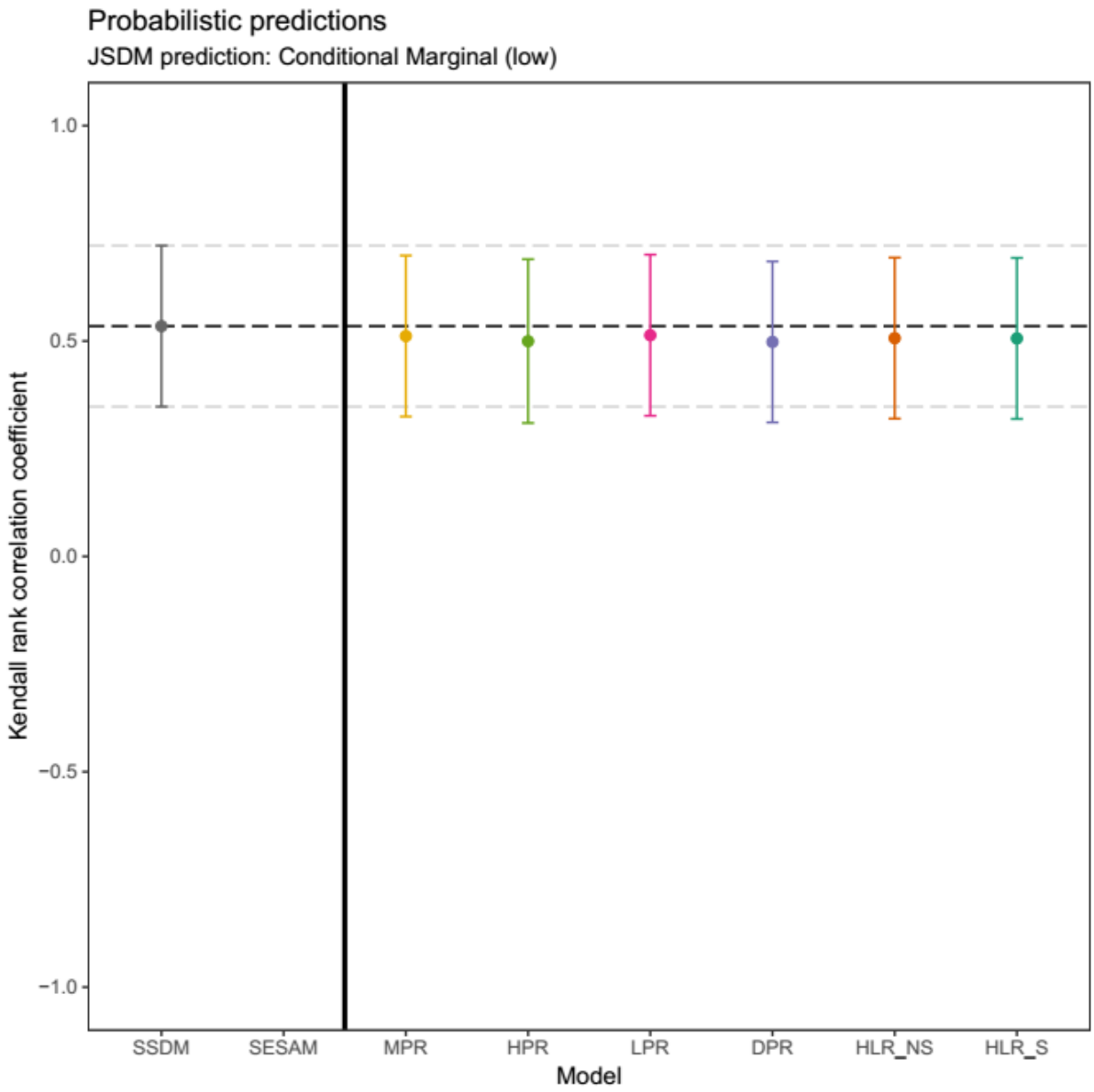


Figure 147: Model performance for the Kendall rank correlation coefficient evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

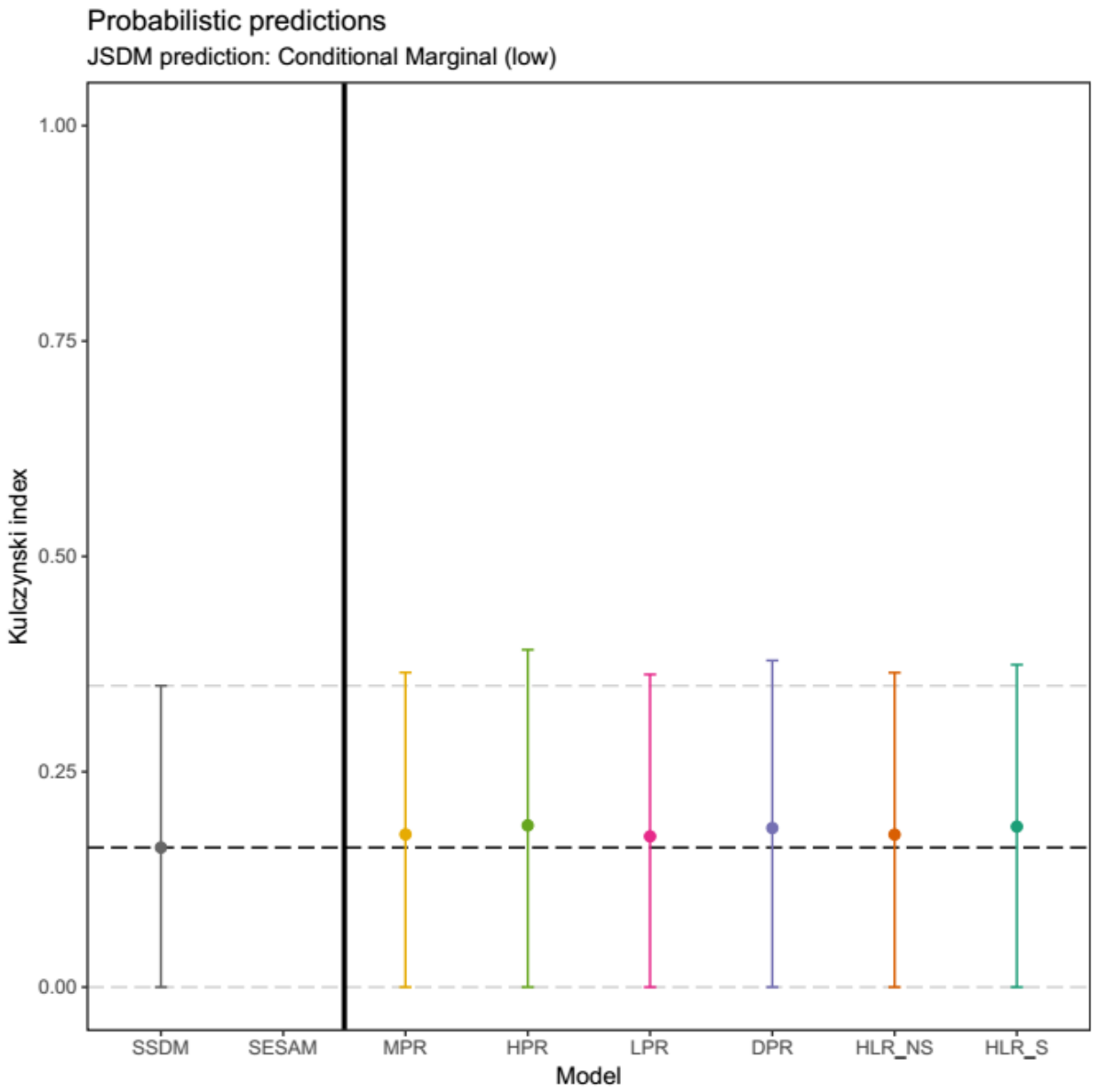


Figure 148: Model performance for the Kulczynski index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

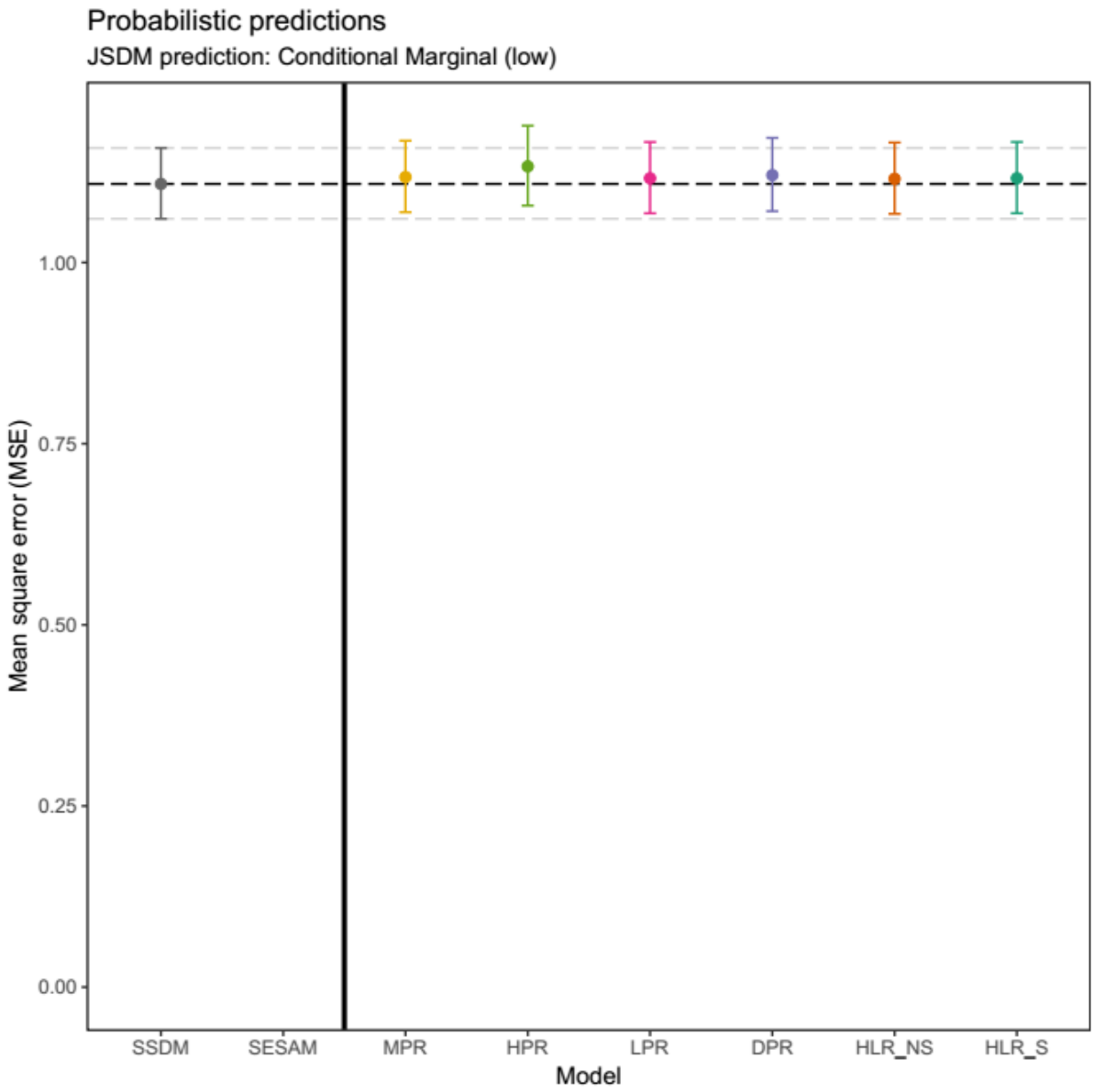


Figure 149: Model performance for the mean square error evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

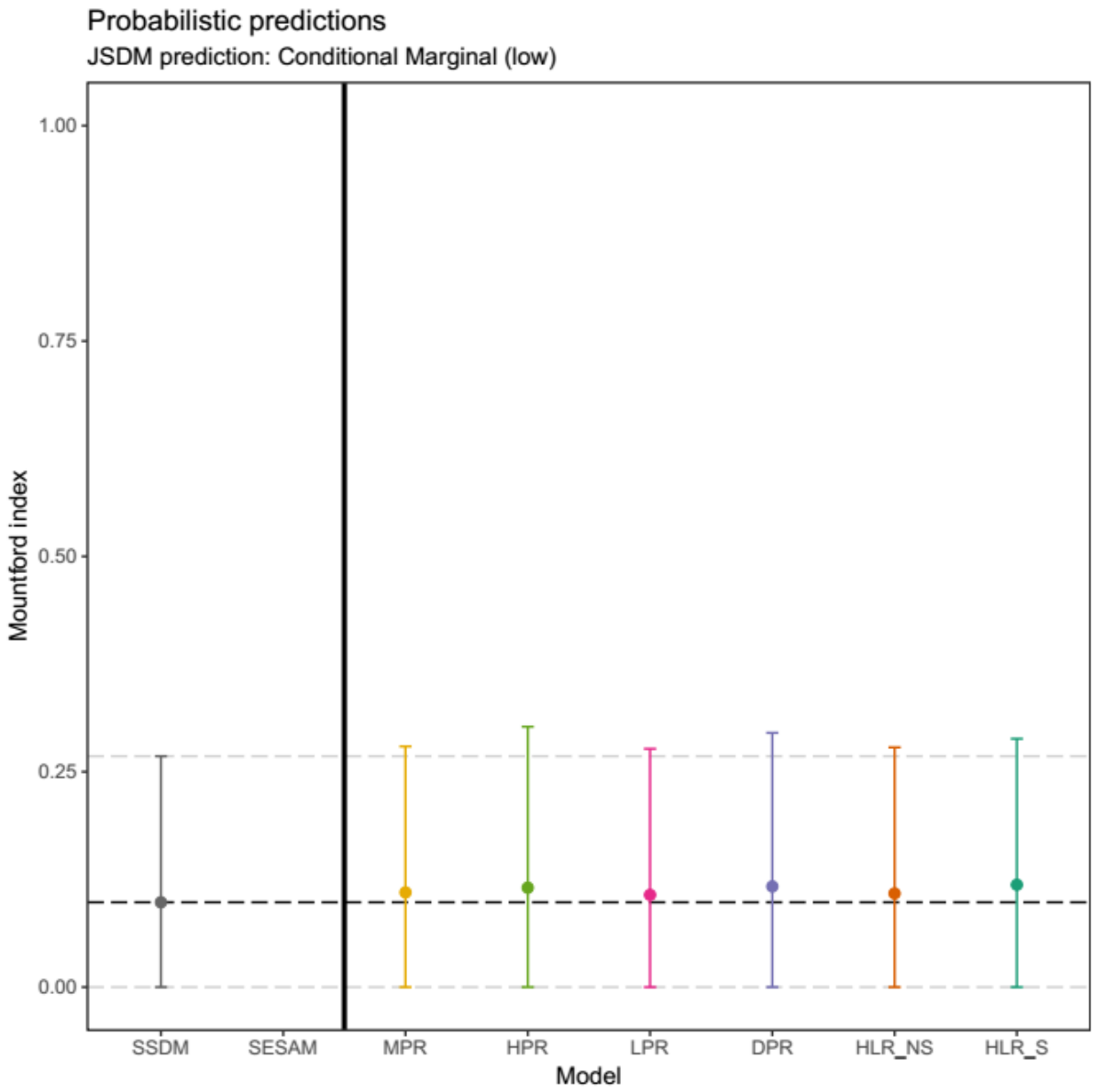


Figure 150: Model performance for the Mountford index evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

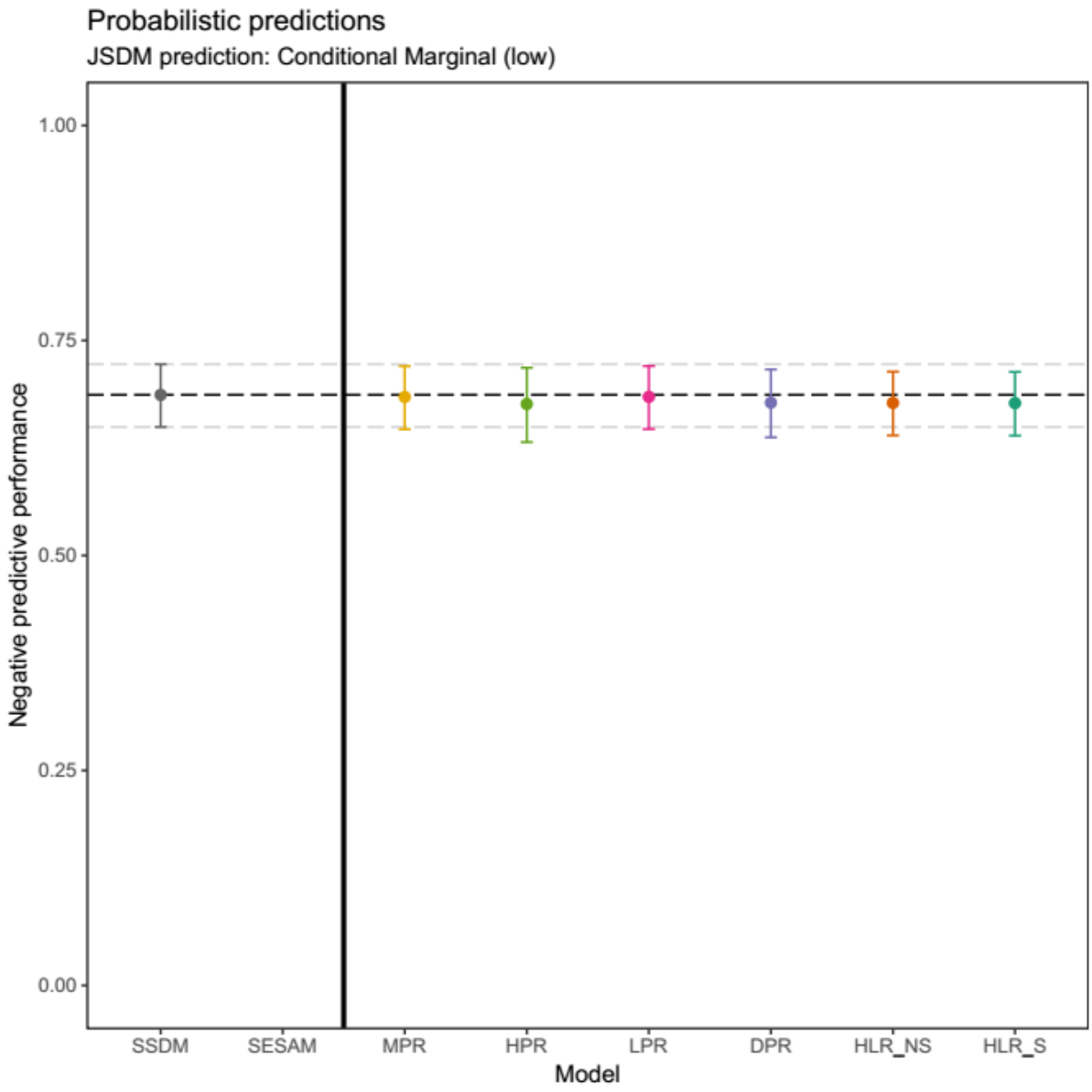


Figure 151: Model performance for the negative predictive performance evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

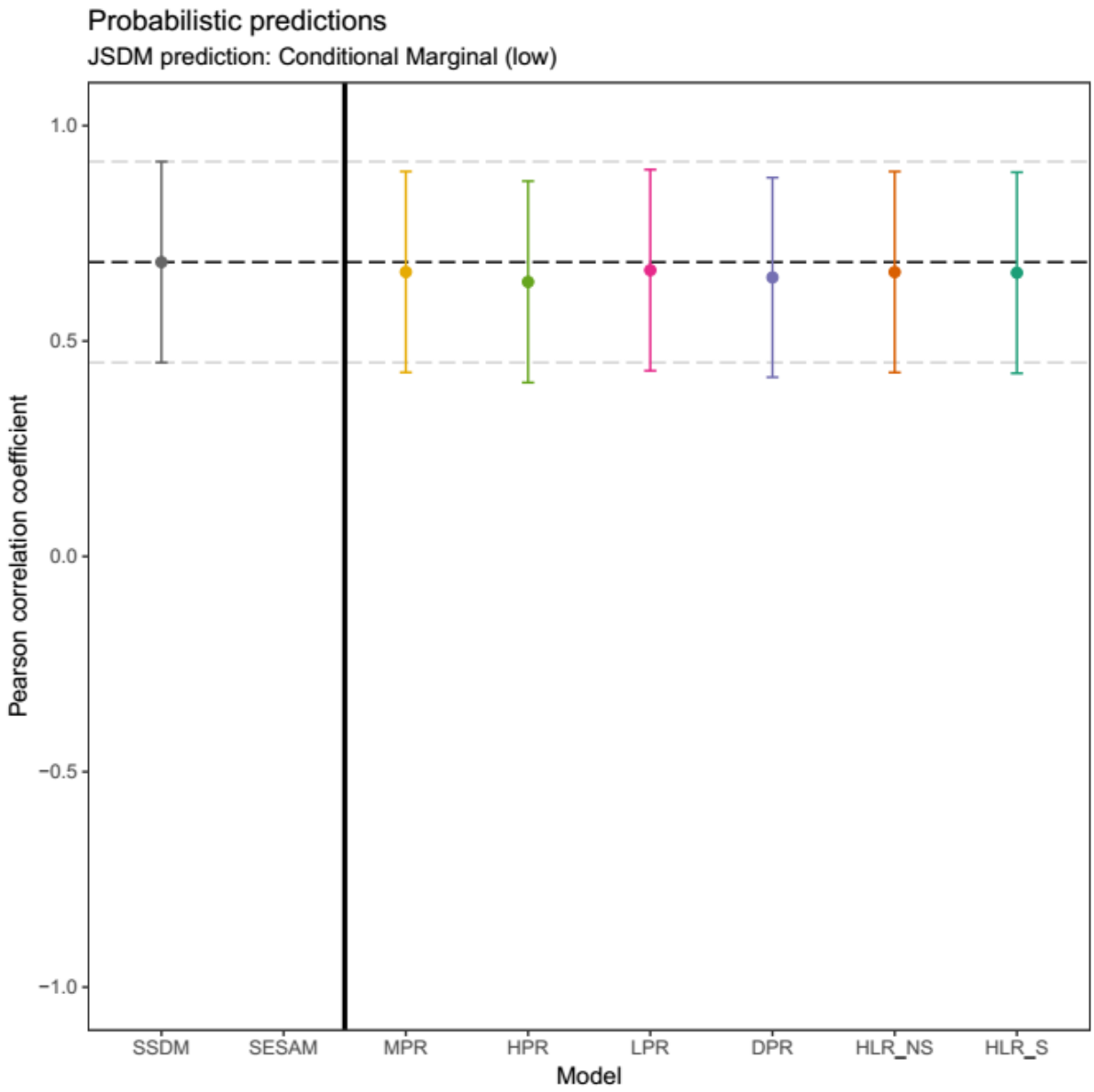


Figure 152: Model performance for the Pearson correlation coefficient evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

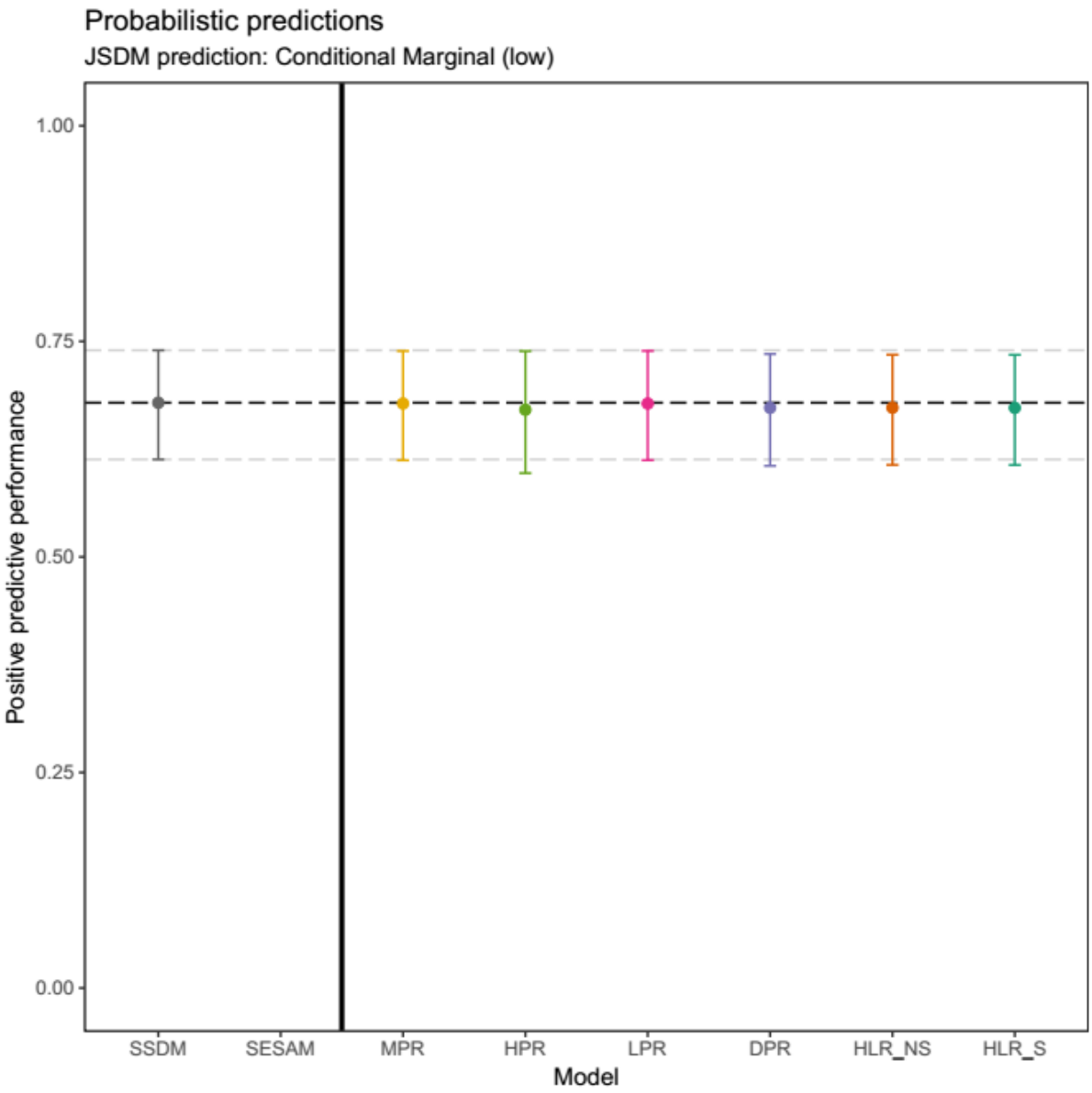


Figure 153: Model performance for the positive predictive performance evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

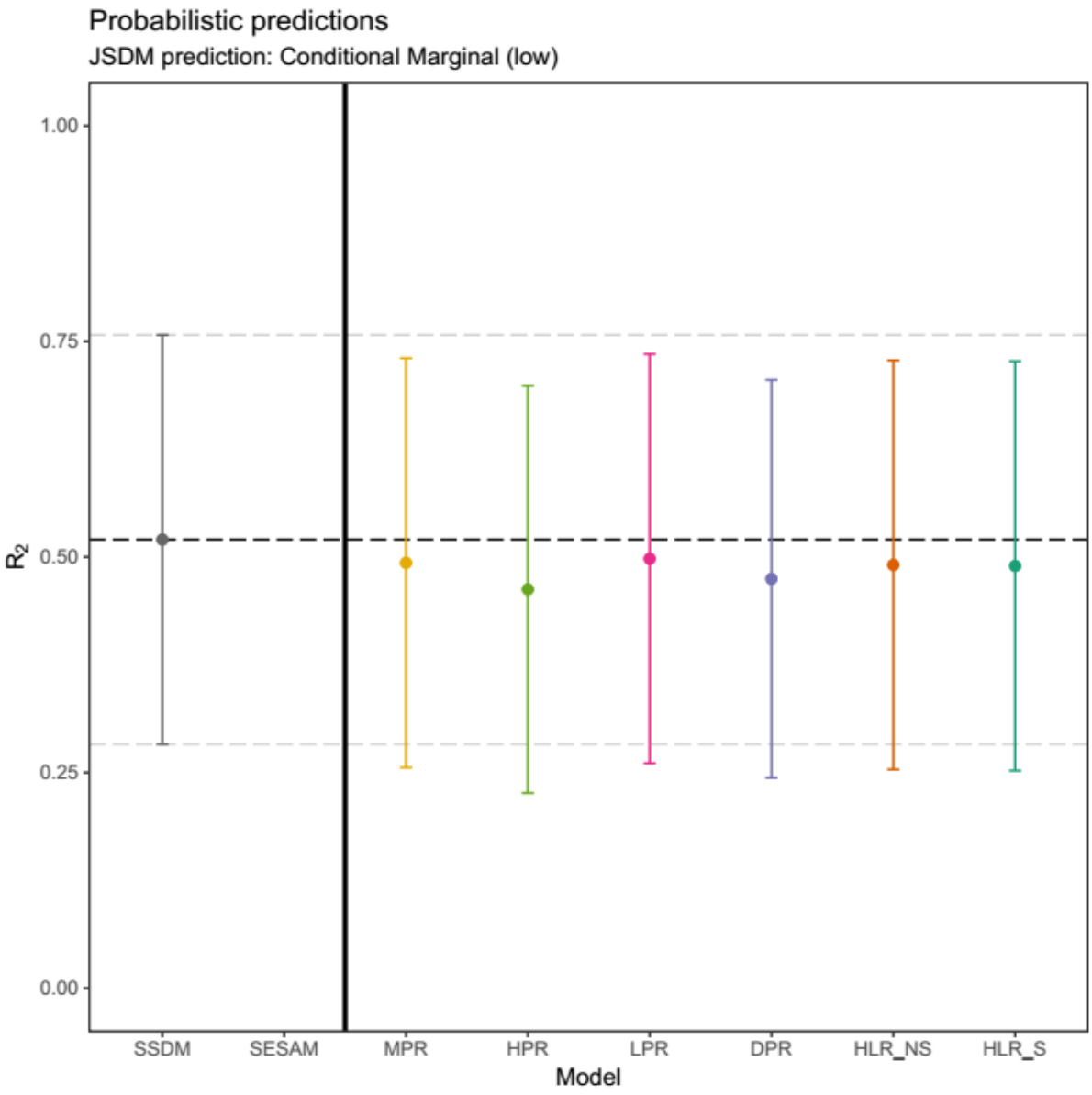


Figure 154: Model performance for the R^2 evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

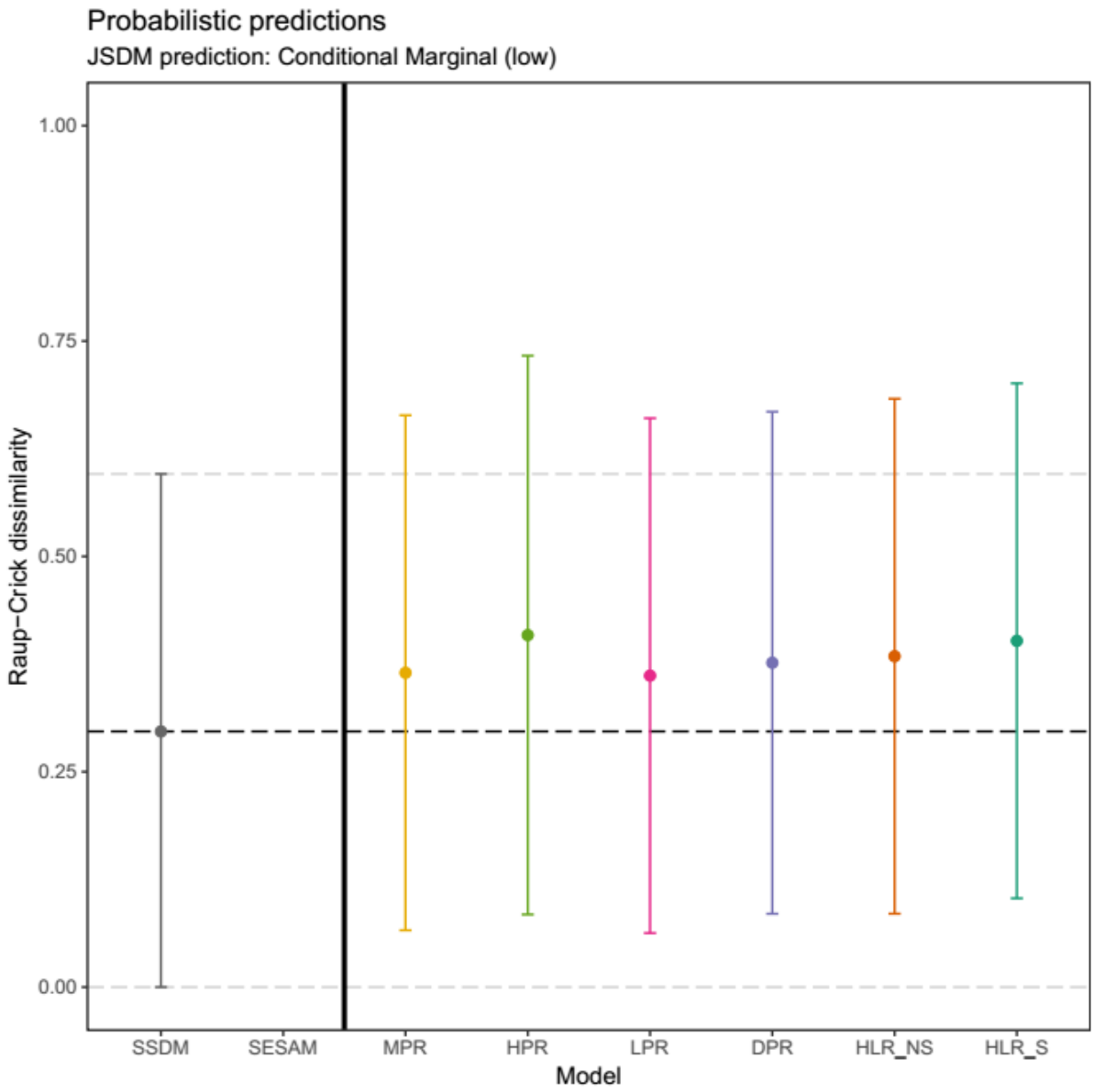


Figure 155: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

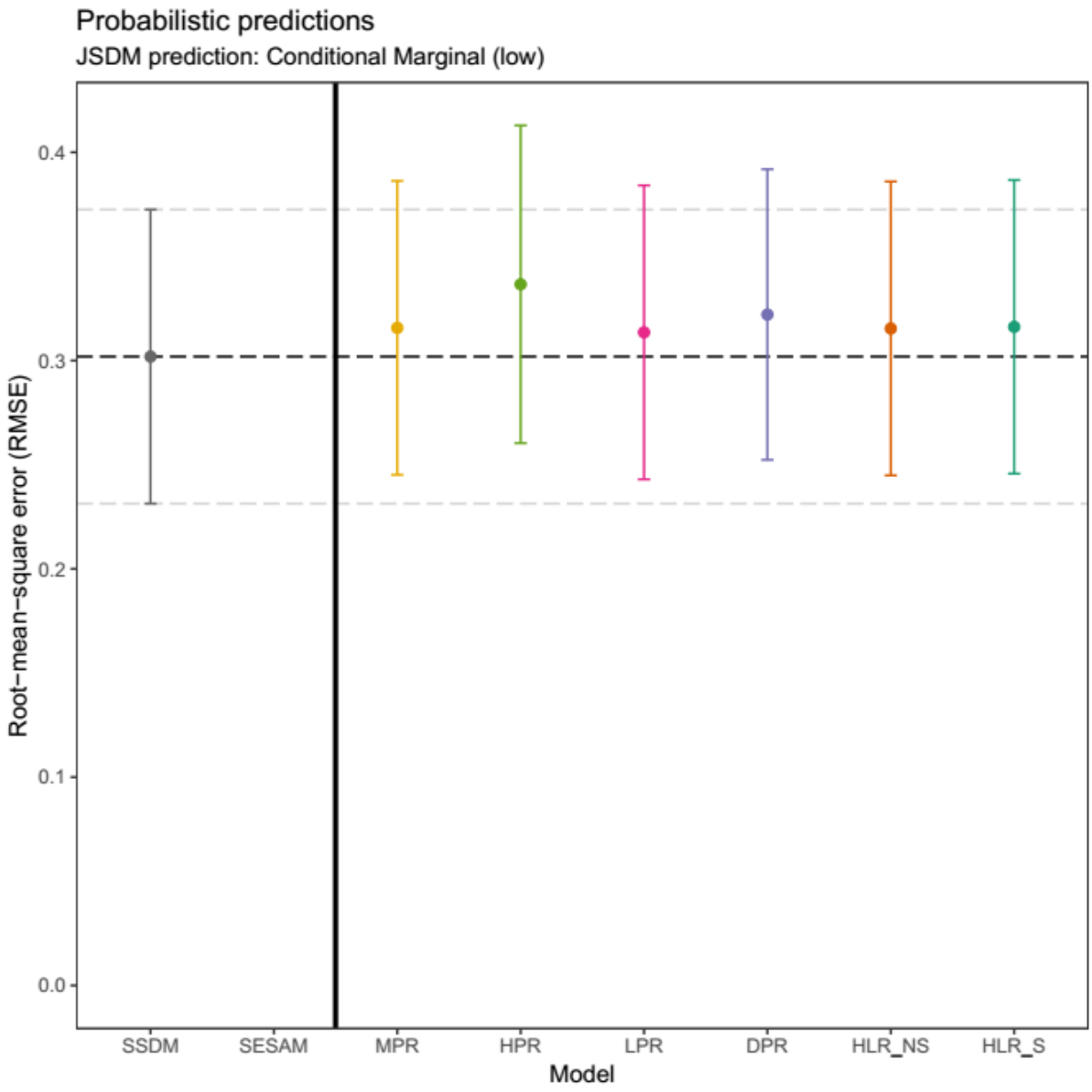


Figure 156: Model performance for the root-mean-square error evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

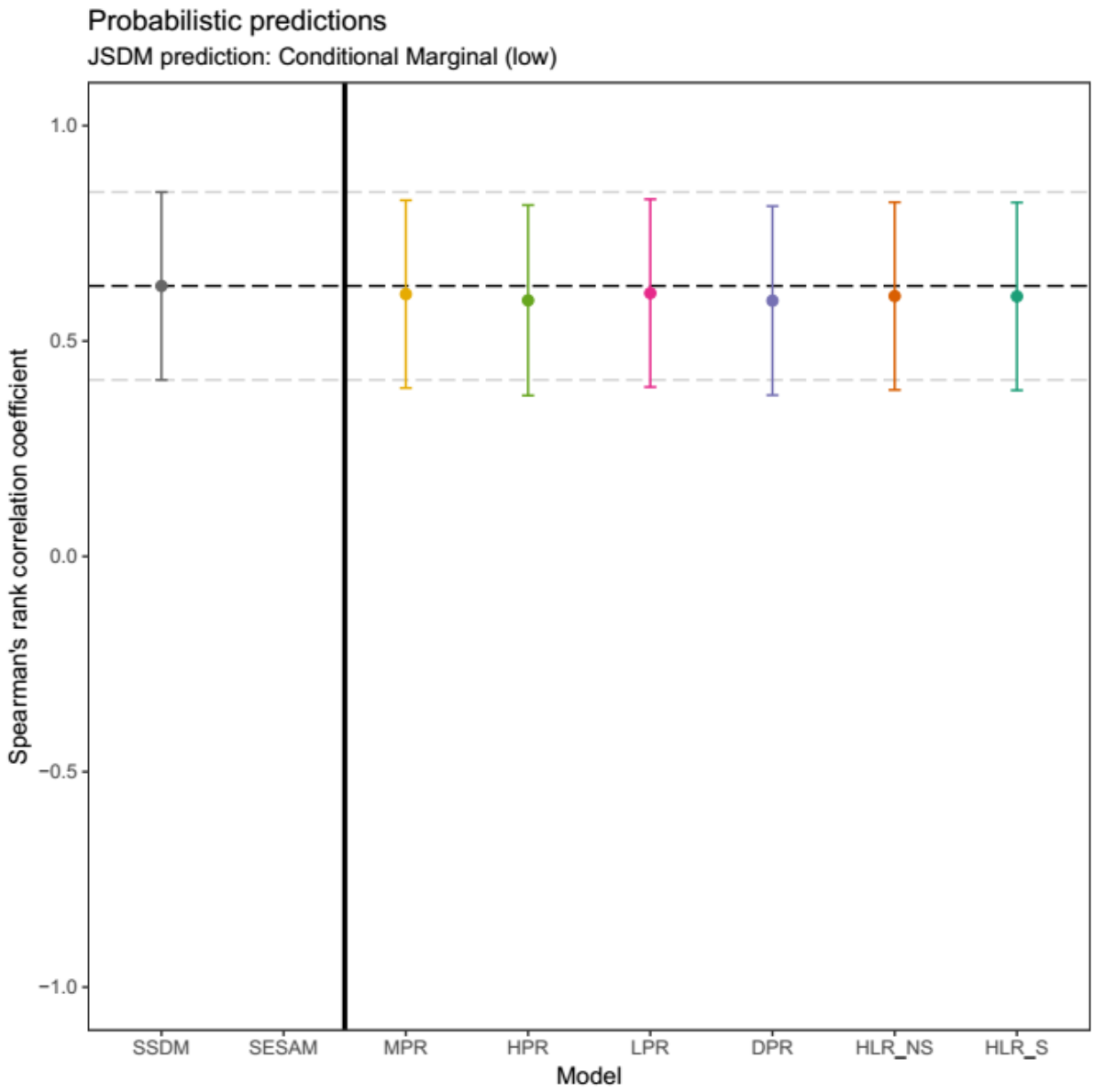


Figure 157: Model performance for the Spearman's rank correlation coefficient evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

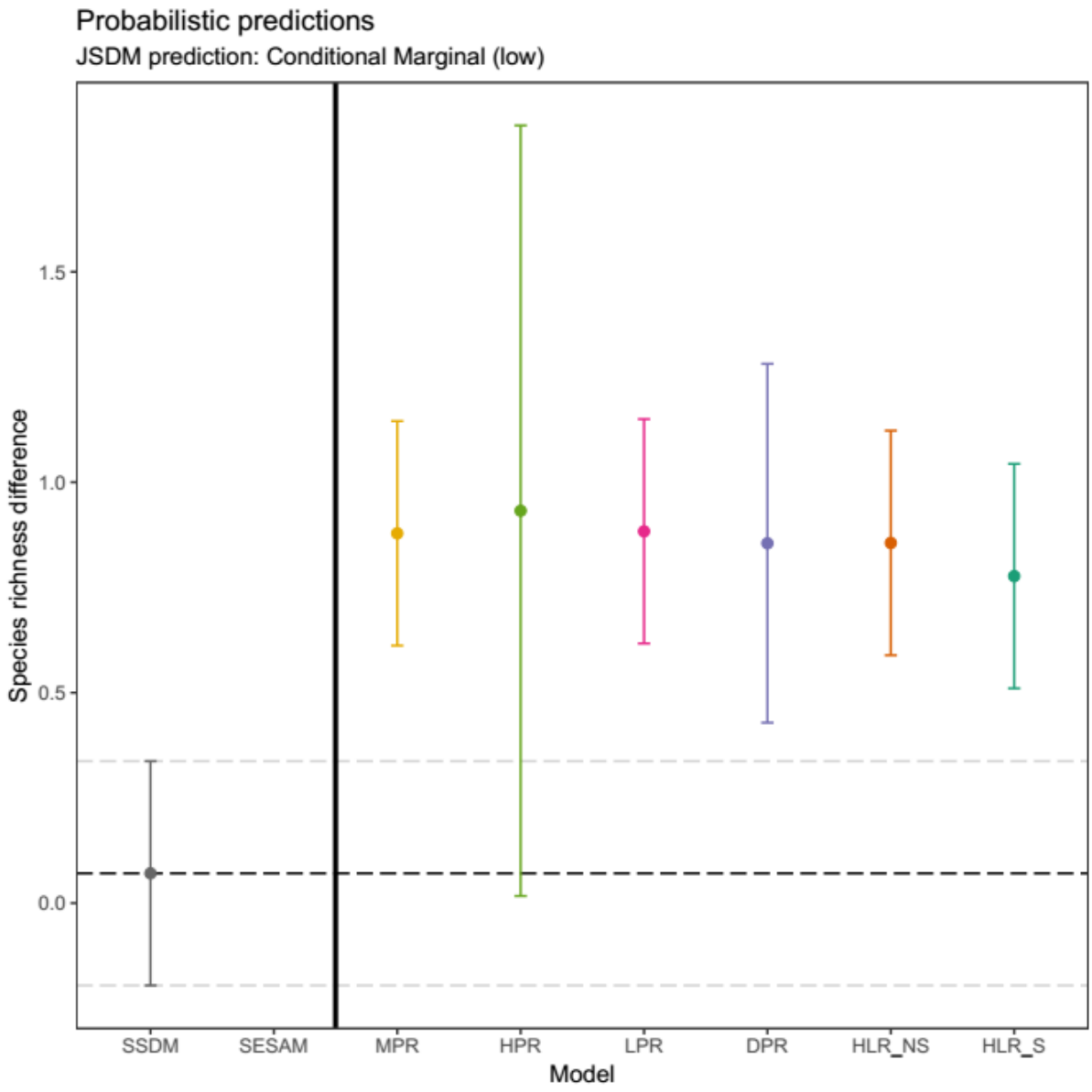


Figure 158: Model performance for the species richness difference evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

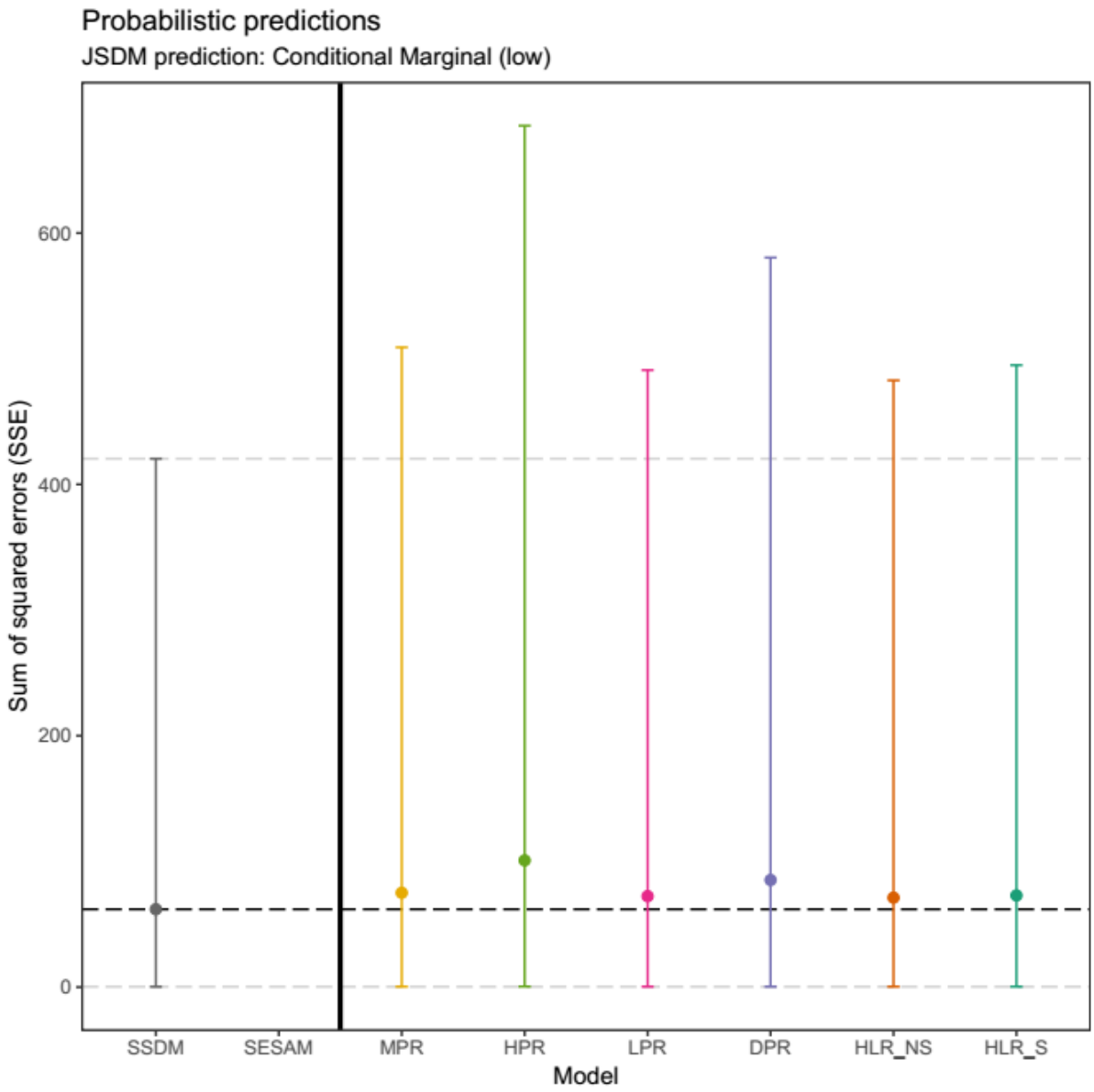


Figure 159: Model performance for the sum of squared errors evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

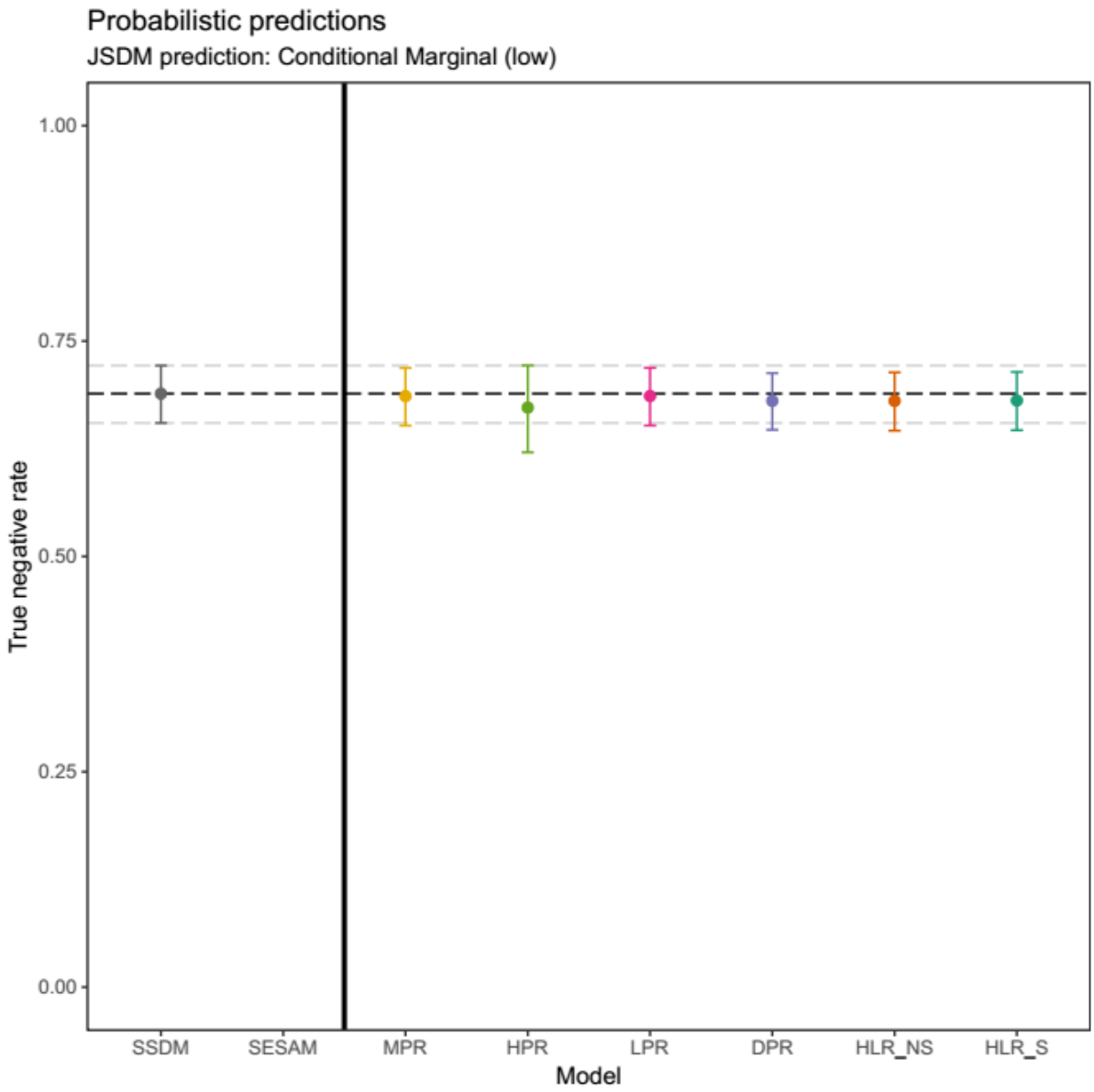


Figure 160: Model performance for the true negative rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

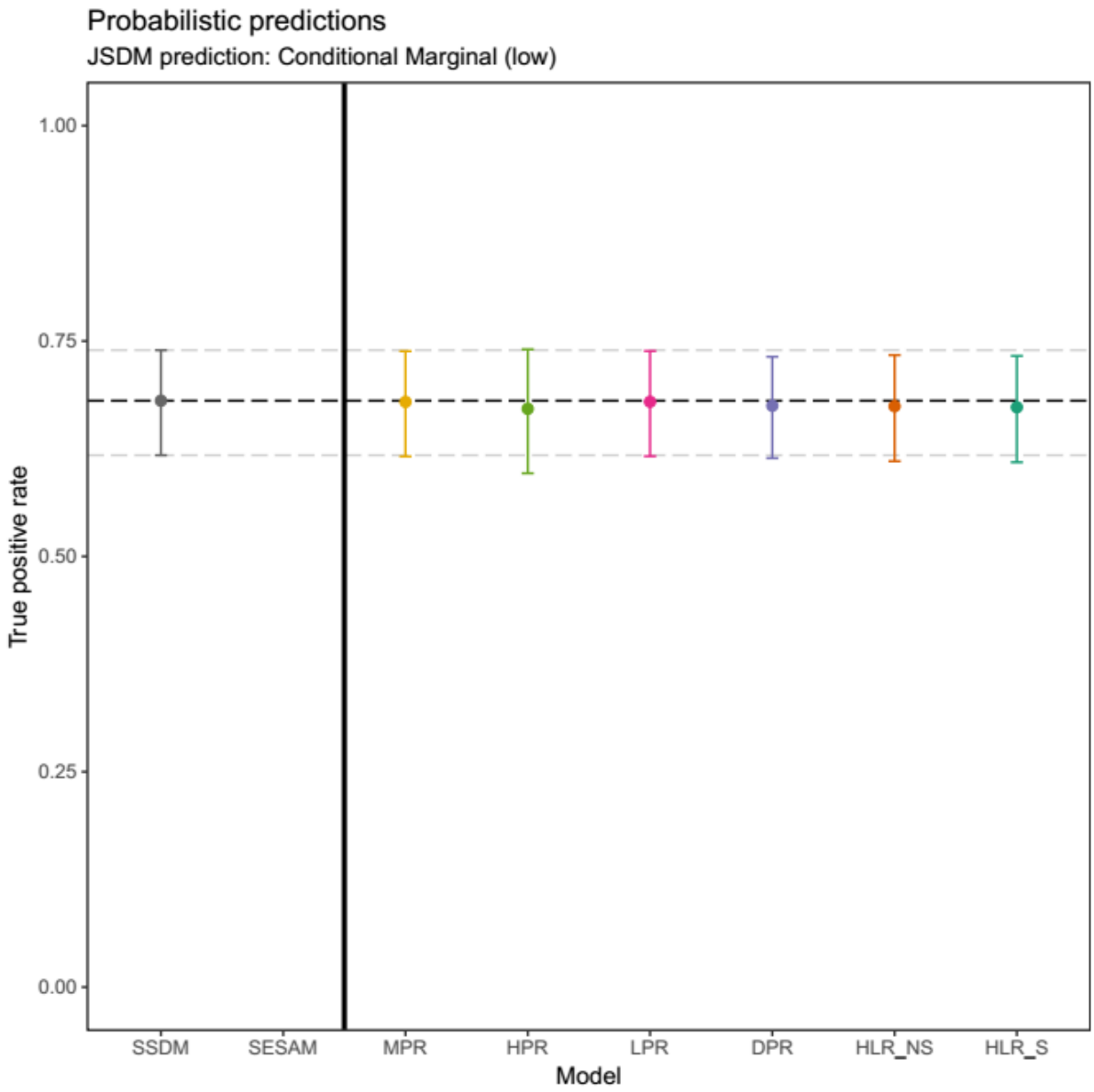


Figure 161: Model performance for the true positive rate evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

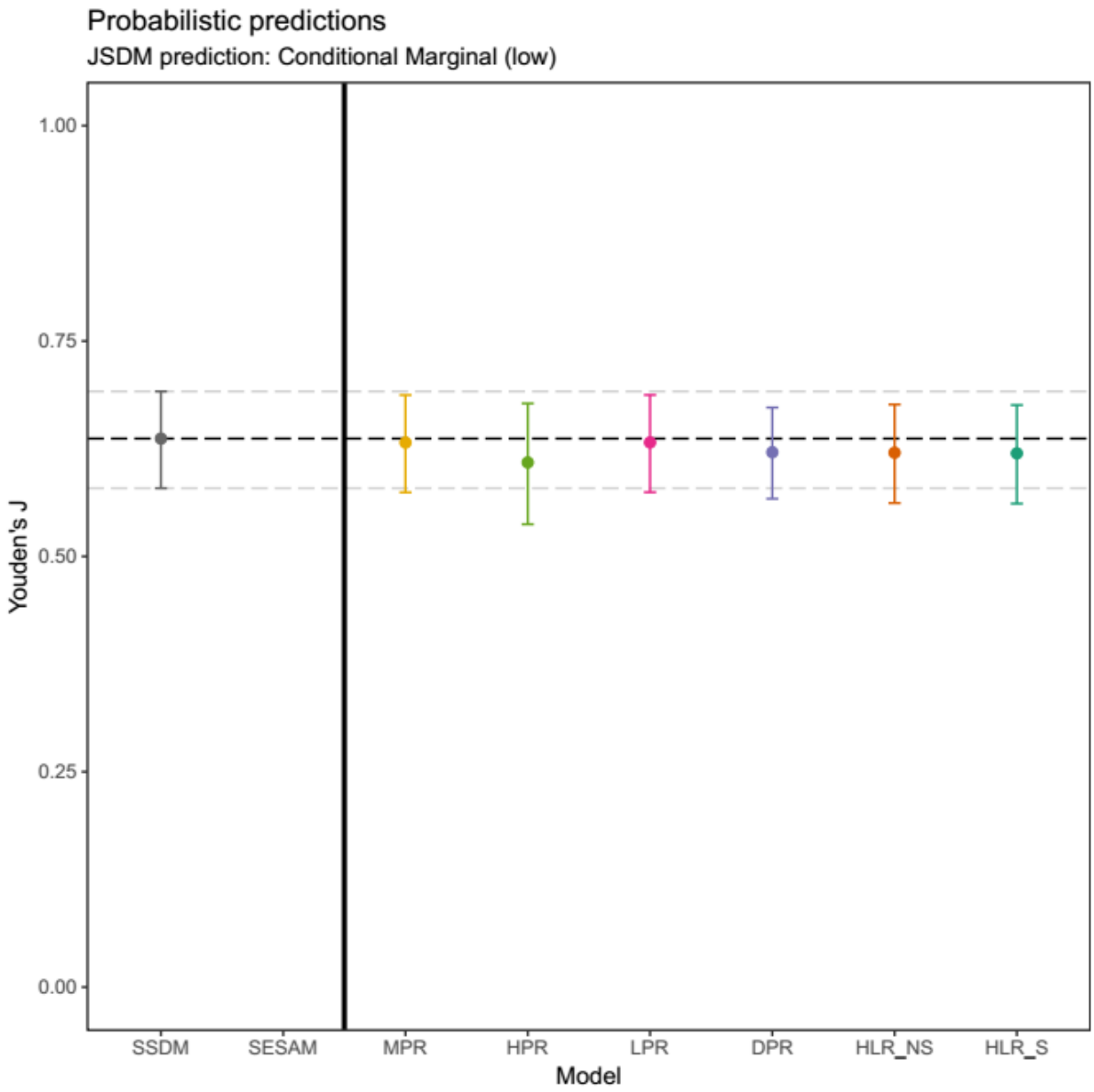


Figure 162: Model performance for the Youden's J evaluation metric for conditional marginal (low-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

8. Conditional marginal (medium-prevalence) predictions

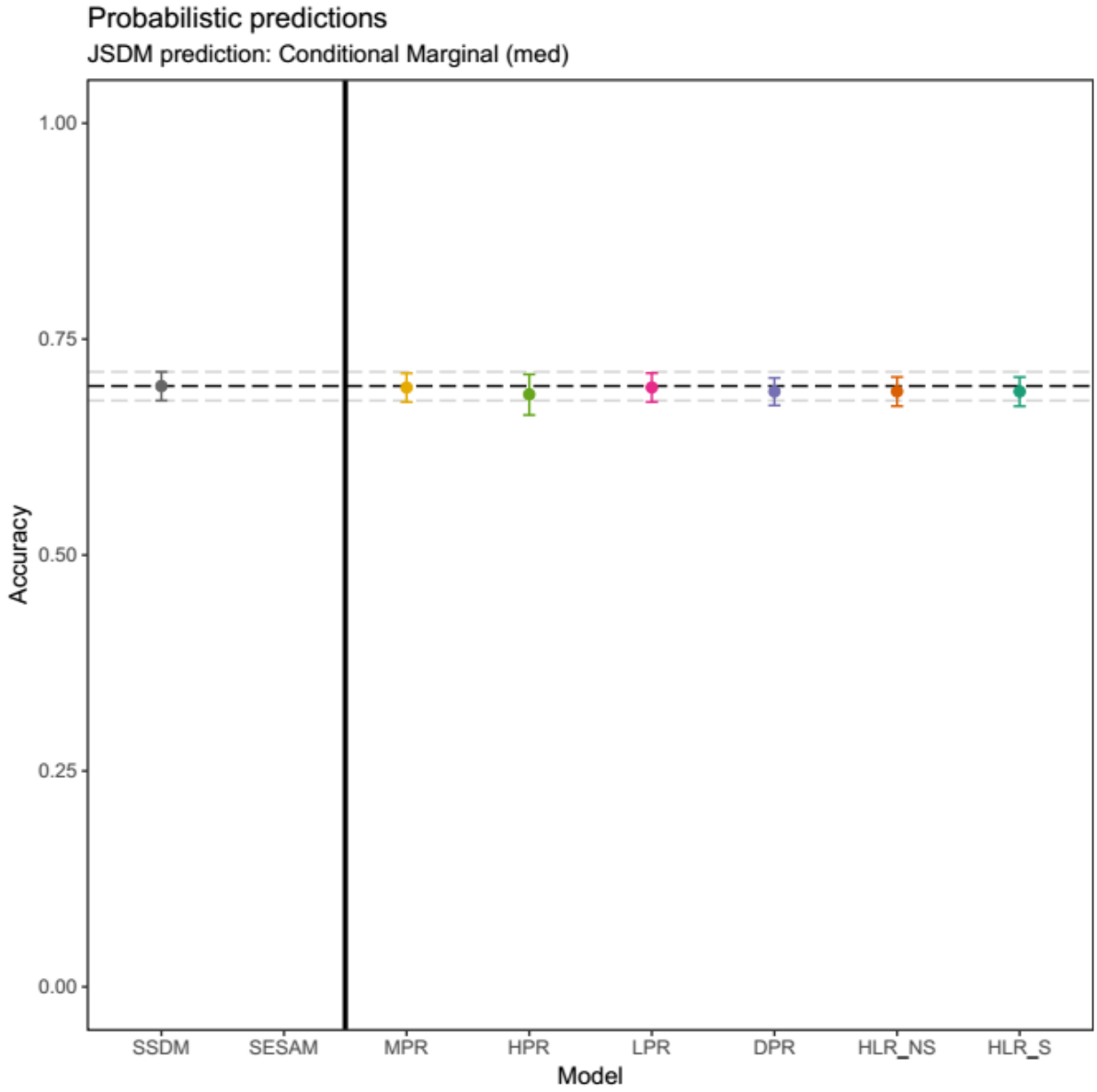


Figure 163: Model performance for the accuracy evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

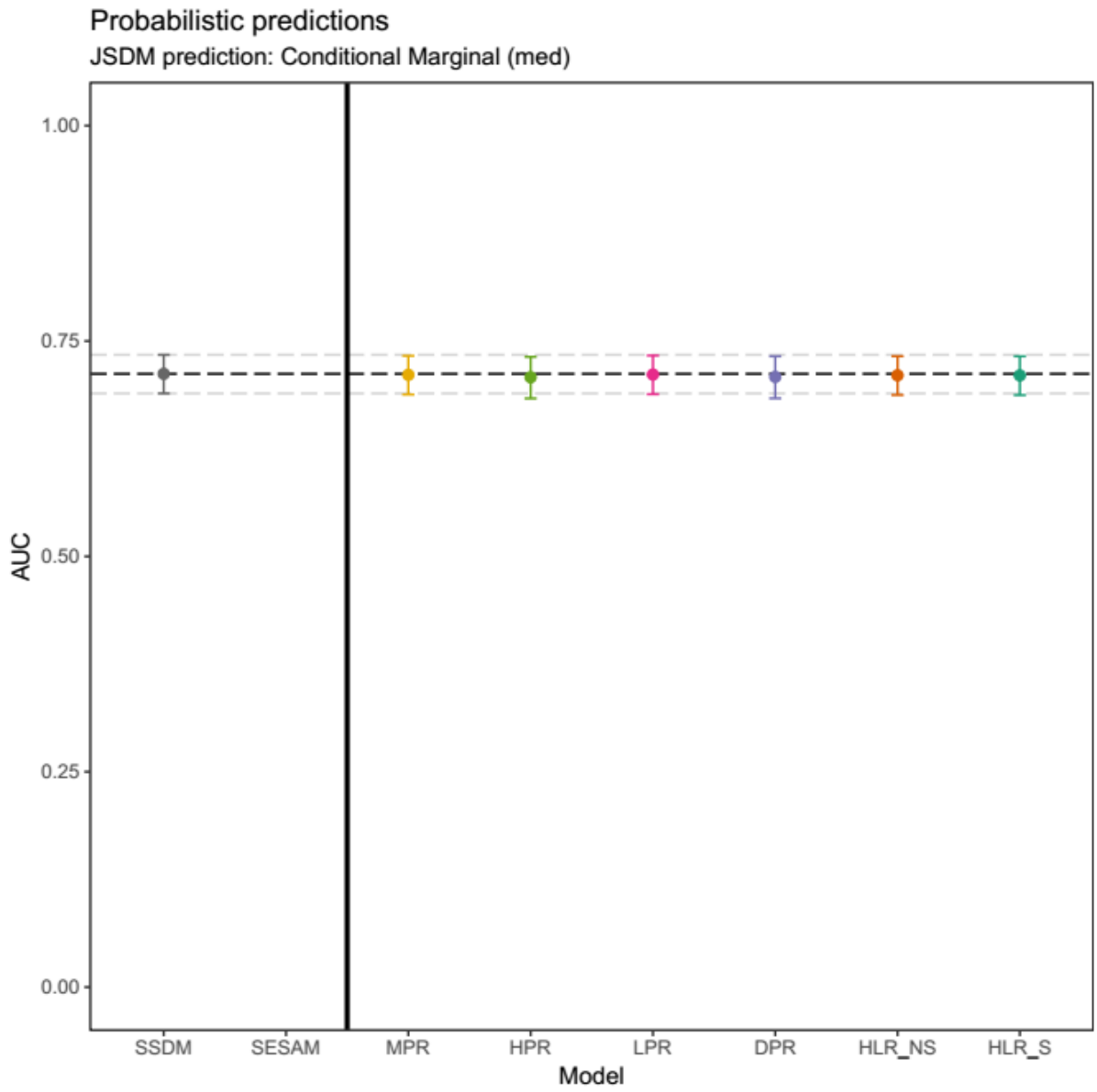


Figure 164: Model performance for the AUC evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

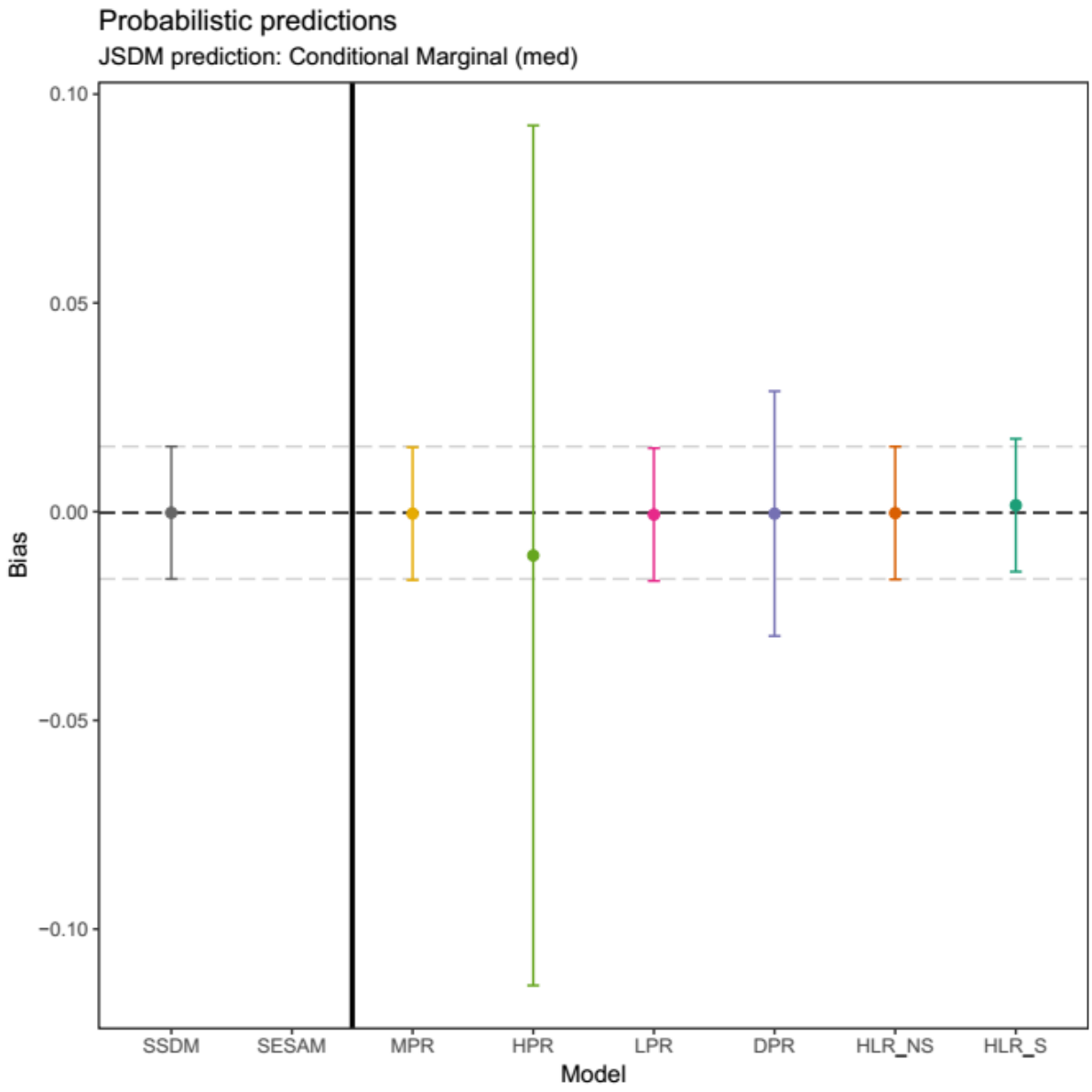


Figure 165: Model performance for the bias (mean error) evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

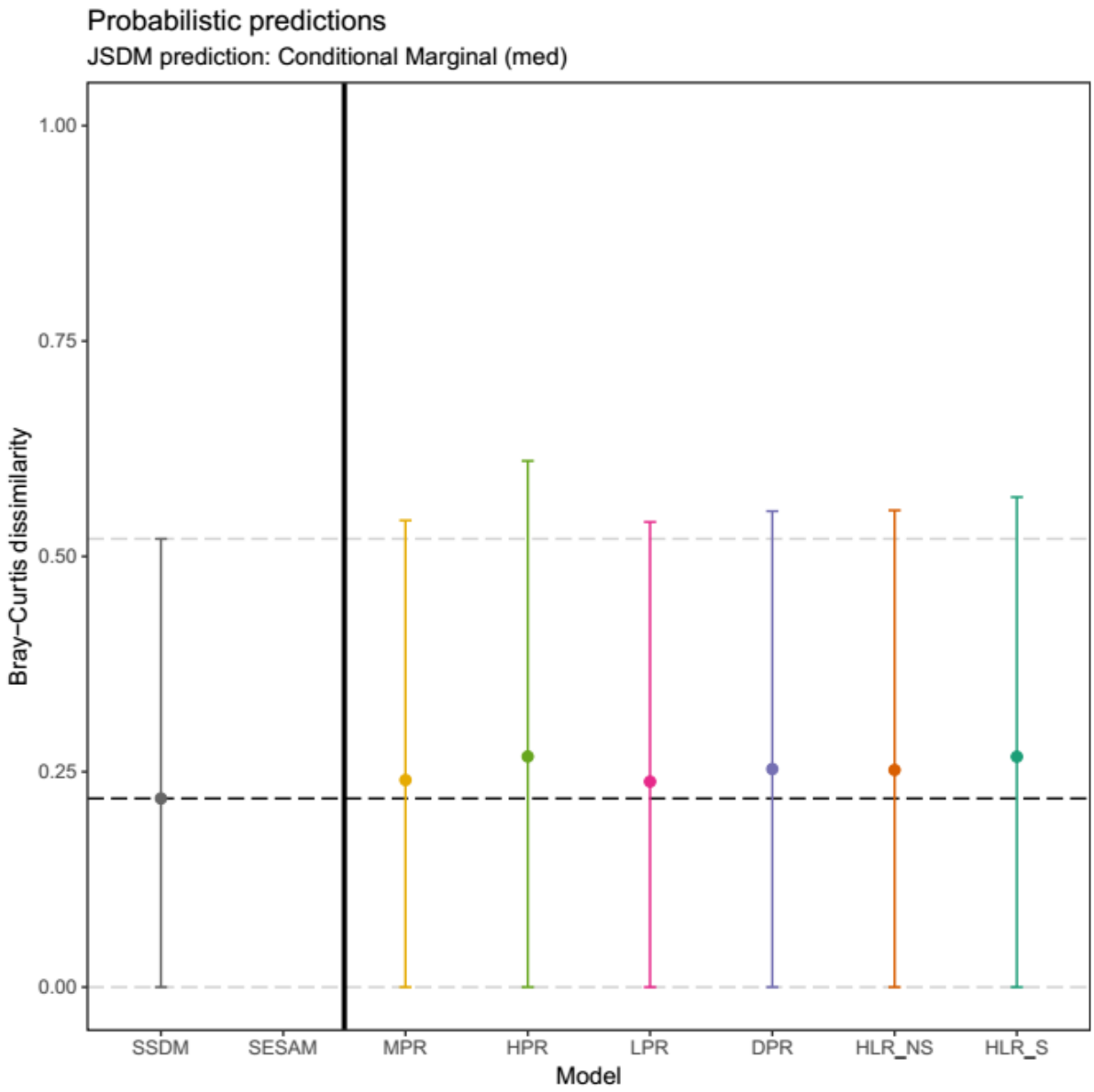


Figure 166: Model performance for the Bray-Curtis dissimilarity evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

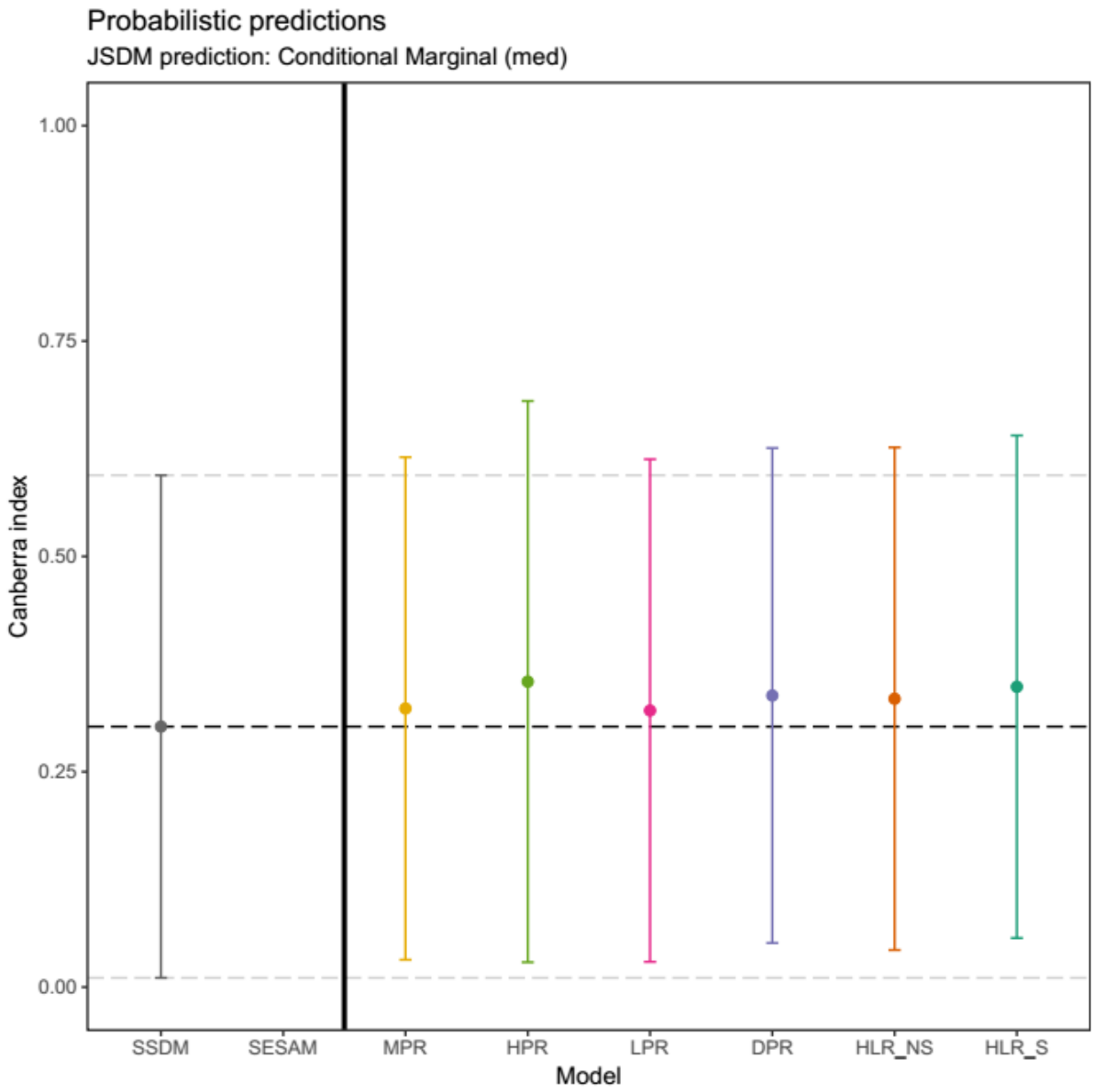


Figure 167: Model performance for the Canberra index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

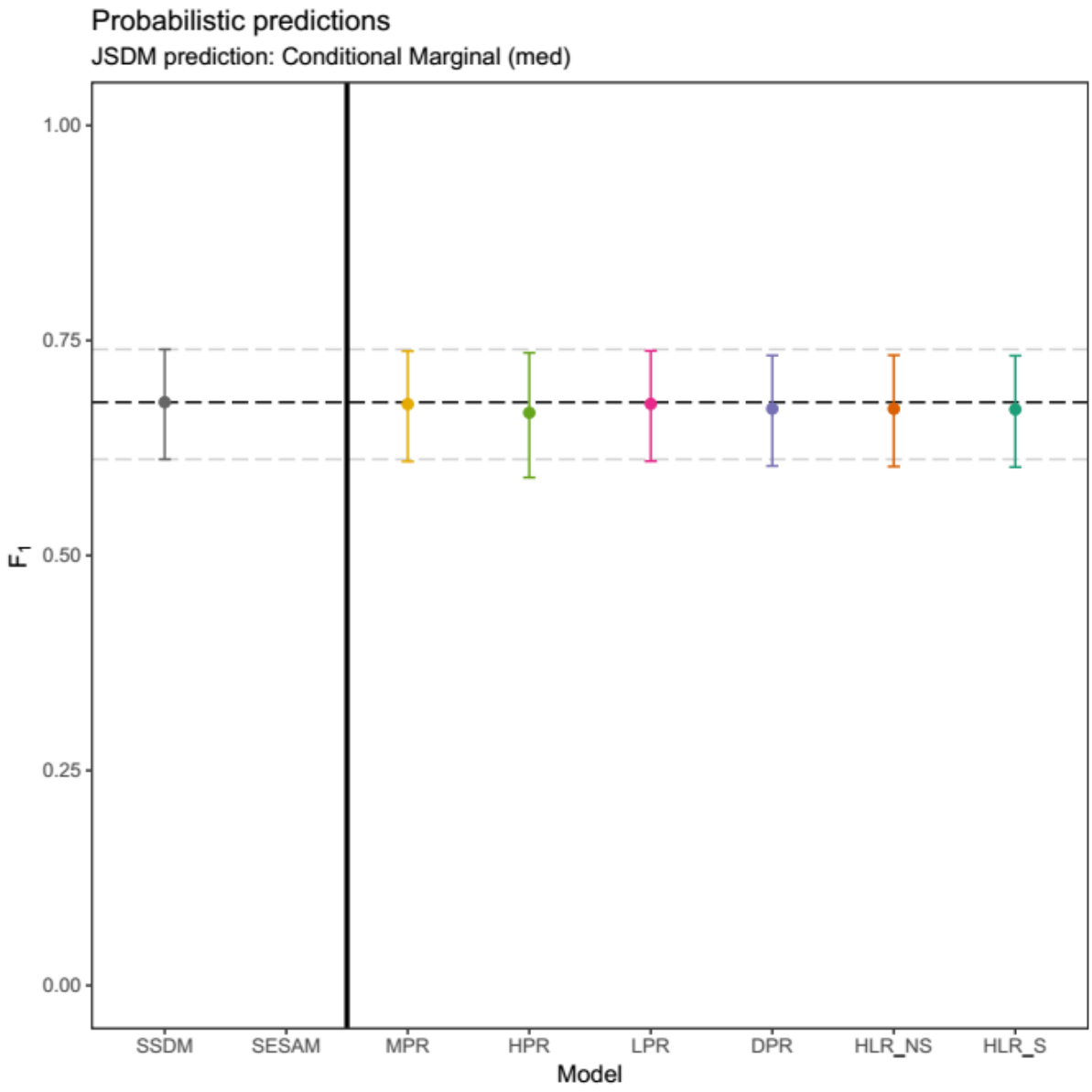


Figure 168: Model performance for the F_1 evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

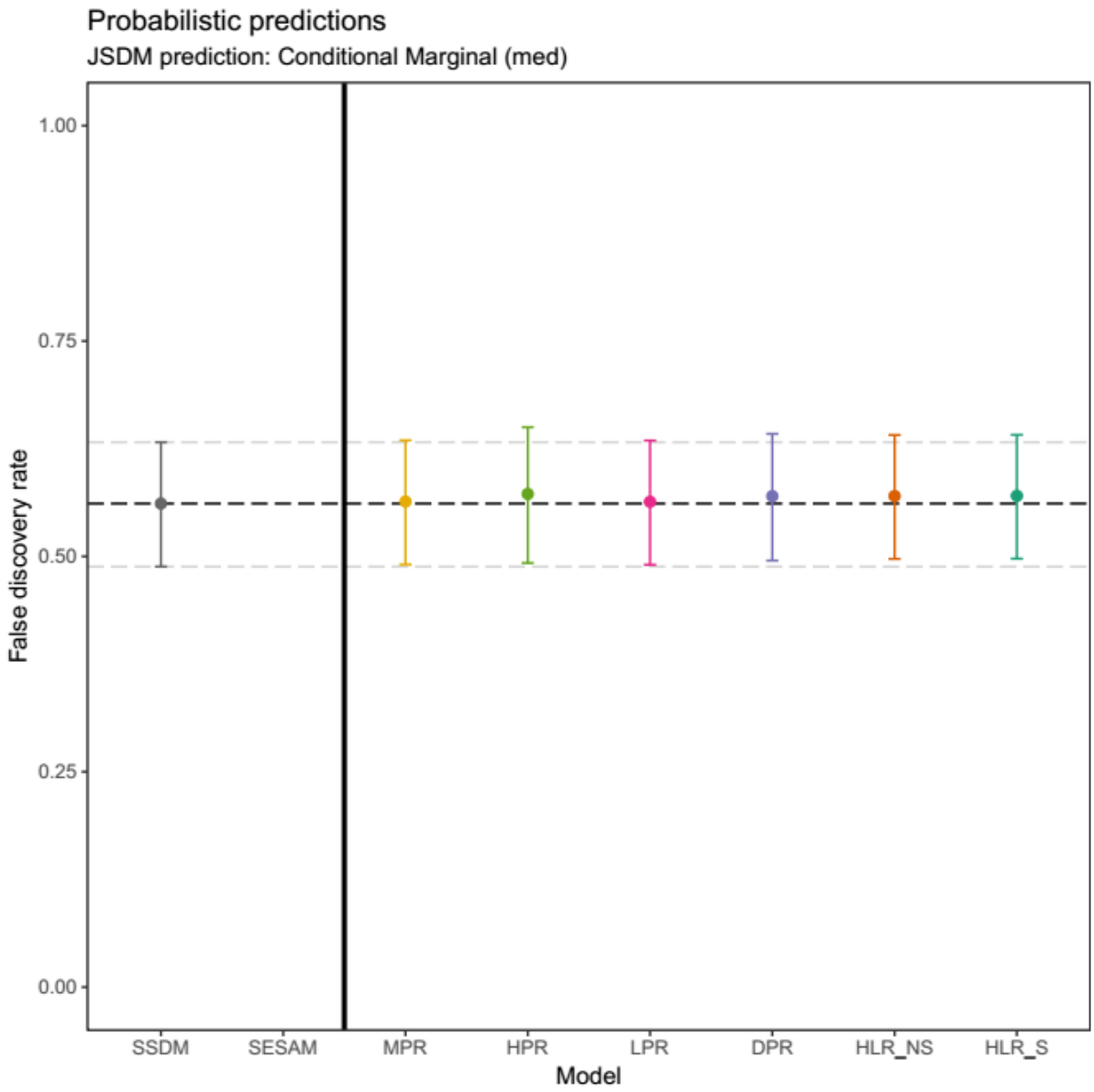


Figure 169: Model performance for the false discovery rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

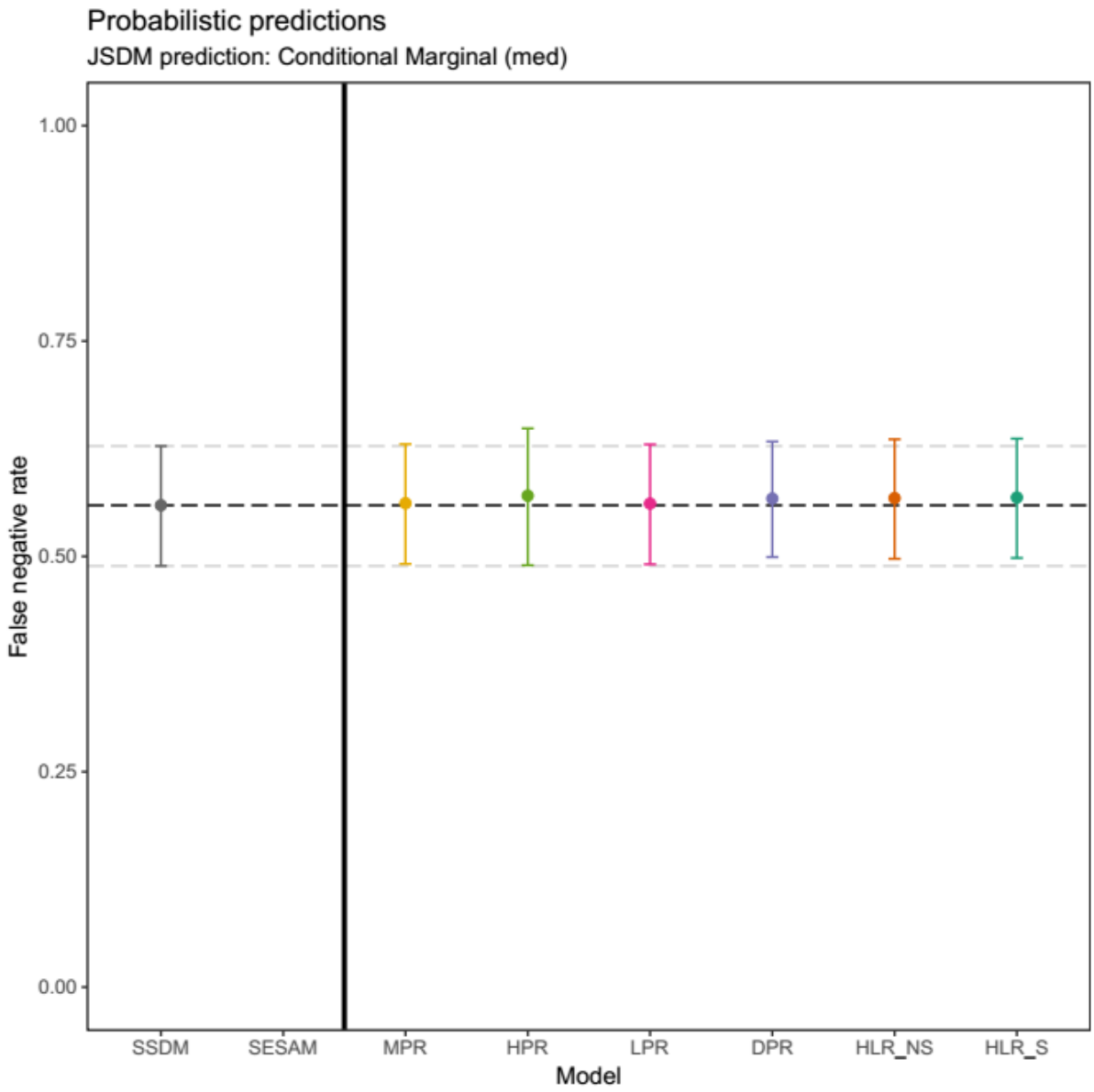


Figure 170: Model performance for the false negative rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

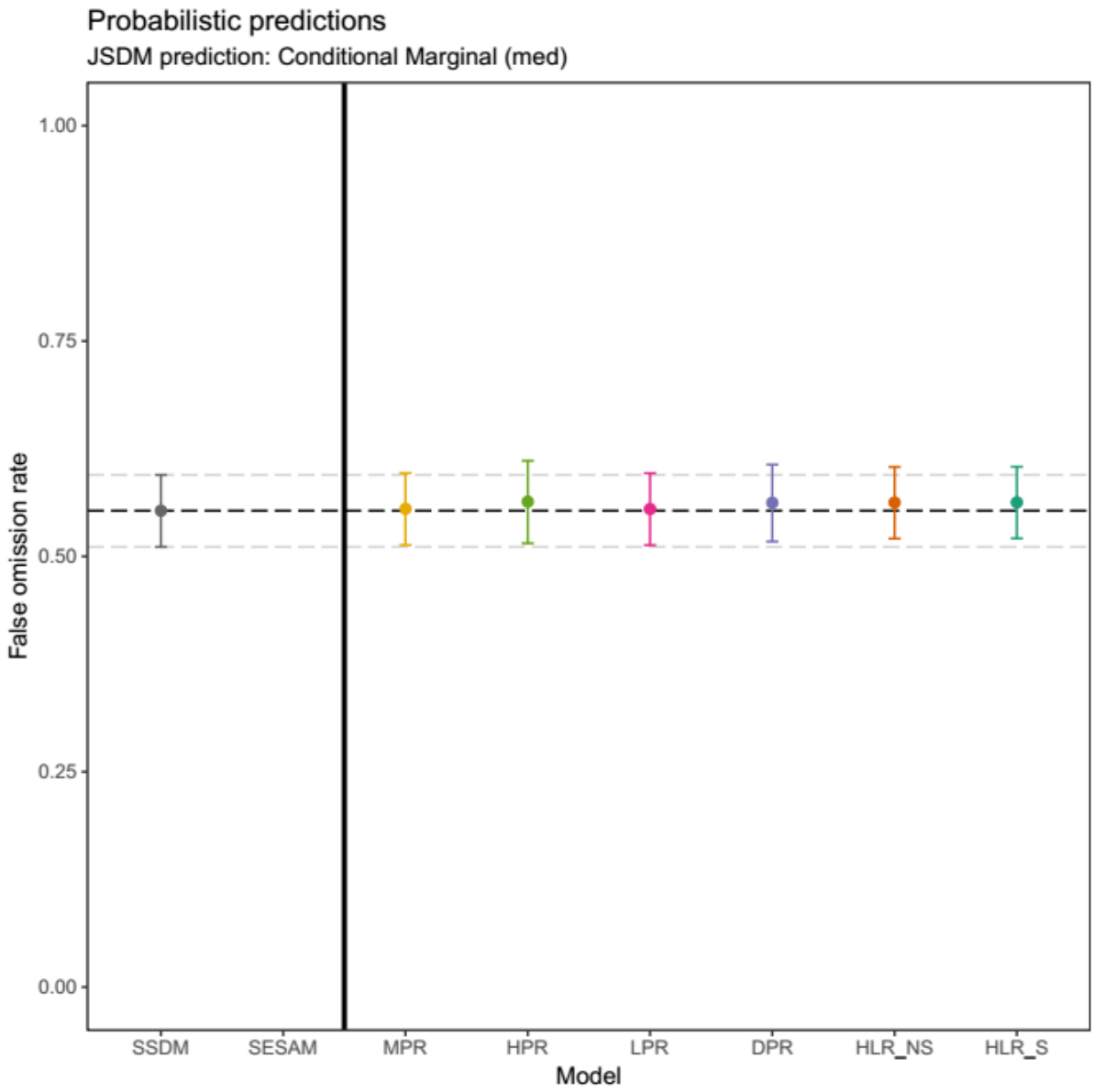


Figure 171: Model performance for the false omission rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

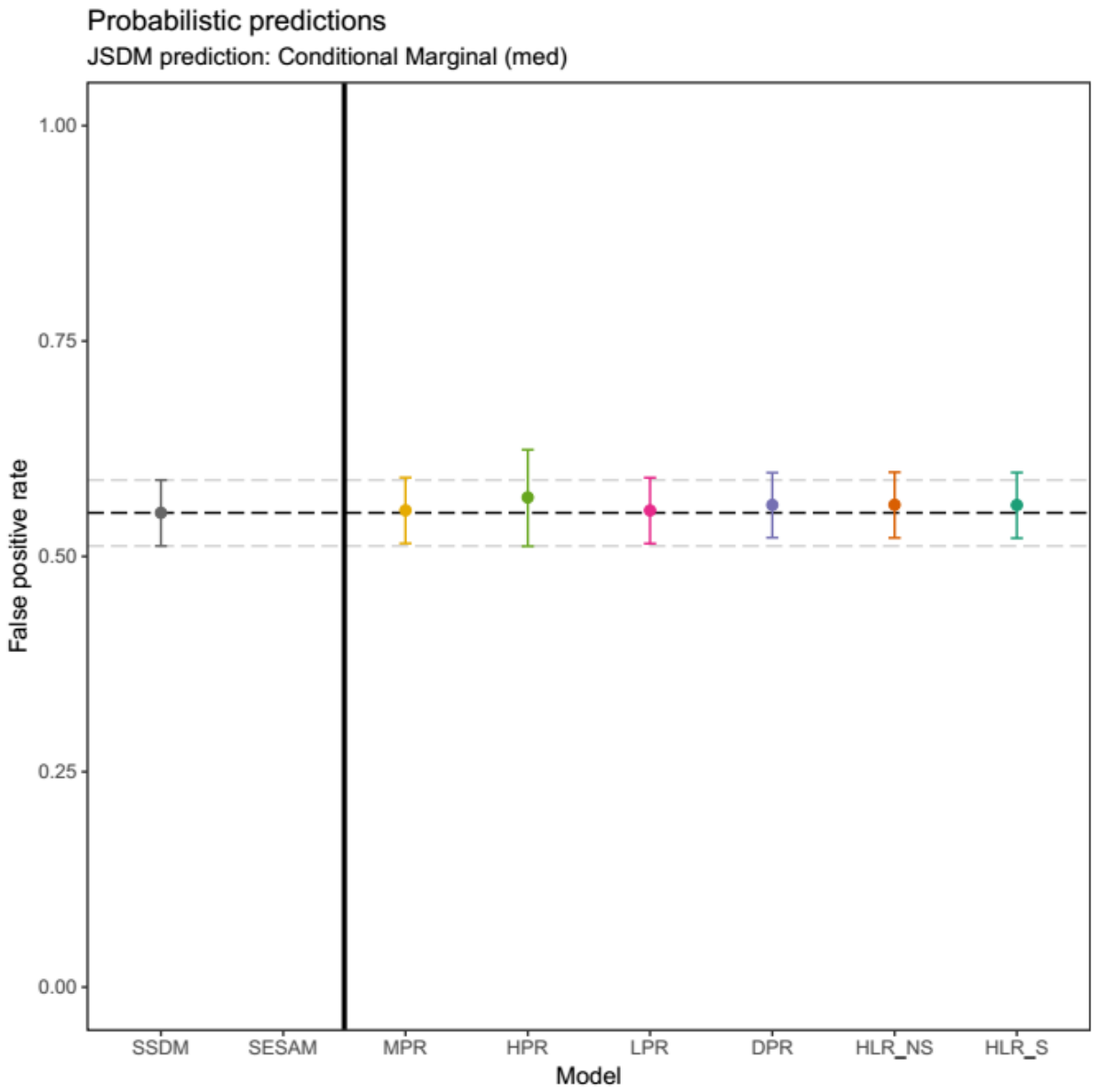


Figure 172: Model performance for the false positive rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

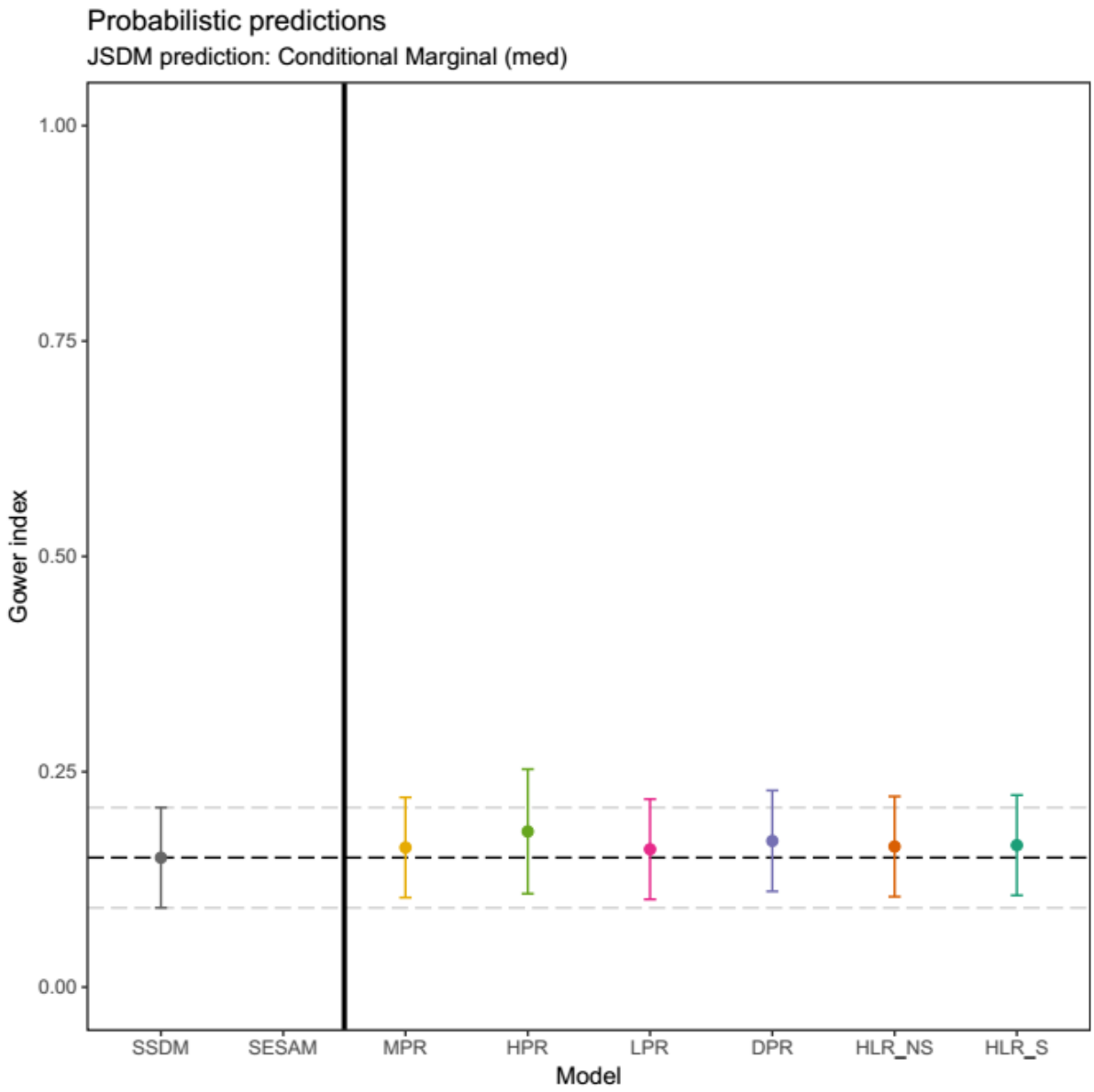


Figure 173: Model performance for the Gower index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

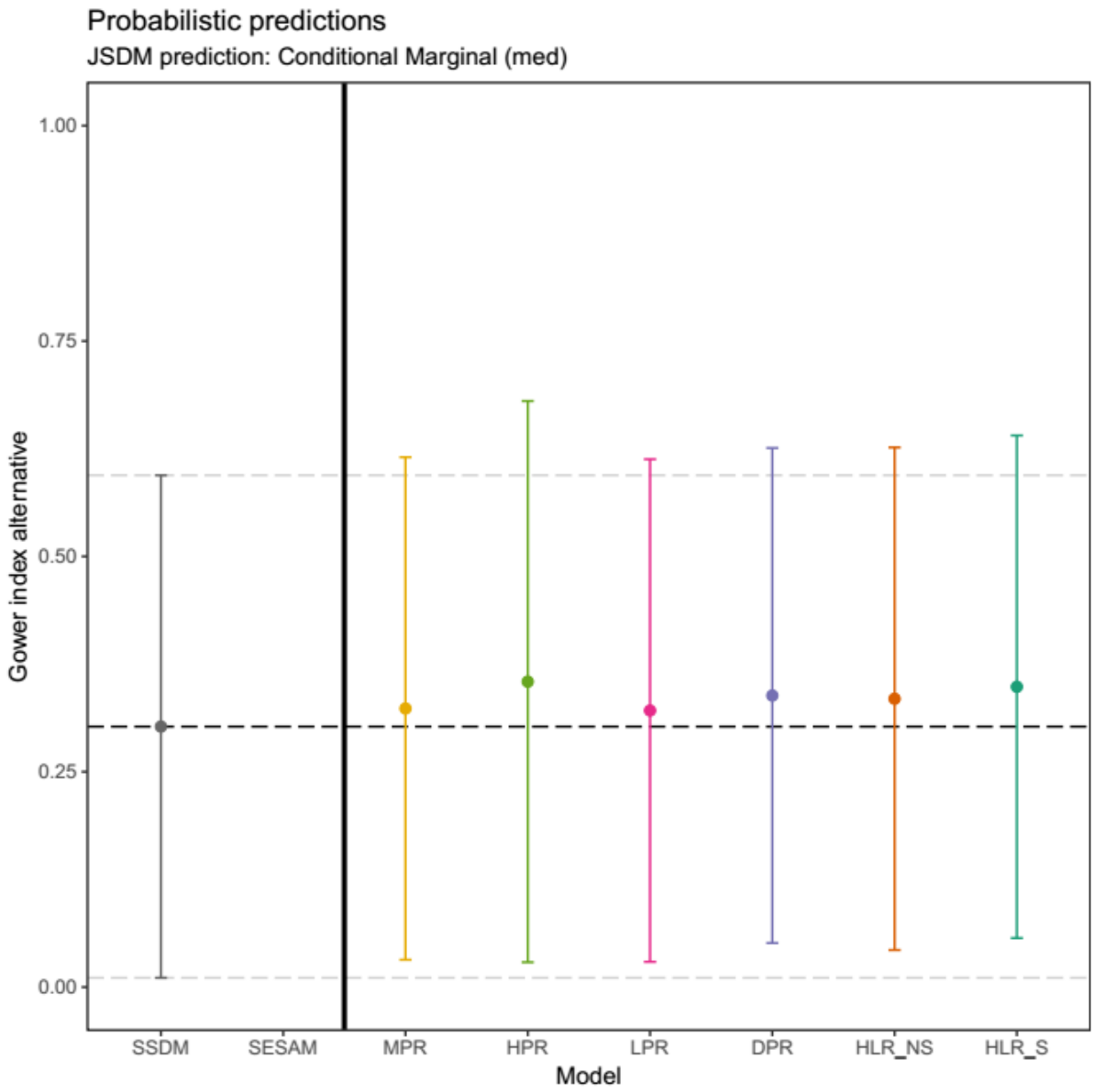


Figure 174: Model performance for the alternative Gower index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

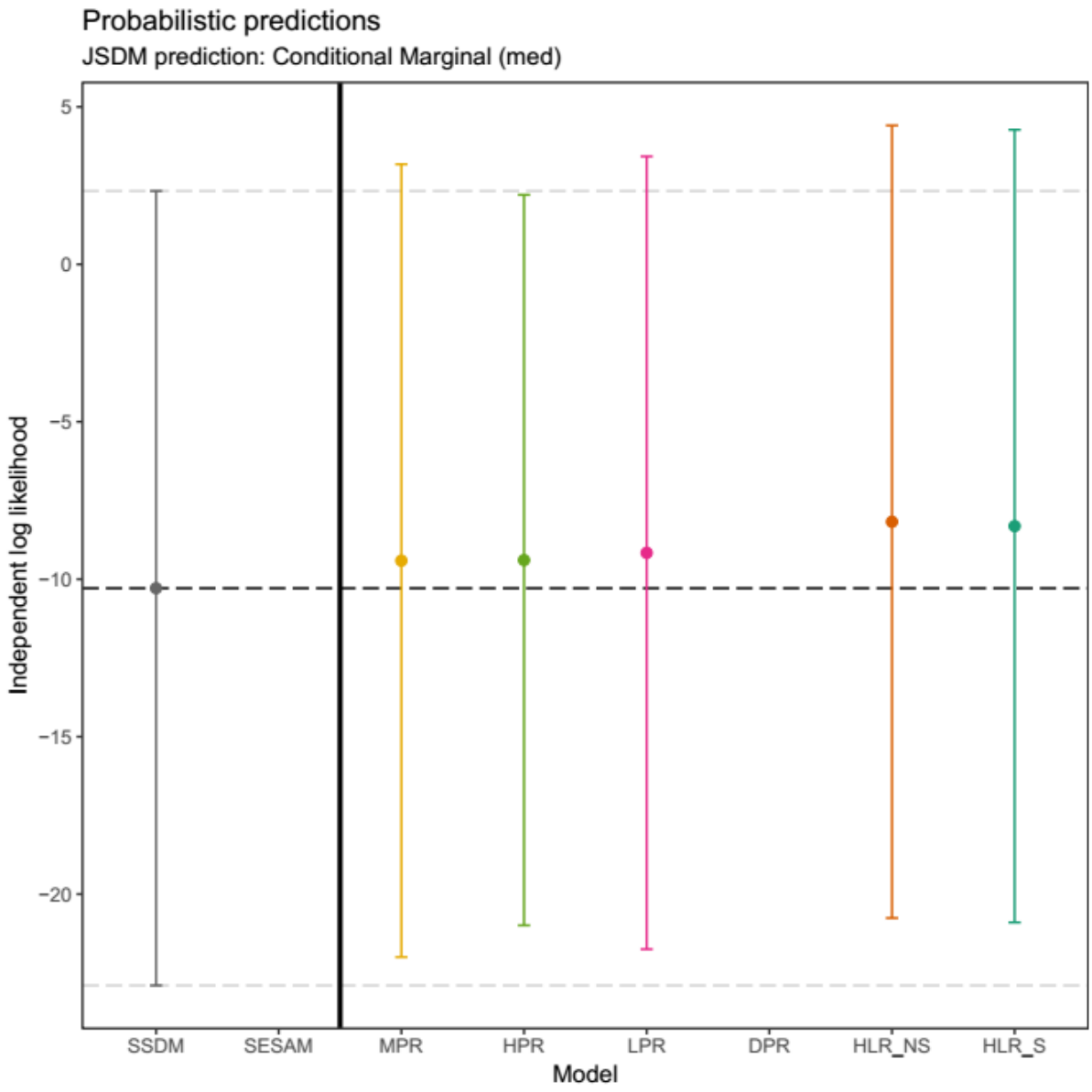


Figure 175: Model performance for the independent log-likelihood evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

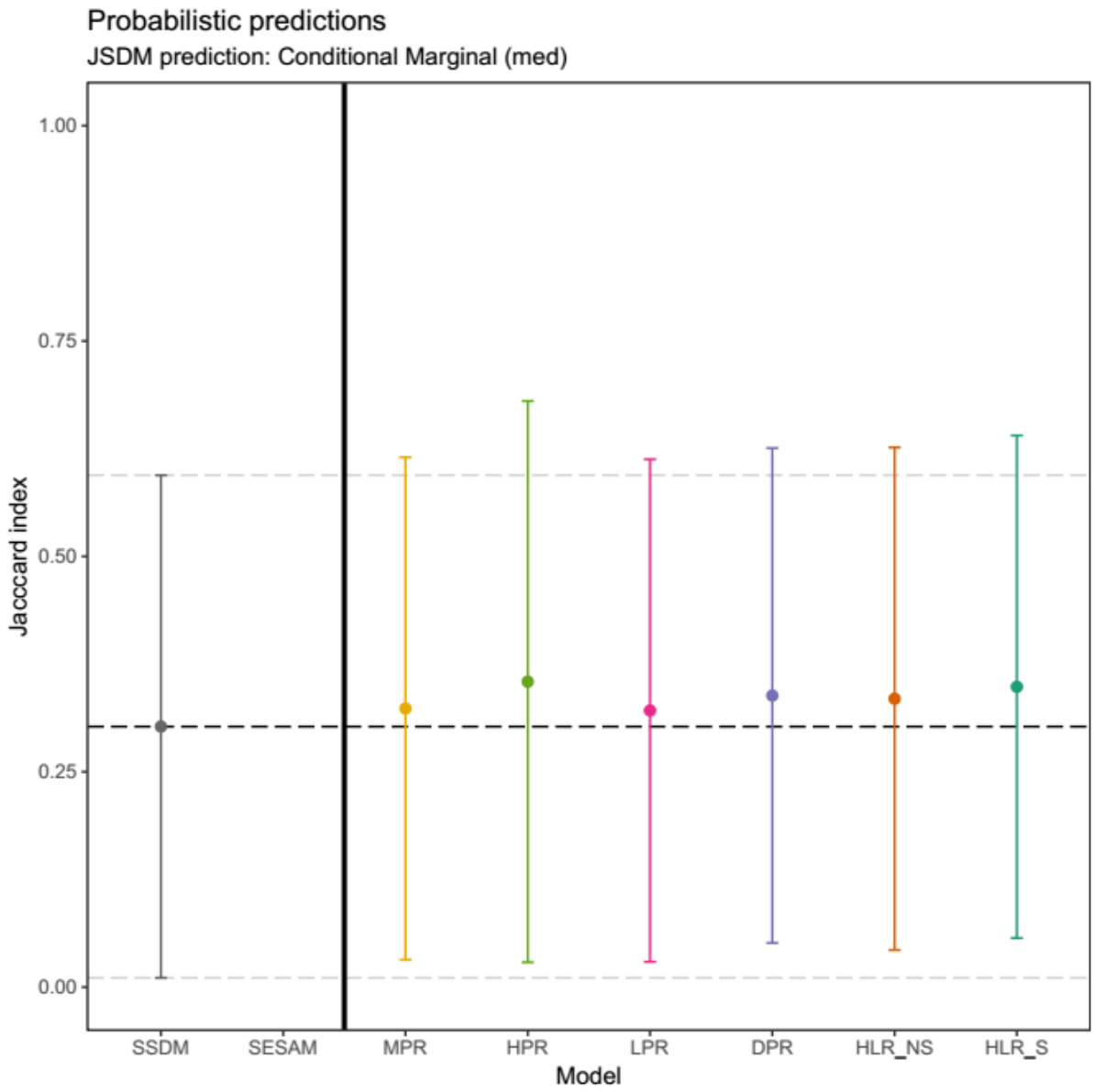


Figure 176: Model performance for the Jaccard index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

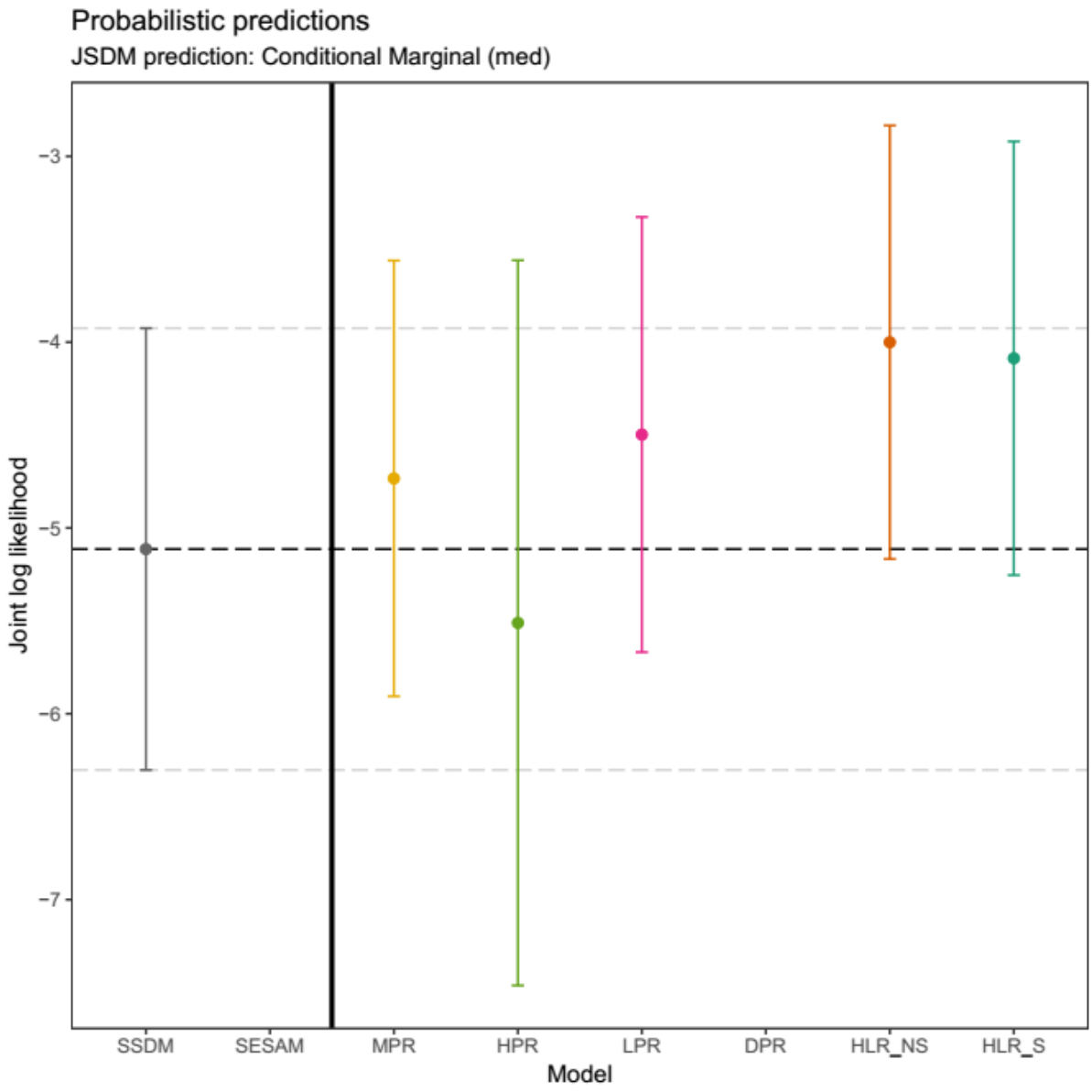


Figure 177: Model performance for the joint log-likelihood evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

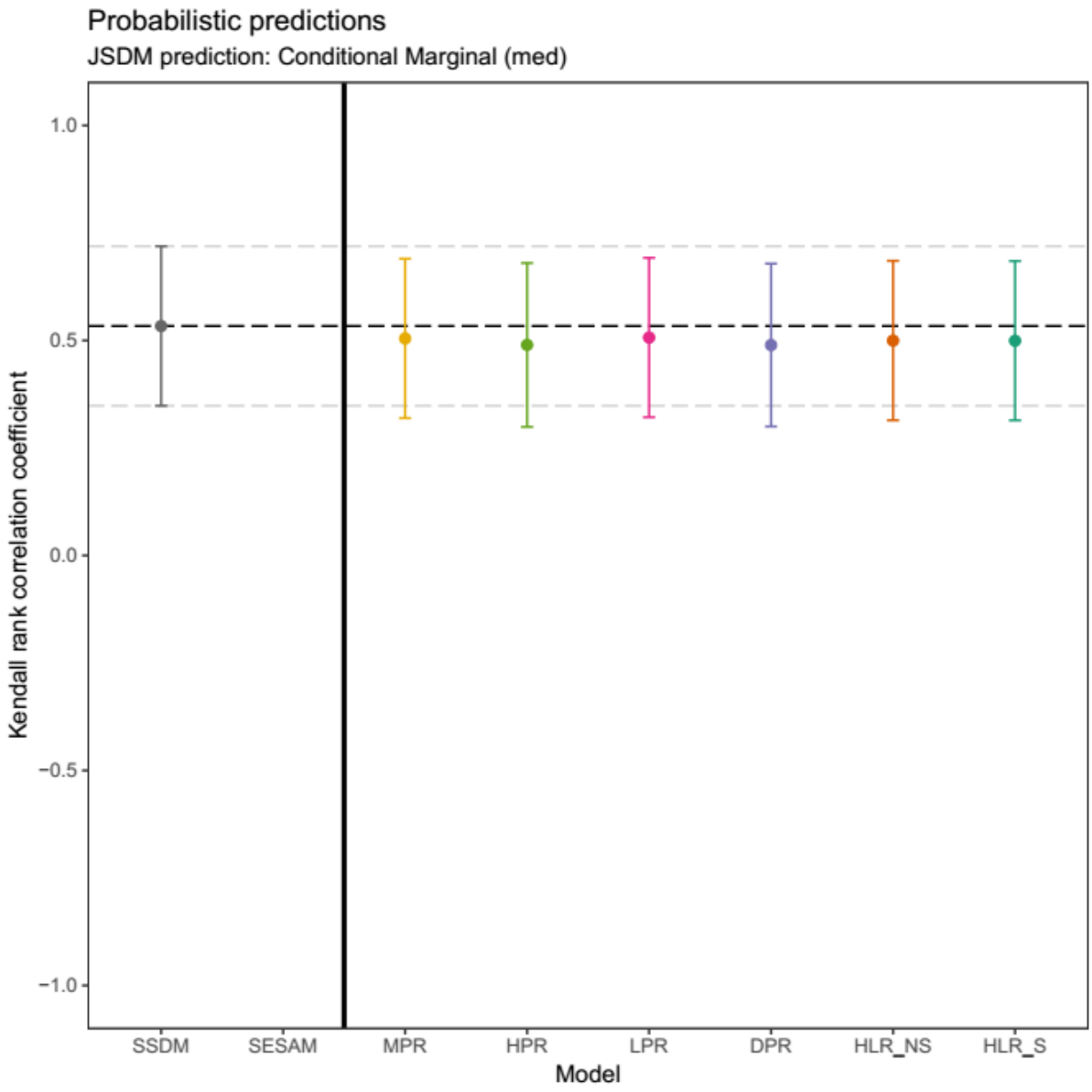


Figure 178: Model performance for the Kendall rank correlation coefficient evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

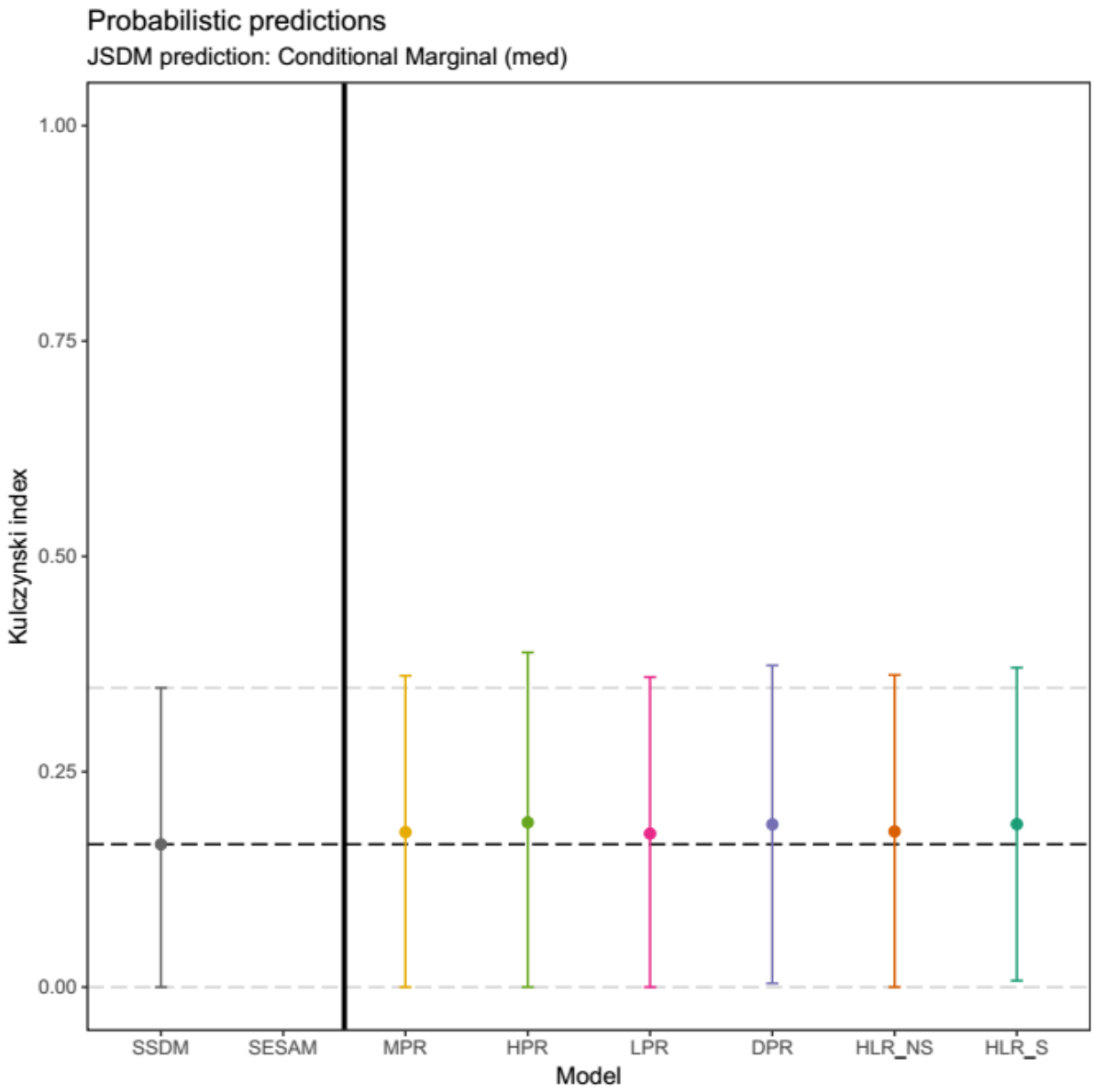


Figure 179: Model performance for the Kulczynski index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

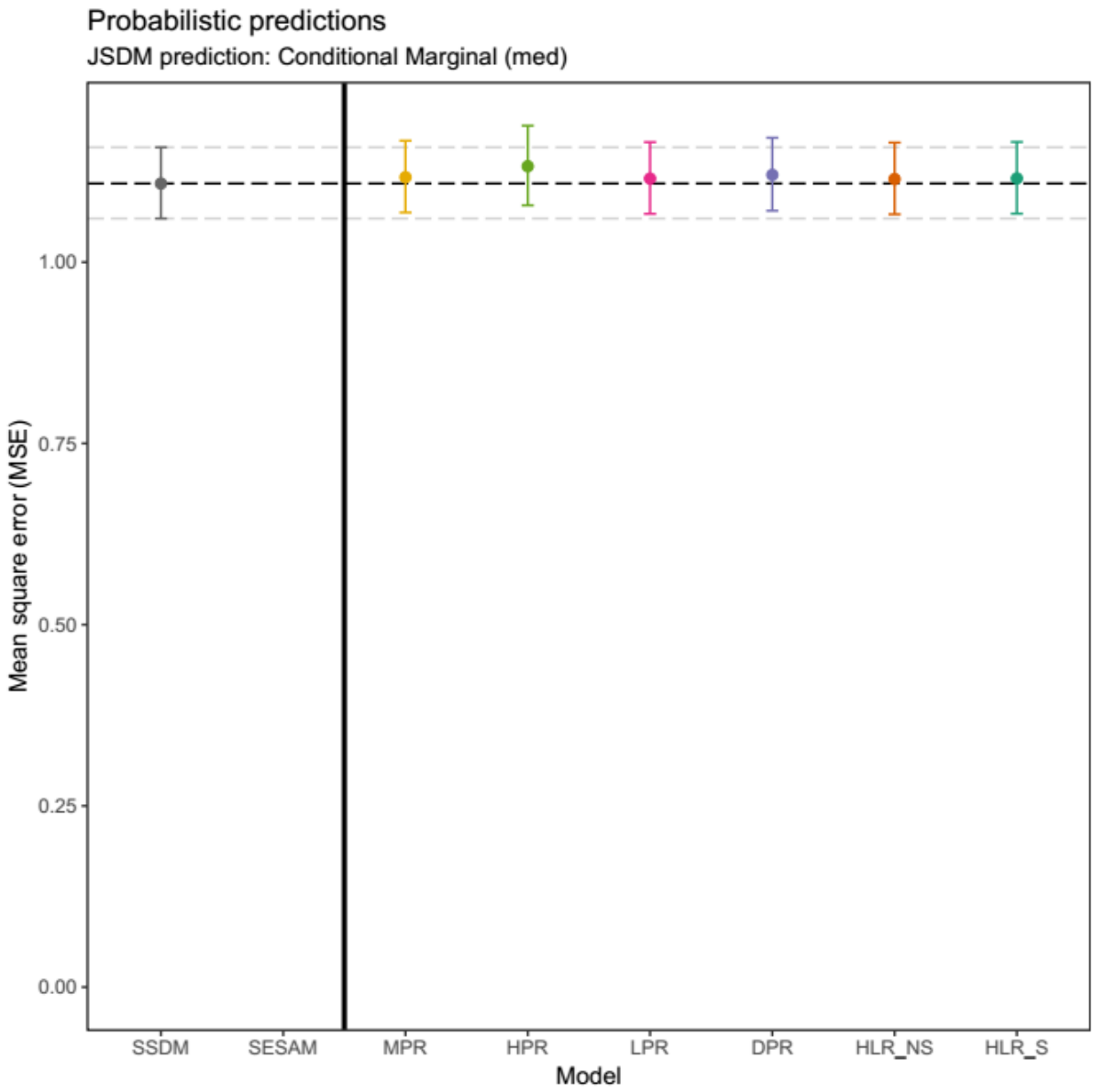


Figure 180: Model performance for the mean square error evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

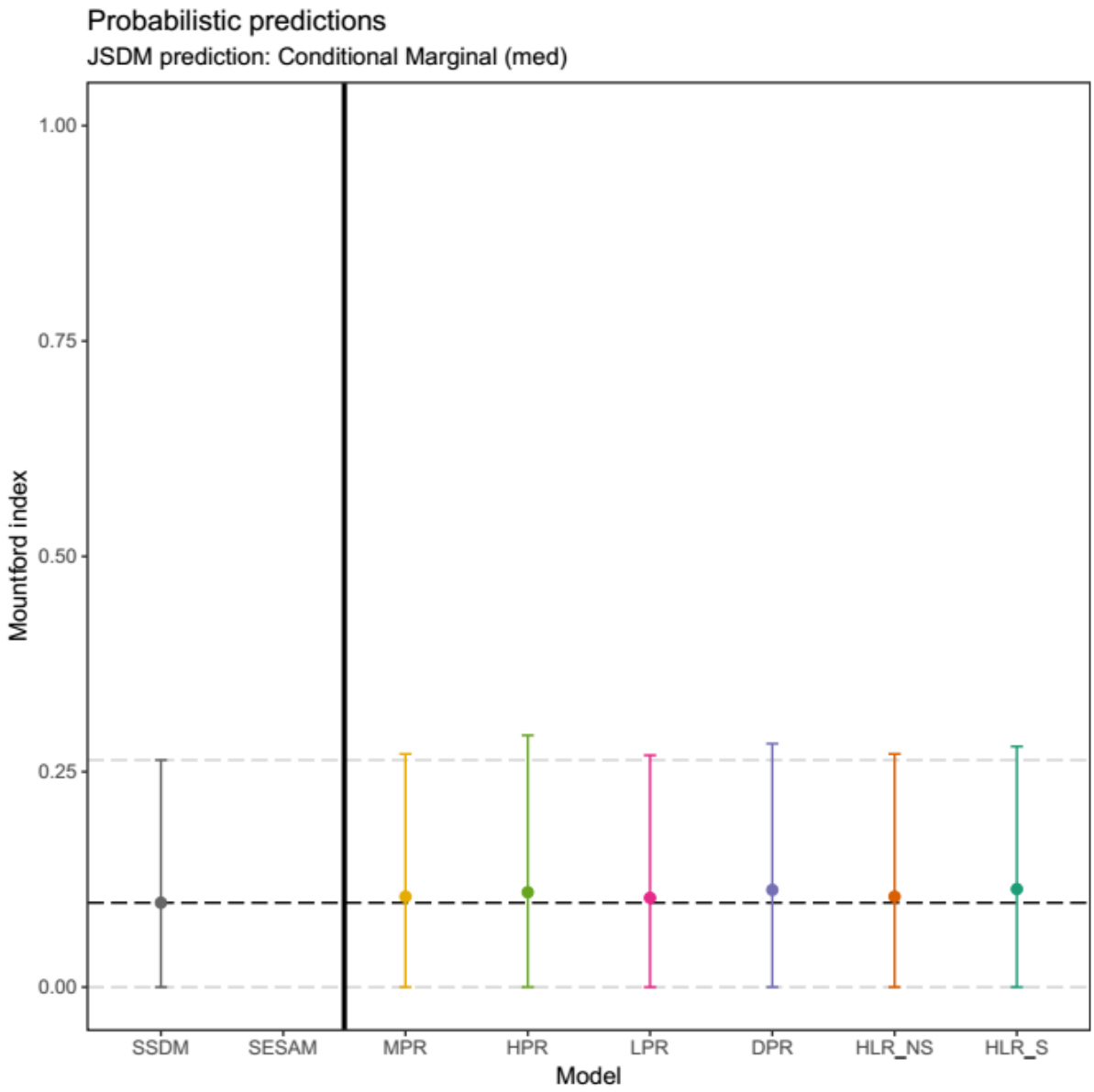


Figure 181: Model performance for the Mountford index evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

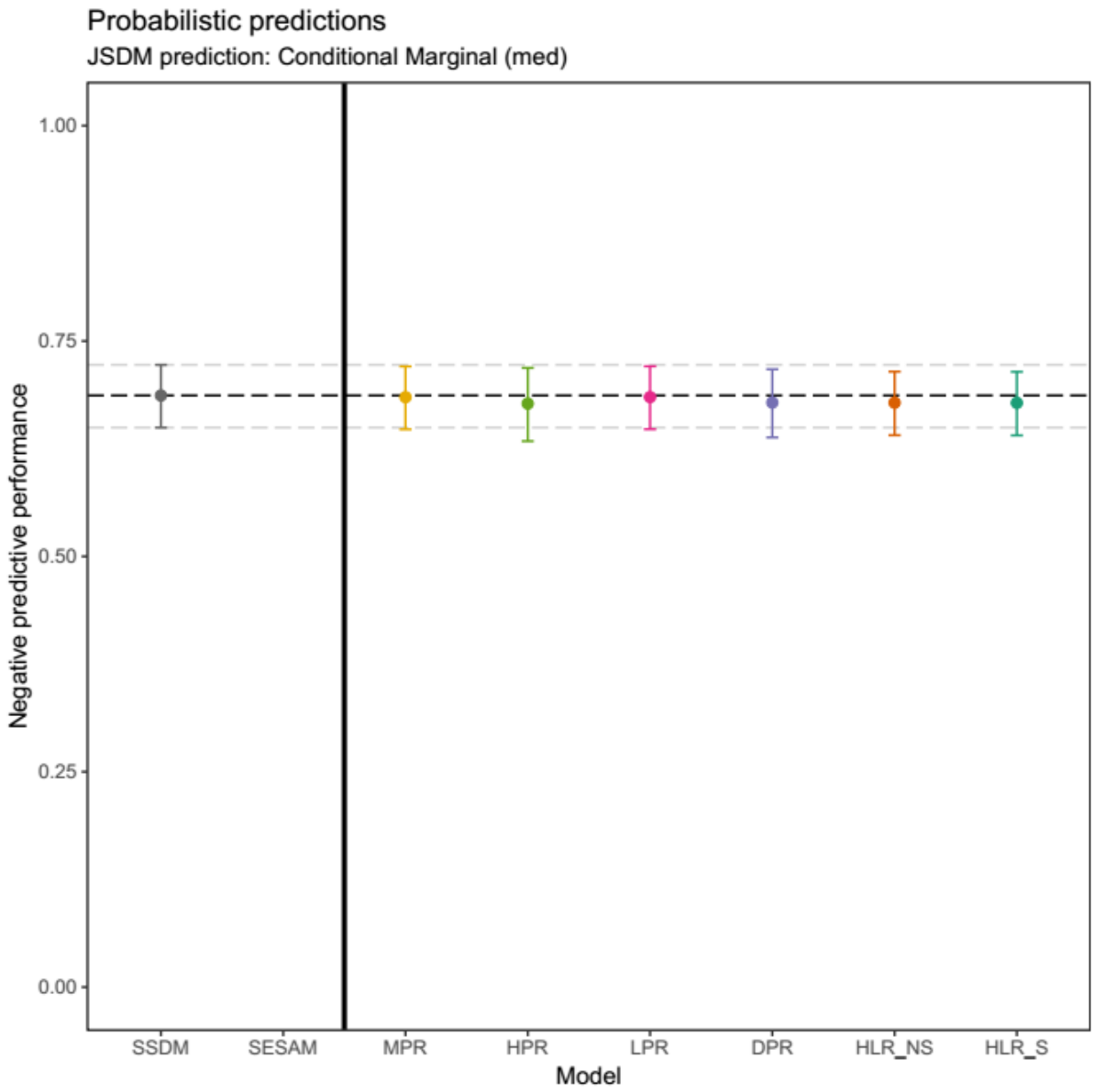


Figure 182: Model performance for the negative predictive performance evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

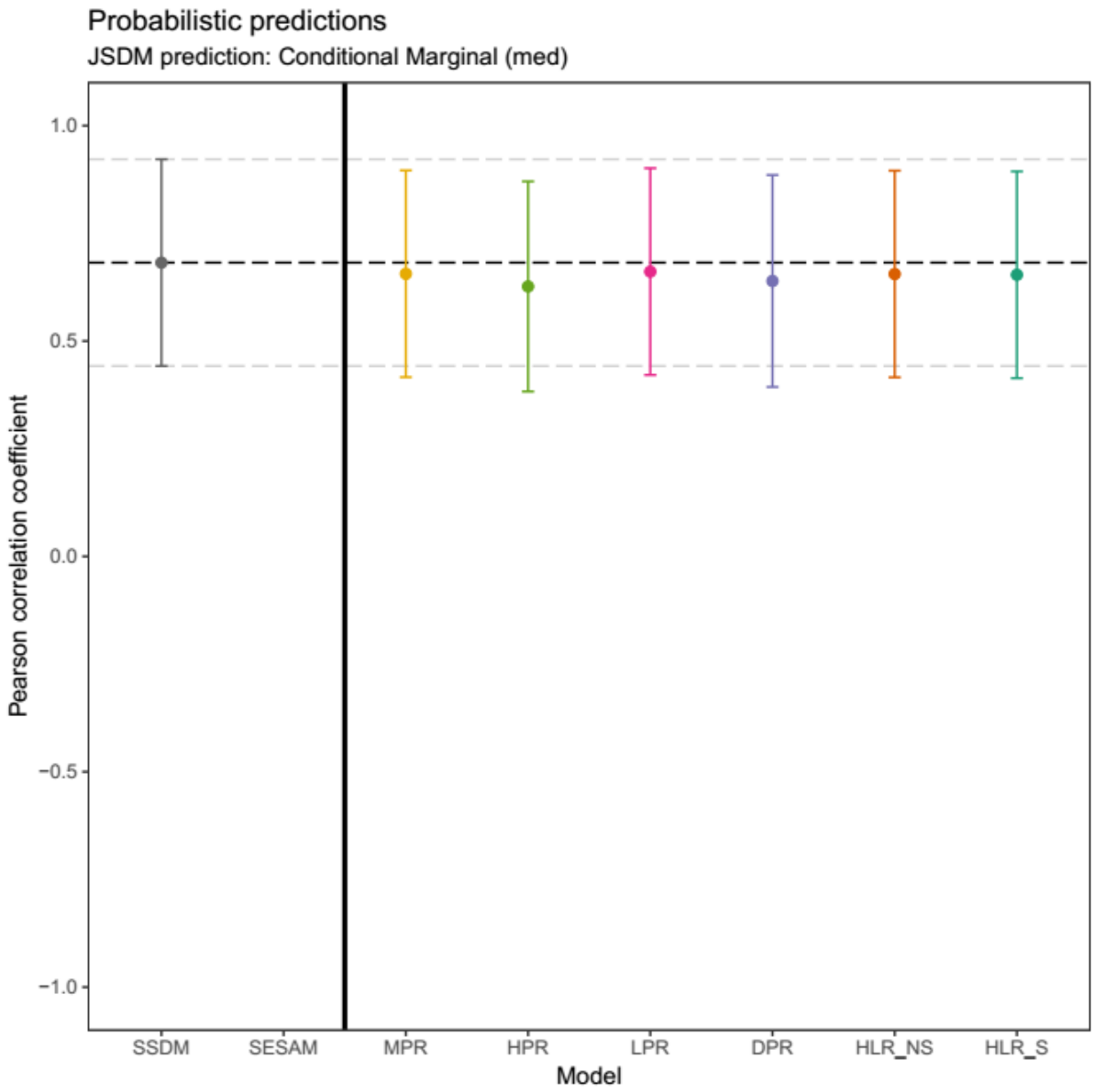


Figure 183: Model performance for the Pearson correlation coefficient evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

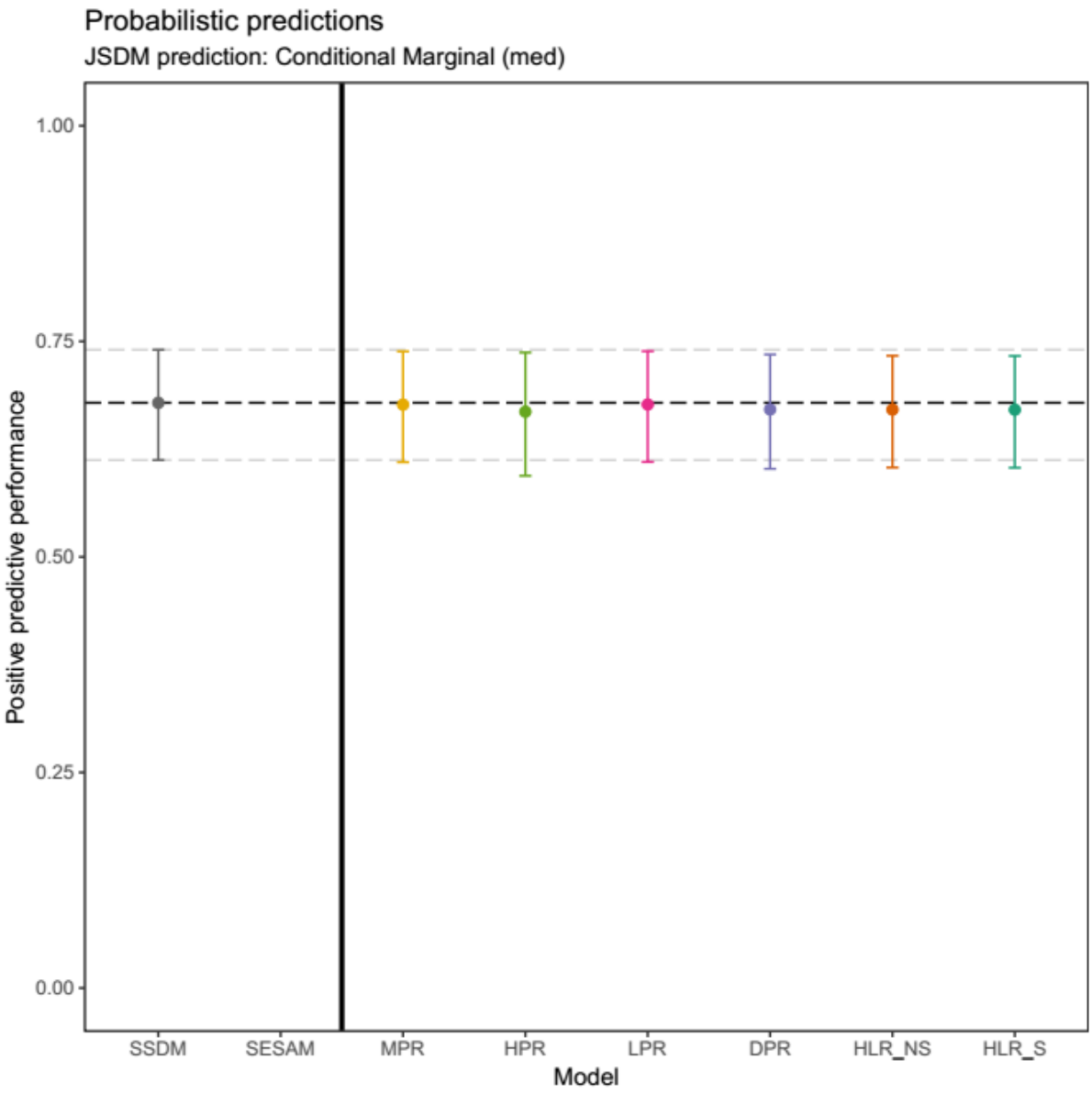


Figure 184: Model performance for the positive predictive performance evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

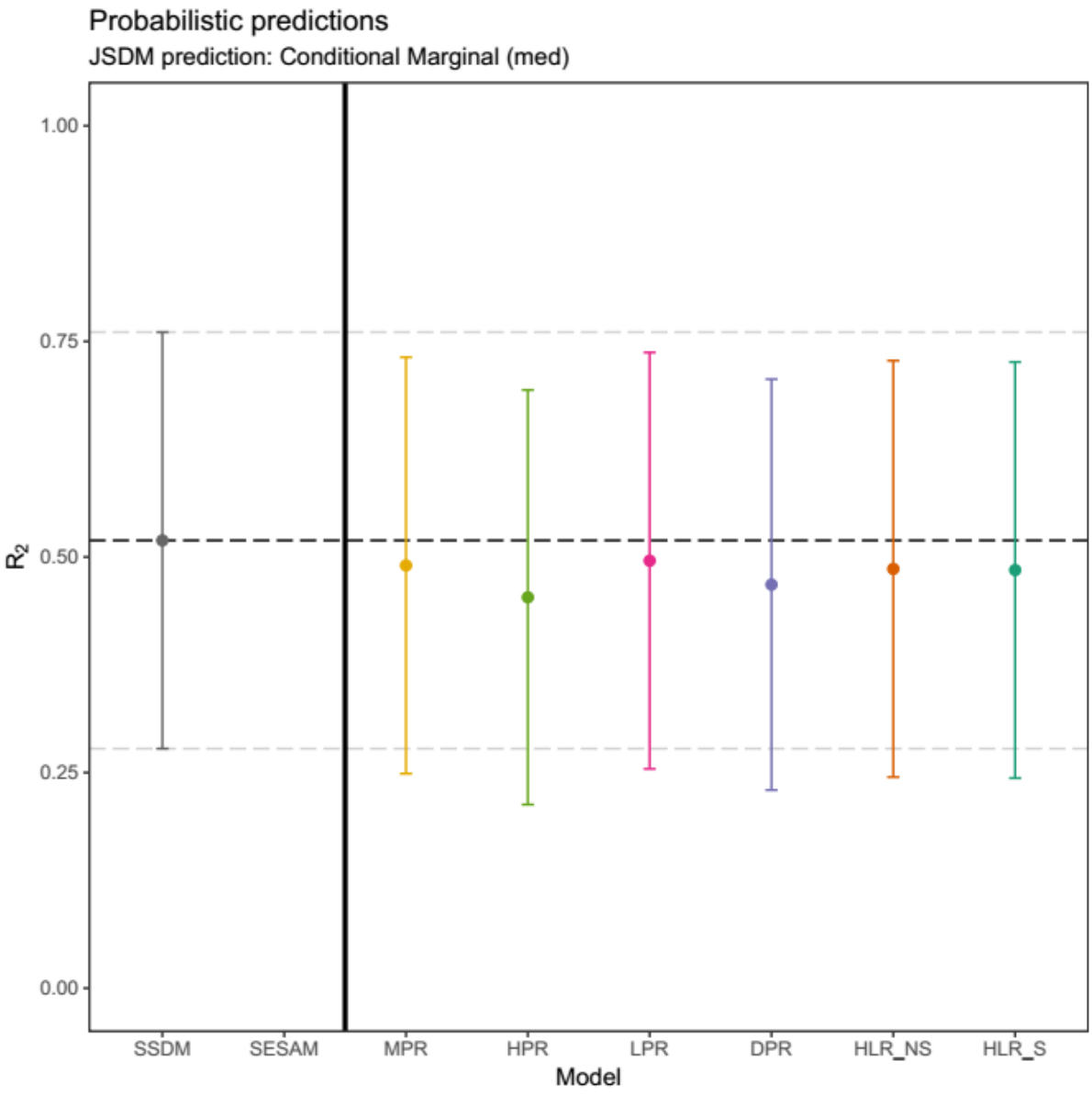


Figure 185: Model performance for the R^2 evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

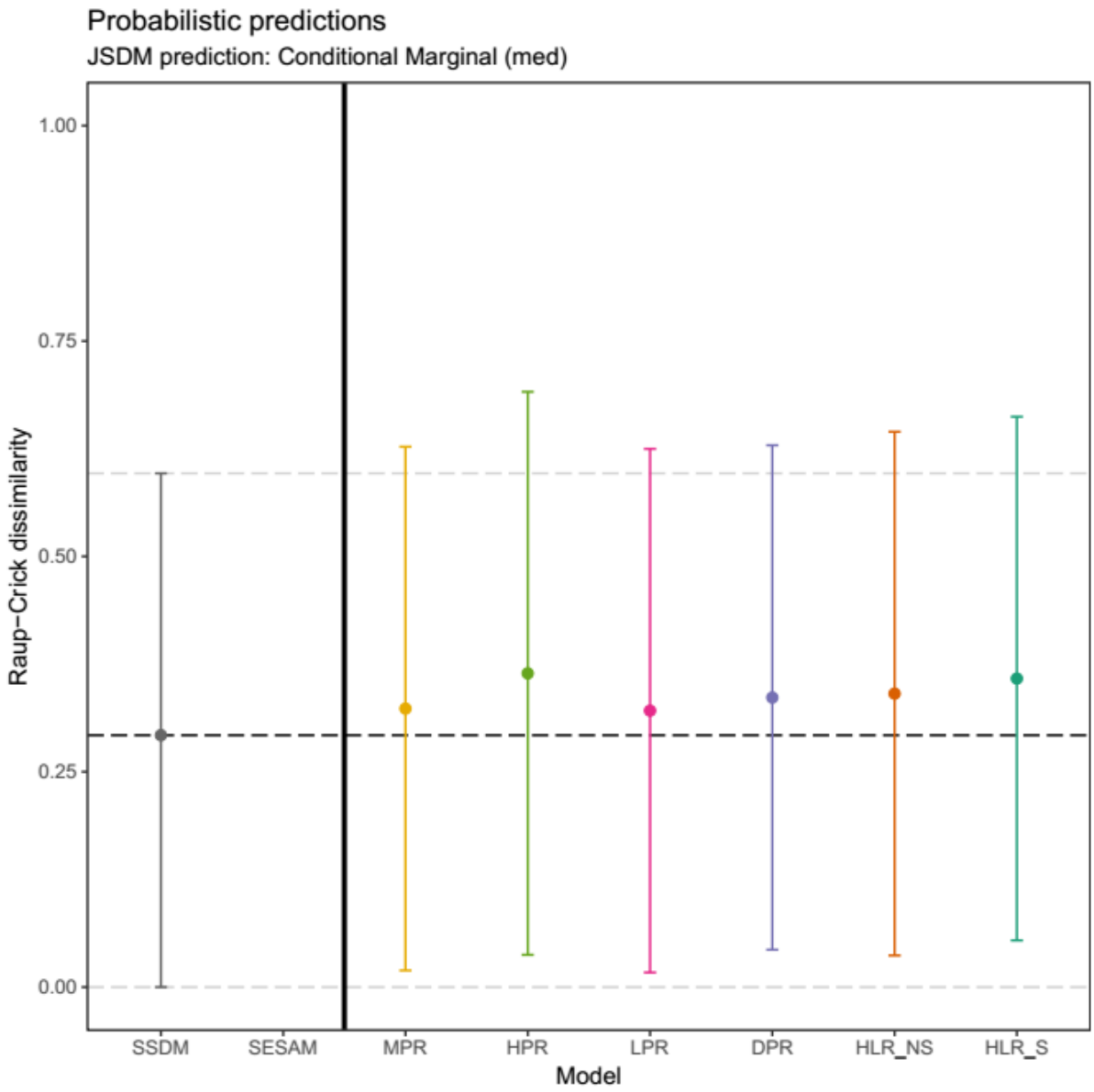


Figure 186: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

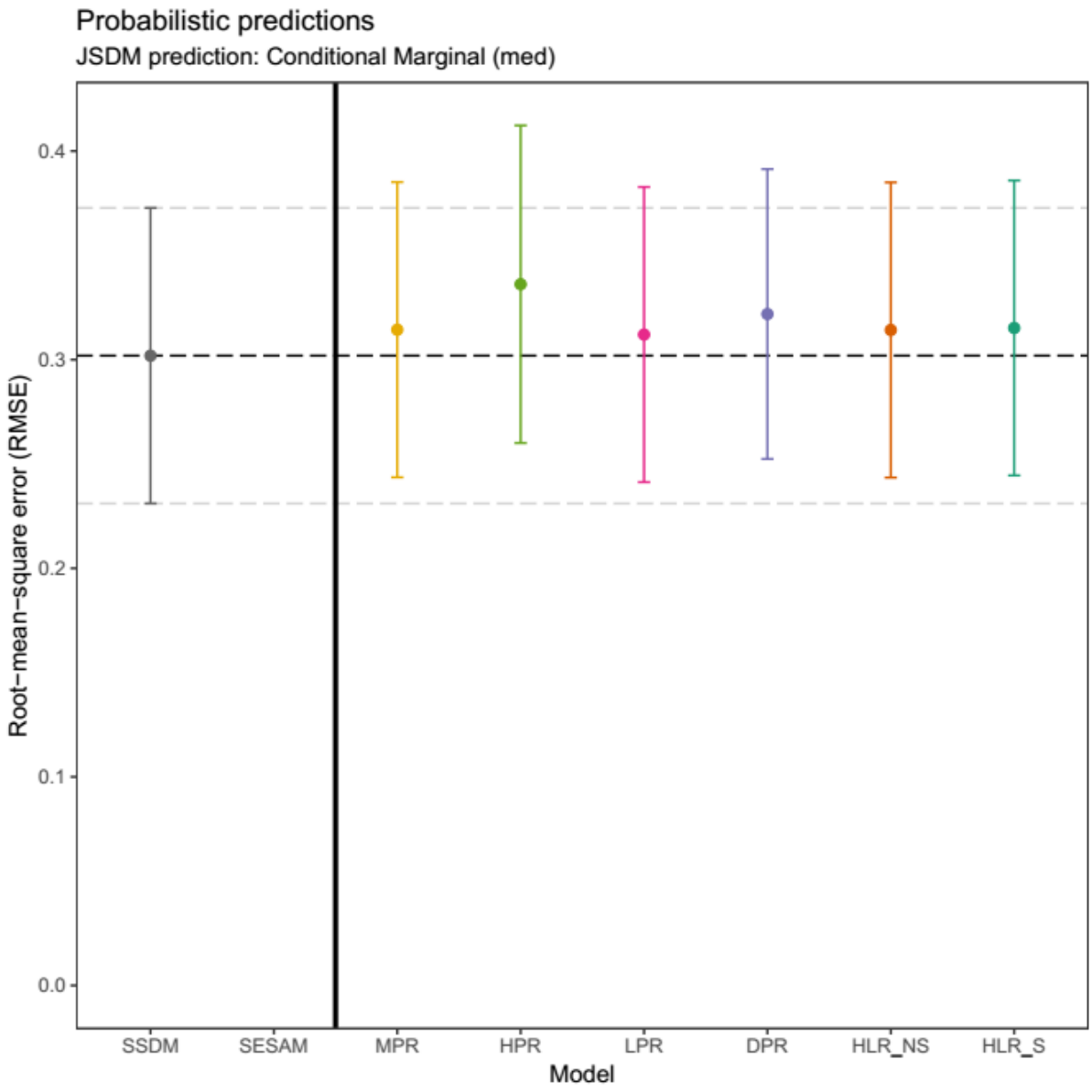


Figure 187: Model performance for the root-mean-square error evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

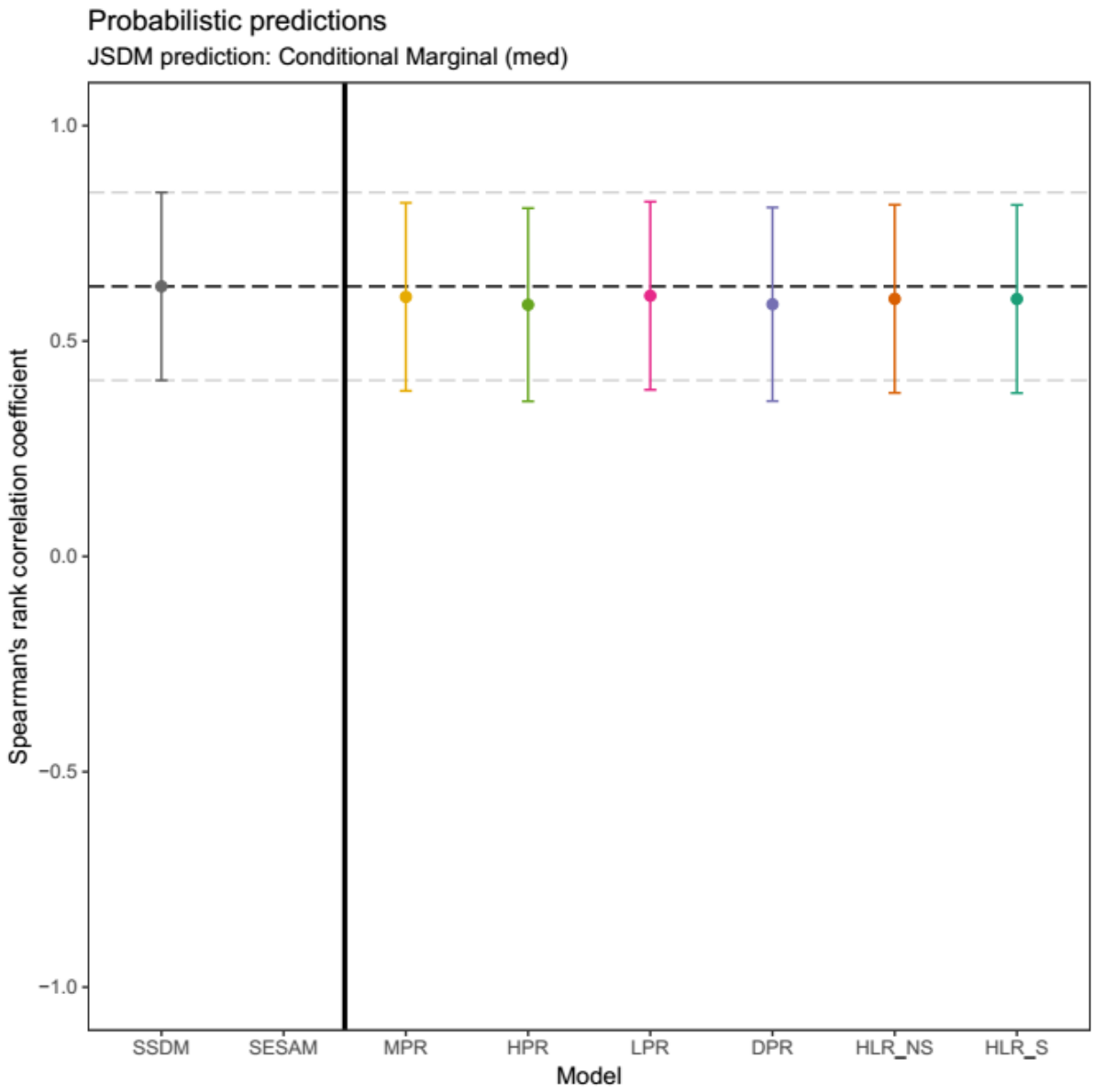


Figure 188: Model performance for the Spearman's rank correlation coefficient evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

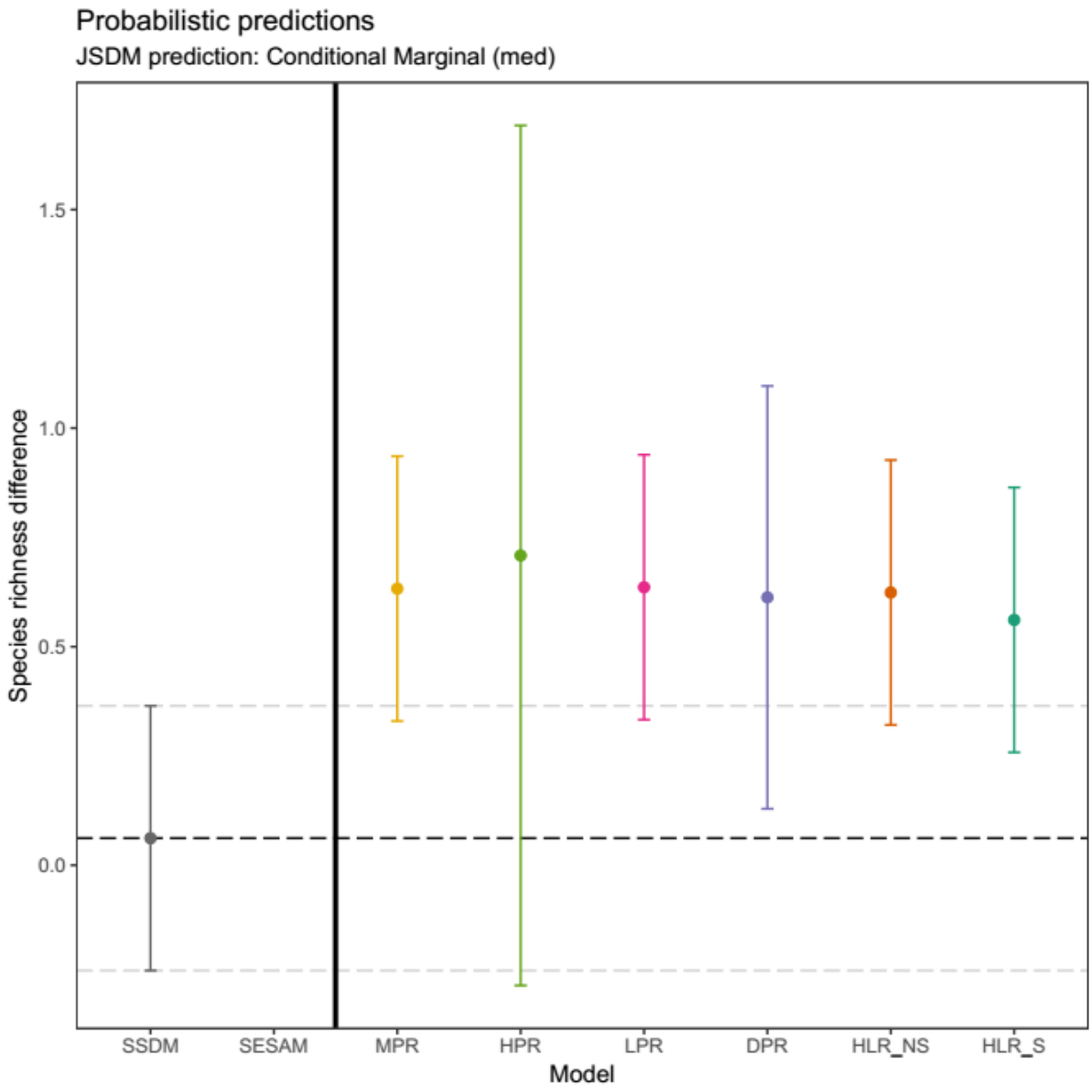


Figure 189: Model performance for the species richness difference evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

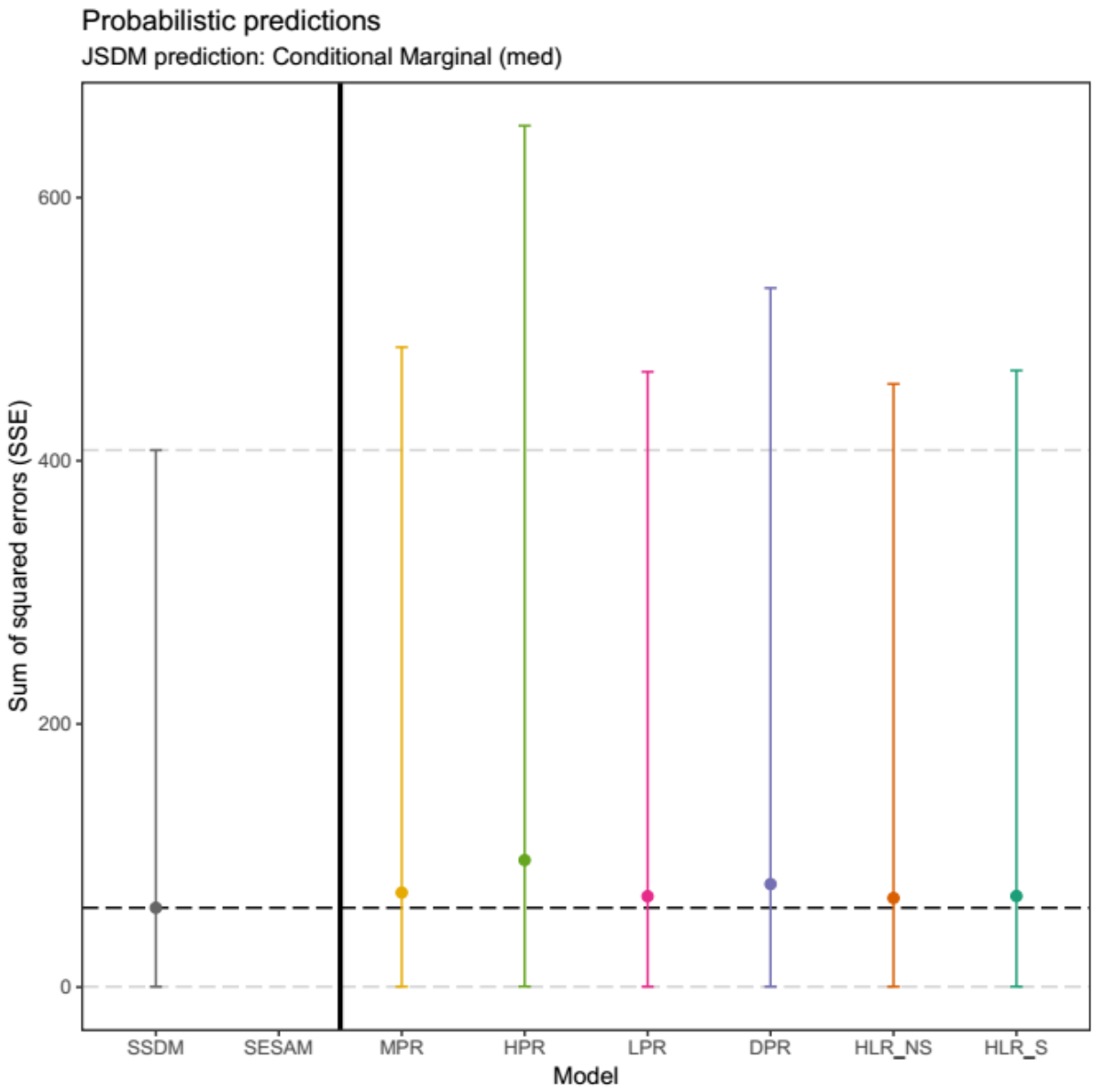


Figure 190: Model performance for the sum of squared errors evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

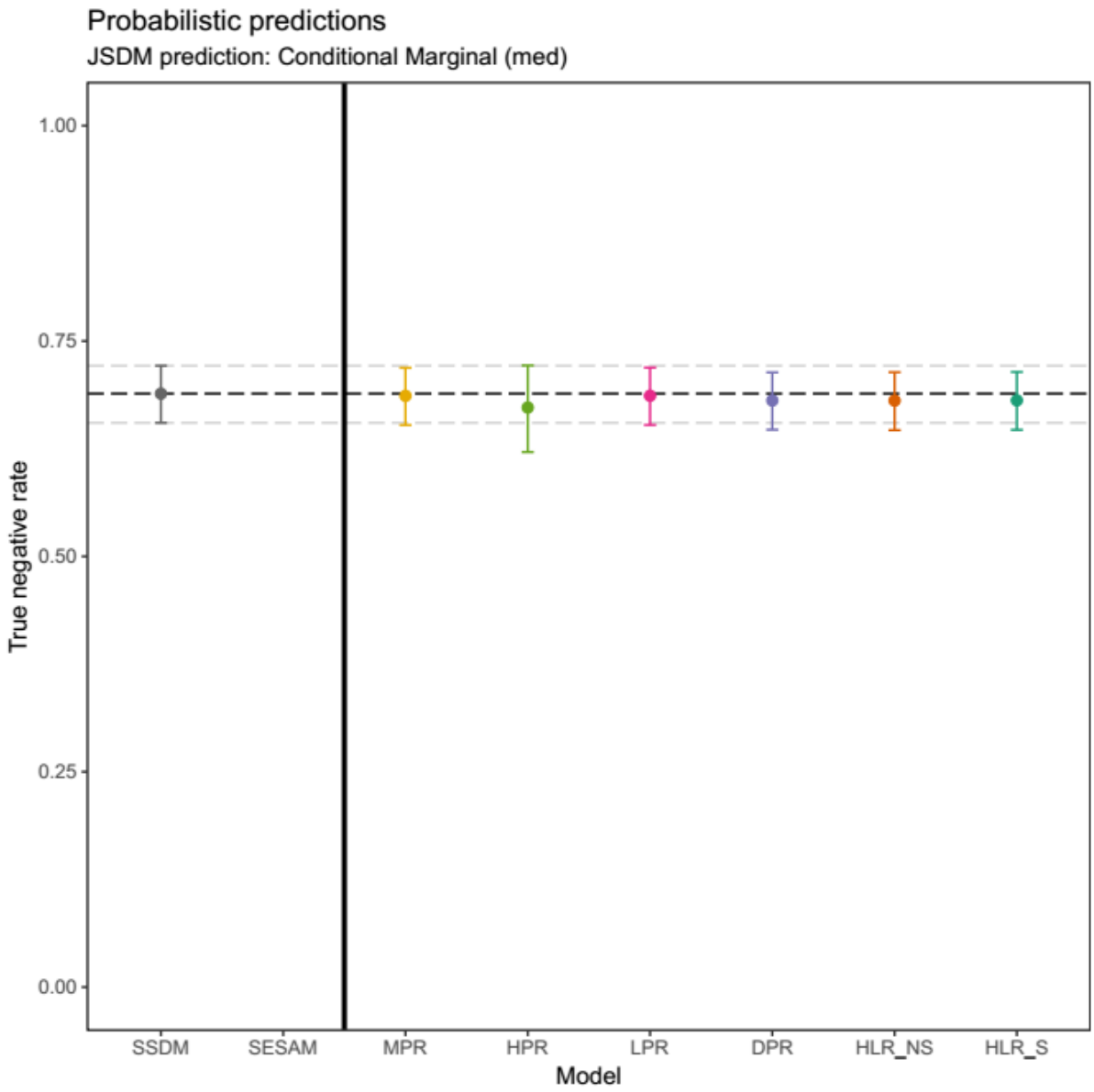


Figure 191: Model performance for the true negative rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

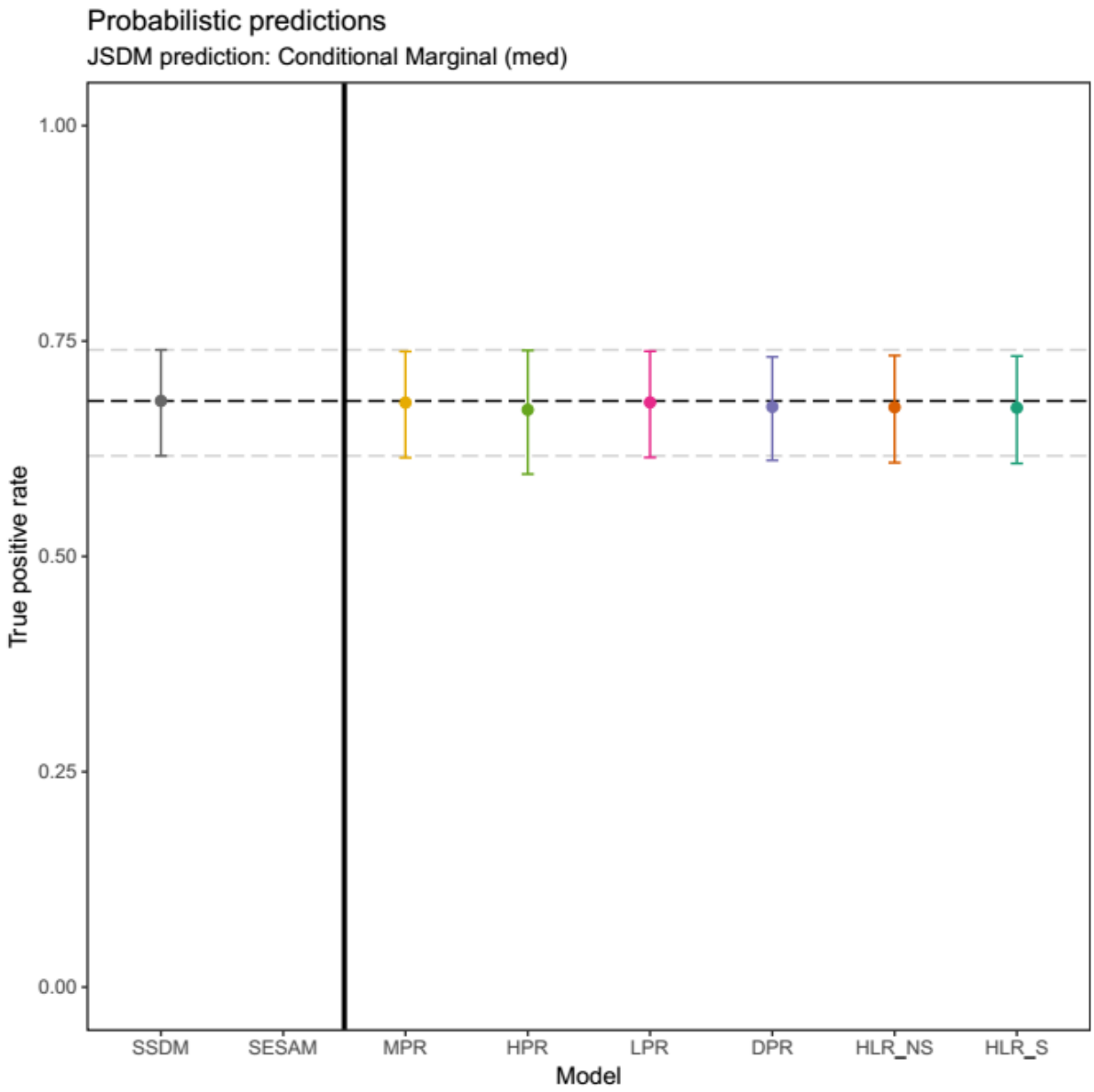


Figure 192: Model performance for the true positive rate evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

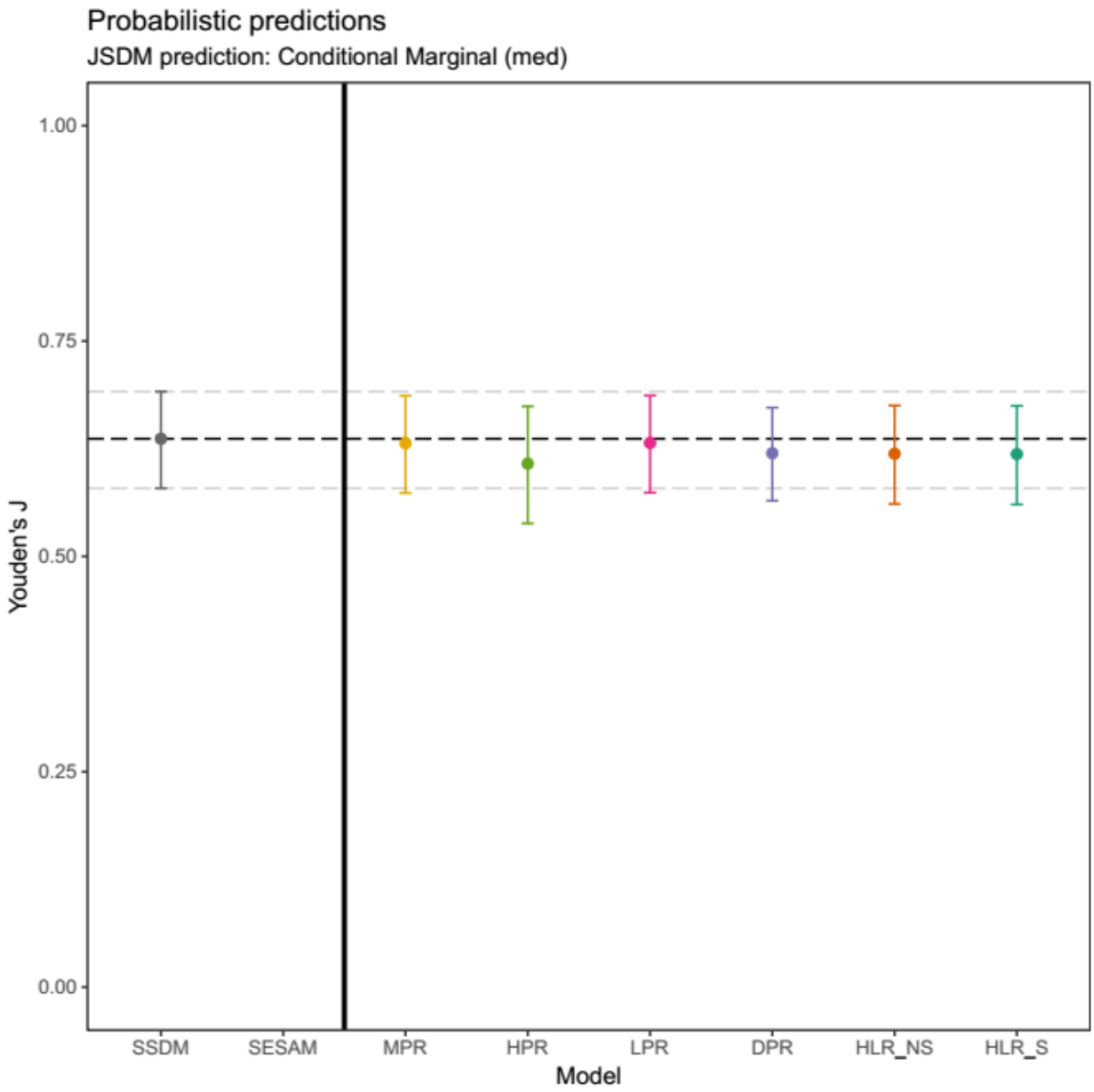


Figure 193: Model performance for the Youden's J evaluation metric for conditional marginal (medium-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

9. Conditional marginal (high-prevalence) predictions

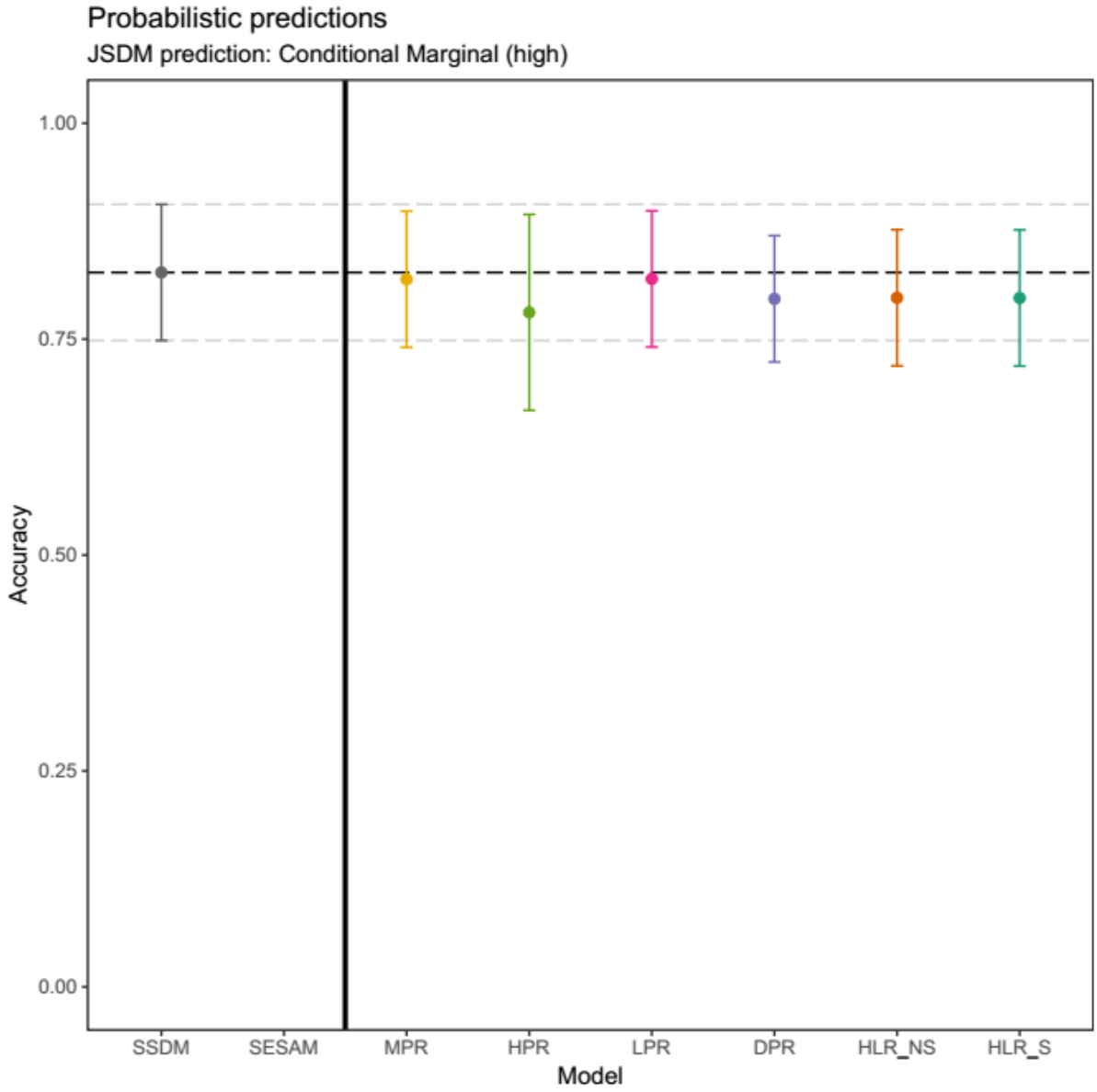


Figure 194: Model performance for the accuracy evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

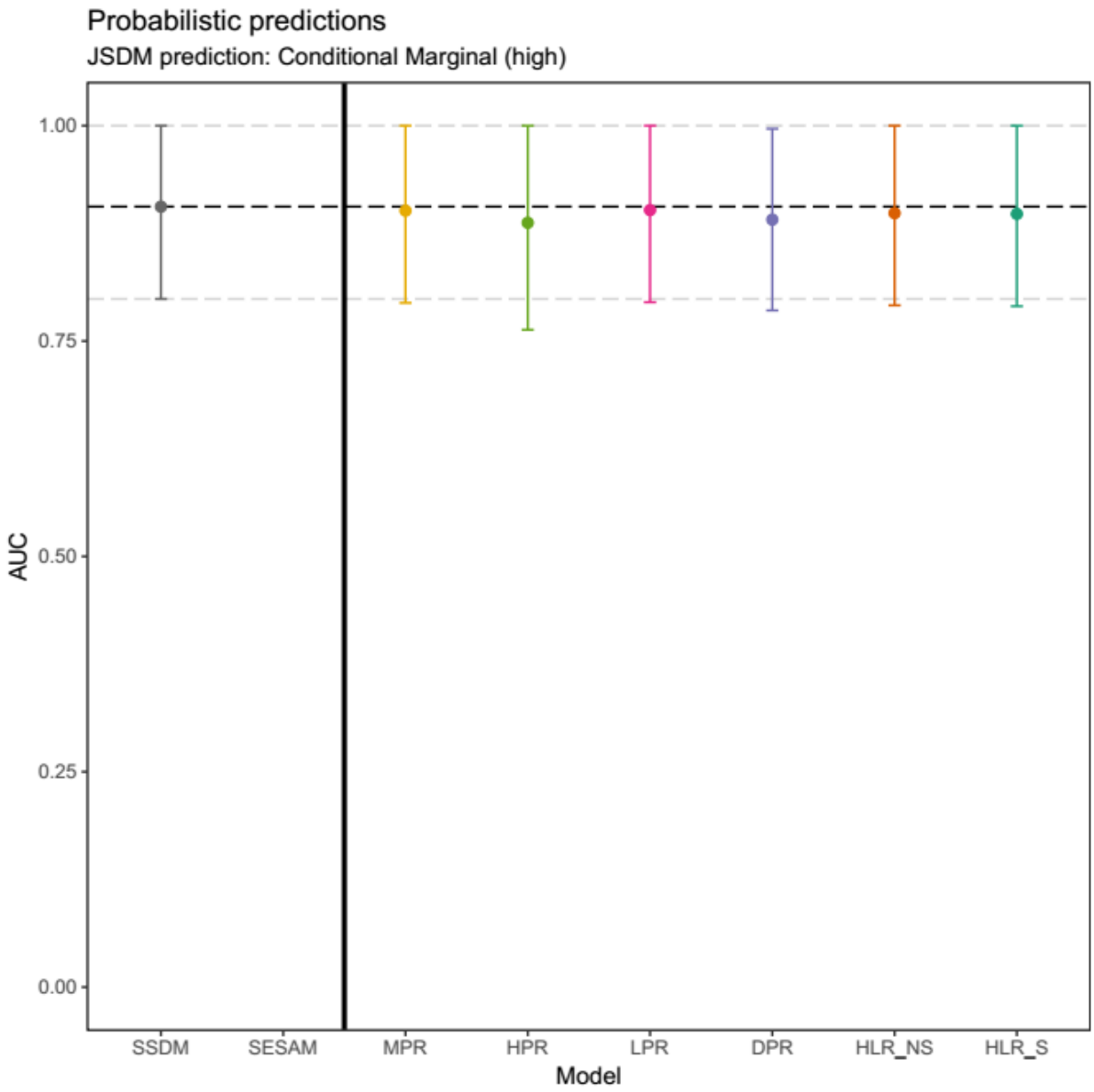


Figure 195: Model performance for the AUC evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

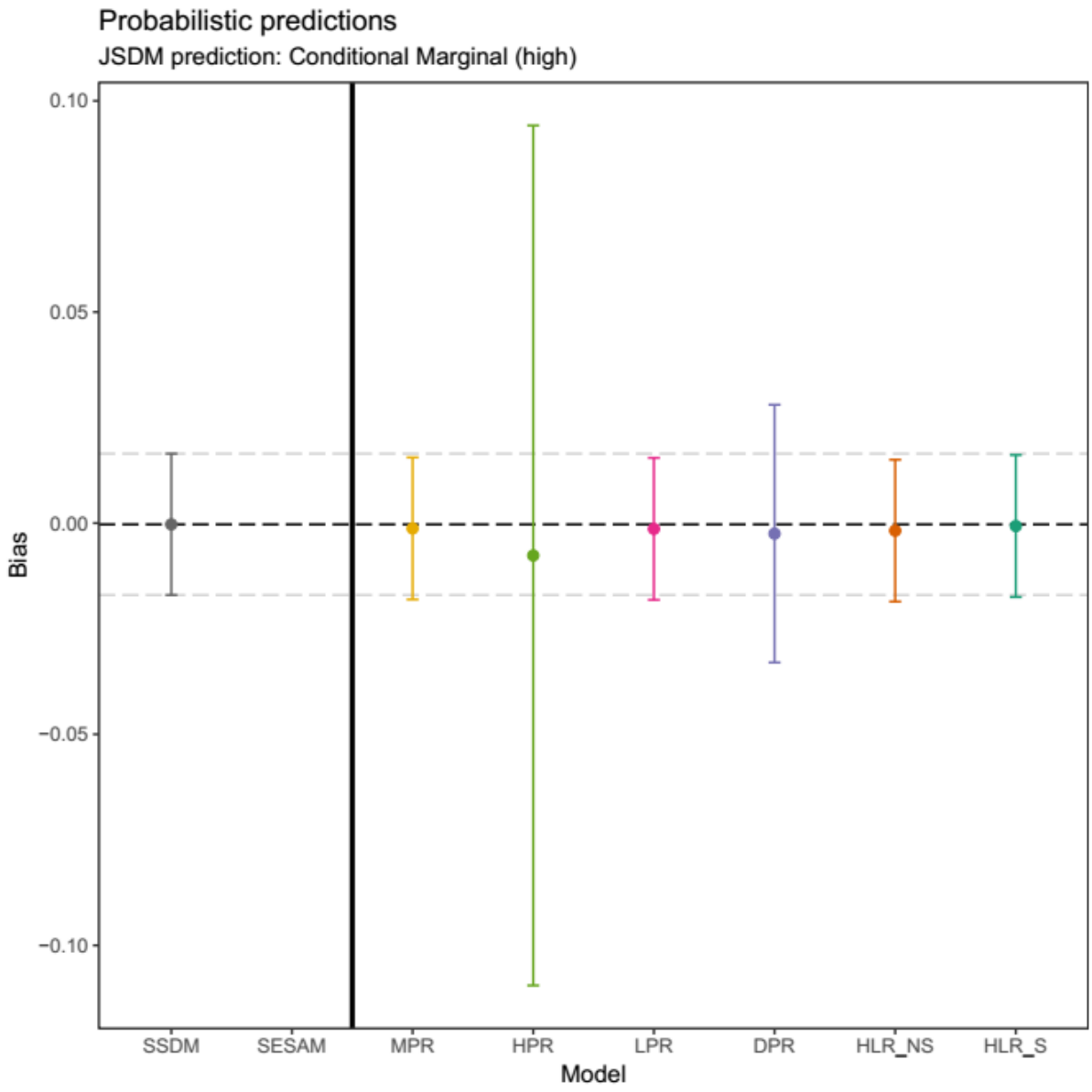


Figure 195: Model performance for the bias (mean error) evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

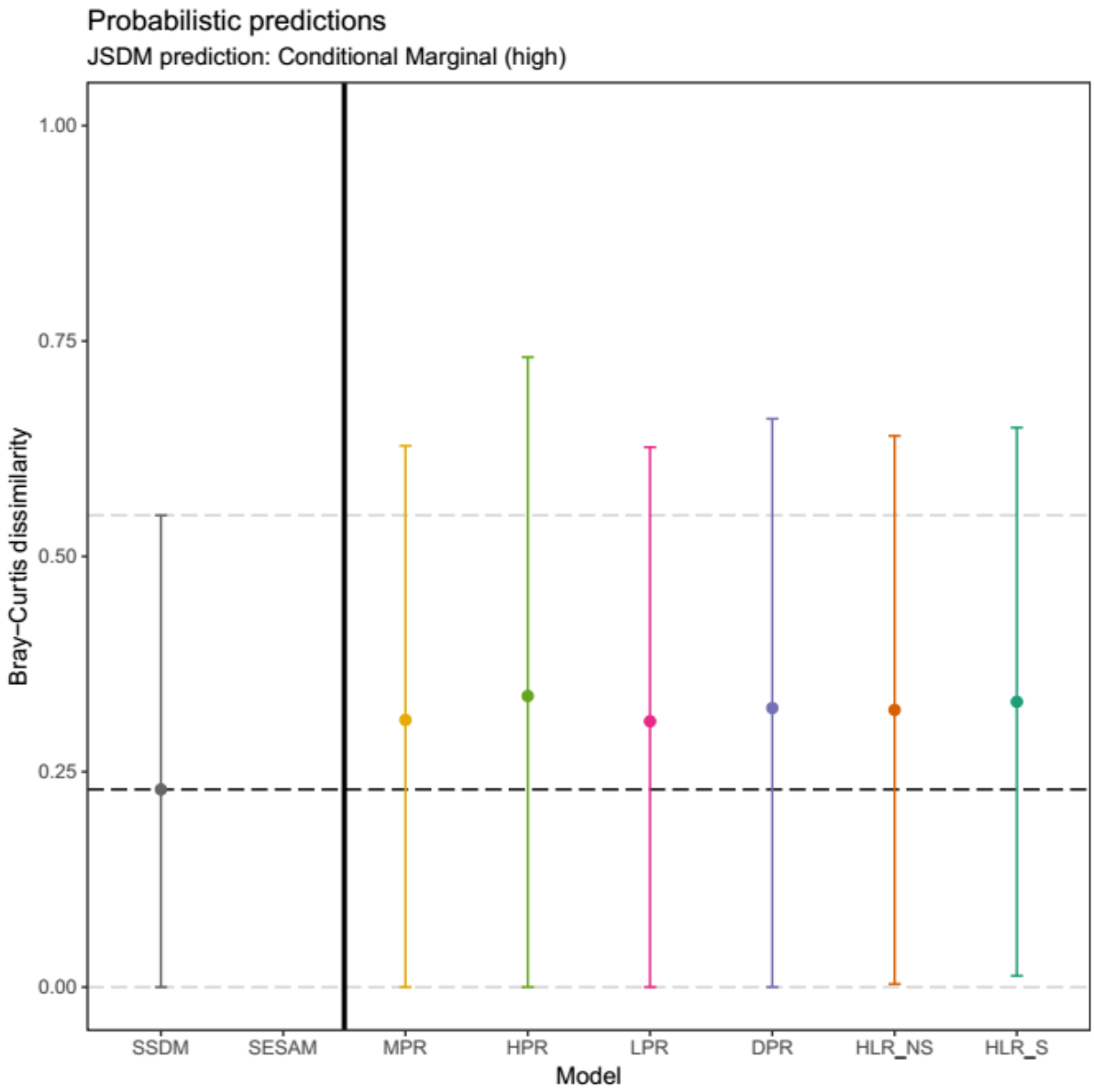


Figure 197: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

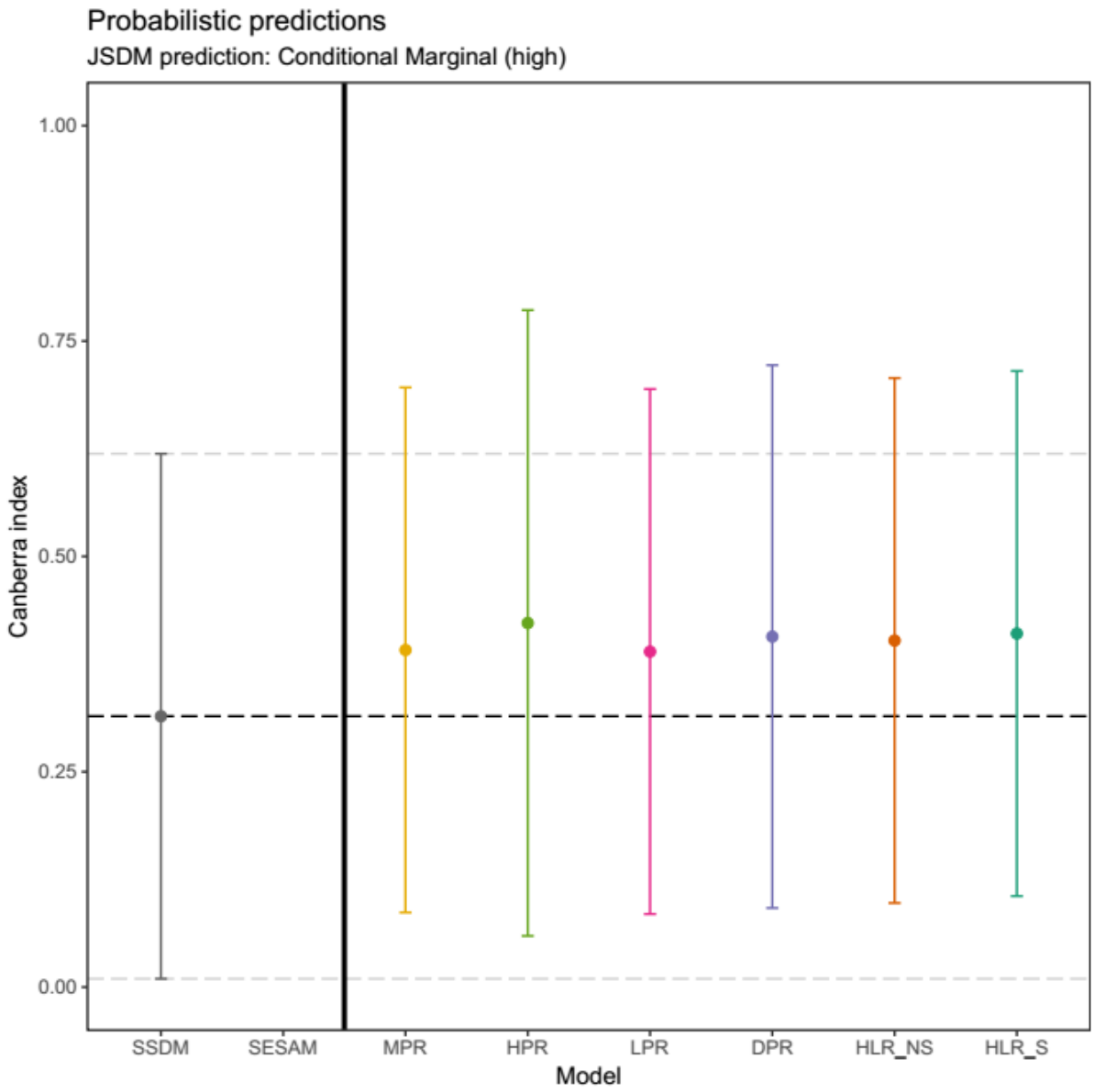


Figure 198: Model performance for the Canberra index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

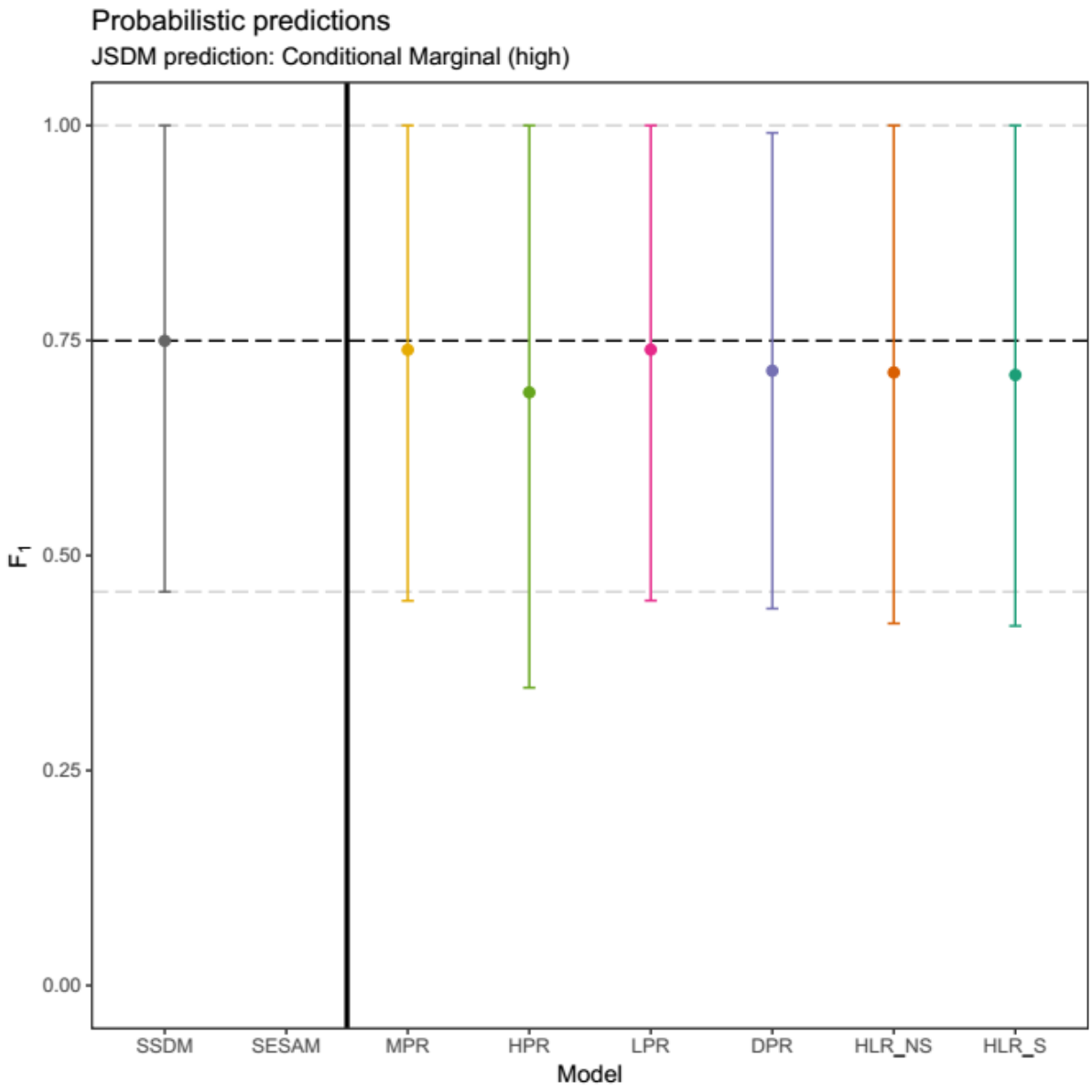


Figure 199: Model performance for the F_1 evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

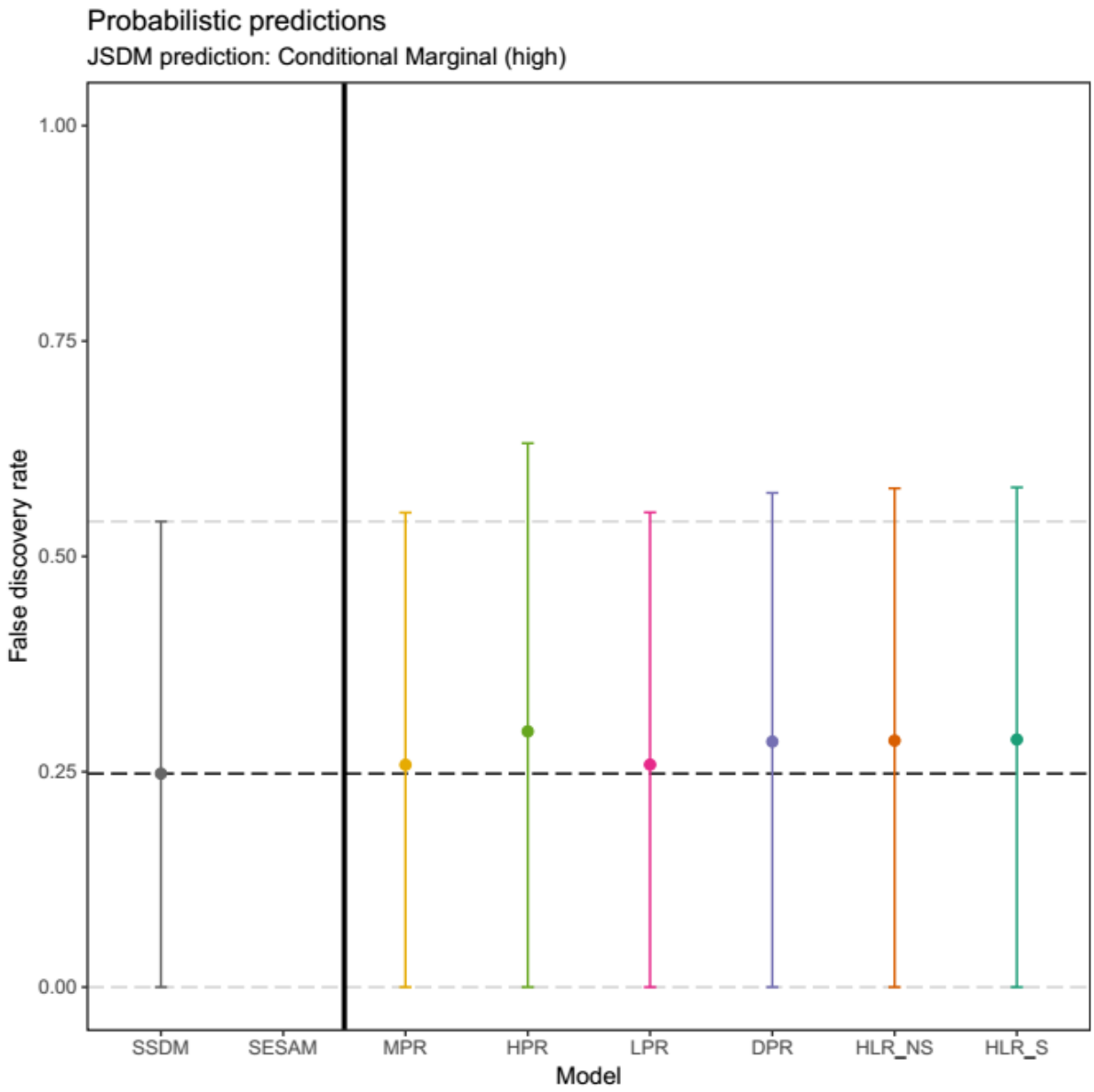


Figure 200: Model performance for the false discovery rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

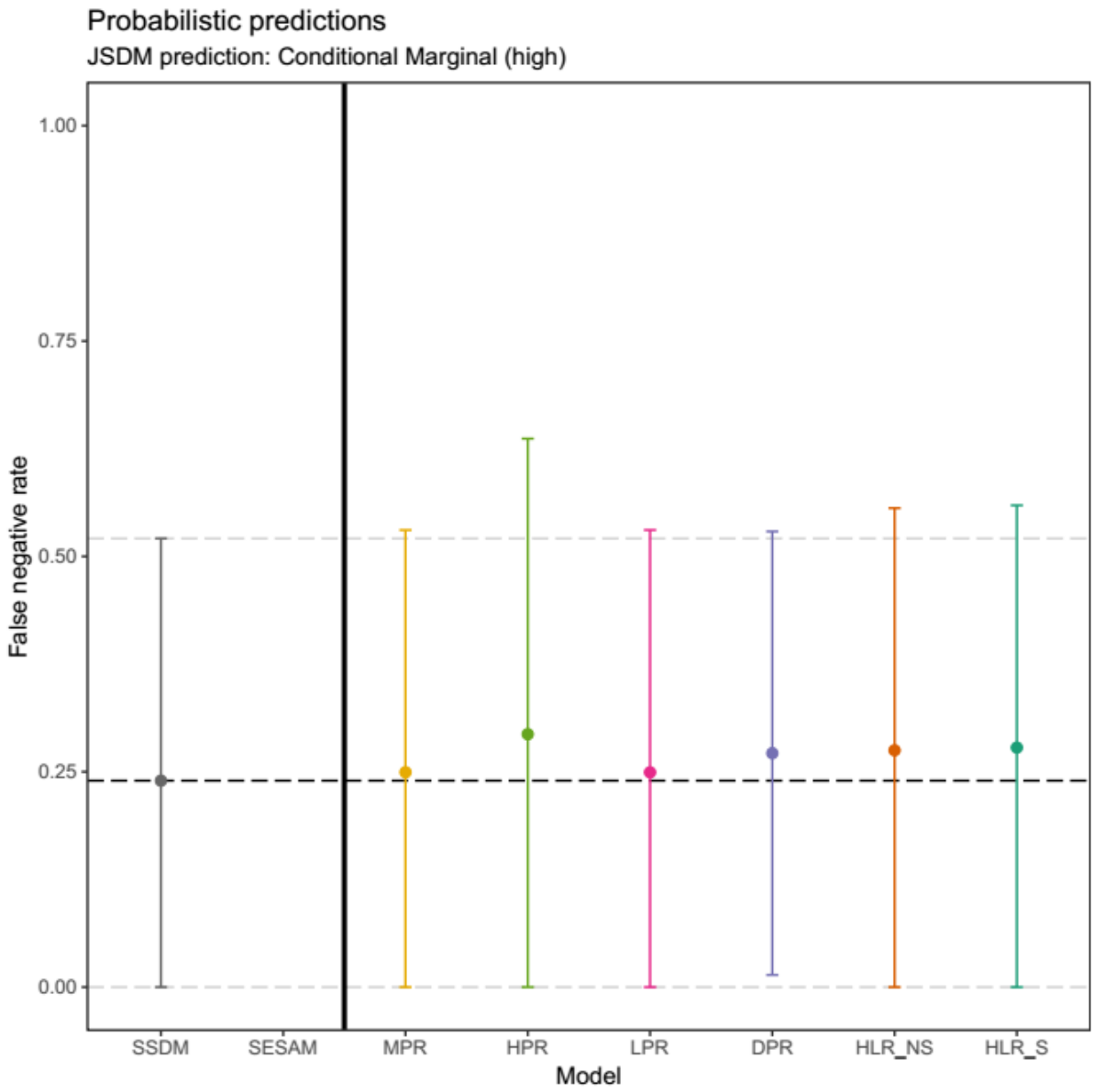


Figure 201: Model performance for the false negative rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

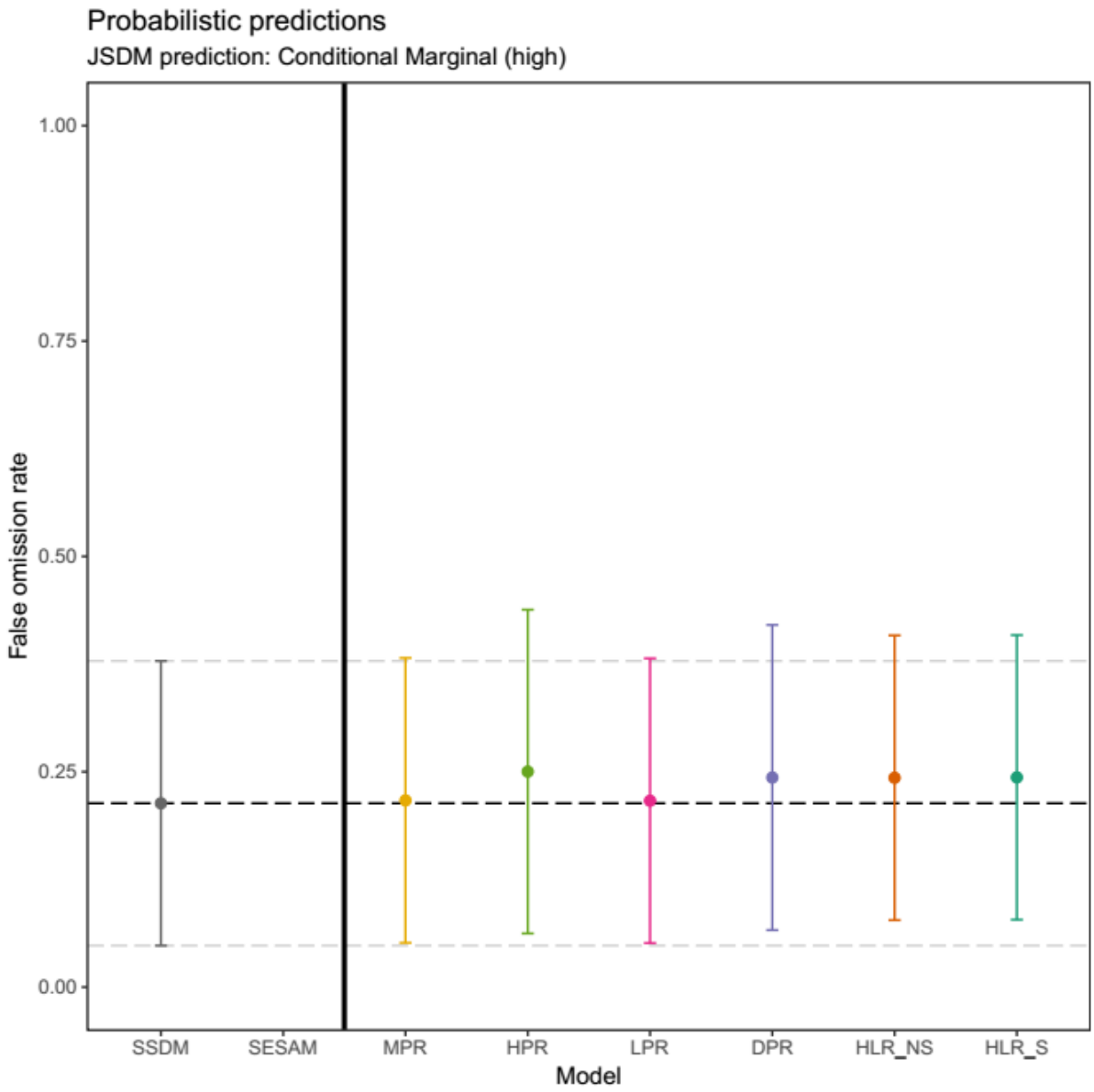


Figure 202: Model performance for the false omission rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

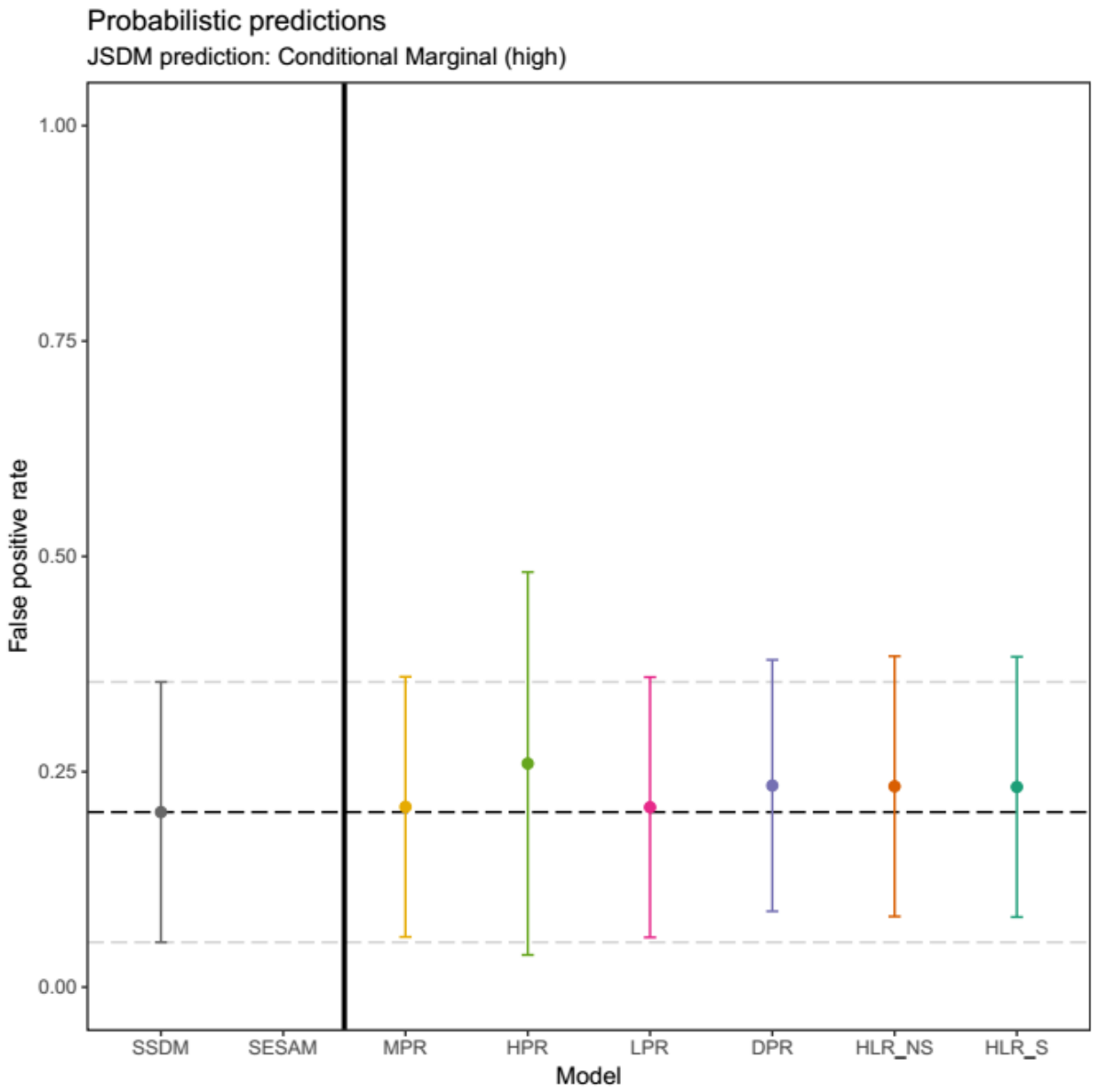


Figure 203: Model performance for the false positive rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

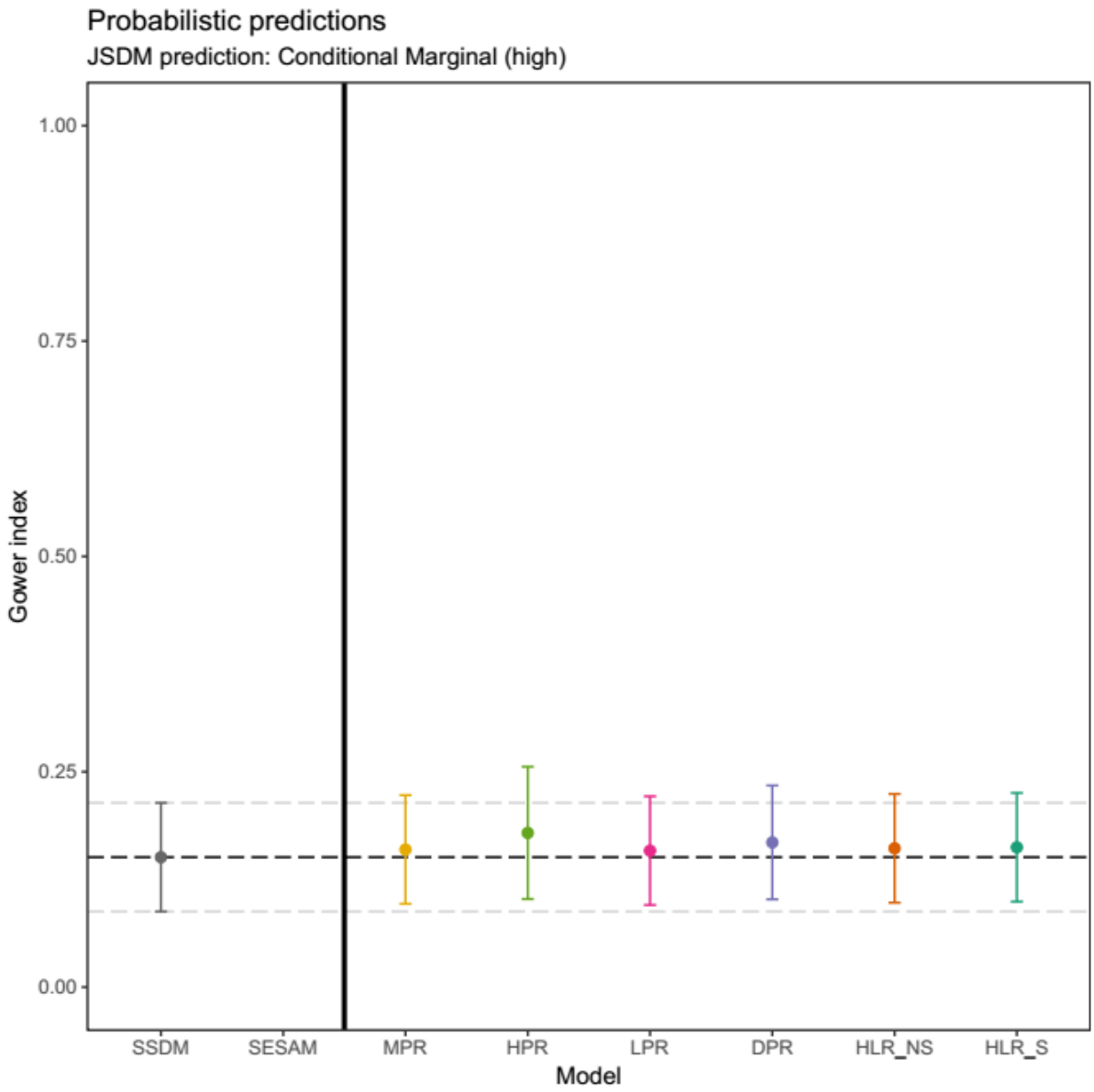


Figure 204: Model performance for the Gower index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

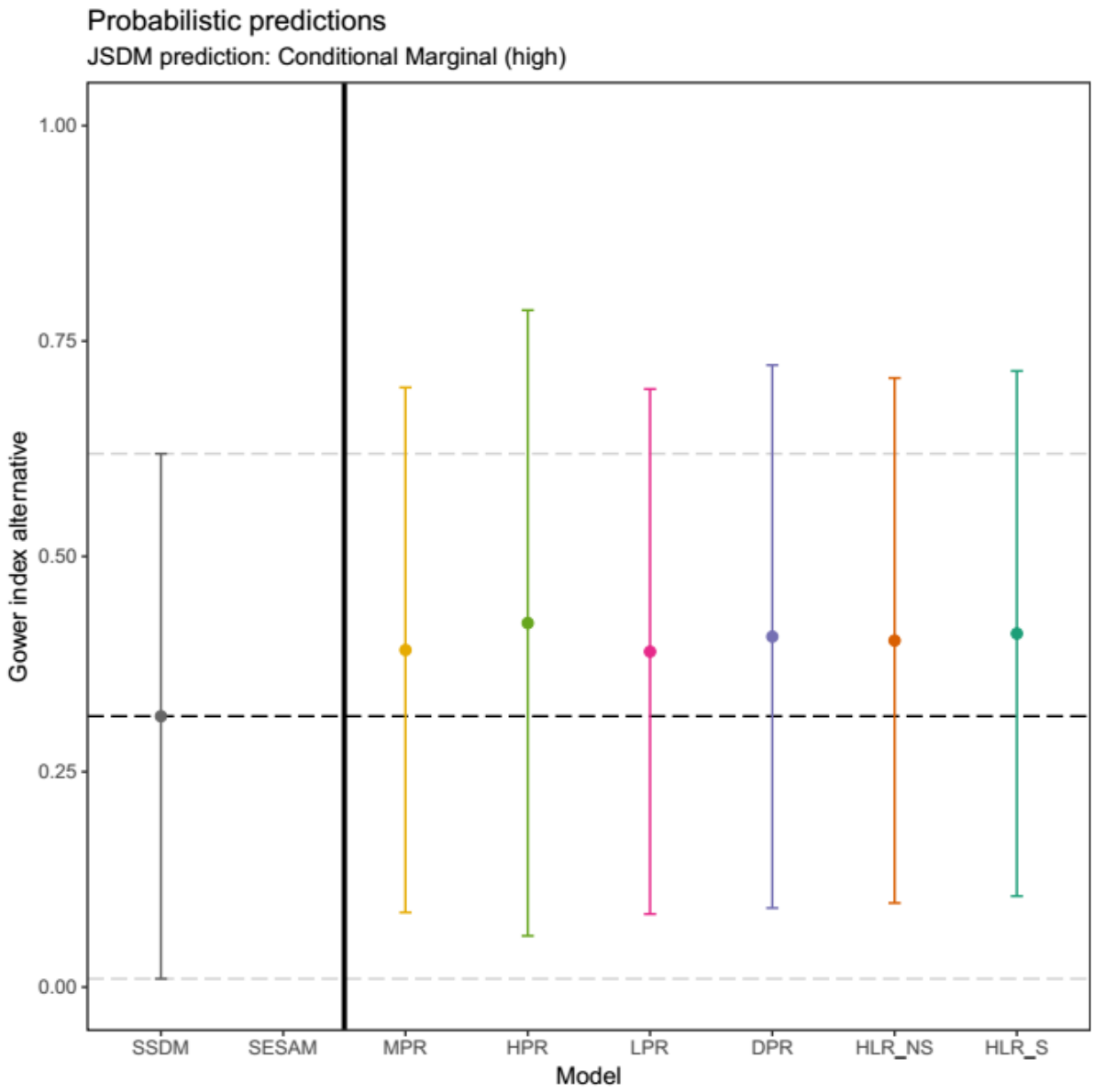


Figure 205: Model performance for the alternative Gower index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

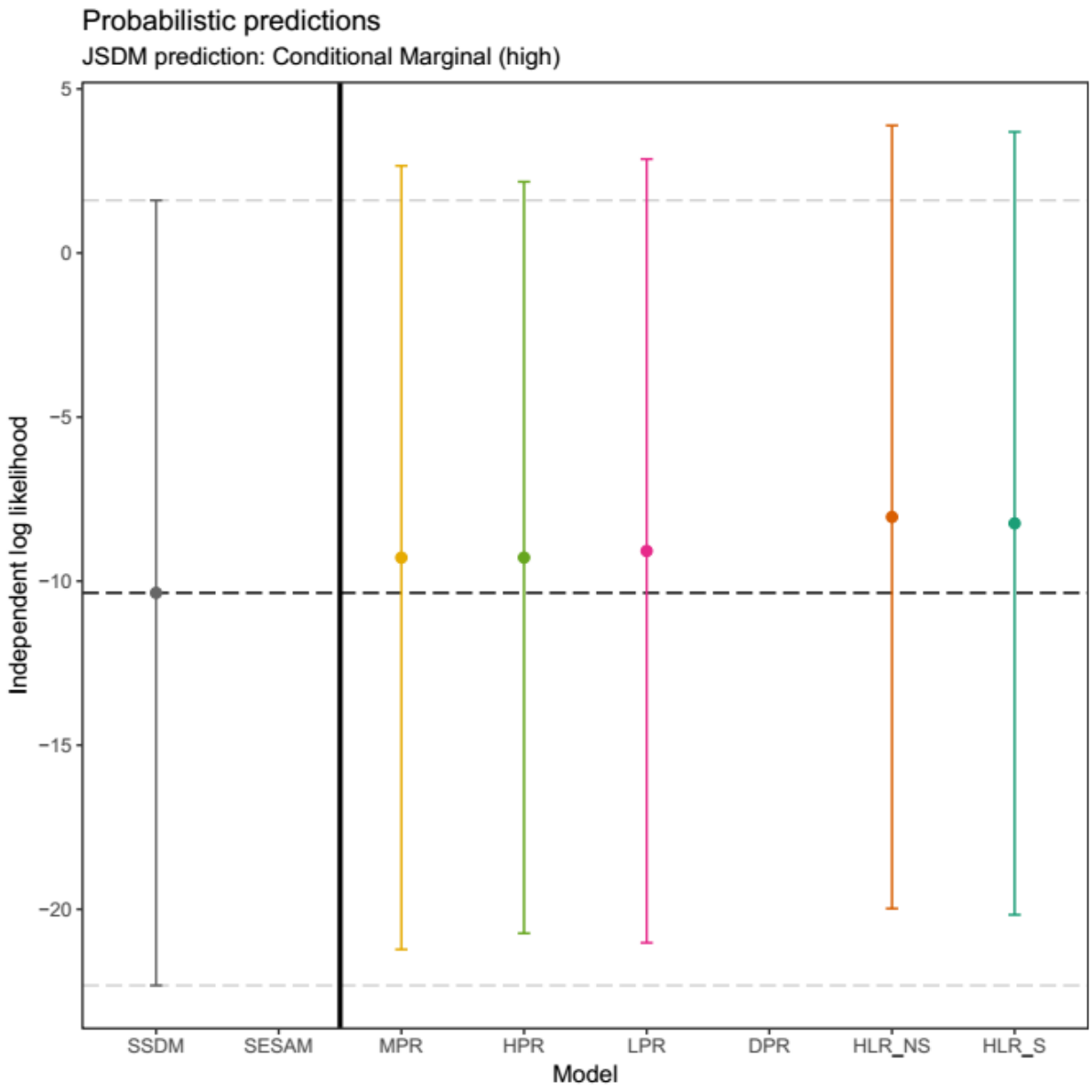


Figure 206: Model performance for the independent log-likelihood evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

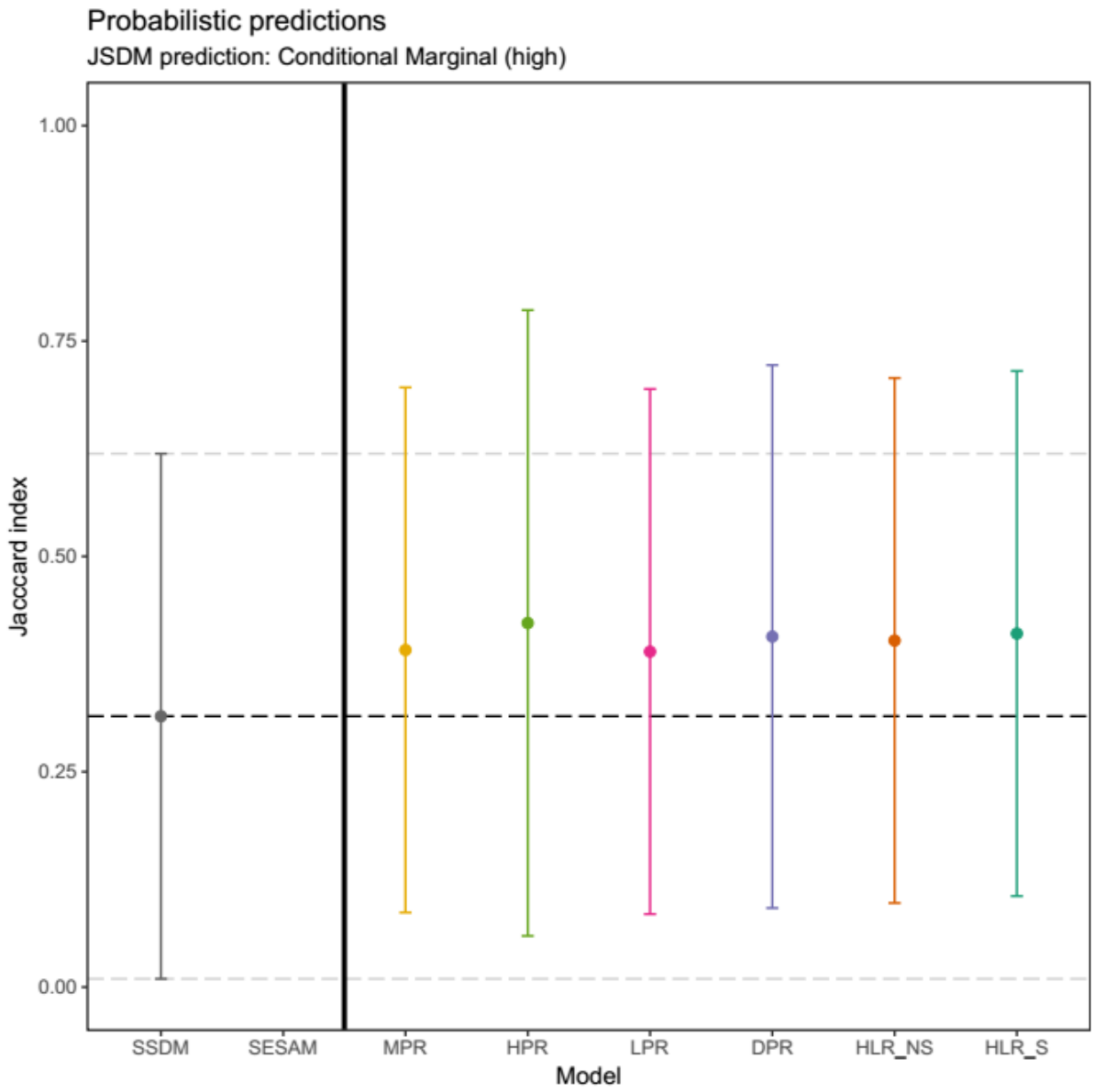


Figure 207: Model performance for the Jaccard index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

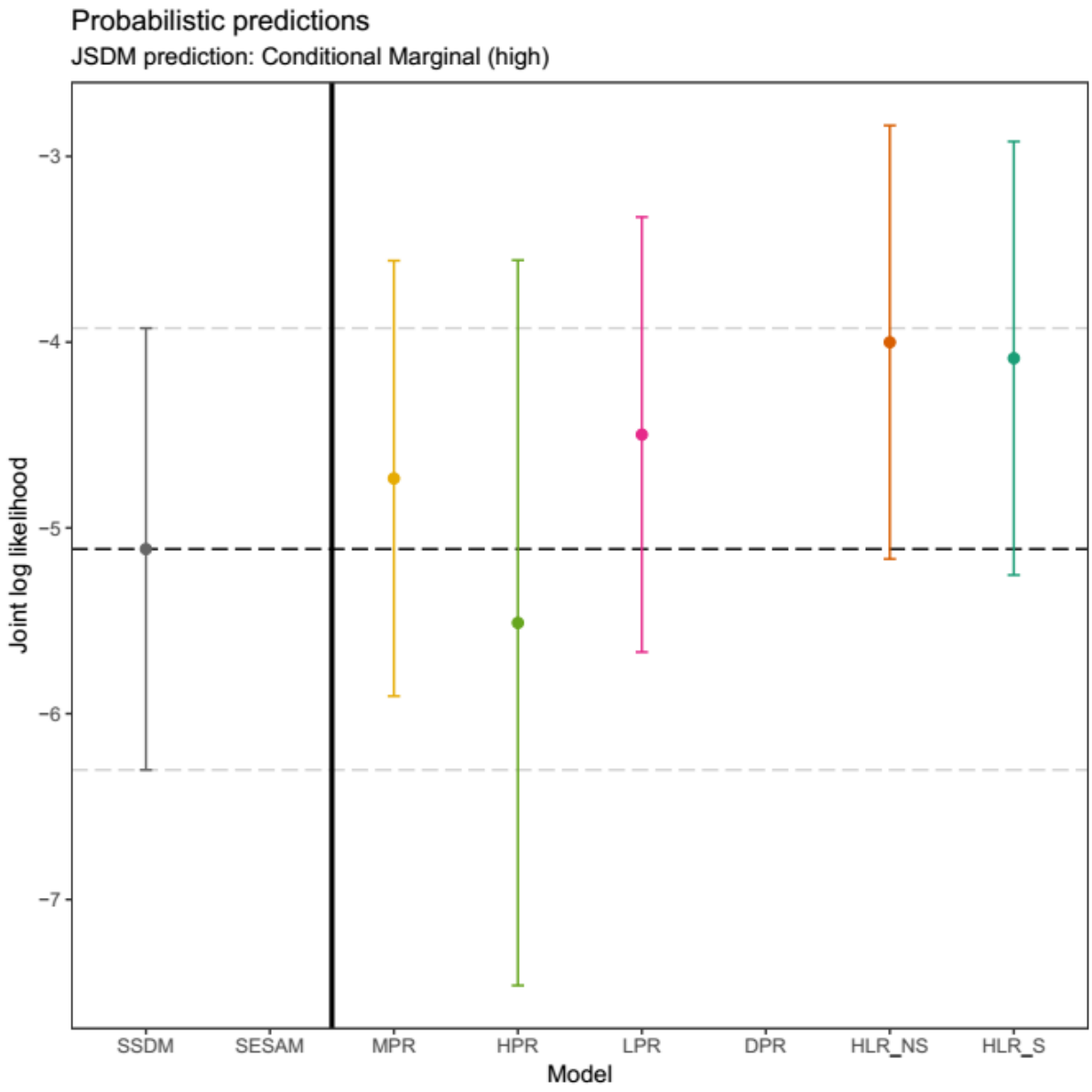


Figure 208: Model performance for the joint log likelihood evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

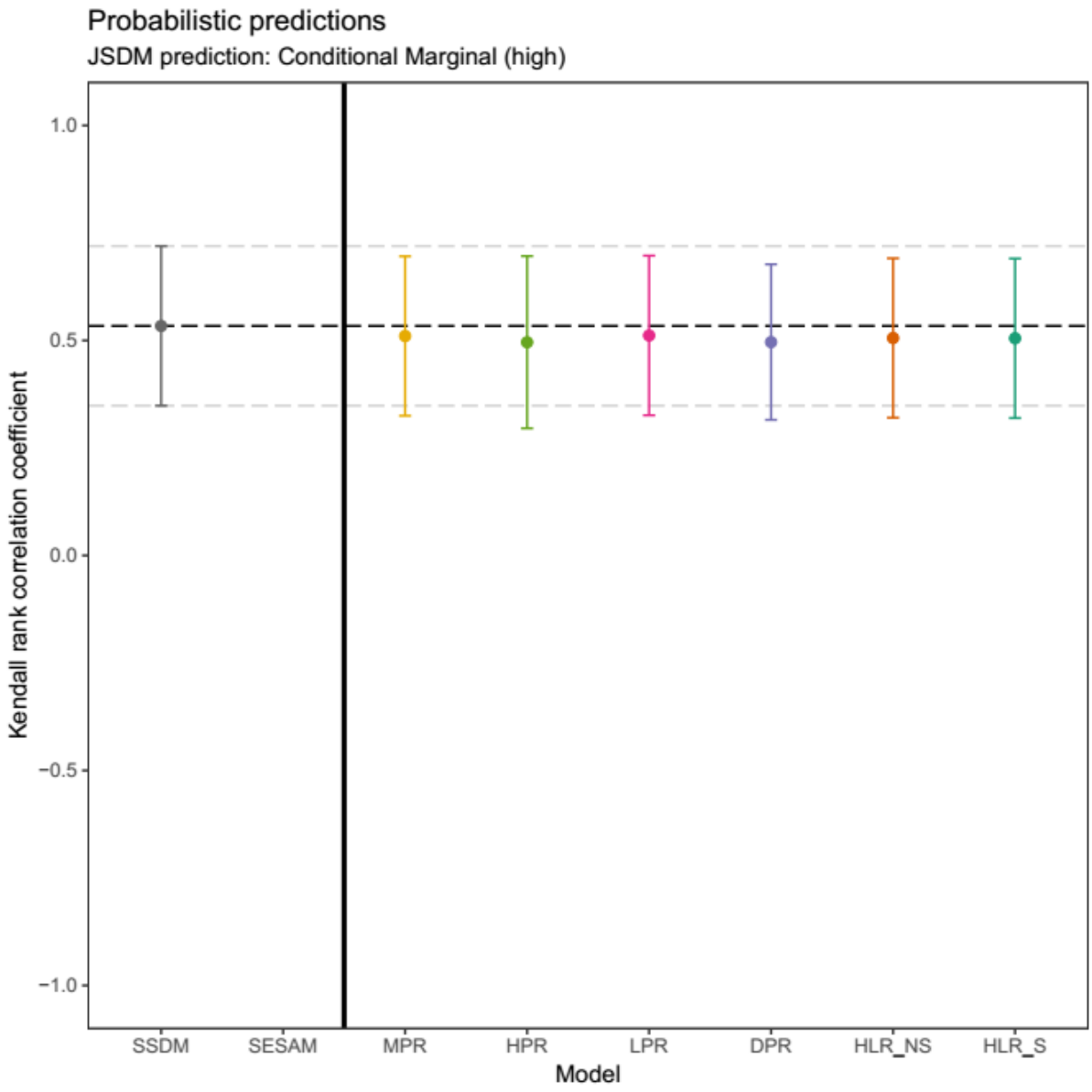


Figure 209: Model performance for the Kendall rank correlation coefficient evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

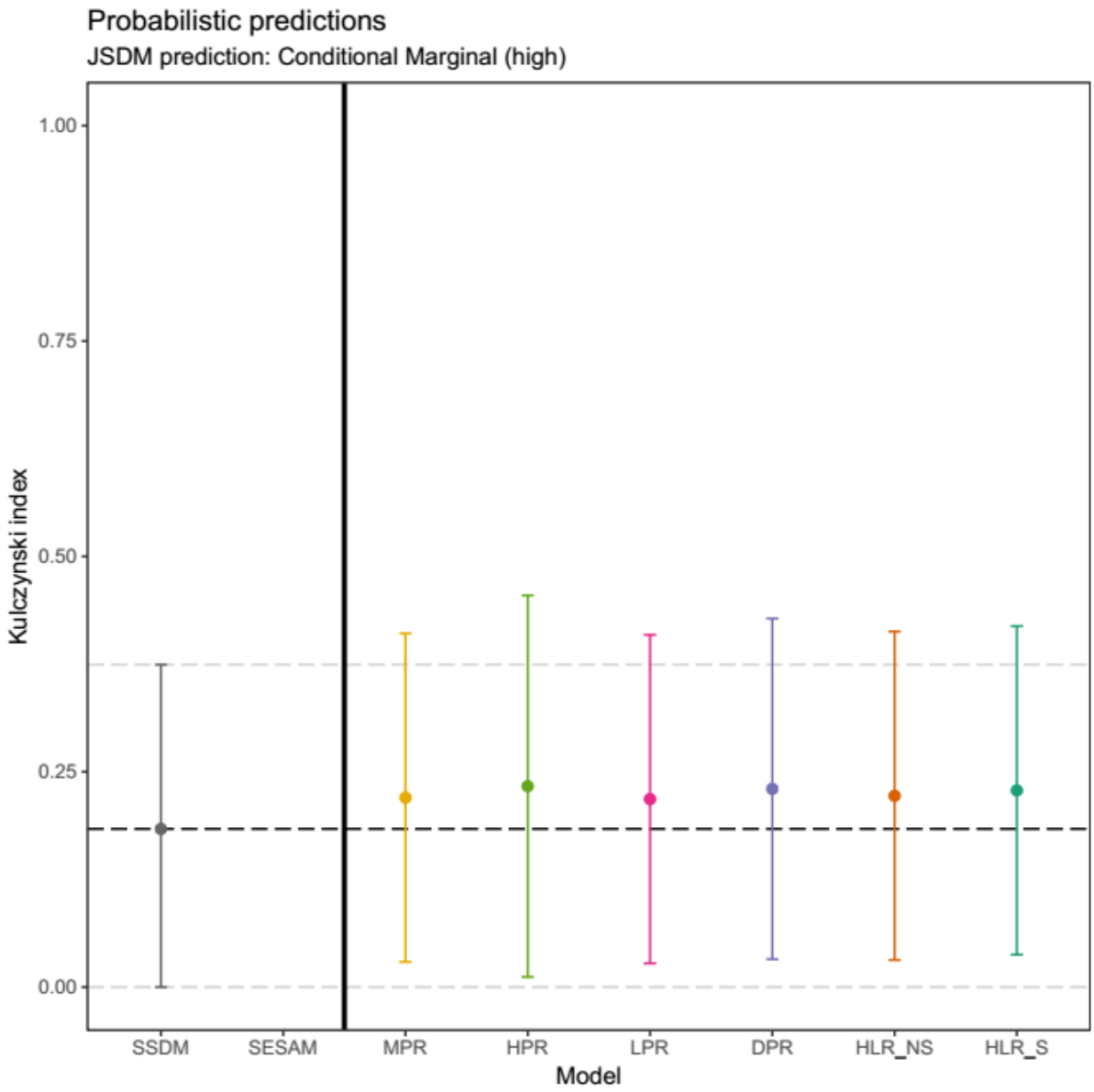


Figure 210: Model performance for the Kulczynski index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

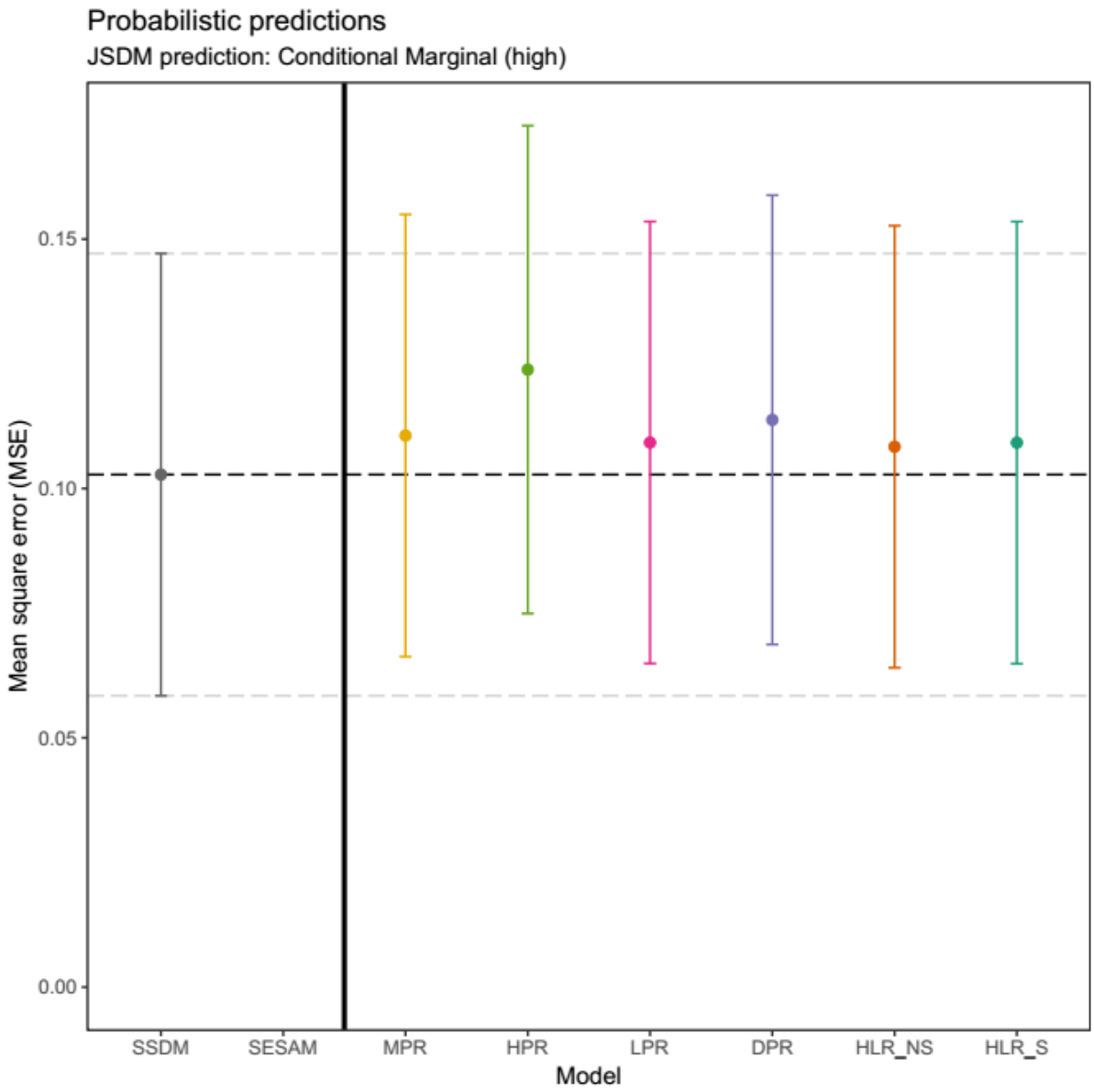


Figure 211: Model performance for the mean square error evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

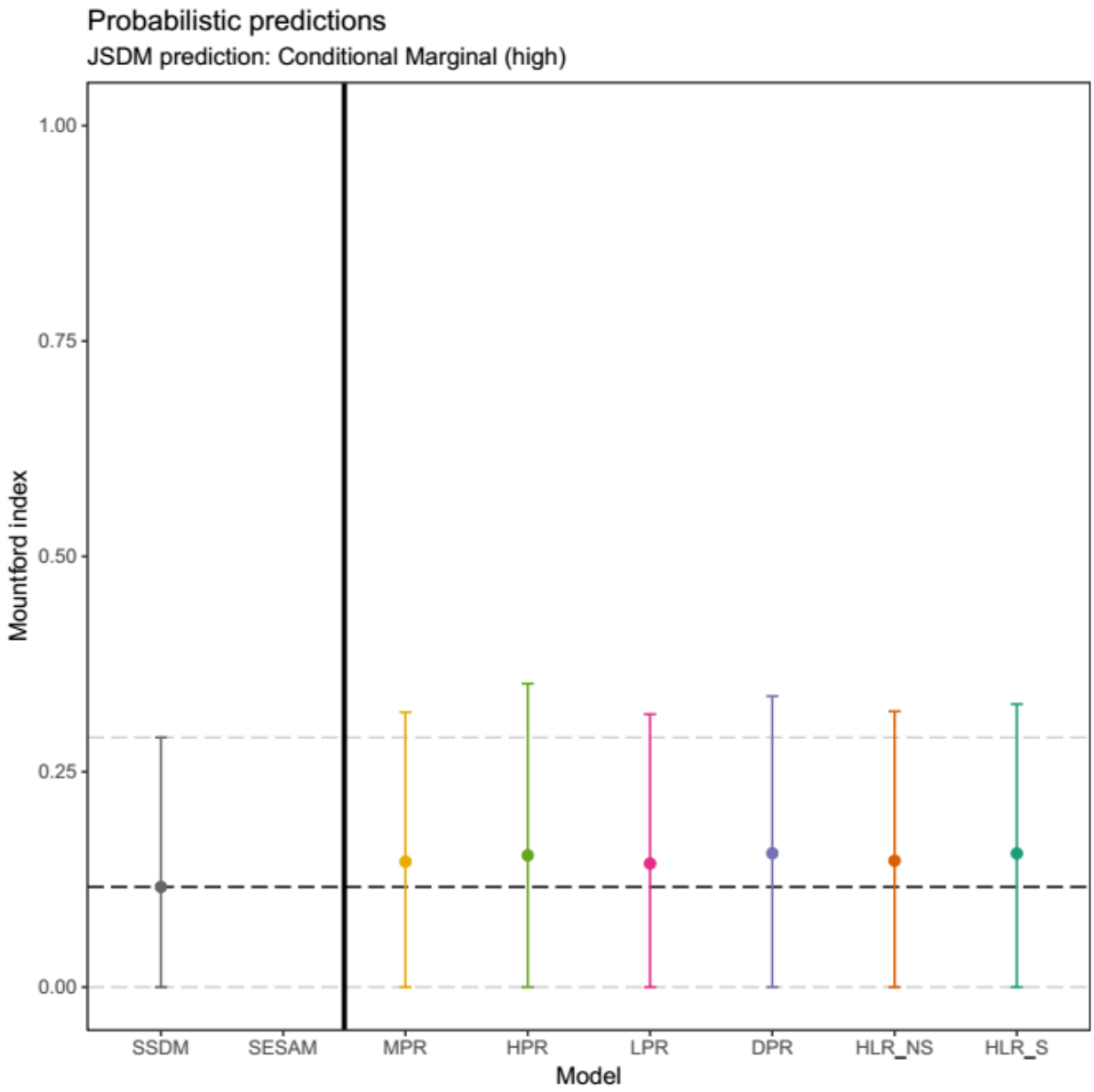


Figure 212: Model performance for the Mountford index evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

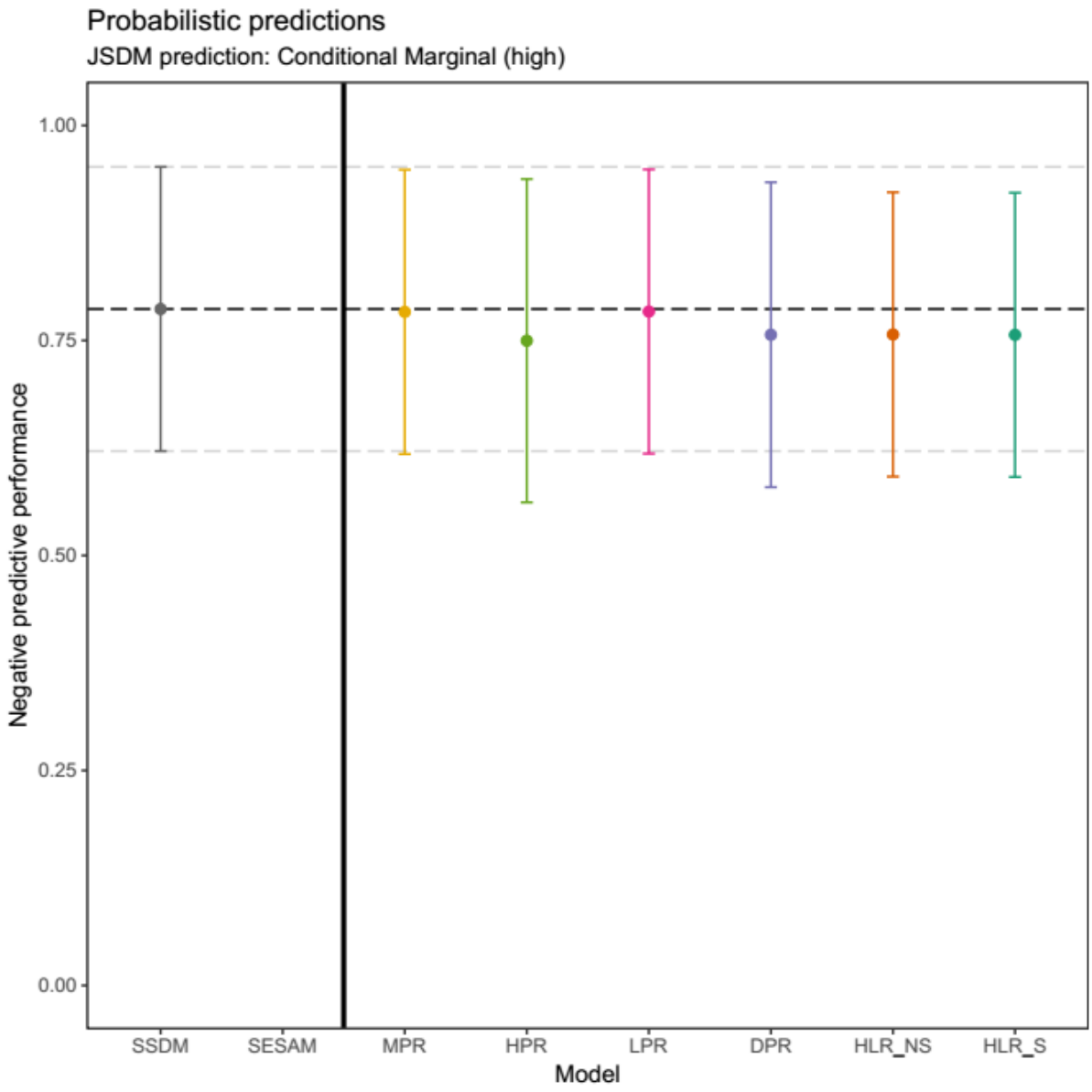


Figure 213: Model performance for the negative predictive performance evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

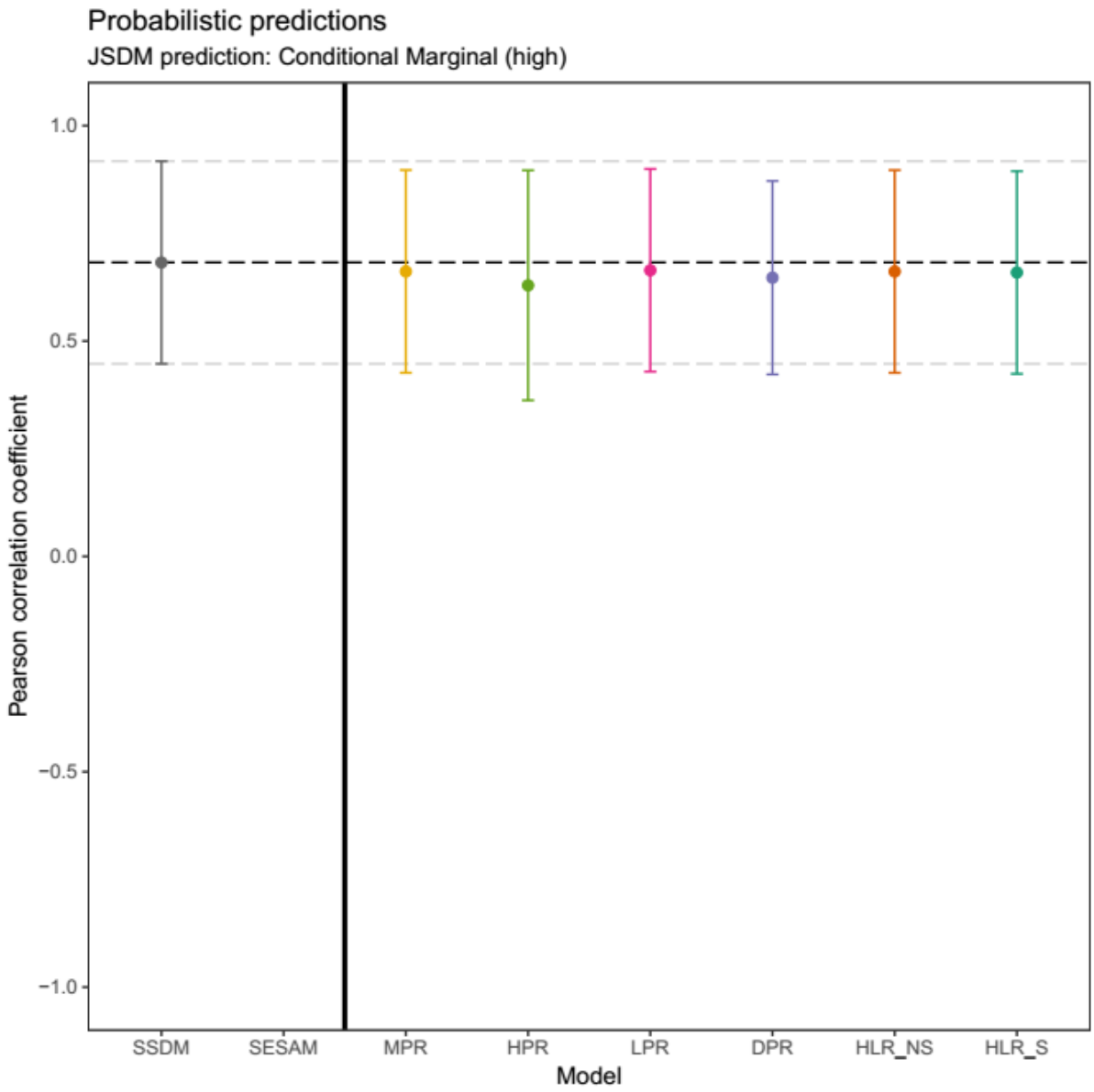


Figure 214: Model performance for the Pearson correlation coefficient evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

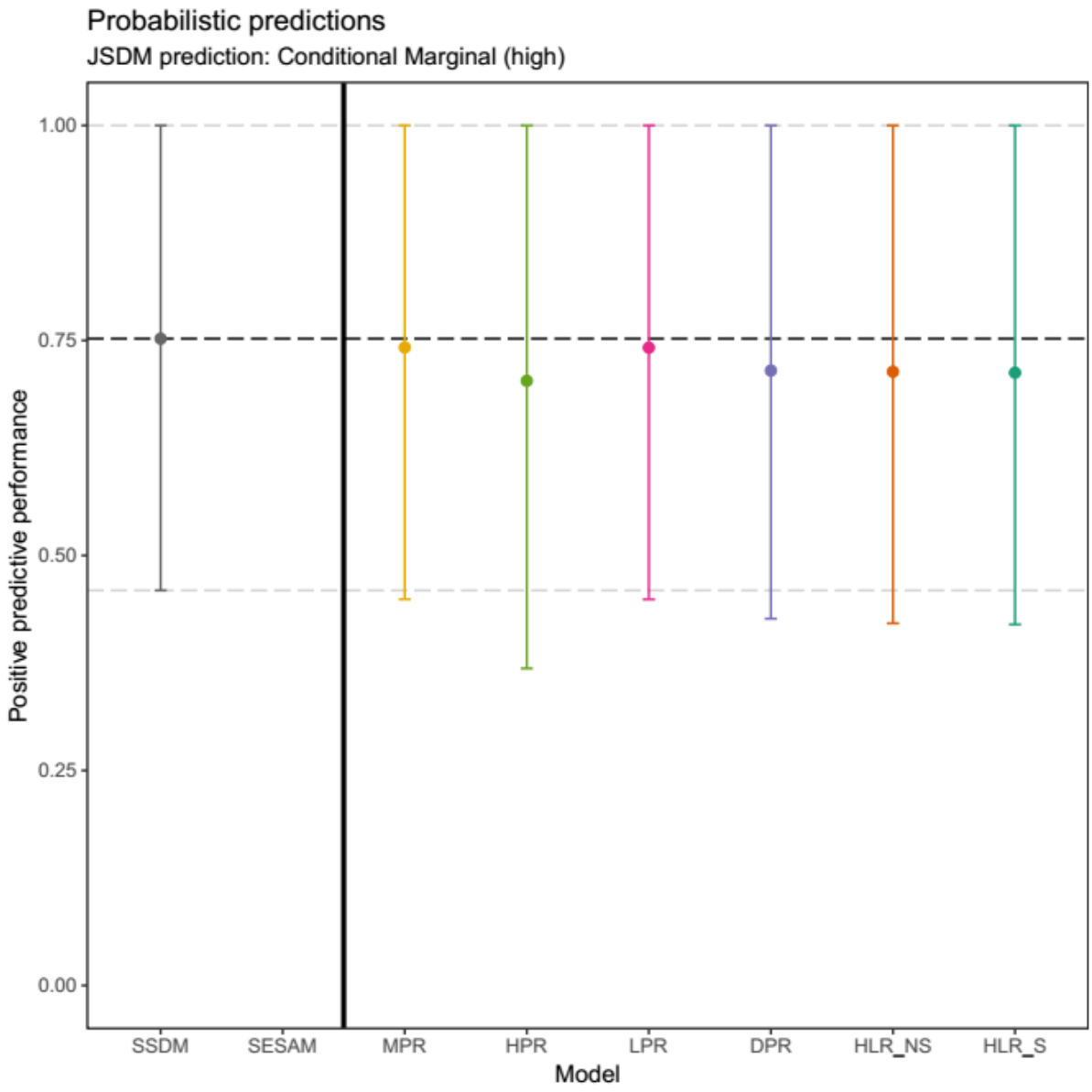


Figure 215: Model performance for the positive predictive performance evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

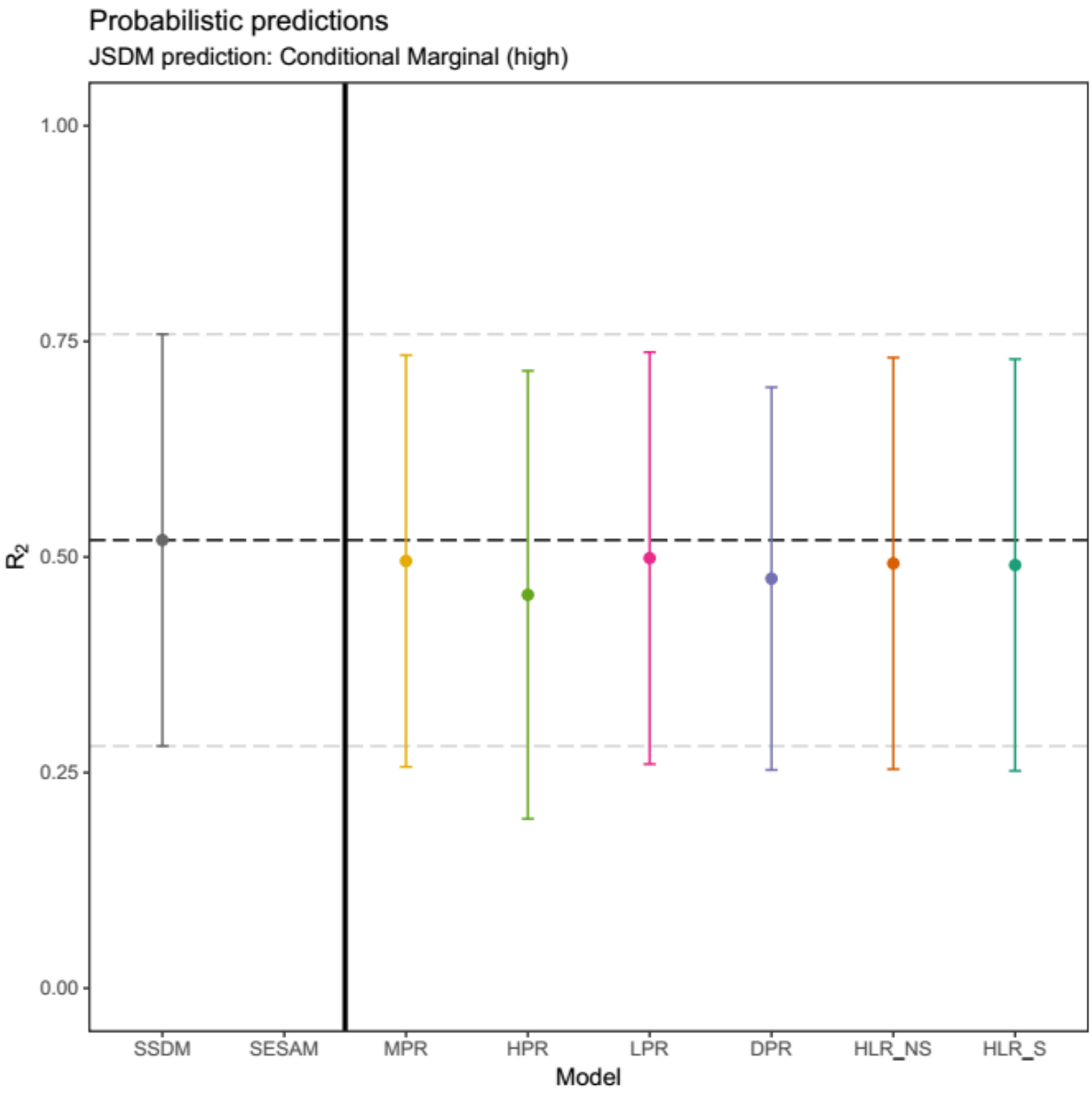


Figure 216: Model performance for the R^2 evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

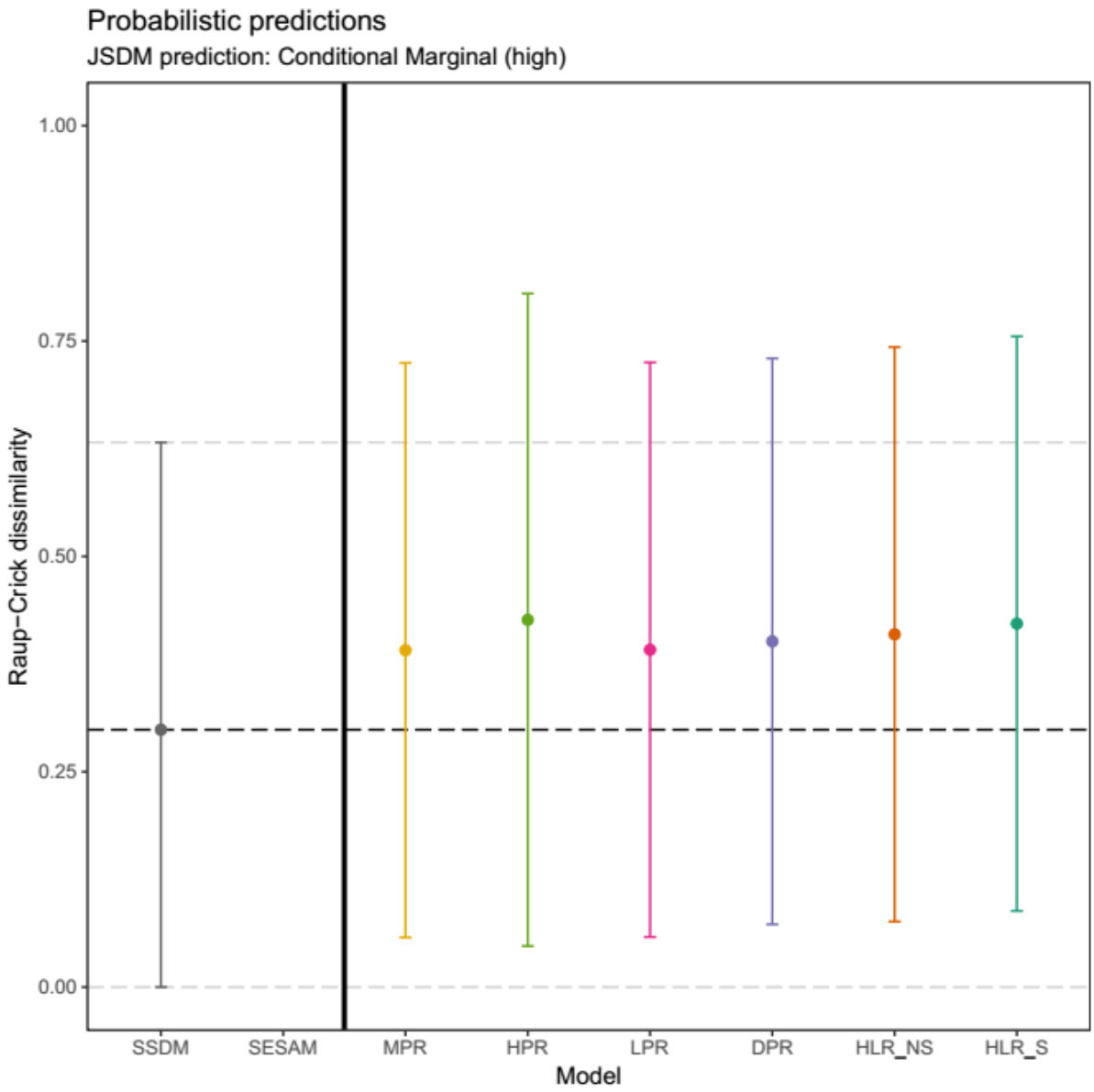


Figure 217: Model performance for the Raup-Crick dissimilarity evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

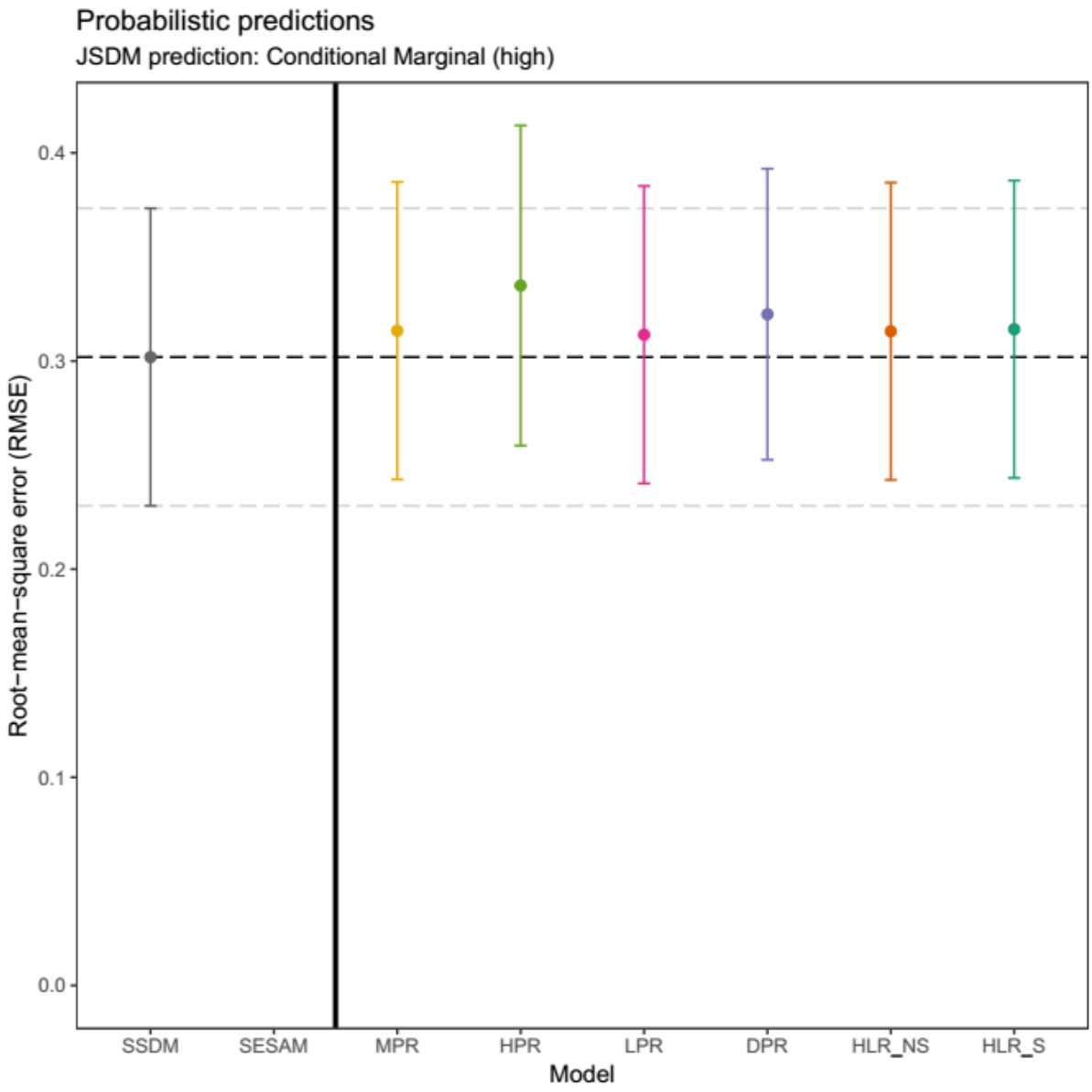


Figure 218: Model performance for the root-mean-square error evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

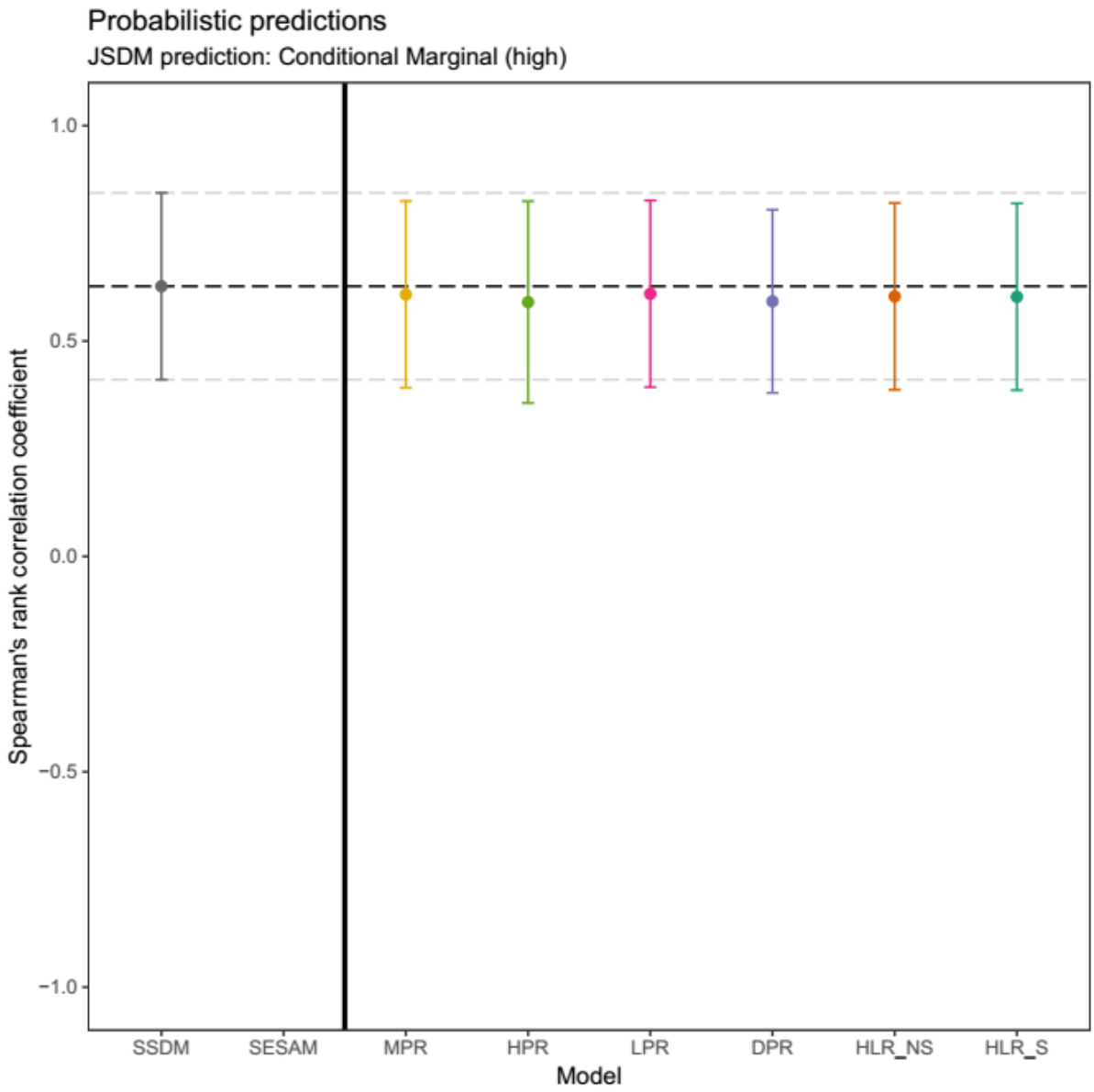


Figure 219: Model performance for the Spearman's rank correlation coefficient evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

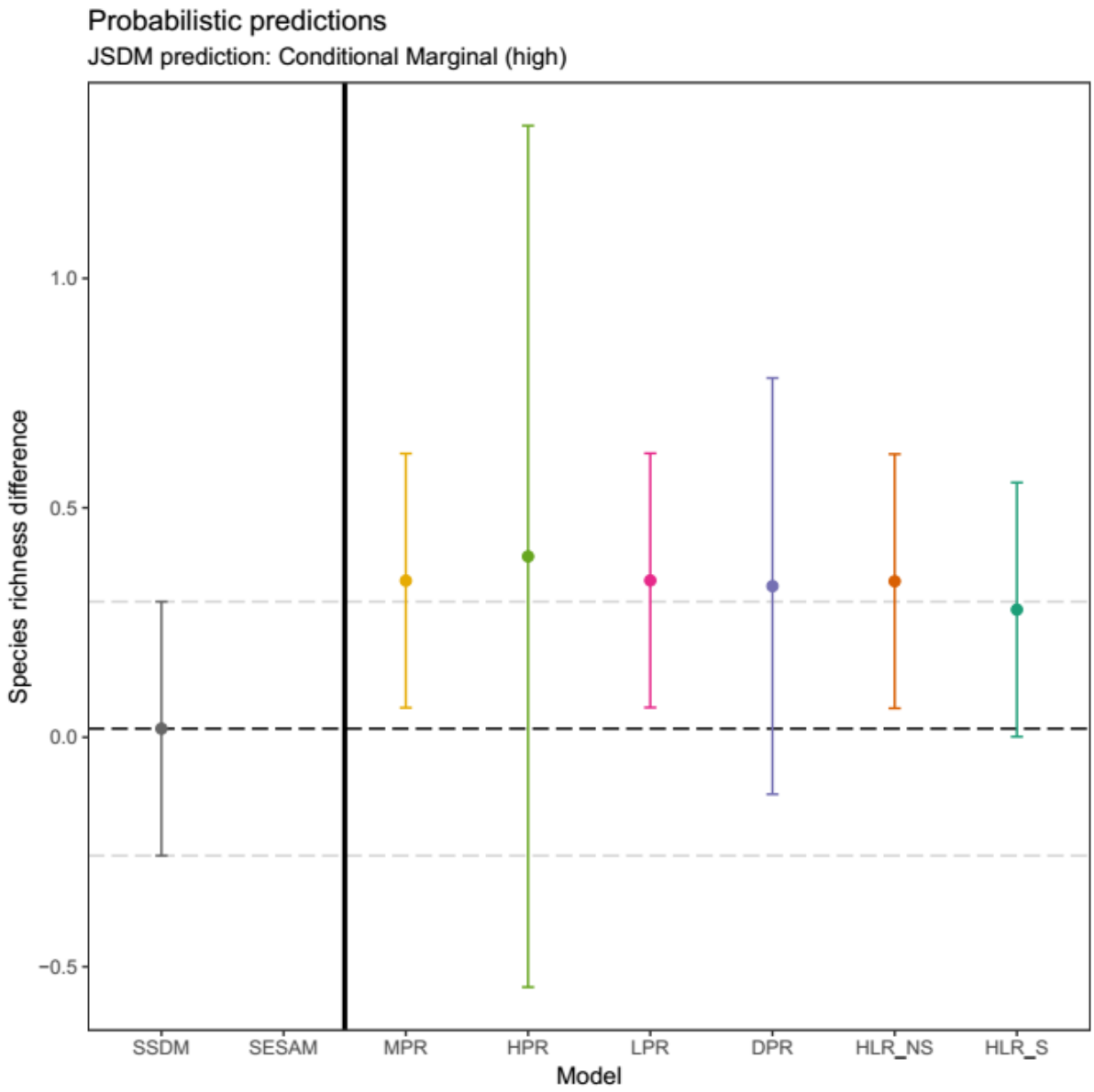


Figure 220: Model performance for the species richness difference evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

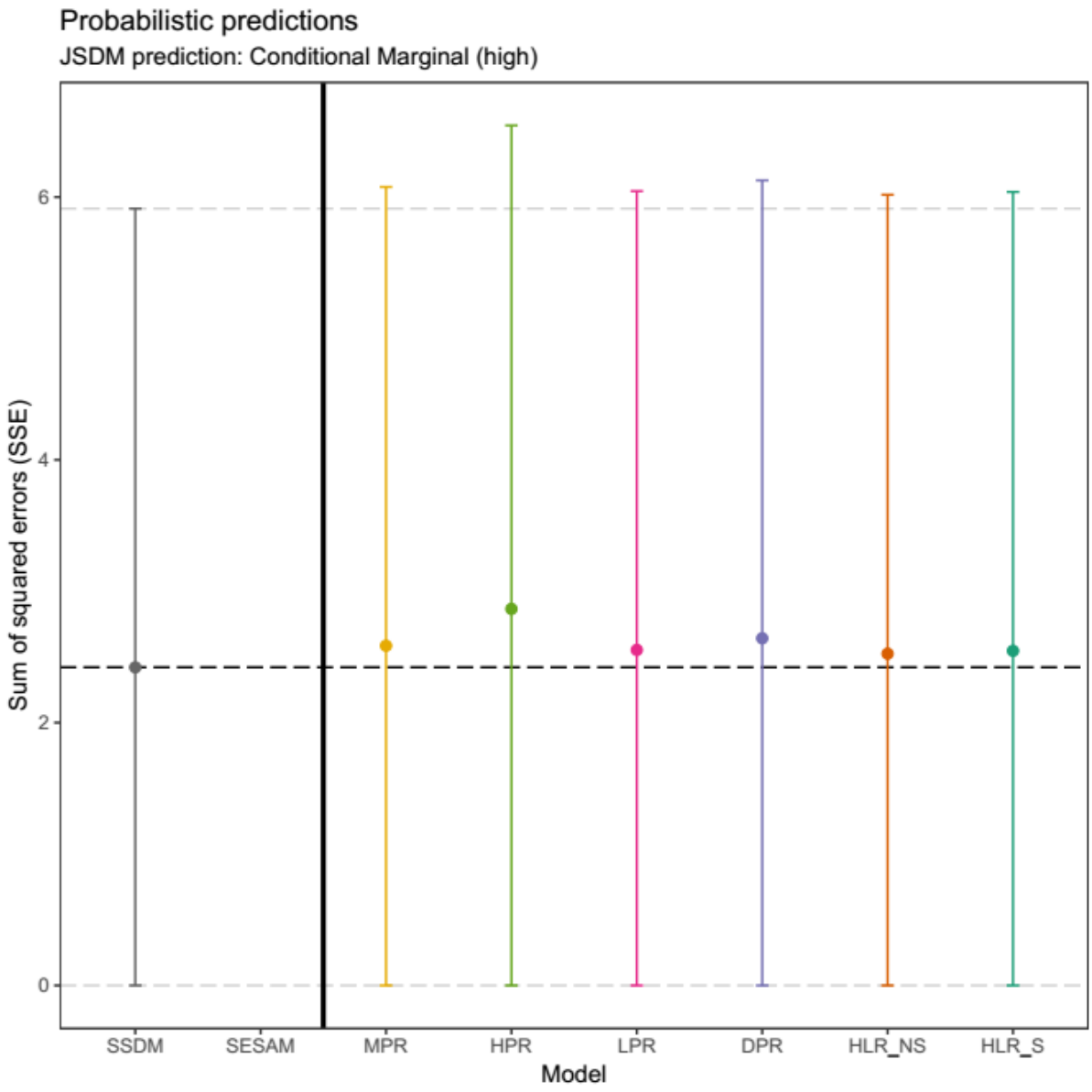


Figure 221: Model performance for the sum of squared errors evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

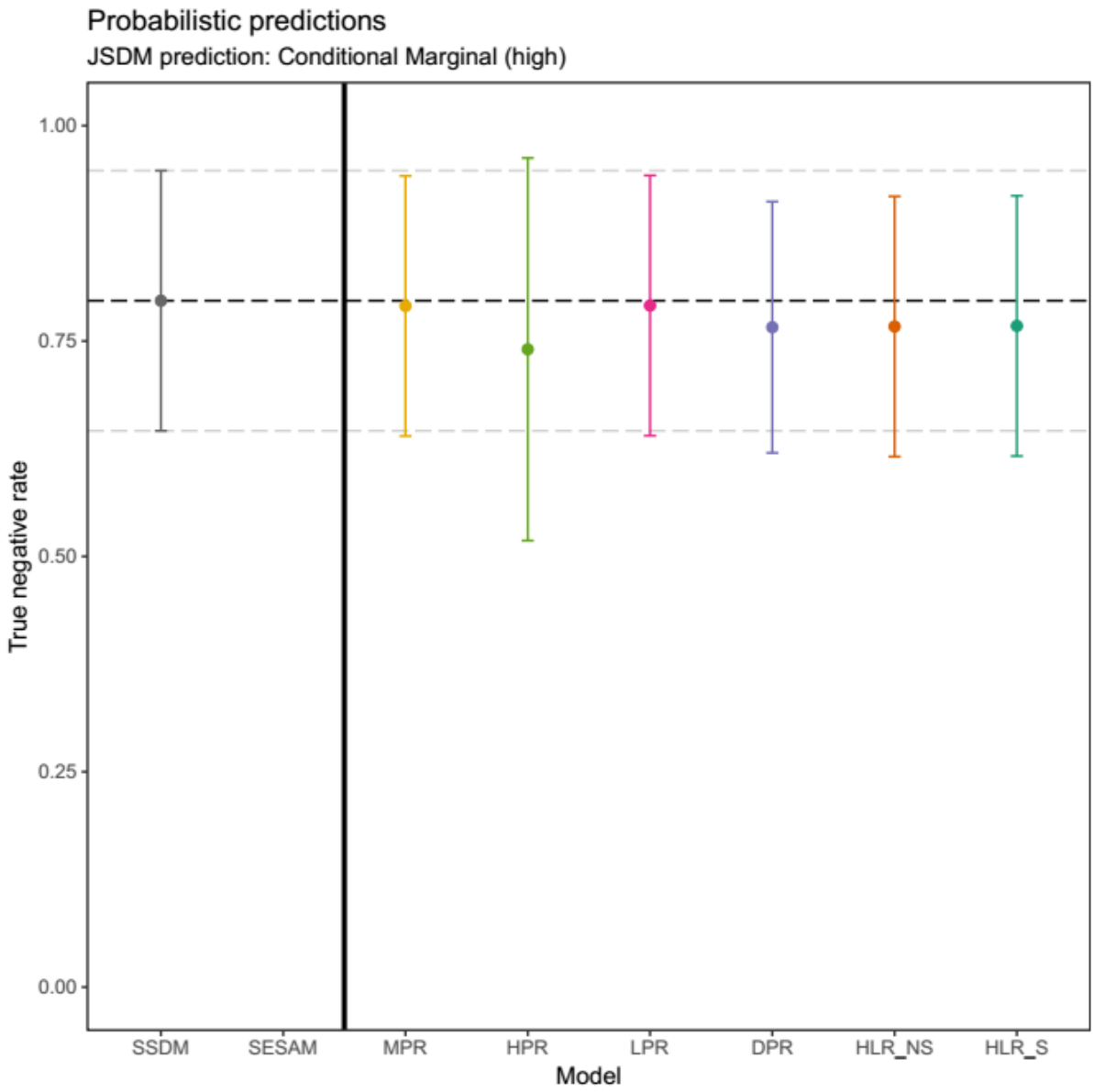


Figure 222: Model performance for the true negative rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

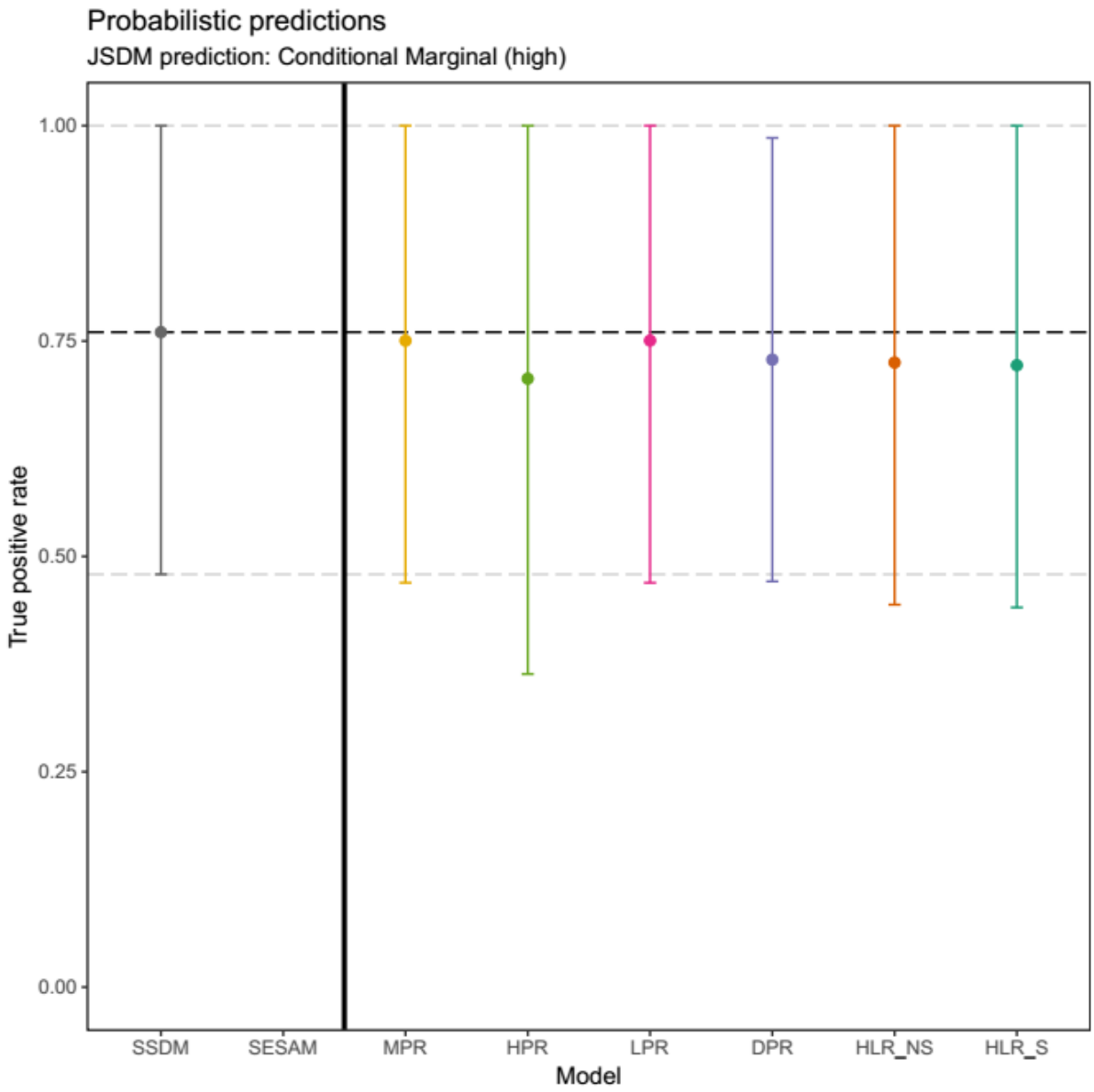


Figure 223: Model performance for the true positive rate evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

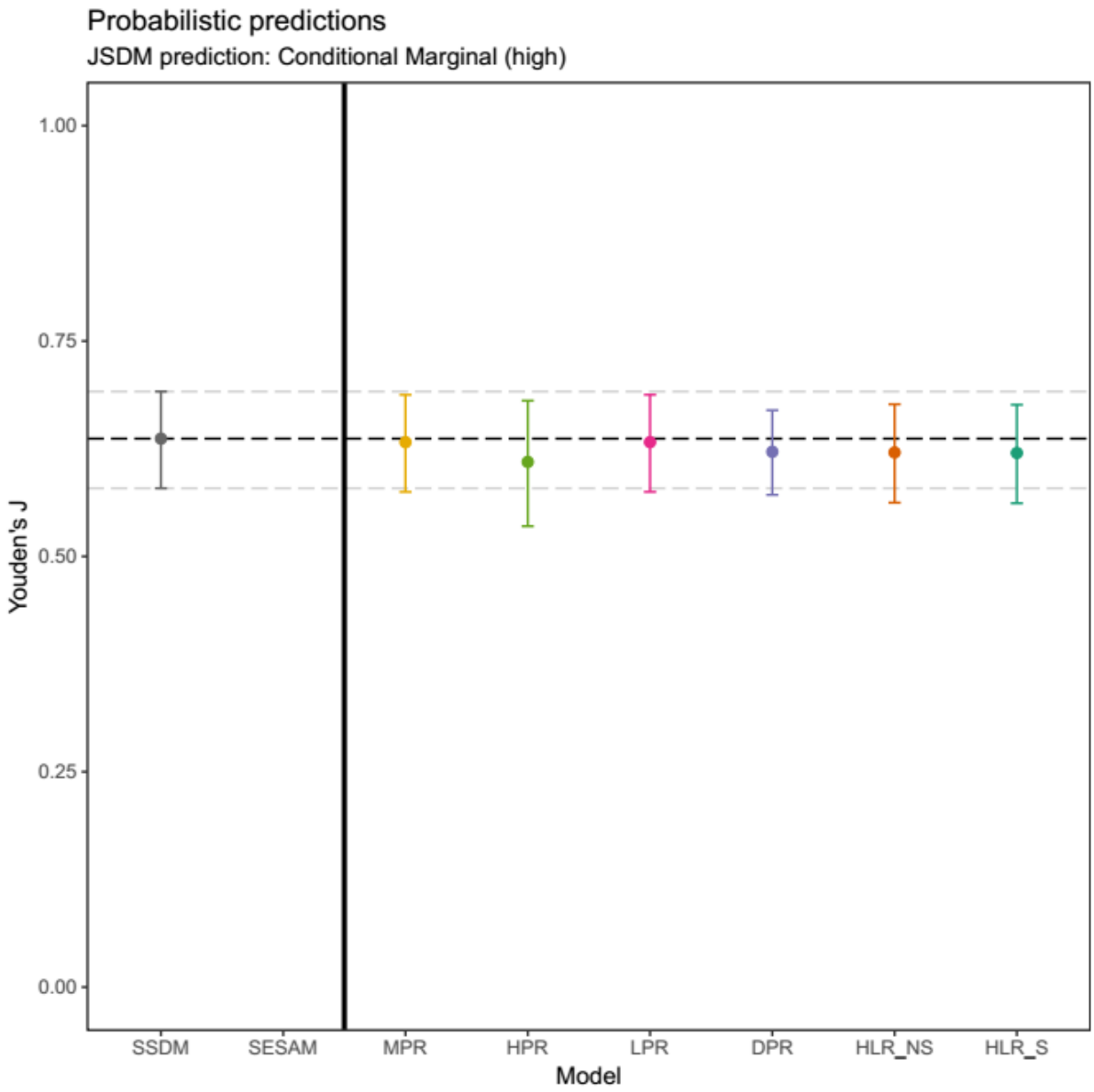


Figure 224: Model performance for the Youden's J evaluation metric for conditional marginal (high-prevalence) predictions. Performance estimates are shown as the mean and 95% confidence intervals calculated with a mixed effect model. Dotted dark grey line corresponds to the SSDM mean estimate, light grey line corresponds to the SSDM 95% confidence interval. Models to the left of the black vertical line are SSDMs, models to the right are JSDMs.

APPENDIX N

RESULTS: RELATIVE PERFORMANCE HEATMAPS

This appendix contains Circos plots for each prediction type showing the relative performance of the models compared to the standard SSDM. The relative performance is calculated as $((JSDM - SSDM)/|SSDM|) * 100$ and expressed as a percentage. A positive relative difference indicates the JSDM outperformed the SSDM, while a negative relative difference indicates that the SSDM outperformed the JSDM.

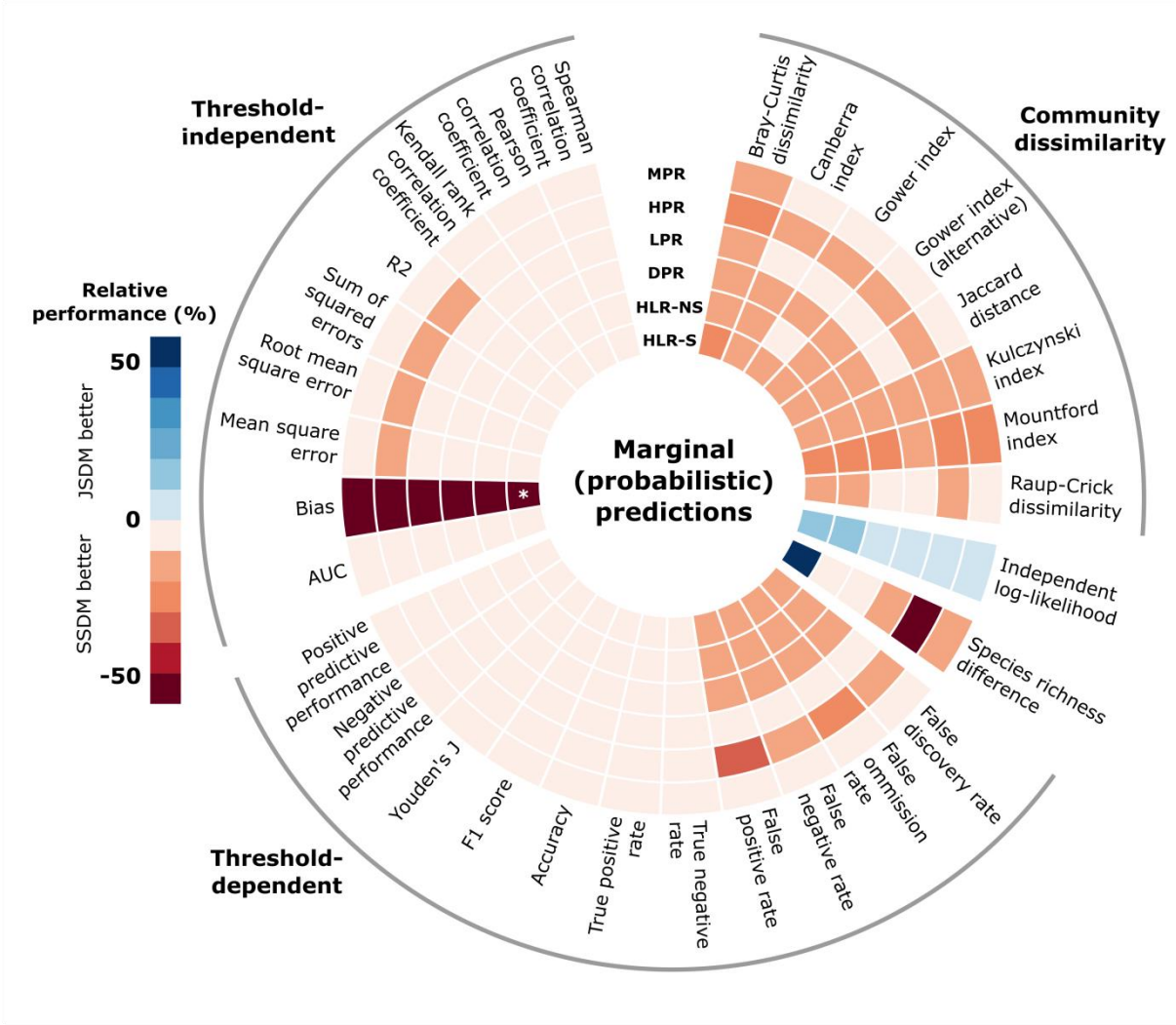


Figure S1: Relative evaluation metric performance of all models compared to the standard SSDM for probabilistic marginal predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

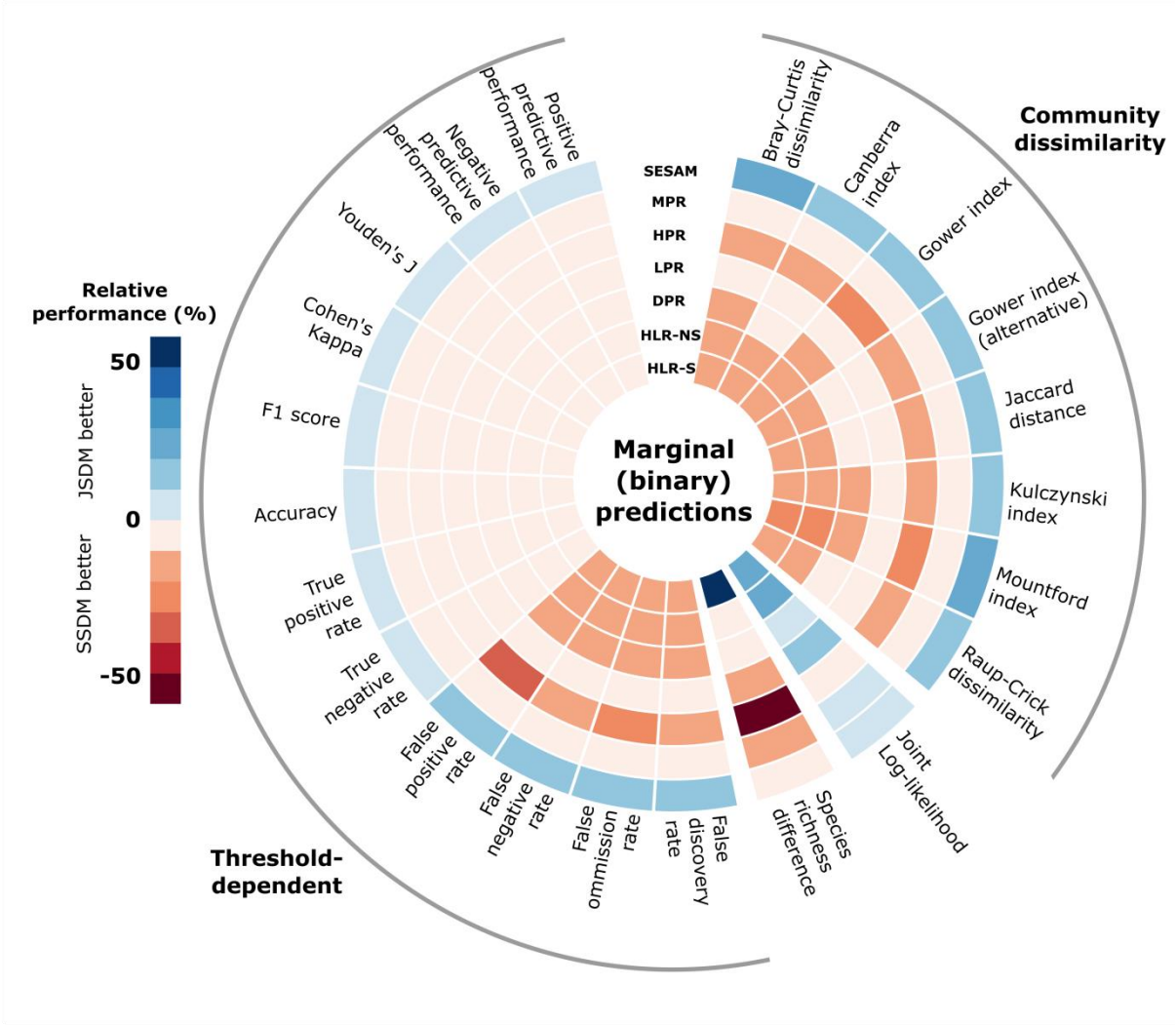


Figure S2: Relative evaluation metric performance of all models compared to the standard SSDM for binary marginal predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

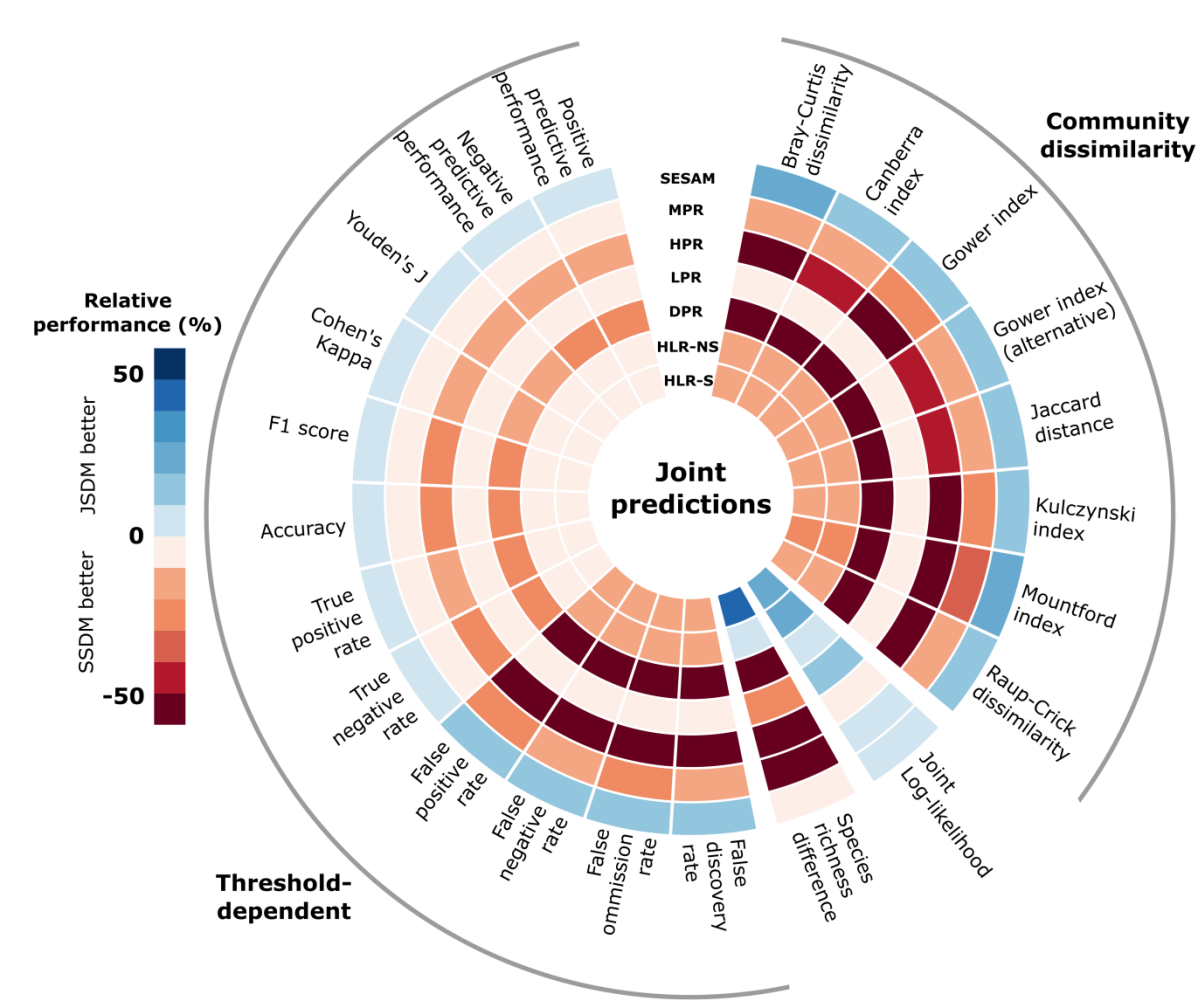


Figure S3: Relative evaluation metric performance of all models compared to the standard SSDM for joint predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

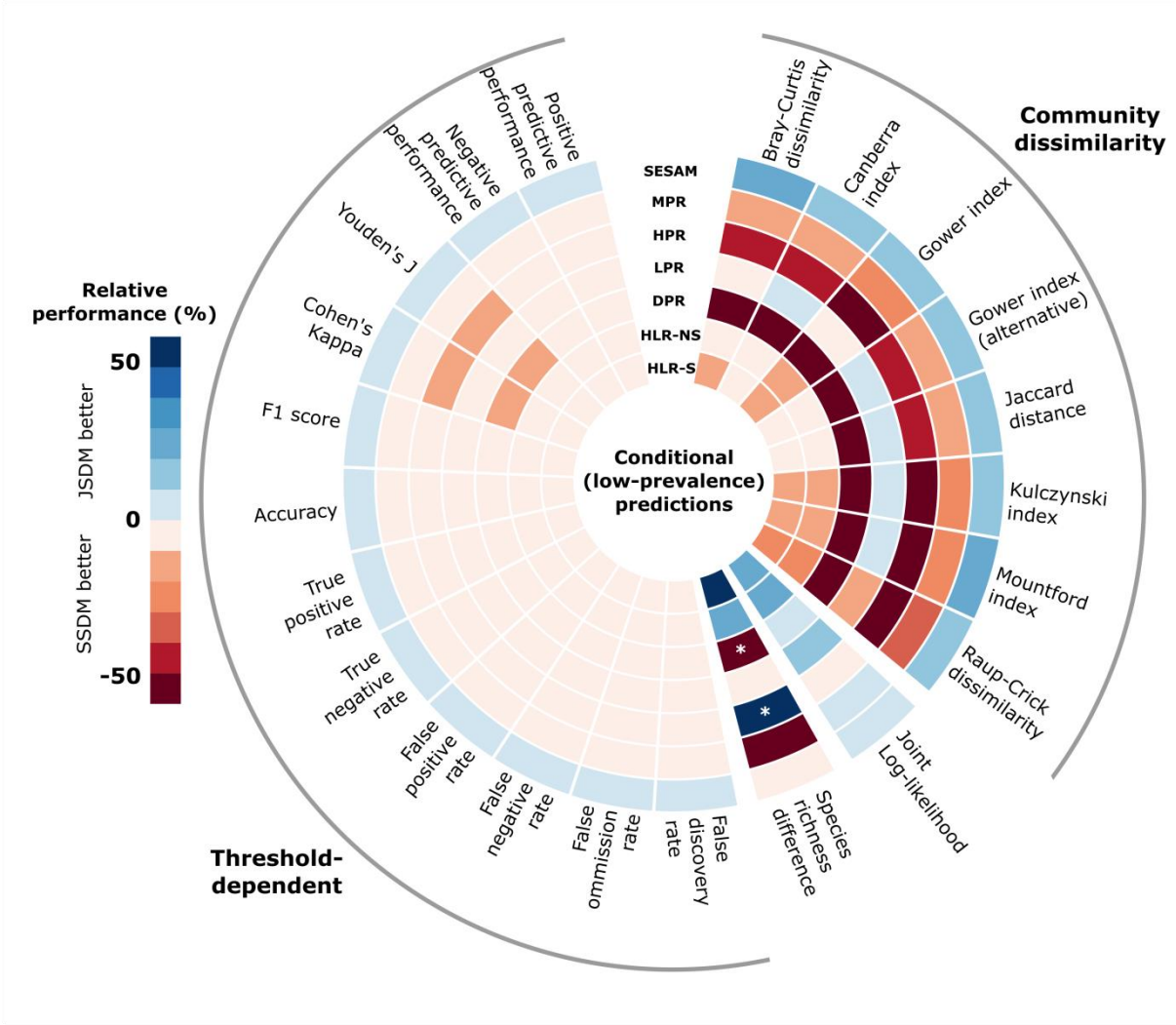


Figure S4: Relative evaluation metric performance of all models compared to the standard SSDM for conditional (low-prevalence) predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

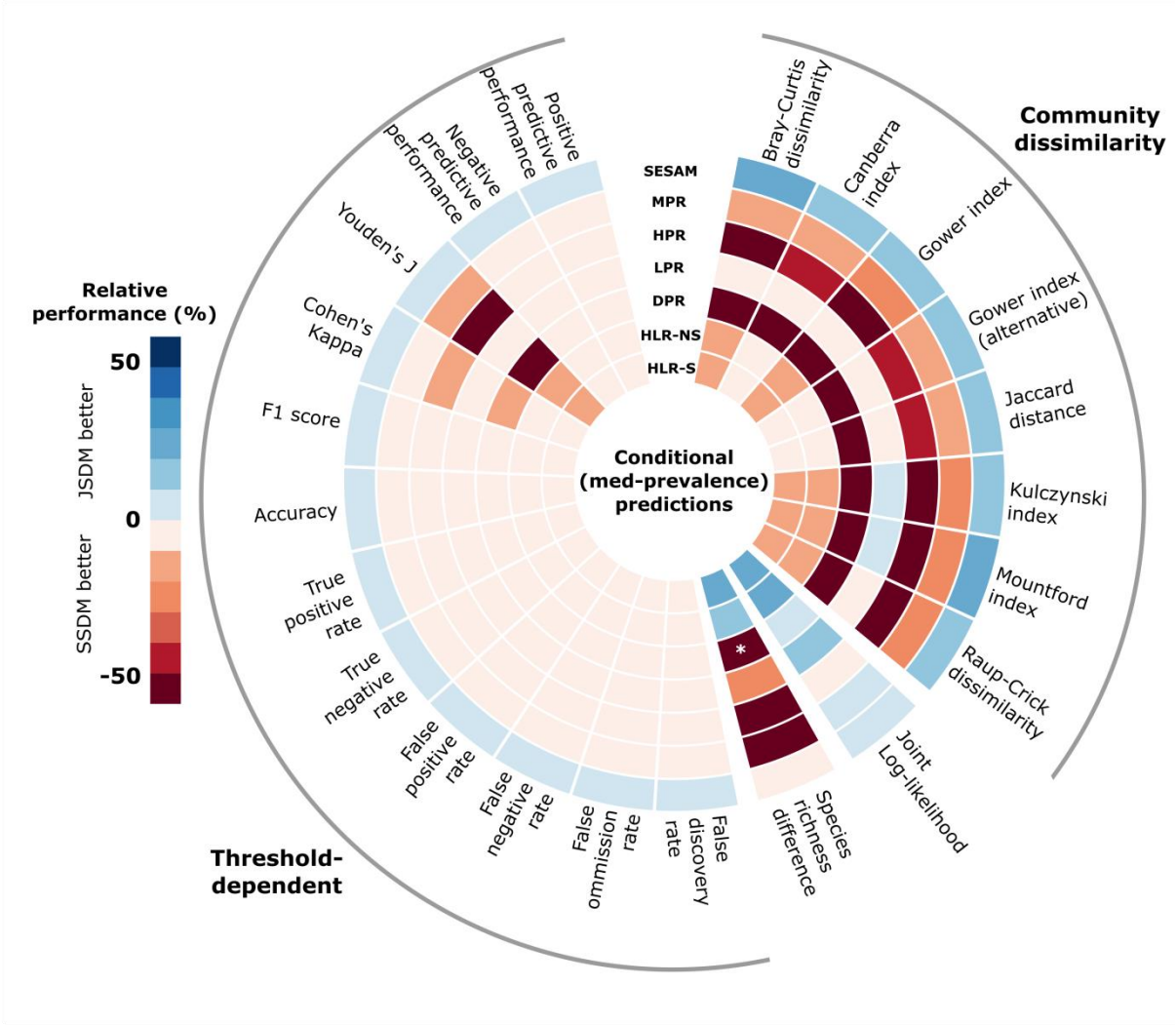


Figure S5: Relative evaluation metric performance of all models compared to the standard SSDM for conditional (medium-prevalence) predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

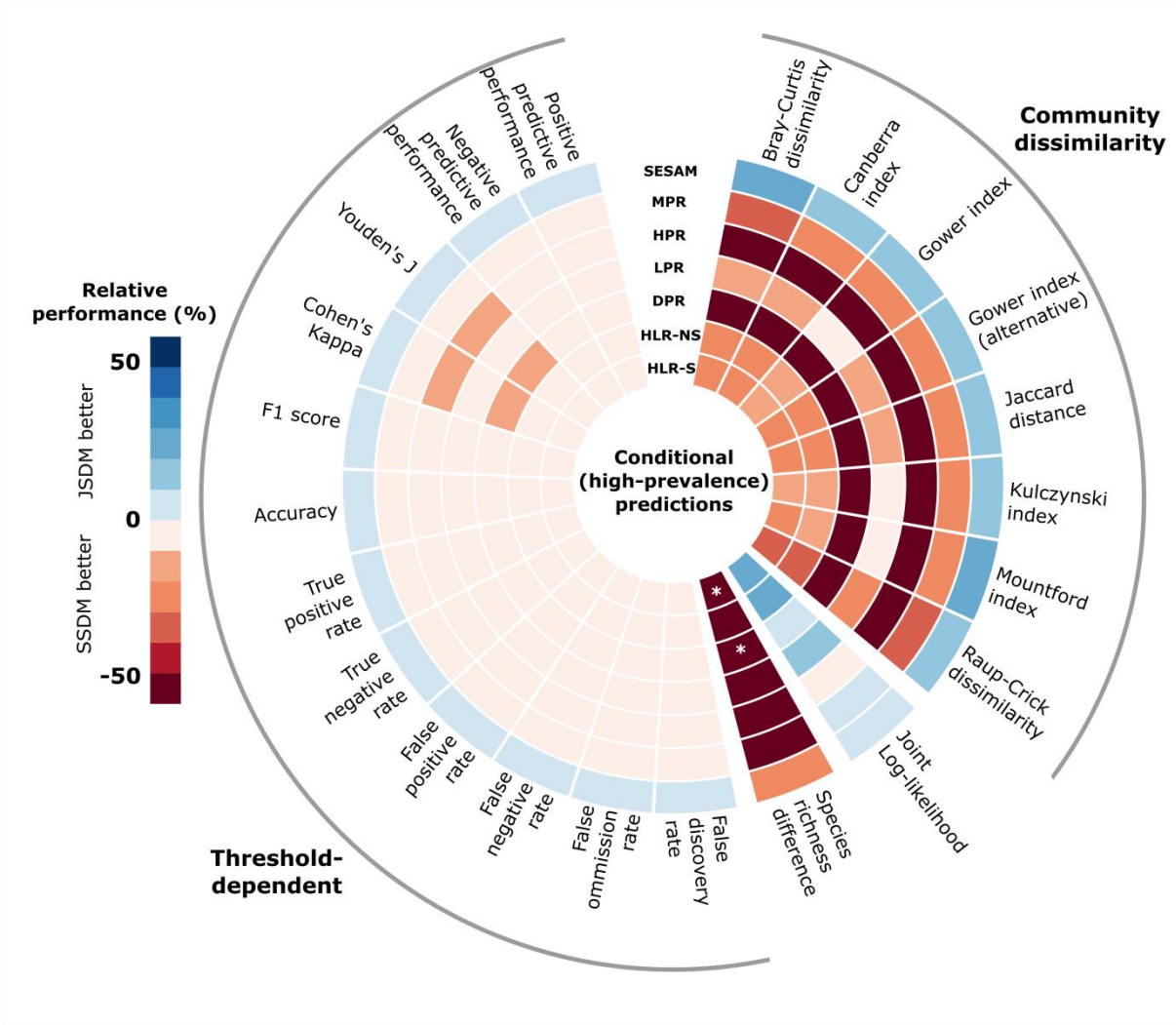


Figure S6: Relative evaluation metric performance of all models compared to the standard SSDM for conditional (high-prevalence) predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

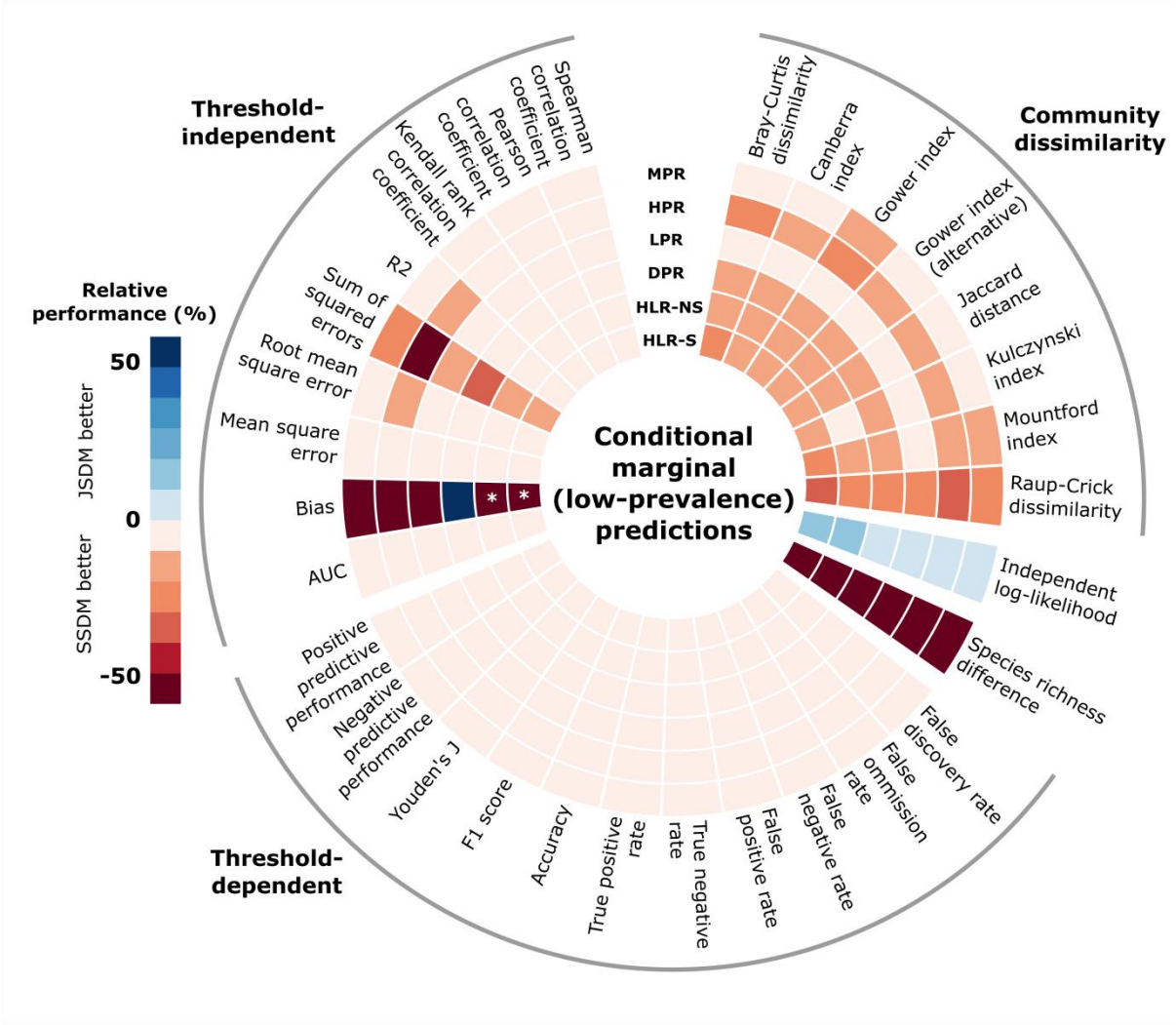


Figure S7: Relative evaluation metric performance of all models compared to the standard SSDM for conditional marginal (low-prevalence) predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

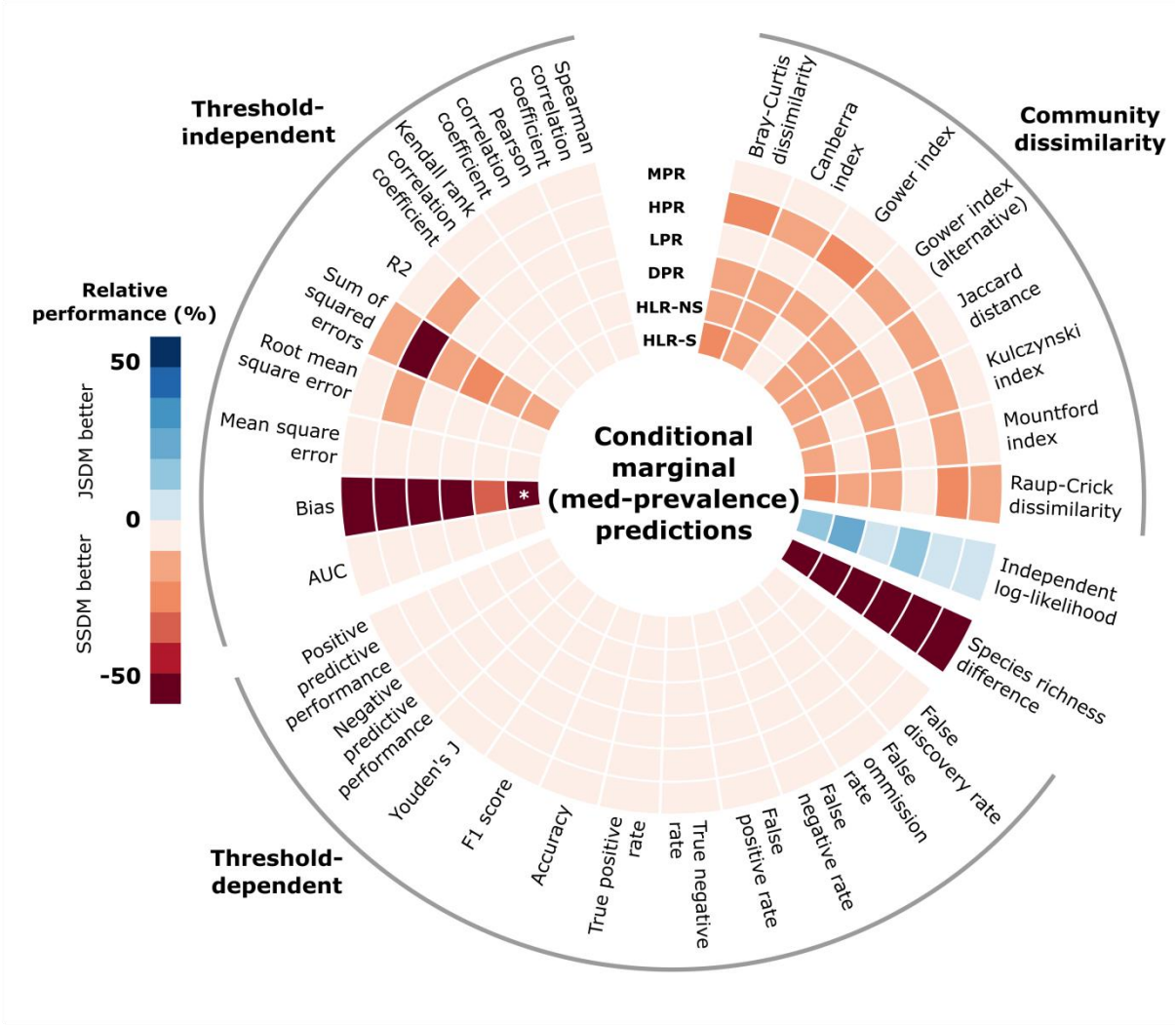


Figure S8: Relative evaluation metric performance of all models compared to the standard SSDM for conditional marginal (medium-prevalence) predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

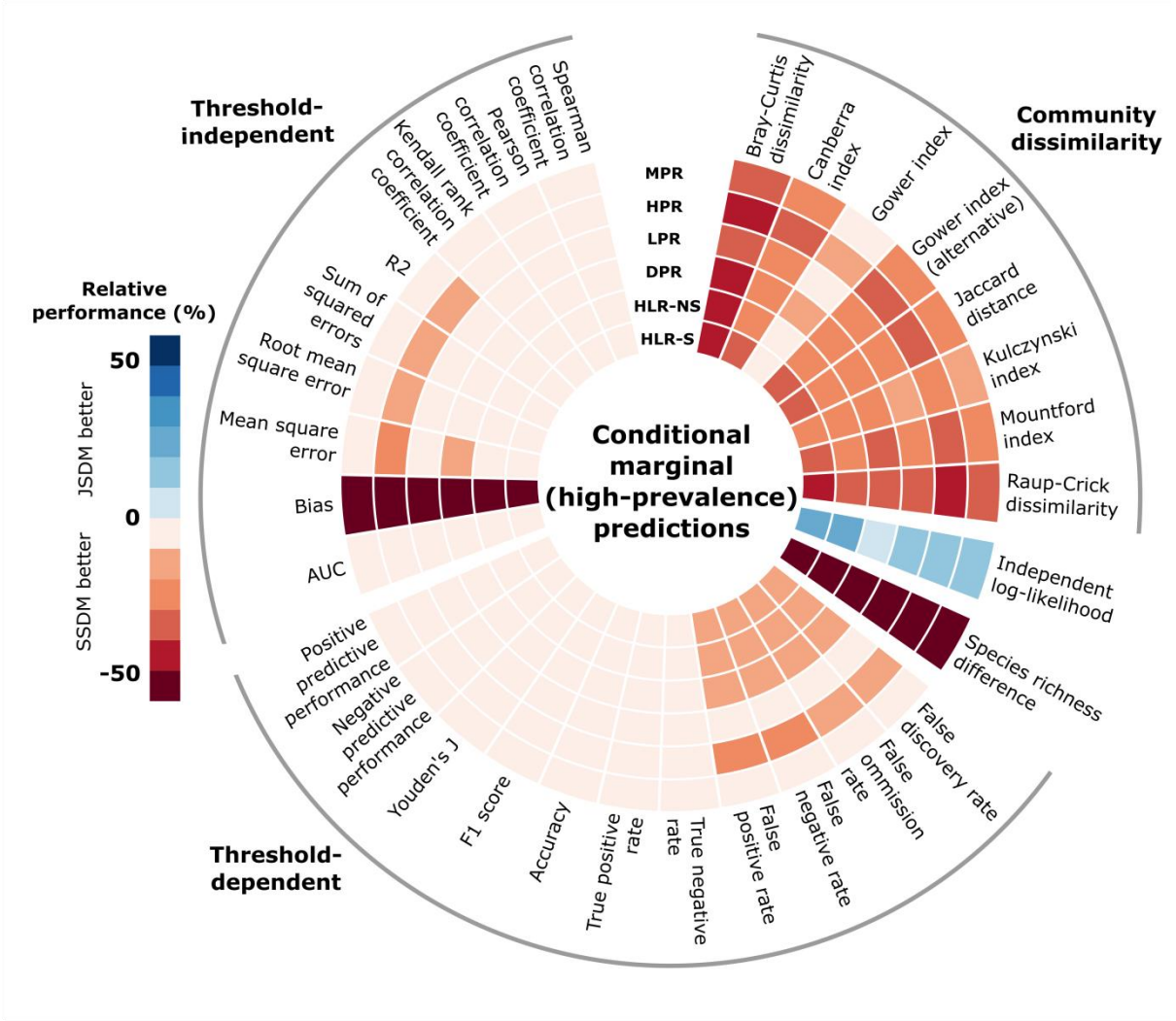


Figure S9: Relative evaluation metric performance of all models compared to the standard SSDM for conditional marginal (high-prevalence) predictions. Each ray is a single evaluation metric, and metrics are grouped according to the five metric classes defined in the main paper. Each ring corresponds to a specific model. Positive values indicate the model has outperformed the SSDM (blue), while negative values indicate the SSDM has outperformed the model (red). Differences are calculated using only the mean effect and do not account for uncertainty in the estimates. The species richness difference and bias metrics are evaluated between $-\infty$ and ∞ with an optimum of 0 so metric performance includes both the distance from 0 as well as if the model and SSDM estimates are in concordance. Models not in concordance with the SSDM estimate are marked with an asterix.

APPENDIX O

TESTING RESULTS FOR ERRORS

Due to the surprising nature of some of our results we endeavoured to determine if they were correct or an artefact of the mixed-effects model fitting process. To check whether the estimates from the MEMs were erroneous we compared them against contrasts calculated directly from the data used to fit them. We computed differences between each model and the SSDM at the level of each data point (site or species), taking the means of these differences across data points (and therefore folds) and then plotting the model-by-dataset combinations and the averages across datasets for each model.

Here we present the results from three of our more surprising results: independent log-likelihood and RMSE for probabilistic marginal predictions, and Bray-Curtis dissimilarity for joint predictions. In the below section we present the calculated contrasts on the left and the absolute values of the metric estimates on the right. Because the results are presented on a different scale the numbers are not directly comparable and we are just looking to identify the same patterns. We found no evidence to suggest that our MEMs were presenting erroneous results.

One key point is that these contrast plots show the spread of mean (across each dataset) differences between each model and SSDM, whereas the forest plots show, in their confidence interval whiskers, the spread of absolute values of the metric across sites or species. So the forest plots look like they are much less certain, but they are showing different things

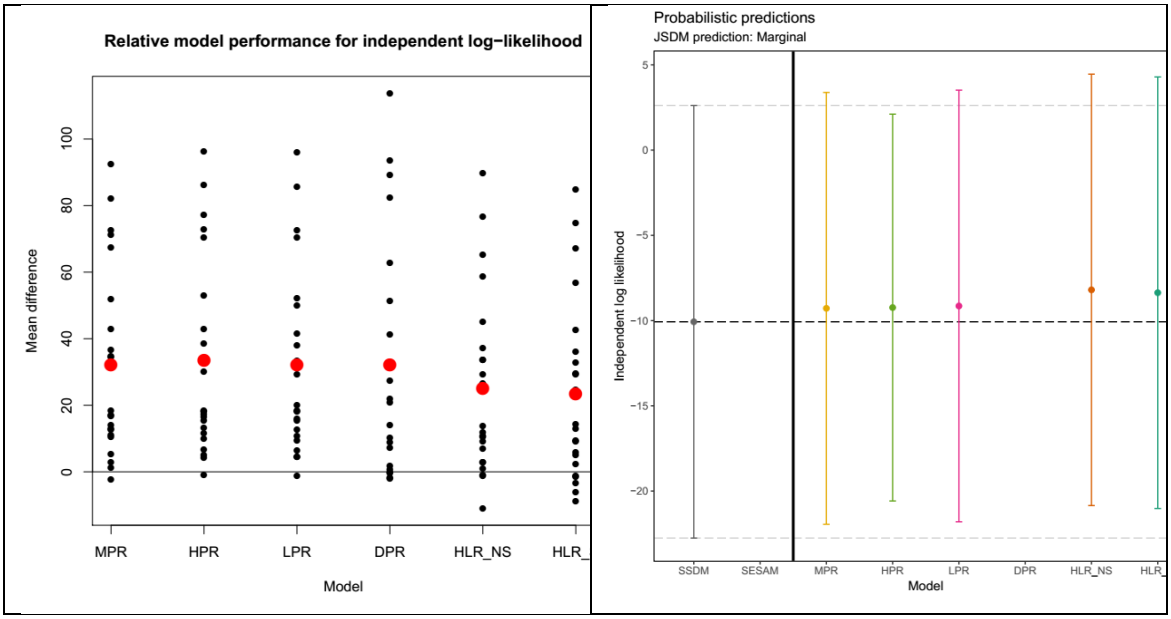


Figure S1: Calculated contrasts of mean difference (left) and mixed-effects model outputs (right) for independent log-likelihood for probabilistic marginal predictions

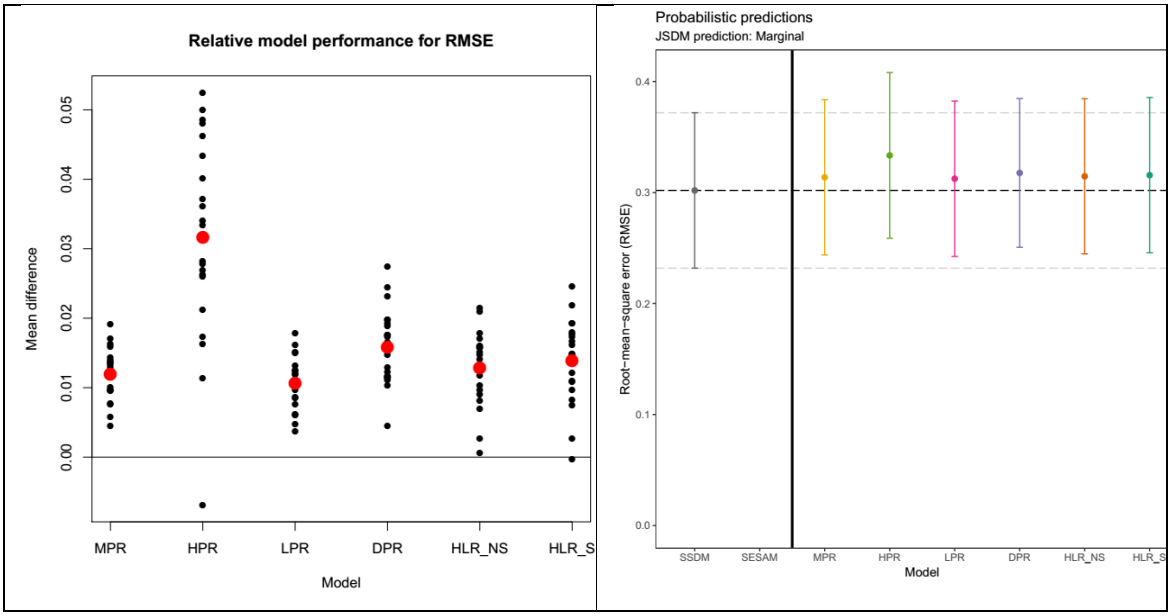


Figure S2: Calculated contrasts of mean difference (left) and mixed-effects model outputs (right) for RMSE for probabilistic marginal predictions

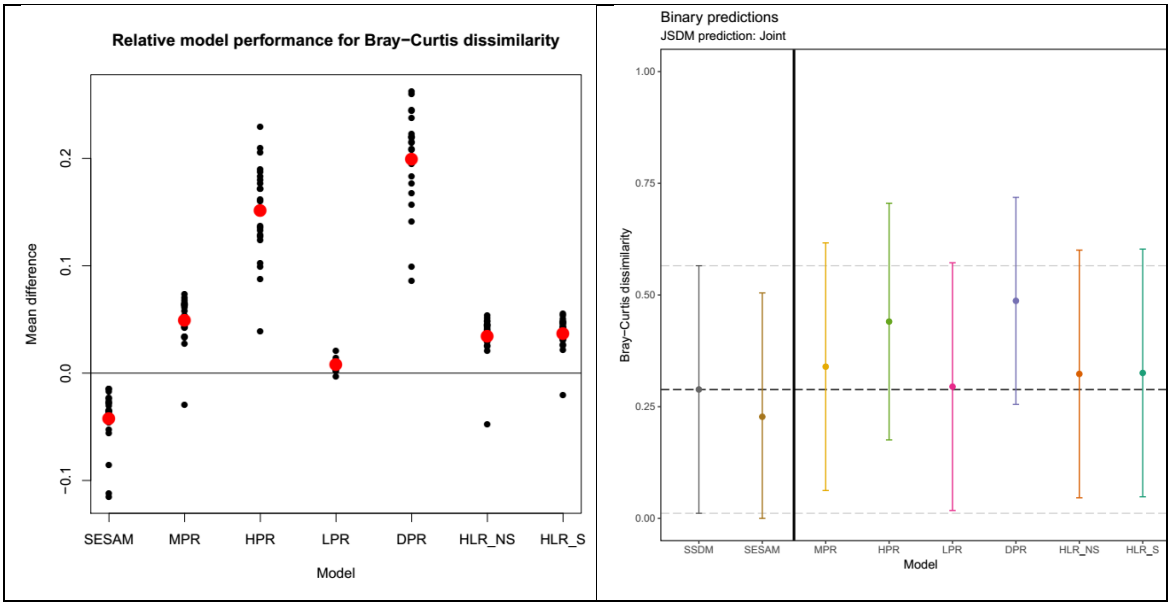


Figure S3: Calculated contrasts of mean difference (left) and mixed-effects model outputs (right) for Bray-Curtis dissimilarity for joint predictions