

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Mahar, Robert

Title:

Statistical modelling of lung function from multiple-breath washout and complex tidal flow waveform data

Date:

2019

Persistent Link:

<https://hdl.handle.net/11343/227058>

Terms and Conditions:

Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.

UNIVERSITY OF MELBOURNE

**Statistical modelling of lung function
from multiple-breath washout and
complex tidal flow waveform data**

by

Robert Mahar

A thesis submitted in total fulfilment of the
requirements for the degree of Doctor of Philosophy

February 2019

Faculty of Medicine, Dentistry, and Health Sciences
Department of Paediatrics

UNIVERSITY OF MELBOURNE

Statistical modelling of lung function from multiple-breath washout and complex tidal flow waveform data

Robert Mahar

 orcid.org/0000-0002-3643-923X

February 2019

Faculty of Medicine, Dentistry and Health Sciences

Department of Paediatrics

Abstract

Multiple-breath washout (MBW) and tidal flow waveform data, widely used to measure lung function in paediatric settings, are typically analysed using simple empirical summaries rather than statistical models. Yet the data have underlying functional forms which suggest model-based approaches may be useful. Novel approaches to modelling such data—from children with and without cystic fibrosis—were developed and applied to estimate various summaries of lung function using Bayesian and spectral analytical methods.

We developed a novel Bayesian statistical model for infant MBW data and applied it to tests from a large cohort of individuals from a major birth cohort study. We focused on the estimation of the lung clearance index (LCI) and functional residual capacity (FRC): commonly used summary measures of lung function. Our results show that the model provides an excellent fit to the data and leads to similar point estimates with comparable within-participant standard deviation compared to the traditional empirical approach. The strength of our approach was that it enabled us to estimate LCI using tests with different degrees of incompleteness, whereas traditional approaches require complete data. Our model therefore allows data that would routinely be discarded to be analysed, an important result because the MBW test is time-consuming, prone to disruption, and repeated measures are typically needed. This new approach leads to more efficient use of collected MBW data when estimating LCI and FRC, which can reduce the need for extensive and potentially costly replication in future studies. This work also provided valuable insight into the proper modelling of growth data with strict monotonicity constraints.

We identified the potential of spectral analytical methods to improve upon current methods for describing individual breath cycles. We developed and tested various estimators of the power-law exponent, which characterises the shape of the tidal flow waveform in the

frequency domain, to improve upon the existing estimation procedures and also to more fully understand the limitations of the method. In doing so we have demonstrated the statistical properties of the estimators which allowed us to apply them confidently to a large amount of data to describe the relationship between complex tidal flow waveforms and epidemiological covariates. This work will serve as both a valuable reference for other researchers using similar methods and as a useful starting point for further methodological refinement.

Building on a growing and topical body of literature, we also considered the effect of both software algorithms and breathing pattern variations on MBW outcomes. This information will allow respiratory scientists to better adapt and refine the MBW method and therefore gain a deeper understanding of lung function.

Our hope is that the findings of the research contained in this thesis will encourage and facilitate more sophisticated statistical analysis of lung function data. Through this we may begin to reveal novel characteristics of the underlying physiology to the ultimate benefit of people living with the consequences of lung diseases.

Declaration of Authorship

I, Robert Mahar, declare that this thesis—*Statistical modelling of lung function from multiple-breath washout and complex tidal flow waveform data*—and the work presented in it are my own. I confirm that:

- the thesis comprises only my original work towards the Doctor of Philosophy except where indicated in the preface,
- due acknowledgement has been made in the text to all other material used, and
- the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed: [Signature withheld]

Date: Wednesday 13th February 2019

Preface

This thesis embodies a program of research undertaken at the Clinical Epidemiology and Biostatistics Unit at the Murdoch Children’s Research Institute (MCRI) and the Department of Paediatrics, Faculty of Medicine, Dentistry, and Health Sciences, University of Melbourne. The seeds of many of the main ideas presented herein were planted upon completing a minor thesis for the partial fulfilment of the Master of Biostatistics, also while at MCRI: *Model-based estimation of the lung clearance index from complete and incomplete multiple breath washout data* (2014). The broad intention of the thesis was to develop and apply novel approaches to modelling lung function data that are typically analysed using simple empirical summaries rather than statistical models. My hope is that the work will encourage and facilitate more sophisticated analysis of lung function data to the ultimate benefit of people living with the consequences of lung diseases.

Several publications and conference presentations related to the research contained in this thesis were made over the course of my candidature. In each case I was the primary author, contributed to more than half of the content and was ultimately responsible for project planning, data management, statistical analysis including programming, initial drafting and contribution to the final manuscripts, with editorial, supervisory, and administrative support from my co-authors. Publications included:

- Mahar, RK, Vukcevic, D, King, L, Carlin, JB, Ranganathan, S. Lack of transparency in software used to analyze multiple breath washout data. *Pediatric Pulmonology* 51: 1108–1110 (2016). (See Section 2.5).
- Mahar, RK, Carlin, JB, Ranganathan, S, Ponsonby, AL, Vuillermin, P, Vukcevic D. Bayesian modelling of repeated measures lung function data from multiple-breath washout tests. *Statistics in Medicine* 37: 2016–2033 (2018). (See Chapter 4).

Conference presentations delivered included:

- Mahar, RK, Carlin, JB, Ranganathan, S, Vukcevic, D. Model-based estimation of the lung clearance index. *Murdoch Children's Research Institute Student Symposium, Parkville*. 15 September 2016.
- Mahar, RK, Carlin, JB, Ranganathan, S, Vukcevic, D. Model-based estimation of the lung clearance index. *Australian Statistical Conference, Canberra*. 7 Dec 2016.

My doctoral research was funded in part from a number of sources including an Australian Government Research Training Program Scholarship (formerly an Australian Postgraduate Award), the Victorian Centre for Biostatistics (ViCBiostat)—a National Health and Medical Research Council (NHMRC)-funded Centre of Research Excellence—and the 2015 MCRI Cystic Fibrosis Research Award.

Acknowledgements

Thank you to my supervisors Prof. John Carlin, Dr Damjan Vukcevic, and Prof. Sarath Ranganathan. Your commitment, guidance, and patience throughout my research journey has been invaluable and I will no doubt benefit from it for many years to come. Thanks also go to my Advisory Committee Chair, Prof. Andrew Davidson, and Senior Advisory Committee Member, Prof. Lyle Gurrin, for your engaging advice and feedback on the big picture of my candidature.

I would also like to thank all of the collaborators and facilitators of my research. In particular, Prof. Pete Vuillermin for opening the doors to the Barwon Infant Study for me and, along with Prof. Anne-Louise Ponsonby, making valuable contributions as co-authors, Dr Lawrence Gray for his crucial input on the influence of breathing patterns on multiple-breath washout outcomes, Louise King and Billy Skoric for getting the data to me and patiently fielding a weighty tome of requests for clarification over the years, Prof. Julie Simpson for giving me the opportunity to teach and to also to get involved in interesting research outside of my thesis, Prof. Urs Frey for his encouraging advice on modelling tidal flow waveforms, Prof. Bob Carpenter for his comments on several of my initial probabilistic models (and, more generally, thanks to the many contributors to the probabilistic programming language Stan who are making complex Bayesian analyses almost fun to do), and Hilianty Tardjono for assistance in automating the data extraction procedure. Importantly, I would like to thank the investigators and participants behind the Australian Respiratory Early Surveillance Team for Cystic Fibrosis and the Barwon Infant Study, my research was only made possible because of your world-class effort and dedication.

I would like express gratitude to my colleagues at the Clinical Epidemiology and Biostatistics Unit—for making my time around the office an enjoyable one—and at the Victorian Centre for Biostatistics—for the many learning and funding opportunities, statistical brainstorming sessions, and fun and supportive environment.

Thank you to all my friends and family both near and far, in particular and most importantly, I am forever indebted to my wife Greer—who despite her own gruelling schedule has shown what can only be described as supernatural support, love, and encouragement throughout what has been a hectic number of years—and my daughter, Asher, and son, Oakley, who both arrived during this journey of discovery and together have been a constant source of inspiration.

Contents

Abstract	i
Declaration of Authorship	iii
Preface	iv
Acknowledgements	vi
List of Figures	x
List of Tables	xii
List of Listings	xv
Abbreviations	xv
1 Introduction	1
1.1 Background	1
1.2 Proposed research	3
1.3 Aims and objectives	4
1.4 Thesis structure	5
2 Data context	7
2.1 Multiple-breath washout (MBW)	8
2.1.1 MBW equipment	8
2.1.2 MBW manoeuvre	9
2.1.3 Lung clearance index (LCI)	12
2.2 Barwon Infant Study (BIS)	14
2.2.1 Data collection and characteristics	16
2.2.1.1 Infant MBW data	16
2.2.1.2 Preschool MBW data	17
2.2.1.3 Tidal flow recordings	18
2.3 Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF)	19
2.3.1 Data collection and characteristics	19
2.4 Software for MBW data processing and its limitations	20
2.5 Publication	22
3 Effect of variations in ideal breathing patterns on MBW outcomes	26

3.1	Background	27
3.1.1	Aims	29
3.2	Methods	29
3.2.1	Data	29
3.2.2	Inclusion rules	29
3.2.3	Statistical methods	36
3.3	Results	36
3.4	Discussion	40
4	Modelling MBW and estimating LCI from incomplete tests	42
4.1	Background	42
4.2	Publication	43
5	Modelling the spectral densities of tidal flow waveforms	63
5.1	Background	64
5.1.1	Aims	68
5.2	Models	68
5.2.1	Time and frequency domains	69
5.2.2	Power-law spectra and scale-invariance	69
5.3	Estimation	71
5.3.1	Time and frequency domains	72
5.3.2	Spectral density estimation	77
5.3.3	Power-law exponent estimation	82
5.4	Simulation	85
5.4.1	Data generation	85
5.4.2	Performance evaluation	88
5.4.3	Results	88
5.5	Discussion	95
6	Epidemiology of tidal flow power-law spectra	97
6.1	Background	98
6.1.1	Aims	99
6.1.2	Limitations	99
6.2	Methods	102
6.2.1	Data	102
6.2.2	Models	110
6.2.3	Estimation	112
6.3	Results	114
6.3.1	Power-law spectra	114
6.3.2	Model of age and cystic fibrosis	115
6.3.3	Model of BIS power-law exponents and covariates	120
6.3.4	MBW and non-MBW agreement analysis	130
6.4	Discussion	130
7	Discussion	135
7.1	Summary of research	135
7.1.1	Data collection procedures	136
7.1.2	Effect of breathing pattern on MBW outcomes	136

7.1.3	Modelling MBW data	139
7.1.4	Modelling the spectral densities of tidal flow waveforms	142
7.1.5	Epidemiology of tidal flow power-law spectra	144
7.2	Conclusion	145
Bibliography		147
A Computational methods		159
A.1	Bayesian data analysis	159
A.1.1	Bayesian principles	160
A.1.2	Motivation	161
A.2	Markov chain Monte Carlo methods	161
A.2.1	Gibbs sampling	162
A.2.2	Metropolis algorithm	163
A.2.3	Hamiltonian Monte Carlo (HMC)	163
A.2.4	Performing MCMC and diagnostics	164
A.2.5	Model checking	166
A.3	Computational tools	167
A.3.1	Stan	167
A.3.2	Ancillary computational resources	169
A.3.2.1	The R programming environment	169
A.3.2.2	Version control system	170
A.3.2.3	Murdoch Children's Research Institute (MCRI) High Performance Cluster	170
A.3.2.4	Unix-based computing	171
B Stan model of MBW outcomes and breathing patterns		172
C Stan models of MBW data and agreement		174
D Power law simulation in R		189
E Stan models for random effects and paired comparisons.		191
F Regression and analysis of variance results from Chapter 6		198

List of Figures

2.1	Example of raw MBW data	10
2.2	Example of end-tidal MBW data	11
3.1	Boxplot of tests satisfying broad validity criteria	38
3.2	Estimates of the variance component model fitted according to MBW validity categories	39
5.1	Examples of tidal flow waveforms	65
5.2	The Koch curve: an example of scale invariance	71
5.3	Periodograms corresponding to tidal flow waveforms	73
5.4	Simulation scheme for time series with an underlying power-law spectrum .	86
5.5	Simulated time signals with an underlying power-law spectrum	91
5.6	Periodogram applied to simulated time signals with an underlying power-law spectrum	92
5.7	Results from estimation of power-law exponent from simulated data	93
5.8	Fitted ordinary least squares estimates of power-law exponents from simulated data	94
6.1	Individual breath cycles identified within waveforms	104
6.2	Power-law exponents estimated from AREST CF and BIS tidal flow data .	109
6.3	Power-law exponents estimated from individual breath cycles	118
6.4	AREST CF data and posterior predictive distributions	121

6.5	BIS data and posterior predictive distributions	122
6.6	Scatterplot of infant and preschool power-law exponents.	123
6.7	Agreement plots comparing estimated power-law exponents	131

List of Tables

2.1	Data collection and patient characteristics	15
3.1	MBW validity criteria	32
6.1	Physiological characteristics of study participants	105
6.2	Characteristics of BIS MBW participants.	106
6.3	Summary of posterior distributions and convergence diagnostics. Sensitivity analysis.	119
6.4	Parameter estimates from agreement model	131
F.1	Regression results of BIS power-law exponents and gender.	199
F.2	Regression results of BIS power-law exponents and gestational age.	200
F.3	Regression results of BIS power-law exponents and birthweight.	201
F.4	Regression results of BIS power-law exponents and mode of delivery.	202
F.5	Regression results of BIS power-law exponents and delivering hospital.	203
F.6	Regression results of BIS power-law exponents and mother's age.	204
F.7	Regression results of BIS power-law exponents and father's age.	205
F.8	Regression results of BIS power-law exponents and mother's highest education.	206
F.9	Regression results of BIS power-law exponents and father's highest education.	207
F.10	Regression results of BIS power-law exponents and number of siblings.	208
F.11	Regression results of BIS power-law exponents and household yearly income.	209

F.12 Regression results of BIS power-law exponents and any maternal cigarette smoking in previous 12 months.	210
F.13 Regression results of BIS power-law exponents and extent of maternal cigarette smoking.	211
F.14 Regression results of BIS power-law exponents and exposure to passive smoke in previous 12 months.	212
F.15 Regression results of BIS power-law exponents and household pets.	213
F.16 Regression results of BIS power-law exponents and household livestock. . .	214
F.17 Regression results of BIS power-law exponents and first degree family history of asthma.	215
F.18 Regression results of BIS power-law exponents and first degree family history of eczema.	216
F.19 Regression results of BIS power-law exponents and first degree family history of hayfever.	217
F.20 Analysis of variance of BIS power-law exponents and covariates	218

List of Listings

A.1	Stan code corresponding to a simple linear regression model	168
B.1	Stan code for variance components model described by (3.1)	173
C.1	Stan code corresponding to <code>mbw_individual_diffuse.stan</code>	177
C.2	Stan code corresponding to <code>mbw_individual_informative.stan</code>	181
C.3	Stan code corresponding to <code>lme_exchangeable_seperate.stan</code>	185
C.4	Stan code corresponding to <code>lme_linked_common.stan</code>	187
D.1	R functions to simulate a noisy signal with a known power law spectrum . .	189
E.1	Stan code for linear random effects model described by (6.1) and (6.2) . . .	194
E.2	Stan code for paired comparison model described by (6.4)	196

Abbreviations

Ar	Argon
AREST CF	Australian Respiratory Early Surveillance Team for Cystic Fibrosis
BIS	Barwon Infant Study
BLUE	Best linear unbiased estimator
BTPS	Body temperature pressure saturated
BUGS	Bayesian inference using Gibbs sampling
CF	Cystic fibrosis
CI	Confidence interval or credible interval
CO₂	Carbon dioxide
CEV	Cumulative expired volume
CEVGM	Cumulative expired volume of gas mixture
CLD	Chronic lung disease
DFA	Detrended fluctuation analysis
CEVTG	Cumulative expired volume of tracer gas
fBm	Fraction Brownian motion
fGn	Fractional Gaussian noise
FRC	Functional residual capacity
GAS	Alveolar SF ₆ gas concentration
GUI	Graphical user interface
He	Helium
HMC	Hamiltonian Monte Carlo
HPC	High performance (computing) cluster
JAGS	Just another Gibbs sampler
LoA	Limits of agreement

LCI	Lung clearance index
LCE	Log chi-square error estimation
MBW	Multiple-breath washout
MCMC	Markov-chain Monte Carlo
MCRI	Murdoch Children's Research Institute
MLE	Maximum likelihood estimation
MM	Molar mass
NHMRC	National Health and Medical Research Council
OLS	Ordinary least squares
O₂	Oxygen
PSD	Power spectral density
RIP	Respiratory inductance plethysmography
RMSE	Root mean squared error
SD	Standard deviation
SF₆	Sulphur hexafluoride
SNR	Signal-to-noise ratio
TB	Tidal breathing
ViCBiostat	Victorian Centre for Biostatistics
VLBW	Very low birthweight

Chapter 1

Introduction

Contents

1.1 Background	1
1.2 Proposed research	3
1.3 Aims and objectives	4
1.4 Thesis structure	5

In this chapter, section 1.1 provides brief background to lung function testing as relevant to this thesis and section 1.2 broadly describes the main research questions. Sections 1.3 and 1.4 outline the specific thesis aims, objectives, and structure.

1.1 Background

Chronic adult diseases often have origins in childhood, including respiratory diseases, and lung function in early childhood is a major determinant of lung function in later life [1–7]. The human lung undergoes rapid development while in utero, continuing up until late infancy and, in some cases, well beyond that. As the consequences of early life exposures on lung growth and development may be lifelong it is important that lung function can be properly assessed in both children, particularly infants and preschool children, and adults to better monitor disease progression and manage responses to therapeutic interventions [8, 9].

Although no single technique is sufficient to fully evaluate lung function, the multiple-breath washout (MBW) test has emerged as a useful way of assessing lung function due

to it being assessed during tidal breathing, particularly in infants and young children, for a number of common and serious lung diseases including cystic fibrosis [10–13]. Broadly speaking, the test simultaneously measures air flow and gas densities during a period of relaxed breathing. Of interest is how some inert tracer gas—which can either be resident in the lungs or not, depending on the form of the test used—is washed-out of the lungs over the course of the test. This information provides a means of detecting the involvement of small airways in ventilation, which is widely understood to be an important aspect of the pathophysiology of chronic lung diseases such as cystic fibrosis [11, 14]. While small airway involvement can be assessed by a number of different calculated parameters, the most commonly used is the lung clearance index (LCI) [11, 15].

Despite being invented more than 60 years ago [16], the MBW has only recently become feasible for routine use due to the development of commercial technologies able to measure gas concentrations accurately over the breathing cycle. It is a particularly useful test for young children as it requires only minimal patient co-operation [11, 14, 17, 18]. Consequently, there has been a considerable amount of recent research devoted to assessing and validating aspects of the MBW such as repeatability, sensitivity, and technical issues related to interpreting the test data [11, 13, 18–32]. Importantly, the widespread adoption of the MBW has culminated in a recent statement of consensus that recommends guidelines for MBW analysis and highlights future research directions [11].

Although LCI is proving to be a useful index of lung function it is a relatively simple measurement of a specific mechanical component of lung function. There is opportunity to complement LCI with additional summary measures of what are known as tidal flow waveforms—respiratory air flows during normal breathing which are typically sampled at very small time intervals—that may provide clinicians with a more complete picture of lung function. A plethora of indices already exist that summarise tidal flow waveforms, for example, the ratio of time to peak tidal expiratory flow (T_{PTEF}) and expiratory time (T_E) (T_{PTEF}/T_E) or the respiratory rate but, similar to LCI, these tend to isolate various components of the breathing pattern in a reductive way [33, 34].

An alternative approach is to take the view of the respiratory system as one with numerous components interacting in complex ways [33, 35]. This complexity is understood to manifest not only in, for example, the physical structure of the lung [36–38], but also its outputs over time [33, 35, 39–44]. Rather than isolating various mechanical components of

the system, methods that summarise system complexity may reveal important information about the robustness of the system as a whole that is not available using standard reductive summaries. This type of insight may broaden the understanding of disease progression in patients with systemic diseases such as cystic fibrosis.

1.2 Proposed research

Over recent years, with the MBW gaining traction as a useful test in the paediatric age range, there has been substantial research and development of some of the more technical aspects of the test. However, there are many unanswered questions about the test procedure and the methods used to convert the raw data into the summary values that measure different aspects of lung function. Many of these questions are documented in the aforementioned consensus statement [11]. The consensus statement highlighted a number of important future research directions, of which two are directly relevant to this thesis. The first is the use of ‘interpolation’ methods to determine an exact end-of-test for LCI, and the second is the flexibility of MBW end-test thresholds and the shortening of tests. Although these ideas are more fully explored in chapter 4, it suffices here to say that we propose to address these questions, in part at least, by developing and applying a novel statistical model to MBW data. Additionally, the consensus statement [11] also clearly recommended that the software algorithms used to analyse MBW data are made transparent to all users. We therefore propose to document the algorithmic transparency of a major piece of commercial software that is routinely used to analyse MBW data.

Beyond the consensus statement we propose to extend the understanding of the effects of variation in ideal breathing patterns on MBW outcomes. Briefly, the ideal breathing pattern is one of passive yet stable respiration. Variations to this ideal may come in the form of breathing anomalies such as sighs, coughs, or periods of hyperventilation, or as equipment malfunctions such as gas leaks. While there have been some recent efforts to assess the effects of imposing various validity criteria upon the test data, the research is still in its early stages as it either is specific to a single criterion, for example sighs [32], or relies on manual visual assessments of tests by MBW operators [45].

We also propose to investigate alternative summaries of lung function, using tidal flow waveforms, based on statistical methods that characterise the behaviour of complex systems. Our focus is primarily on methods that rely on the assumption that a complex system, in the form of a physiological output over time, can be characterised as exhibiting scale invariance or, equivalently, self-similarity. The degree of scale invariance in a system may be associated with different states of robustness, for example lack of disease [39]. This scale invariance is typically modelled by a so-called ‘power-law exponent’ or similar [39, 46, 47]. Although this approach to assessing lung function has been described as promising in the past [48], there has been only limited exploration of it in a respiratory context [39].

This thesis is structured such that further review of the current literature is provided in the chapters that address the proposed research questions.

1.3 Aims and objectives

The overarching aim of the thesis is to develop and apply novel statistical approaches to modelling lung function data. This aim is broken into several specific objectives:

1. To investigate the effect of software algorithms on MBW outcomes.
2. To investigate the effects of variations in ideal breathing patterns on MBW outcomes.
3. To develop an appropriate statistical model for per-breath MBW data that allows for a model-based estimation of MBW outcomes.
4. To investigate how well we can estimate the model-based MBW outcomes using incomplete MBW test data.
5. To evaluate the use of spectral analytical methods to model the shape of tidal flow waveforms using a power-law exponent.
6. To explore the potential value of the power-law exponent model for understanding the epidemiology of tidal flow power-law spectra and lung function.

1.4 Thesis structure

Having introduced the lung function setting in which this research took place, and outlined the main aims and objectives, the remaining chapters of this thesis take the following structure:

- Chapter 2 provides background to the data used in this thesis along with characteristics of the studies from which it was sourced and details of the data management procedures. This chapter includes a publication that addresses the first research objective.
- Chapter 3 addresses the second research objective by defining and applying a set of broad ‘validity criteria’ to a large number of MBW tests and modelling the effect of enforcing each validity criterion on subsequent MBW outcomes. The analysis in this chapter also provides a means for extracting a working dataset for use in later chapters in the absence of manual review by an MBW trained physician.
- Chapter 4 addresses the third and fourth research objectives by developing a novel model for end-tidal MBW data, assessing its statistical properties using a Bayesian variance components model, and investigating its use with shortened MBW tests. This work is in the form of a published manuscript, with some additional background material also provided.
- Chapter 5 addresses the fifth research objective by providing a review of common methods to estimate the power-law spectra of physiological systems, identifying and further developing a particular method suitable for application to tidal flow waveforms corresponding to a single breath, and empirically assessing the performance of the method via simulation.
- Chapter 6 addresses the sixth research objective by applying the methods of chapter 5 to estimate power-law exponents from a large number of tidal flow waveforms, and developing and estimating a number statistical models to explore relationships between the power-law spectra of tidal flow waveforms, age, disease, and other co-variates.

- Chapter 7 concludes the thesis by summarising the main findings, reflecting on the statistical challenges encountered, and suggesting potential avenues for future research.

In addition to the above chapters, this thesis also includes a number of appendices that details some of the more technical aspects of the work, notably, Appendix A provides an overview of the main statistical methodologies used throughout the thesis, focusing on the practicalities of the computational methods used rather than an exhaustive theoretical explanation.

Chapter 2

Data context

Contents

2.1 Multiple-breath washout (MBW)	8
2.1.1 MBW equipment	8
2.1.2 MBW manoeuvre	9
2.1.3 Lung clearance index (LCI)	12
2.2 Barwon Infant Study (BIS)	14
2.2.1 Data collection and characteristics	16
2.3 Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF)	19
2.3.1 Data collection and characteristics	19
2.4 Software for MBW data processing and its limitations	20
2.5 Publication	22

This chapter provides a brief overview of the data used in this thesis, and the data collection and management methods. The data on which this research was based were collected by the Barwon Infant Study (BIS) and the Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF). It is from these two studies, which have a substantial respiratory testing component, that the MBW and associated data used in this thesis were sourced.

We begin with a brief background to the MBW testing procedure in section 2.1 and detail the data characteristics and collection methods of relevant studies in sections 2.2 and 2.3. The discussion is not complete without a background on the commercial software used

to collect, assess, and process MBW tests, which is found in section 2.4. A number of issues arose from the use of this software, which were formally documented in the included research letter published in the journal *Pediatric Pulmonology* provided in section 2.5.

2.1 Multiple-breath washout (MBW)

The MBW test is technically complicated, and the choice of which equipment and gases are used to perform it has implications for what data are collected and for ‘downstream’ data management and analysis [49]. Therefore, a sound practical understanding of how the MBW is performed is crucial to successful modelling and methodological development using the test data. The following sections describe the main equipment and techniques used to perform the MBW.

2.1.1 MBW equipment

There are a number of variations of the MBW depending on the washout gas, measurement device, and software that is used. Measurement of the flow and inert gas concentration can be achieved using either direct measurement, for example with a respiratory mass spectrometer, or indirect measurement, for example with ultrasonic technology, which measures gas densities based on changes in the molar mass in the air stream. Although the respiratory mass spectrometer is very accurate and considered the ‘gold standard’ of MBW testing, it is expensive and requires the customised hardware and software to enable it to be used in a working MBW system [18, 50, 51]. Because of this, MBW systems using a respiratory mass spectrometer have not been successfully commercialised and are therefore not feasible outside of highly specialised respiratory centres. Ultrasonic technology, on the other hand, is relatively inexpensive and more practical, has recently been successfully commercialised, and is currently in use in research centres world-wide [14, 45]. Throughout this thesis we focus on MBW performed using ultrasonic technology. Comparisons of the ultrasonic and mass spectrometry methods in lung function testing are described in detail elsewhere [51, 52].

Ultrasonic technology measures the molar mass of respired gases (g/mol) and air flow (L/s) at high frequency (typically 200 Hz, or 5 ms intervals) from the transit times of an

ultrasound emitted by two opposing transducers (comprising an ultrasonic ‘flowmeter’) mounted on either side of an air flow tube [21, 53, 54].

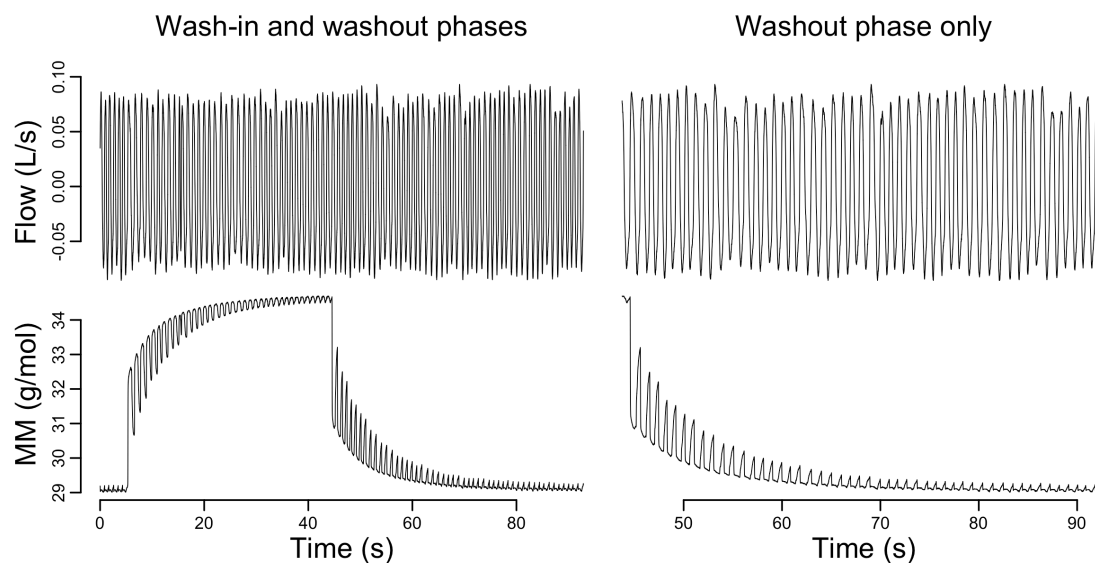
Ultrasonic flowmeters come in two forms: mainstream and sidestream [11, 14]. Mainstream ultrasonic flowmeters measure directly from the patient air flow, and are consequently affected by changes to the temperature and humidity of the respired flow, requiring substantial corrections to the raw molar mass measurements. Furthermore, the mainstream technique measures the molar mass of all respired gases—including oxygen (O_2), nitrogen (N_2), carbon dioxide (CO_2), and the tracer gas of interest—and does not directly distinguish between the molar masses of different gases. Because of this, approximations are necessary to derive the density of the tracer gas, and it is not always clear what those approximations are (see later sections of this chapter for further discussion). In contrast, sidestream ultrasonic flowmeters measure the molar mass of the respired gases ‘off to the side’ (i.e. not in communication with the main air flow) using Nafion tubing to establish uniform temperature and humidity levels at the gas sampling point (therefore not requiring any temperature and humidity corrections), however flow is still measured using a mainstream flowmeter. Again in contrast to mainstream measurement, the sidestream method also uses an infrared CO_2 analyser to correct the sidestream molar mass signal, an aggregate measure of all respired gases, for the molar mass of CO_2 , allowing for the relatively straightforward derivation of each separate gas density [55, 56].

2.1.2 MBW manoeuvre

To illustrate how the MBW is performed in practice, we can consider the test as it is typically performed on infants using the Exhalyzer D apparatus (Ecomedics, Duernten, Switzerland), that is, using the inert gas sulphur hexafluoride (SF_6) as the tracer gas along with a mainstream flowmeter. A face mask is connected to the measurement apparatus containing a flowmeter and is fitted to the patient. After ensuring that the patient’s passive breathing pattern and expiratory lung volume measurements are stable while breathing medical air, the patient begins to breathe a mixture of medical air and 4% SF_6 until the concentration of SF_6 at the end of each breathing cycle reaches an approximate equilibrium. This is known as the ‘wash-in’ phase of the test and is clearly visible in the left-hand side of Figure 2.1. Once equilibrium is achieved, the patient is switched back to room air, commencing the ‘washout’ phase of the test. As the washout progresses, the concentration

of the tracer gas at the end of a normal tidal breath (derived from the molar mass signal) declines toward zero over the course of the washout, which is also clearly visible in the right-hand side of Figure 2.1.

FIGURE 2.1: Example of raw MBW data: molar mass (MM) (g/mol) and flow (L/s) over time (s).

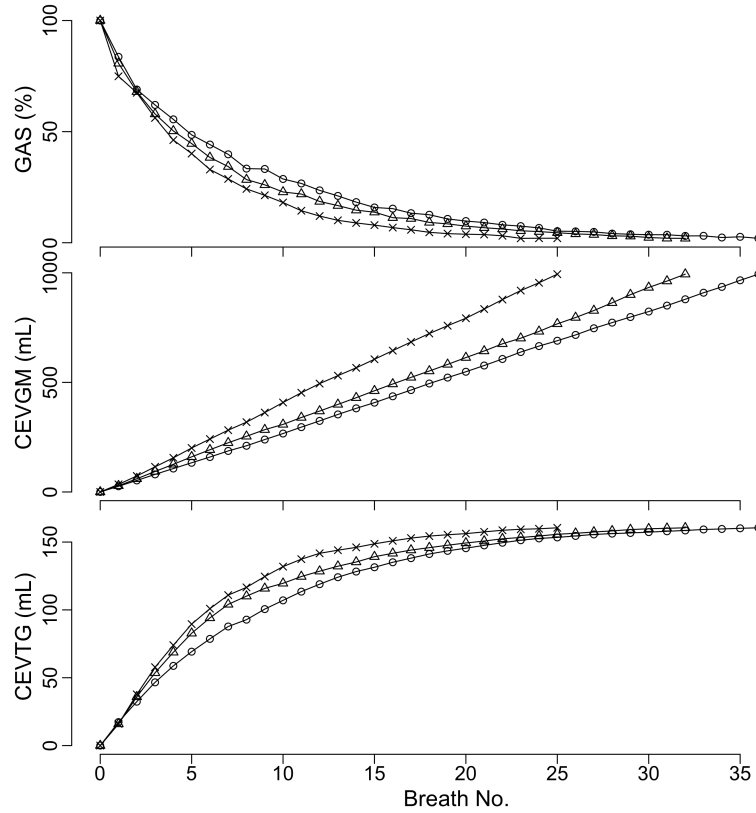


Once the SF_6 concentration at the end of a normal tidal breath is sufficiently washed out, typically to 1/40th of its concentration at the start of the washout, the test is stopped. Specialised commercial software uses the molar mass and air flow data to detect individual breathing cycles, and derive quantities such as the cumulative expired volume of all gas mixtures (CEVGM), the expired volume of the SF_6 tracer gas (CEVTG), and the alveolar SF_6 gas concentration (GAS) at the end of each breathing cycle (see Figure 2.2). These are in turn used to derive summaries of lung function such as functional residual capacity (FRC), cumulative expired volume (CEV), and LCI (which are discussed in detail in the next section).

There are several variants of MBW based on what is the washout gas of interest. The two most common gases are the previously mentioned SF_6 , and N_2 . Less commonly, Argon (Ar) and Helium (He) are also used. N_2 -based MBW involves a constant intake of 100% O_2 into the patient's lungs whereby the washout of the resident N_2 is measured, meaning there is no need for a wash-in phase and equalisation of some non-resident tracer gas, making it arguably the conceptually simplest form of MBW. Tests using 100% O_2 are

often known as multiple breath N_2 washout (MBNW) or, more simply, nitrogen washout, but throughout this thesis we will refer to all variants of the test simply as MBW.

FIGURE 2.2: Example of end-tidal MBW data for three individuals: GAS (%), cumulative expired volume of gas mixture (CEVGM) (ml), and cumulative expired volume of tracer gas (CEVTG) (ml).



Ideally, the MBW occurs over a period of time in which the patient is in a natural physiological state of undisturbed regular breathing, also known as ‘tidal breathing’. A tidal breath can be defined to have an inspiratory phase and an expiratory phase. The exact moment where expiration ends and inspiration begins is defined as the ‘end-tidal’ point, as it represents the end-point of a complete tidal breath cycle.

Although stable tidal breathing is the ideal, in practice it is not always feasible. Whereas in adults stable breathing can be achieved by instructing the subject to breath at a fixed tidal volume, in children stable tidal breathing is far more difficult to achieve, particularly in young children. Although infants are often sleeping or sedated when tested, helping to achieve a passive breathing pattern, young children are typically awake for the manoeuvre, unable to breath under instruction, and more easily distracted than older children.

One way to stabilise passive breathing patterns is to show the child a film or similar distraction in an effort to calm and focus them. However, this process is far from foolproof, and variation in the breathing pattern or breathing artefacts—due to, for example, sighs, panting, or mask leaks—often occur. These artefacts are often considered undesirable to the point of abandoning the test data and re-testing, although there is evidence to suggest that at least some of these artefacts, such as sighing, do not materially affect some MBW outcomes [32].

2.1.3 Lung clearance index (LCI)

As described previously, LCI is the most commonly used MBW outcome. Other MBW outcomes exist and are useful [11], but our primary focus is on LCI and its constituents CEV and FRC.

The standard definition of LCI is the number of FRC lung volumes, also known as turnovers (TO), that are required to reduce the alveolar gas concentration at the end of expiration to some pre-defined end-test threshold (typically 1/40th of the original concentration at the start of the washout). This is calculated by the ratio of CEV to FRC assessed empirically at the end of the expiration corresponding to the end-of-test. CEV is a measure of the expired volume of all respired gases over the MBW washout period, including the tracer gas. FRC can be understood as the volume of the lungs at the end of a passive expiration, or equivalently, at the end of a tidal breath. It is therefore a natural measure of passive lung volume, and consequently acts as a good yardstick by which to measure other associated lung volumes. FRC is typically calculated from MBW data as a function of the observed CEVTG and the end-tidal gas concentrations at the start and end of the washout. Ventilation inhomogeneity is an unevenness of ventilation distribution which may be generated by either small or large airway pathologies but, as the bulk of gas mixing occurs in the small airways, it is often reflective of small airway involvement in ventilation. It occurs in health as a function of the natural airway branching morphology but is often increased in individuals with lung disease. In patients with more small airway involvement, the clearance of inert tracer gas is delayed, and thus LCI is typically increased [11, 14]. The LCI and FRC are therefore useful summary measures of very specific mechanical aspects of the lung and airway involvement.

Calculating CEV and FRC from the raw molar mass and flow data collected during an MBW test is relatively straightforward in theory, if not always in practice. Generally, over the washout, for sampling increment $i = (1, \dots, I)$ —in our case regularly spaced 5 ms intervals as determined by a 200 Hz sampling frequency—such that we define a time vector $t = (0, 5, \dots, T)$ and $k = (0, \dots, K)$ indexing the end of each breathing cycle, we have vectors of measured molar mass $\mathbf{MM} = (MM_1, \dots, MM_I)$ and flow $\mathbf{V}' = (V'_1, \dots, V'_I)$ of length I . We use the notation i_k to denote the unique sampling increment where the end of breathing cycle k occurs. Volume, $\mathbf{V}$, is derived simply by numerically integrating the flow, typically by using the trapezoidal rule when working with rapid sampling frequencies such as 200 Hz, corrected for the initial volume which is zero where $i = 1$ [57]:

$$V_i = V_1 + \sum_{i=1}^{I-1} (t_{i+1} - t_i) \frac{V'_{i+1} - V'_i}{2}. \quad (2.1)$$

Instantaneous gas concentration $\mathbf{C} = (C_1, \dots, C_I)$ is derived via an algorithm specific to the device, gas, and software used for the washout.

Central to the standard derivation of the MBW outcomes—that is, CEV, FRC, and LCI—is the determination of an ‘end-of-test’ breath corresponding to a given gas threshold, in our case 1/40th of the starting end-tidal gas concentration. We denote this ‘end-of-test’ breath as¹:

$$k^{(40)} = \min\{k : \{C_{i_k}, C_{i_{k+1}}, C_{i_{k+2}} \leq 1/40\}\}. \quad (2.2)$$

We index sampling increments corresponding to an increase in volume (i.e. during expiration) as

$$i^* = \{i \leq i_{k^{(40)}} : V'_i - V'_{i-1} > 0\}. \quad (2.3)$$

CEV can be expressed as a function of $\mathbf{V}$, $k^{(40)}$, and the average deadspace volume present in the respiratory system (V_D), which reflects the volume of inhaled air that does not take part in gas exchange:

$$\text{CEV}(\mathbf{V}, k^{(40)}, V_D) = \sum_{i \in i^*} (V_i - V_{i-1}) - k^{(40)} V_D. \quad (2.4)$$

¹Note that $k^{(40)}$ is in fact a function of $\mathbf{C}$, however we use the stated notation for ease of exposition.

Similarly, where $V_{D,GAS}$ refers to the average washout gas deadspace present in the system, FRC can be expressed as:

$$\text{FRC}(\mathbf{C}, k^{(40)}, \mathbf{V}, V_{D,GAS}) = \frac{1}{C_0 - C_{i_{k(40)}}} \sum_{i \in i^*} (V_i - V_{i-1}) C_i - k^{(40)} V_{D,GAS}. \quad (2.5)$$

With these quantities defined we can state a more formal definition of LCI:

$$\text{LCI}(\mathbf{C}, k^{(40)}, \mathbf{V}, V_D, V_{D,GAS}) = \frac{\text{CEV}(k^{(40)}, \mathbf{V}, V_D)}{\text{FRC}(\mathbf{C}, k^{(40)}, \mathbf{V}, V_D, V_{D,GAS})} \quad (2.6)$$

In practice, there are a number of variations in which CEV and FRC can be determined based on, for example, different deadspace definitions, whether deadspace is included or excluded, and alternative algorithms used to determine what is a complete tidal breath. Any variation may subsequently affect how the values at each k , $k^{(40)}$, and $\mathbf{C}$ are identified from the raw data. A full discussion of how these inputs are determined would digress from the main focus of this section. Differences in method depend on the exact measurement technique and algorithms employed, often with a number of manual user inputs [11, 19, 28, 49, 58]. Because the commercial measurement technology is still evolving and not necessarily subject to rigorous scientific review, many of the exact details behind the algorithms to derive the gas density, for example, are not immediately clear. This limits the extent that we can explore different modelling approaches using the raw data. While we make this point later in Section 2.4 along with an accompanying research letter [49] in Section 2.5, it is worth stating here that MBW results will differ depending on what software, hardware, or inputs are used. The descriptions of the method presented here provide a more mathematically explicit definition of the key MBW outcomes than is common in the clinical literature.

2.2 Barwon Infant Study (BIS)

BIS is a birth cohort study in the Barwon region of south-eastern Australia, designed to investigate the impact of early-life exposures on immune, allergic, cardiovascular, respiratory, and neurodevelopmental outcomes, and is a joint initiative of the MCRI, Barwon Health, and Deakin University [59]. Women who satisfied the eligibility criteria—broadly speaking, permanent residents of Australia and the Barwon region who were planning to give birth at either Geelong University hospital (public) or St John of God Hospital

(private)—were invited to take part in BIS at a routine antenatal appointment approximately 15 weeks into their pregnancy. Infants born before 32 weeks gestation or with either a postnatal diagnosis of serious illness, known major congenital malformation, or genetically determined disease were excluded.

A total of 1,158 women were recruited over a 3-year period ending in June 2013, which resulted in a total of 1,074 participating infants. Biological samples including blood, faeces and urine were collected from the participating mothers during pregnancy and from infants during subsequent follow-up reviews. Infant participants were reviewed at birth, and again at approximately 1, 6, 9, 12, and 24 months, with 4-year reviews in progress having commenced in late 2014. In addition, cardiovascular health and lung function were assessed at 1-month and 4 years, allergy testing was performed at 1 year, and neurodevelopmental assessment was performed at 9 and 24 months, and 4 years.

The objective of the respiratory component of BIS is to investigate the relationship between postnatal immune development, lower respiratory tract illness, aeroallergen sensitisation and determination in lung function during the first 3 years of life [59]. Lung function data were collected using MBW and tidal flow recordings at the 1-month (otherwise known as the ‘infant’) review, and only MBW at the 4-year (otherwise known as the ‘preschool’) review. The number of tests available for use along with baseline characteristics is shown in Table 2.1. The specific collection methods and data characteristics are described in detail over the next subsections.

TABLE 2.1: Data collection and patient characteristics. Numbers correspond to successful tests. SD: standard deviation.

Origin	Type	Follow-up	Gas	Participants (n)	Tests (n)	Male (%)	Age (SD)
AREST-CF	MBW	12 w	SF ₆	45	141	50.4	0.3 (0.08)
		1 yr	SF ₆	47	142	44.4	1.04 (0.11)
		2 yr	SF ₆	40	124	58.1	2.01 (0.1)
		3 yr	N ₂	5	13	38.5	3.24 (0.21)
		4 yr	N ₂	16	48	52.1	4 (0.17)
		5 yr	N ₂	17	58	51.7	4.98 (0.25)
		6+ yr	N ₂	19	67	47.8	5.97 (0.22)
BIS	TB	4 w	n.a.	465	465	52.7	0.11 (0.03)
	MBW	4 w	SF ₆	433	1213	51.9	0.11 (0.03)
		4 yr	N ₂	249	474	54.0	4.15 (0.27)

2.2.1 Data collection and characteristics

2.2.1.1 Infant MBW data

Infant MBW data were collected on appointment from infants aged between 4 and 12 weeks between February 2011 and December 2013 by trained MBW technicians at the BIS offices in Geelong, Victoria. Participants were tested during natural sleep in a supine position. MBW tests were performed in accordance with current guidelines [11]. All tests were performed with 4% SF₆ using a mainstream ultrasonic flowmeter via a facemask interface to measure molar mass and flow (Exhalyzer D, Ecomedics, Duernten, Switzerland).

MBW data were processed with WBreath software (version 3.19, ndd Medizintechnik AG, Zurich, Switzerland). A breakdown of the data processing workflow within WBreath is provided in the attached research letter (see Section 2.5). In brief, post-testing, a WBreath file (`.brw`) that included the raw data, along with various patient and environment meta-data, was saved into a location on the MCRI server. This file was then opened in WBreath for review by a trained MBW operator. Quality characteristics, along with whether the test was deemed acceptable or not, were documented in a spreadsheet by the technician, which was then uploaded to a central database. Acceptable tests exhibited a stable breathing pattern with no artefacts such as sighs, sucking, snoring, or mask leaks in line with guidelines at the time [11]. In the event that a test was deemed acceptable, an AutoIt script [60], initially written by a colleague at MCRI and which we later modified substantially, was run to manipulate the WBreath graphical user-interface to perform several modifications to the raw data before exporting a number of comma-separated values (`.csv`) files containing data specific to that MBW test. The files exported included:

- A ‘trial’ file, which includes processed MBW data including time (ms), air flow (L/s), and molar mass (g/mol) measured over the washout. Each record represents a sampling increment—in our case 200 Hz or 5 ms intervals—over 1 to 2 mins.
- A ‘breath table’ file, which includes various end-tidal quantities, such as tidal volume (V_t) and the derived SF₆ gas concentration. Each record represents a single tidal breath.
- A ‘trial table’ file, which includes, among other things, various MBW outcome parameters such as LCI, FRC, and starting SF₆ concentration.

These files, along with the spreadsheet indicating which tests were deemed acceptable, were imported into R for further cleaning and subsequent analysis.

A total of 4,121 tests were attempted on 568 infants. Of these attempts, 1213 tests from 433 infants were deemed acceptable by a trained MBW operator. Characteristics of these tests are described in Table 2.1.

Of these infants, 89 (21%) had only a single test available, while 115 (27%), 98 (23%), 81 (19%), 35 (8%), 13 (3%), and 1 (<1%) participants had 2, 3, 4, 5, 6, and 7 replicate measurements available, respectively.

2.2.1.2 Preschool MBW data

Preschool MBW data were collected on appointment from children aged between 3.9 and 5.5 years between February 2015 and February 2018 by trained MBW operators at the BIS offices in Geelong, Victoria. In contrast to the infant cohort, the preschool participants were tested while awake, sitting upright and distracted using on-screen content (for example, an animated television show). All tests were performed using N₂ MBW using a sidestream ultrasonic flowmeter via a mouthpiece interface (Exhalyzer D, Ecomedics, Duernten, Switzerland). Again, MBW tests were performed in accordance with current guidelines [11].

Also contrasting the infant MBW collection workflow, the preschool MBW data used the Spiroware software suite (Spiroware 3.1.6, Ecomedics AG, Duernten, Switzerland) rather than WBreath for data processing. Thus the MBW operator exported a number of `.csv` files from Spiroware for each test, albeit in a slightly different format and containing different data to that of the WBreath export. In particular:

- A ‘trial’ file, which includes air flow (L/s), molar mass (g/mol), and O₂ and CO₂ concentration data (%) measured over the washout. Each record represents a sampling increment—in our case 200 Hz or 5 ms intervals—over 1 to 2 mins.
- A ‘breath table’ file, which includes various end-tidal parameters, such as tidal volume (V_t) and N₂ concentration (%). Each record represents a single tidal breath.
- A ‘trial table’ file, which includes, among other things, various MBW outcome parameters such as LCI, FRC, and S-III, starting N₂ concentration, and trial number

(because a single Spiroware trial table may summarise multiple trials, in contrast to WBreath’s output).

These files were imported into R. At the time, the MBW tests had yet to undergo post-hoc quality assessment. Thus, with the assistance of Dr Lawrence Gray we automated a number of validity criteria to approximate this manual quality assessment and determine which tests could reasonably be considered to be of sufficient quality for analysis (see Chapter 3).

A total of 1,658 MBW tests were attempted² from 415 preschool children. Of these attempts, 474 tests from 249 preschool children were deemed acceptable by our validity criteria algorithm (see Chapter 3). Characteristics of these tests are described in Table 2.1.

Of these infants, 99 (40%) had only a single test available, while 90 (36%), 48 (19%), 10 (4%), 1 (<1%), and 1 (<1%) participants had 2, 3, 4, 5, and 6 replicate measurements available, respectively.

2.2.1.3 Tidal flow recordings

At the time of MBW testing at the 1-month review, separate tidal flow data were also recorded using the mainstream ultrasonic flowmeter via a facemask interface (Exhalyzer D, Ecomedics, Duernten, Switzerland) and processed with WBreath (version 3.19, ndd Medizintechnik AG, Zurich, Switzerland). Although MBW tests also record tidal flow data as a normal part of the test, the separate tidal flow data were collected as the patient breathed normal room air prior to commencing the first SF₆ wash-in. The MBW operator was instructed that a valid tidal flow recording should contain a minimum of 30 stable breath cycles. Tidal flow recordings are typically used to derive more traditional measures of lung function such as T_{TPEF}/T_E and respiratory rate. The resulting exported .csv files were similar to those from the MBW.

Processing of the raw data in WBreath was similar to the infant MBW data. The tidal breathing traces were examined by a trained operator for acceptability. A total of 1,402 tests were attempted on 535 infants. Of these attempts, 465 tests from as many infants were deemed acceptable by a trained MBW operator. Characteristics of these tests are described in Table 2.1.

²Defined here as at least five recorded breaths post switch to 100% O₂.

2.3 Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF)

The Australian Respiratory Early Surveillance Team for Cystic Fibrosis (AREST CF) is a collaboration of over 60 medical doctors and researchers from specialist paediatric cystic fibrosis centres located in Perth, Western Australia, and Melbourne, Victoria (<https://arestcf.telethonkids.org.au>).

AREST CF runs an early surveillance program that follows infants diagnosed with cystic fibrosis after newborn screening. The program aims to assess, treat, and monitor cystic fibrosis in young children. Early surveillance started in Melbourne in 1994, with Perth commencing in 1999 in conjunction with the introduction of newborn screening for cystic fibrosis in Western Australia. A grant from the United States Cystic Fibrosis Foundation then paved the way for the foundation of AREST CF in 2005. Participation involves a comprehensive assessment at birth and every year thereafter until the age of six. Broadly, measurements taken include cases of infection, chest scans, and lung function, including MBW. Tests were performed at 12 weeks, and at subsequent 6-monthly reviews where possible. Additional tests were performed on a more ad hoc basis where participants presented for eradication of *Pseudomonas aeruginosa*. Baseline characteristics of the participating children are shown in Table 2.1.

2.3.1 Data collection and characteristics

MBW data were collected on appointment over the period January 2009 to March 2017 by trained MBW technicians at the Royal Children’s Hospital, Melbourne. A total of 82 children had at least one successful MBW test over the study period.

For ages younger than 3 years, all tests were performed and processed in line with BIS infant protocols with 4% SF₆ using a mainstream ultrasonic flowmeter via a facemask interface to measure both molar mass and flow (Exhalyzer D, Ecomedics, Duernten, Switzerland). In contrast to the BIS infant testing protocol, participants were tested while sedated in a supine position. The processing and exporting of the data in WBreath was identical to that used for the BIS infant MBW data (see Section 2.2.1.1). Characteristics of the infants and tests are described in Table 2.1.

From ages 3 years and older, all tests were performed and processed in line with BIS preschool protocols with 100% O₂ using a sidestream ultrasonic flowmeter via a mouth-piece interface (Exhalyzer D, Ecomedics, Duernten, Switzerland) (see Section 2.2.1.2). In contrast to the BIS preschool tests, a trained MBW technician rather than an automated set of validity criteria determined whether a test was of acceptable quality. Characteristics of these tests are described in Table 2.1.

2.4 Software for MBW data processing and its limitations

WBreath is the major commercial software package used to analyse inert gas MBW using molar mass as the gas measurement framework. In brief, WBreath contains a collection of algorithms to detect individual breaths, perform adjustments to the high-frequency flow and molar mass data based on certain user input and environmental metadata, calculate gas concentrations, and combine the data to calculate key quantities such as LCI and FRC. It is important that this type of software is available to researchers who may have neither the time nor expertise to code and implement these algorithms and derivations themselves. However, for research purposes, many of the software algorithms used within MBW analysers were not documented, which limited the extent we could fully control the data pre-processing via our own custom-written software. We summarised our concerns in a research letter that we published in the journal *Pediatric Pulmonology*; this is included in Section 2.5.

An illustrative example of an undocumented WBreath algorithm is the one used to determine a complete breath cycle from the data. We now explain this in some detail.

There are well-documented and validated breath detection algorithms available for use with tidal flow data [57, 61]. As an example, one of these algorithms examines the behaviour of the flow and rejects breathing cycles that are determined due to either minor flow rate artefacts or unreasonably high respiratory frequencies. The breath identification algorithm starts by finding all zero crossings of the flow trace. The magnitude of the peak flow between consecutive crossings is then identified and, using a forward search algorithm, if the flow magnitude does not exceed a given threshold between two crossings they are then discarded. As an additional check, the sign of the flow signal is also monitored. If two

zero crossings are found to occur within a time period less than a given threshold, then they are discarded.

While no single algorithm works correctly all of the time, the algorithm described above has been shown to be particularly robust in the presence of severely disturbed respiratory signals or irregular breathing patterns [57, 61]. Importantly, adjusting the thresholds of these algorithms affects how a complete breath cycle is defined. Thus, in the event that a similar algorithm is used to assess MBW outcomes in commercial software packages, the threshold values would ultimately affect the final MBW outcomes, including LCI. Ideally, for transparency and reproducibility, this type of algorithm and its inputs (which are peak flow and time thresholds) would be fully documented prior to using the data in subsequent analysis.

As we describe in our research letter [49], WBreath uses a number of algorithms similar to that just described to process the raw MBW data. While some of these are well known and documented, such as correcting flow to account for differences in humidity and pressure (known as the body temperature pressure saturated correction or BTPS), others are not immediately clear from either the official documentation [62], relevant patents [63], or other sources [53, 64, 65]. Furthermore, because the software is proprietary, there is no way of accessing the underlying source code. As a consequence, we were forced to abandon our efforts to fully reproduce the processing of the raw data, and instead rely on WBreath's more-or-less 'black box' workflow. As we state in our research letter, these issues would be easily addressed by fully documenting the processing algorithms and, ideally, also making the source code openly available.

Beyond processing algorithms, we also encountered more practical roadblocks to performing fully reproducible analyses using the WBreath software. Central to this was that WBreath had no native mass export function, and requires a substantial amount of user input to export the processed data for even a single test. This would be a relatively minor issue if a researcher were interested in only a handful of tests. However, with several hundred tests to process, and exports taking several minutes for each test using WBreath's graphical user interface, the need for an automated export procedure is clear to ensure both reproducibility and economy.

Our approach to exporting the data was to automate, to the extent that it was possible, the use of WBreath's graphical user interface. With the help of a colleague at MCRI, this

was accomplished using AutoIt, a scripting language [60]. AutoIt allows most elements of a graphical user interface in current versions of Windows to be used programmatically. Functions were written to perform the actions of the MBW operator, using a standard internal protocol for user input and data export. Although broadly successful, there were some components of the WBreath data processing workflow that could not be automated using the AutoIt software. In particular, some configuration parameters were not able to be modified, such as ambient temperature conditions, although it was not clear whether this was needed since these parameters were automatically populated upon opening a `.brw` file for each test.

While our approach made processing the raw MBW data much more transparent and reproducible, it is clearly not without its limitations. Further, our concern was that the lack of software transparency may stall future methodological development of tests using MBW data, hinder overall reproducibility, and mask potential inconsistencies in WBreath's algorithms. This concern was the primary motivation for the attached research letter.

2.5 Publication

The included research letter was accepted for publication in January 2016. After several months of trying to understand how the raw data were being processed, mostly to no avail, we decided it was important to summarise and share what we understood to be the data processing workflow of a major commercial software package and its deficiencies.

Lack of Transparency in Software Used to Analyze Multiple Breath Washout Data

To the Editor:

The multiple breath washout (MBW) is an exciting new technique that is increasingly used to detect lung function abnormalities, and to monitor progress and response to interventions in people with respiratory conditions. For example, in cystic fibrosis an increase in the lung clearance index (LCI) is associated with the identification of bronchiectasis on chest computed tomography and the technique has been used to study changes in response to treatment with mucolytics.¹ Although the technique is used increasingly in multi-centre studies there are few commercial devices available and many technical aspects in relation to validity and performance remain undocumented.²

WBreath (ndd Medizintechnik AG, Zurich, Switzerland) is a major commercial software package used to analyze raw inert gas MBW data for mainly research but also clinical purposes. Methodological development of tests using sulfur hexafluoride (SF₆) MBW data collected via mainstream gas analyzer relies heavily on WBreath to transform the raw data into usable form and produce measures of lung function. This is problematic because the algorithms encoded in WBreath's software are not publicly available (see Fig. 1). It seems likely that other software has a similar lack of transparency although our direct experience is limited to WBreath.

A typical MBW analysis using WBreath begins by loading a file containing data including flow rate, measured assuming “ambient temperature pressure saturated” (ATPS) conditions, and native molar mass, both sampled at 200 Hz. Additional patient and environmental metadata are also included in this file. The metadata are used along with various user inputs to run algorithms that modify the data to represent “body temperature pressure saturation” (BTPS) conditions, and to correct for changes in temperature and response in the ultrasonic device. These modifications are intended to account for the complexities in measuring molar mass and flow. The resulting data used for analysis—BTPS flow rate and a final molar mass signal measured at 200 Hz—are typically substantially different to the input raw data.

The analysis that follows these data transformations uses algorithms to detect individual breaths and to calculate end-tidal quantities such as gas concentration and expired gas and air volumes. These quantities are in turn used in various algorithms to derive indices of lung function such as functional residual capacity (FRC) and LCI. Whether WBreath calculates instantaneous gas concentration as part of this analysis, or if such a calculation is indeed possible with SF₆ data collected using a mainstream gas analyzer, is unclear.

As experienced users of WBreath, it remains difficult for us to establish clear connections between WBreath's inputs and outputs. While there is information available regarding some aspects of WBreath's operation, including the user manual and a number of related application notes, the software remains proprietary and many of the aforementioned algorithms are either undocumented, or not documented in sufficient detail to allow results to be reproduced from the raw SF₆ MBW data with user-written software.

Furthermore, questions remain about the internal accuracy of the software. Recent studies have found that normal use of WBreath with SF₆ MBW data may provide either inaccurate LCI values in infants with cystic fibrosis,³ or significantly different LCI values when using different versions.⁴ Measurement bias introduced by minor changes to MBW software settings has also been demonstrated in software similar to WBreath (Spiroware 3.1.6, Ecomedics AG, Duernten, Switzerland), which will have implications for long-term and multi-centre studies and further reinforces the need for greater transparency.⁵

We are concerned that the lack of software transparency may stall methodological development of tests using SF₆ MBW data, hinder overall reproducibility, and mask

Conflict of interest: None.

Correspondence to: Sarath Ranganathan, MB, ChB, PhD, Department of Respiratory Medicine, Royal Childrens Hospital, Flemington Road, Parkville 3052, Australia. E-mail: sarath.ranganathan@rch.org.au

Received 20 January 2016; Accepted 31 January 2016.

DOI 10.1002/ppul.23420

Published online 6 June 2016 in Wiley Online Library (wileyonlinelibrary.com).

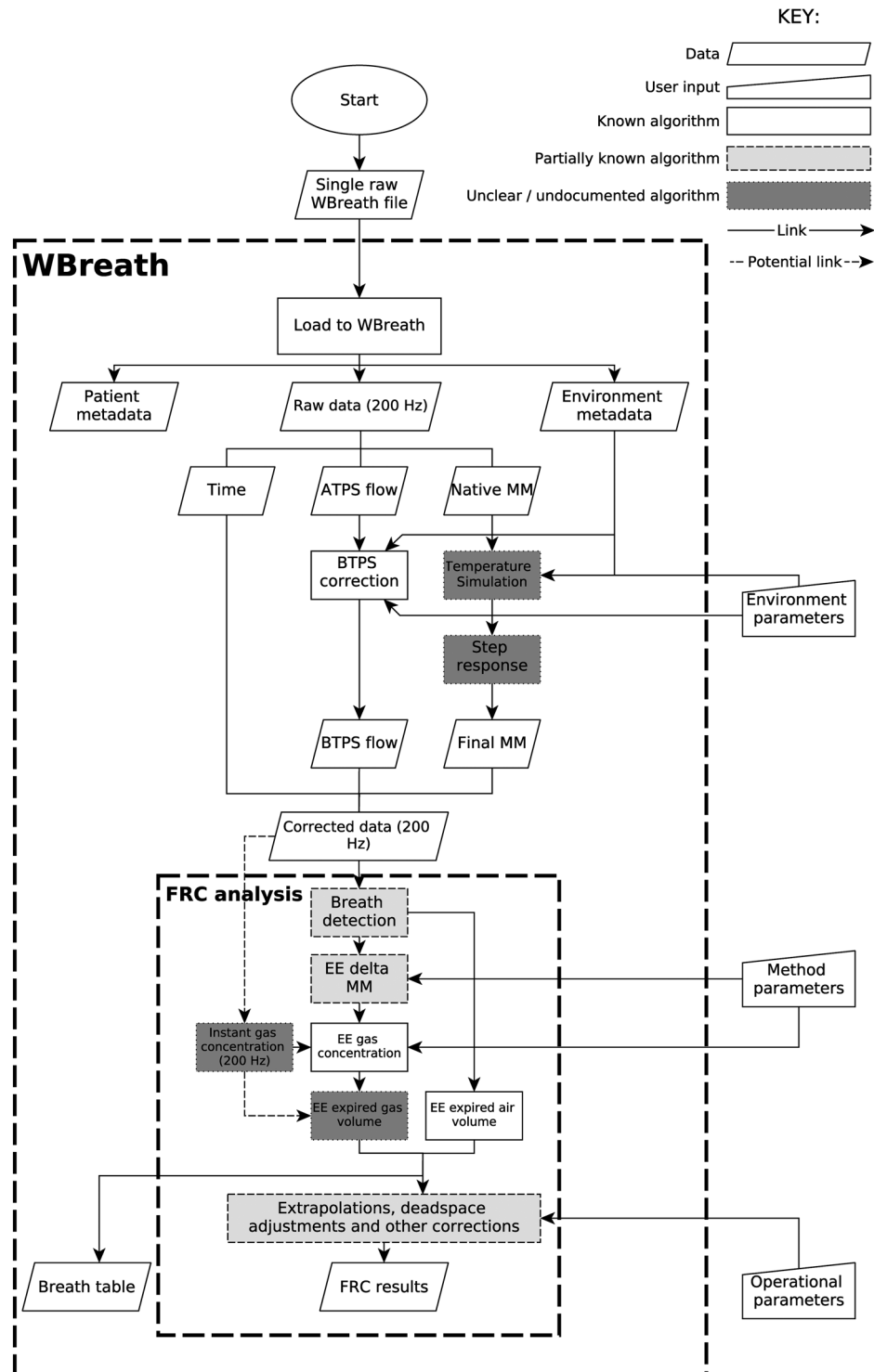


Fig. 1. Flowchart of WBreath data and algorithms for a single SF6 MBW performed with a mainstream gas analyzer. Algorithms are identified as known, partially known in other contexts but not necessarily used by WBreath, or unclear/undocumented. ATPS, ambient temperature pressure saturated; BTPS, barometric temperature pressure saturated; EE, end-expiration; FRC, functional residual capacity; MM, molar mass.

potential inconsistencies in WBreath's algorithms. It is relevant to note that our concern is reflected more generally in the recent ERS/ATS consensus statement.² It is therefore our hope that WBreath's algorithms will be fully documented, or alternatively made available as open source code, to address these transparency issues and potential inconsistencies.

—ROBERT K. MAHAR, MBIOSTAT

*Data Science
Murdoch Childrens Research Institute
Parkville, Australia
Department of Paediatrics
Faculty of Medicine, Dentistry and Health Services
University of Melbourne
Parkville, Australia*

DAMJAN VUKCEVIC, DPHIL

*Data Science
Murdoch Childrens Research Institute
Parkville, Australia
Faculty of Science
School of Mathematics and Statistics
University of Melbourne
Parkville, Australia*

LOUISE KING, BSC

*Infection and Immunity
Murdoch Childrens Research Institute
Parkville, Australia
Department of Respiratory Medicine
Royal Childrens Hospital
Parkville, Australia*

JOHN B. CARLIN, PHD

*Data Science
Murdoch Childrens Research Institute
Parkville, Australia
Department of Paediatrics*

*Faculty of Medicine, Dentistry and Health Services
University of Melbourne
Parkville, Australia
Faculty of Medicine, Dentistry and Health Services
Melbourne School of Population and Global Health
University of Melbourne
Parkville, Australia*

SARATH RANGANATHAN, MB, CHB, PHD

*Department of Paediatrics
Faculty of Medicine, Dentistry and Health Services
University of Melbourne
Parkville, Australia
Infection and Immunity
Murdoch Childrens Research Institute
Parkville, Australia
Department of Respiratory Medicine
Royal Childrens Hospital
Parkville, Australia*

REFERENCES

1. Ramsey KA, Rosenow T, Turkovic L, Skoric B, Banton G, Adams AM, Simpson SJ, Murray C, Ranganathan SC, Stick SM, et al. Lung clearance index and structural lung disease on computed tomography in early cystic fibrosis. *Am J Respir Crit Care Med* 2016;193:60–67.
2. Robinson PD, Latzin P, Verbanck S, Hall GL, Horsley A, Gappa M, Thamrin C, Arets HG, Aurora P, Fuchs SI, et al. Consensus statement for inert gas washout measurements using multiple- and single-breath tests. *Eur Respir J* 2013;41:507–522.
3. Anagnostopoulou P, Yammine S, Schmidt A, Korten I, Kieninger E, Mack I, Trachsel D, Hafen G, Moeller A, Casaulta C, et al. False normal lung clearance index in infants with cystic fibrosis due to software algorithms. *Pediatr Pulmonol* 2015;50:970–977.
4. Ebdon AM, Skoric B, Robertson C, Ranganathan S. Comparing analysis software versions in infant multiple breath washout. *Eur Respir J* 2013;42:57.
5. Summermatter S, Singer F, Latzin P, Yammine S. Impact of software settings on multiple-breath washout outcomes. *PLoS ONE* 2015;10:e01250.

Chapter 3

Effect of variations in ideal breathing patterns on MBW outcomes

Contents

3.1	Background	27
3.1.1	Aims	29
3.2	Methods	29
3.2.1	Data	29
3.2.2	Inclusion rules	29
3.2.3	Statistical methods	36
3.3	Results	36
3.4	Discussion	40

This chapter investigates the effect of variations in ideal breathing patterns on MBW outcomes. We begin with a brief background and description of the aims in Section 3.1, followed by a description of the data, methods, and models used in the analysis in Section 3.2. Results and discussion are provided in Sections 3.3 and 3.4.

3.1 Background

Collecting high-quality MBW tests is typically a demanding process that involves a substantial amount of researcher time and effort [18, 66]. It is generally accepted that a stable breathing pattern is needed to achieve accurate estimates of widely used MBW outcomes such as LCI and FRC. As described in the previous chapter, it is considerably easier to obtain stable breathing measurements in adults compared to children, for whom breathing artefacts such as sighs are common. Often data containing breathing artefacts goes unused as per recommendations in the recent statement of consensus [11]. Aside from a small number of studies that have examined the influence of sighs [32] or various validity criteria [45] on MBW outcomes, there is little existing work that examines the influence of a broader set of variations in ideal breathing patterns on these outcomes. Because of this, many MBW tests may be excluded from analyses unnecessarily.

The MBW consensus statement is perhaps the definitive source of criteria for assessing validity of an MBW test based on breathing pattern [11]. More recently, Jensen et al. [45] proposed and assessed additional validity criteria to that of the consensus statement, in the form of a set of ‘checklist items’, in an effort to more systemically evaluate test quality. They found that although their additional criteria improved inter-rater agreement on test quality, they had little effect on the overall mean and variance of LCI, but also substantially reduced the sample size by unnecessarily rejecting otherwise acceptable tests. However, it appears they applied their criteria to tests that had already satisfied the consensus statement’s criteria for repeatability of measurement of FRC which, in addition to some of the checklist items being somewhat ambiguous in their implementation, makes it difficult to directly understand the effects of using their various criteria on MBW outcomes. Nevertheless, it appears to be the first study that has attempted to systematically assess the influence of multiple validity criteria, thus implicitly assesses the effect of variation in ideal breathing patterns, and therefore it serves as a good basis for further investigation.

We contend that it is possible that only a limited set of validity criteria are needed to provide accurate and repeatable estimates of MBW outcomes while still preserving sample size. For example, there is substantial overlap between criteria, such that only a subset may be required to indicate a valid test. Validity criteria based on a reduced set of checklist items might adequately capture the majority of problems that occur during MBW testing, particularly if the MBW operator has attempted to identify issues at the time of testing.

Furthermore, a major issue with the implementation of these validity criteria is that they are typically made via manual derivation of various indices from the software output and visual assessment of the MBW signals by an experienced MBW operator. Because the manual assessment relies heavily on visual interpretation of the test data it is highly subjective. Further, manual assessment is a time-consuming process that is susceptible to operator error. It is easy therefore to see how the manual approach may not be optimal, particularly with large amounts of MBW data.

A more automated approach to assessing validity criteria could have a number of benefits: save researcher time and effort; provide a more objective evaluation of test quality; enable a higher degree of reproducibility in MBW analyses; and be easily adapted and rerun according to the researchers' specific needs or as new insights into the process are gained, with little loss of overhead (something clearly impractical with a more subjective, less quantified, and manual approach). These advantages are all particularly relevant given that a number of clinical trials and cohort studies are adopting the MBW as a primary measure of lung function in young children.

We assessed the influence of variations in ideal breathing patterns using a set of validity criteria, based on the checklist provided by Jensen et al. [45], each item of which we attempted to frame as a simple binary inclusion rule based on quantifiable parameters. To the extent that some of the checklist items could not be easily framed as a quantitative inclusion rule, we dropped such checklist items from the analysis.

The trade-off in using a reduced set of checklist items is between increasing sample size and potentially allowing inferior quality MBW tests into subsequent analyses. This may result in either biased or imprecise MBW outcomes. It is therefore necessary that we examine the statistical properties of the MBW outcomes under a range of different validity criteria. Ideally we would compare the validity criteria outcomes against outcomes obtained using an accepted 'gold standard' for inclusion. The current gold standard is manual review by a trained respiratory scientist but, at the time of writing, this has not been done for the MBW tests included in the analysis reported in this chapter. We therefore took the approach of Vukcevic et al. [32], who assessed the influence of sighing respirations on MBW outcomes using a variance components model. Because within-participant repeatability is perhaps the most desirable property for MBW test outcomes, a natural aim is to

minimise the within-participant variation while maintaining between-participant variation and clinically realistic mean values for MBW outcomes.

3.1.1 Aims

The overarching aim of this study was to explore the influence of variations in ideal breathing pattern on MBW outcomes through various quantifiable validity criteria. This aim was broken into the following specific objectives:

- To create specific binary inclusion rules from an appropriate subset of Jensen et al's [45] checklist items and group them into broader 'validity criteria'.
- To compare the means and within- and between-participant variances of MBW outcomes resulting from application of the various validity criteria for MBW tests.

3.2 Methods

3.2.1 Data

This study used MBW data from the BIS preschool wave which was extensive, required analyses in the later chapters of the thesis, and had not been assessed for quality at the time (see Chapter 2.2). Spiroware's 'breath table' (Spiroware 3.1.6, Ecomedics AG, Duernten, Switzerland) file was used as the basis for assessing many of the various criteria, although it only included end-tidal data for the washout phase of the test. Because some of the checklist items relate to the wash-in phase of the test, we implemented a breath detection algorithm [61, 67] on the high-frequency MBW data contained in the 'trial table' file to allow us to derive end-tidal quantities for this phase of the test.

3.2.2 Inclusion rules

We used the checklist provided in the online supplement of Jensen et al. [45], herein referred to as the 'checklist', as the basis for our assessment. See Table 3.1 for the complete list. We categorised each specific checklist item into one of the following broader 'validity criteria':

- *Technical*: whether the MBW test was acceptable on minimal technical grounds (i.e. N₂ returned to baseline, decayed appropriately over the washout, and the test ended).
- *Stable wash-in*: whether the tidal volumes over the five breaths immediately prior to the washout commencing were generally stable and realistic.
- *Stable washout*: whether the tidal volumes over the washout period were generally stable and realistic.
- *Regular switchover*: whether the last wash-in and first washout breaths (i.e. ‘switchover’ breaths) were regular tidal breaths.
- *Sigh/pants*: whether there was evidence of sighing or panting across the washout that may have affected test outcomes.
- *Ventilation*: whether ventilation was stable across the washout according to the end-tidal CO₂ measurements.
- *Flow*: whether respiratory flow was stable and without major excursion from a realistic rate across the washout.

For clarity, in Table 3.1, we also indicate whether a particular checklist item relates to the wash-in or washout ‘phase’ of the test.

For each checklist item, we defined an ‘inclusion rule’, which is a specific, unambiguous and quantifiable rule that can be evaluated algorithmically. Where the checklist item was already worded in such an unambiguous form, we used it verbatim for its inclusion rule. Where the method for assessing a particular checklist item was not clear from either the consensus statement or the checklist, we defined one with the advice of Dr Lawrence Gray—an experienced MBW-trained paediatrician.

Assessing some of the checklist items requires a degree of subjective observer judgement, for example the broad requirement for the non-existence of a leak. Codifying such checklist items would require developing software that is outside the scope of this validity study. Therefore, we did not define inclusion rules for such items (we noted these with ‘Not included’ in the Table). Additionally, some of the checklist items were judged to likely be sufficient correlated with others to justify not including them (similarly noted as ‘Not

included' in the Table). For example, evidence of a major leak is likely to be captured by a loss of decay in the end-tidal N₂ concentration.

TABLE 3.1: MBW validity criteria. Cet: (end-tidal) concentration; CDC: Centers for Disease Control and Prevention; CO₂: carbon dioxide; CV: coefficient of variation; EELV: end-expiratory lung volume; et: end-tidal; N₂: nitrogen; PEF: peak expiratory flow; PIF: peak inspiratory flow; SD: standard deviation; Vt: tidal volume.

Validity criteria	Phase	Checklist item [45]	Inclusion rule
Technical	Wash-in	Is starting end-tidal concentration of N ₂ at baseline (first trial of test session) $\leq 77\%$?	Verbatim.
Technical	Washout	Does N ₂ decay throughout the washout?	Cet N ₂ values above 1/4 of the starting Cet N ₂ always fall unless rise coincides with breath $< 0.5 \times$ mean Vt.
Technical	Washout	Does the N ₂ concentration reach 2.5% for 3 consecutive breaths?	Cet N ₂ $\leq 2.5\%$ of starting Cet N ₂ for 3 consecutive breaths exists.
Technical	Washout	On closer inspection, is there evidence of a leak?	Not included.
Stable wash-in	Wash-in	Is Vt appropriate for subject size (10–15 ml/kg ideal body weight)?	Mean Vt within 10–15 ml/kg for ideal body weight given age and sex (according to CDC growth charts [68]).
Stable wash-in	Wash-in	Is Vt stable (consistent volume breath to breath) for 5 breaths immediately prior to start of washout?	CV of Vt for these 5 breaths $\leq 25\%$.
Stable washout	Washout	Is Vt appropriate for subject size (10–15 ml/kg ideal body weight)?	Mean Vt within 10–15 ml/kg for ideal body weight given age and sex (according to CDC growth charts [68]).

Stable washout	Washout	Is Vt stable throughout the washout?	CV of Vt is $\leq 25\%$ over the washout.
Regular switchover	Wash-in	Is the last breath of the wash-in irregular (i.e. swallow, very small or very large)?	Vt of this breath is within 0.5 to $1.5 \times$ mean Vt.
Regular switchover	Washout	Is the first breath of the washout irregular (i.e. swallow, very small or very large)?	Vt of this breath is within 0.5 to $1.5 \times$ mean Vt.
Sighs/pants	Washout	Are there any episodes of panting ($< 0.5 \times$ mean Vt) or a sighing ($> 1.5 \times$ mean Vt)?	Not included.
Sighs/pants	Washout	If there is an episode of panting has the software wrongly assumed end of test? (ie. 3 small breaths where Cet N ₂ % $<$ target)	All 3 breaths that cause the test to end $\geq 0.5 \times$ mean Vt.
Sighs/pants	Washout	If there is a sigh is there any evidence of trapped gas release?	No more than a single breath where Vt $\geq 1.5 \times$ mean Vt. Cet N ₂ for the same breath must not increase at all.
Flow	Washout	Does inspiratory flow ever go above 1000 ml/sec? If yes – is there a leak?	Absolute value of flow < 1000 ml/sec.
Flow	Washout	Is the flow signal stable (PIF/PEF) with passive expiratory phase (no forced exhalation)?	Not included.
Flow	Washout	Based on the flow signal is there a stable per-breath respiratory rate (breaths per minute) across the washout?	Moving 3-breath CV of per-breath respiratory rate $< 25\%$ across test.

Flow	Washout	Is there evidence of breath holding, swallow or other deviation from tidal breathing?	Not included.
Flow	Washout	Is there evidence of a cough?	Not included.
Ventilation	Washout	Is the Cet CO ₂ between 4-6% over the washout?	Across washout, 3.9% < Cet CO ₂ < 6.1% (to avoid boundary cases).
Ventilation	Washout	Is the CO ₂ tracing stable?	Not included.
Ventilation	Washout	Is the Cet CO ₂ variable? SD of Cet CO ₂ is running calc. of 3 breaths and will trigger error when SD > 0.25%?	Verbatim. (i.e. Moving 3-breath CV of Cet CO ₂ < 25% across test)
Ventilation	Washout	Is the Cet CO ₂ continuously increasing or decreasing across the washout?	Not included.
n.a.	Wash-in	Is EELV stable?	Not included.
n.a.	Wash-in	Is volume time trace visually stable?	Not included.
n.a.	Wash-in	Is deviation in EELV at start of washout within 10% of mean Vt of 5 breaths immediately preceding the start of washout?	Not included.
n.a.	Washout	Is EELV (according to volume time trace) stable throughout the washout?	Not included.

n.a.	Washout	Is there any evidence of breath stacking or incomplete exhalation in the volume trace?	Not included.
------	---------	--	---------------

3.2.3 Statistical methods

The statistical properties of MBW outcomes were assessed using a Bayesian variance components model:

$$\begin{aligned} y_{ij} &\sim N(\mu_i, \sigma_W^2) \\ \mu_i &\sim N(\theta, \sigma_B^2) \\ \theta &\sim N(0, 10^2), \quad \sigma_B, \sigma_W \sim \text{half-Cauchy}(0, 2.5), \end{aligned} \tag{3.1}$$

where y_{ij} is the j -th MBW outcome of interest for participant i , μ_i is the mean MBW outcome for participant i with within-participant standard deviation σ_W , and θ is the mean MBW outcome for all participants with between-participant standard deviation σ_B . The model assumes that each patient and test vary independently, and that the within-participant variation is exchangeable across participants. A diffuse normal prior distribution was used on the mean parameter and half-Cauchy priors for the scale parameters [69, 70].

We sampled from the posterior distribution of each parameter estimated in this chapter using Stan [71], called from the R software environment [72] (see Appendix A). Stan was chosen for the variance components modelling as it is well-suited to the analysis of repeated measures data and provides distributions for variance parameters of the model. In each case, four MCMC chains were run in parallel comprising a discarded ‘warm-up’ phase and a ‘sampling’ phase, the latter being used for parameter estimation. The parameter chains were run for 1,000 warm-up iterations followed by 1,000 sampling iterations resulting in 4,000 samples from the joint posterior distribution of the parameter. Convergence of the Markov chains was assessed both graphically, via visual inspection of the MCMC chains, and statistically, by ensuring that the scale reduction factor $\hat{R}$ was near one for each parameter [73]. Posterior medians and 95% credible intervals were chosen to describe the posterior distributions. All other data management and analyses were performed using R.

3.3 Results

Each included checklist item was applied to the 1,658 MBW tests where washout had definitively commenced (defined as at least five recorded breaths post switch to 100% O₂).

At a minimum, we are interested in tests that are technically acceptable. Therefore, for a given test, after confirming that the patient indeed commenced the washout phase of the test, technical acceptability—i.e. satisfying the included checklist items referred to as ‘Technical’ in Table 3.1—was established before the application of any other checklist item. Upon being technically acceptable, each checklist item was applied independently of one another, and then intersections of various checklist items were made to form the more general validity criteria shown in Figure 3.1.

The variance components model was then fitted to the repeated measures MBW outcomes from the successful tests under each validity criterion. Parameter convergence was sufficient in most cases, with values close to one, and effective sample sizes were sufficient given the number of chains used (see Appendix A for definitions of sufficiency).

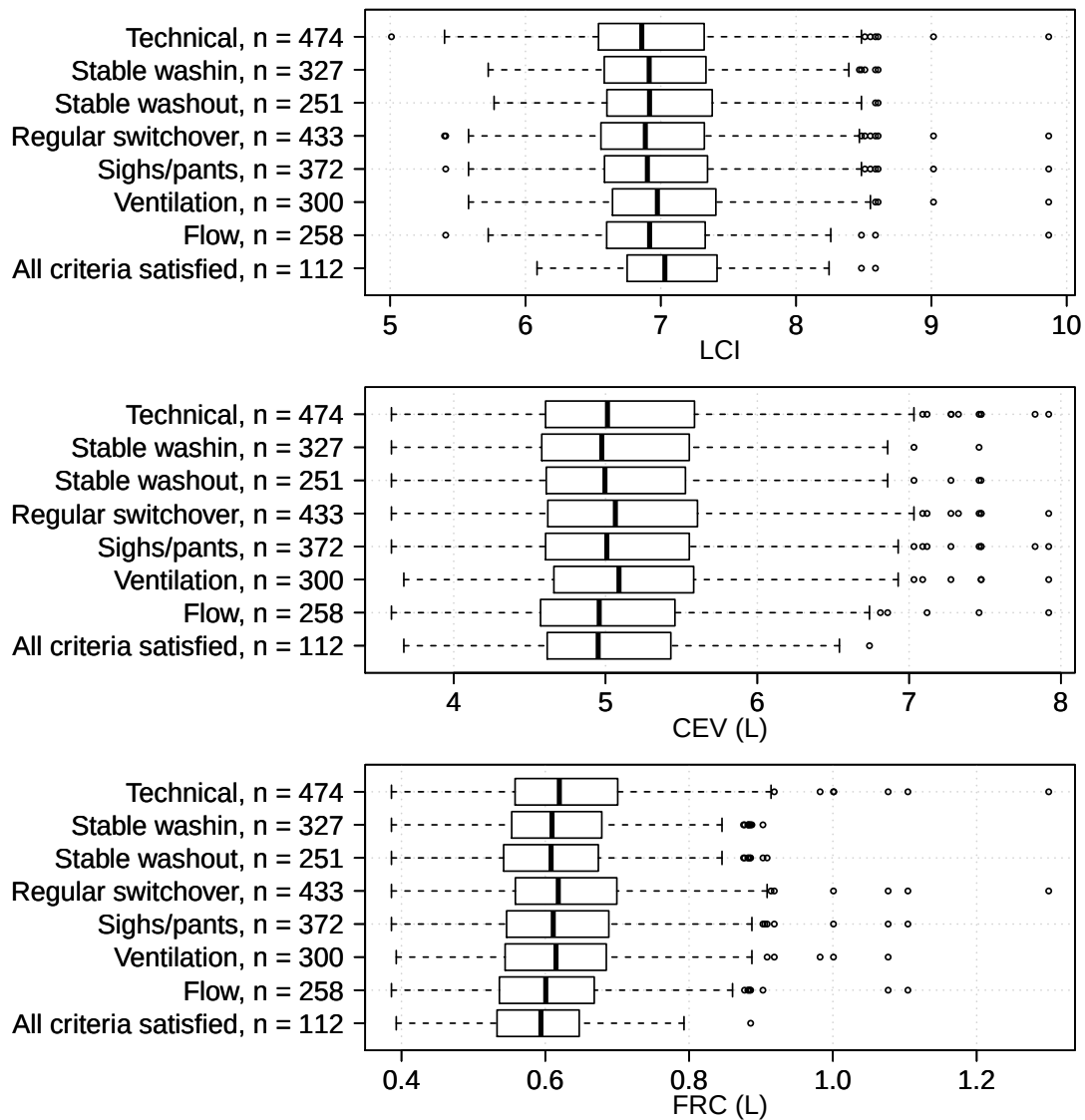
Figure 3.1 summarises each validity criteria in terms of number of tests that satisfied the inclusion rules contained therein and the subsequent distribution of the MBW outcomes.

Only 474 (28.6%) of 1,168 tests where the N₂ washout was initiated were considered technically acceptable. Some validity criteria defined a substantially reduced number of tests considered successful without a major change in either the median or extreme MBW outcomes. Combining all criteria reduces the number of successful tests dramatically, such that only 112 (6.8%) tests were considered to satisfy the ‘All validity criteria’ category. Overall, from Figure 3.1, we can observe that although there is a substantial drop in sample size when certain validity criteria are applied, in particular ‘Stable wash-in’ and ‘Flow’, the distribution of outcomes remains similar. The exception is when all criteria are applied in unison, which reduced the range of the outcomes and eliminated a small number of outliers, albeit at a dramatic cost to sample size.

As stated in the background, it would be ideal to use criteria that did not result in substantially different mean values of the MBW outcomes, while minimising within-participant variation and preserving between-participant variation. The estimates obtained for the variance components model, shown in Figure 3.2, show posterior distributions of the mean and within/between participant standard deviations.

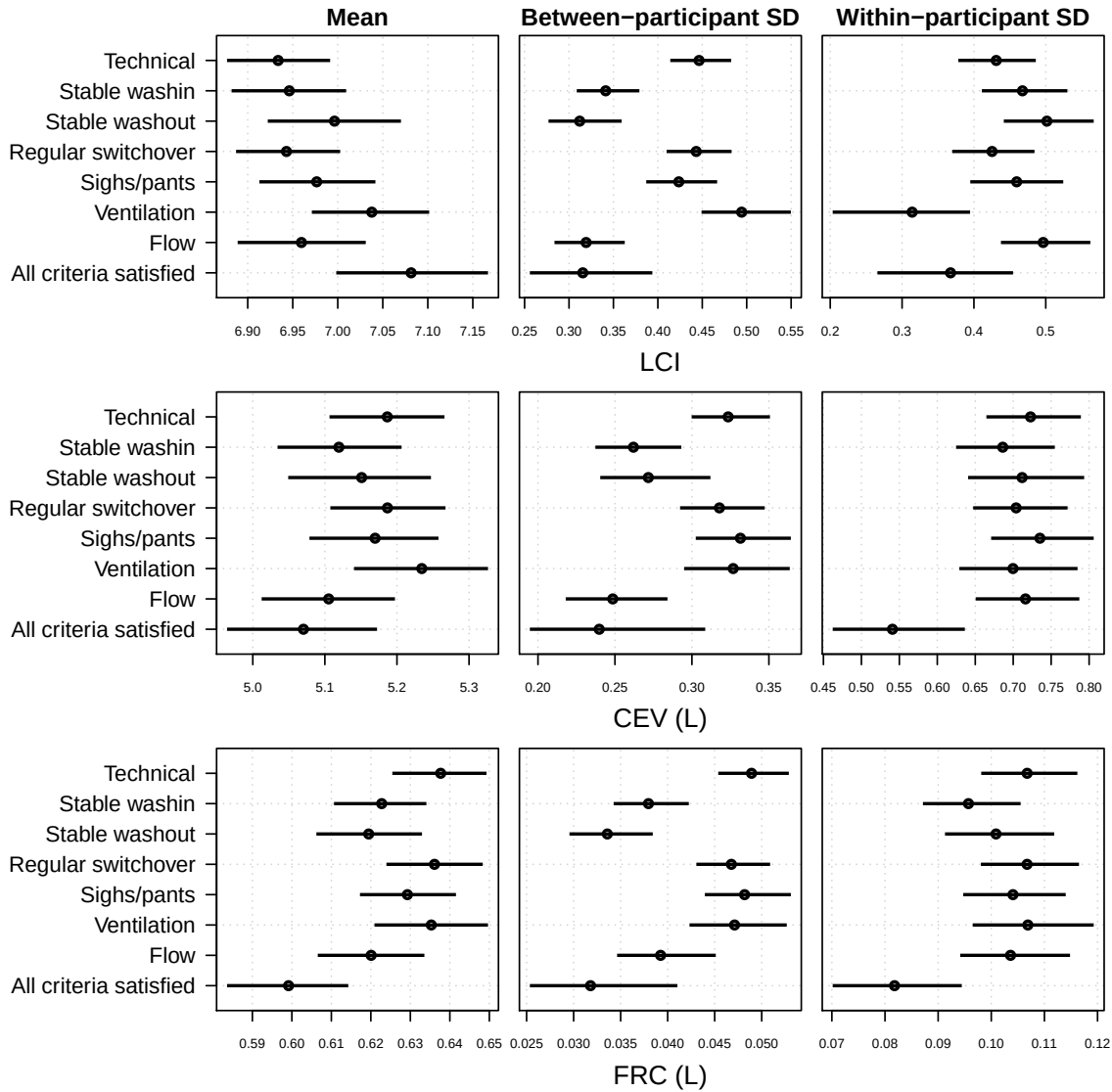
The estimated mean of the MBW outcomes was similar across the criteria and, with a few exceptions, showed considerable overlap among the posterior distributions for most of the validity criteria. When all criteria were satisfied, the mean CEV and FRC appeared

FIGURE 3.1: Boxplot of tests satisfying broad validity criteria. LCI: Lung clearance index; CEV (L): Cumulative expired lung volume (litres); FRC (L): Functional residual capacity (litres).



slightly lower than other categories, and together resulted in a higher mean LCI, perhaps suggesting that children with smaller passive lung volumes and/or those with mildly increased ventilation inhomogeneity tend to have more stable breathing patterns. An alternative explanation may be that the tests obtained when failing to fulfil all the validity criteria contain errors or artefacts that result in overestimation of FRC and therefore underestimation of the LCI. Nevertheless the differences, where apparent, were relatively small in an absolute sense, particularly in the case of LCI, suggesting they are unlikely to be clinically meaningful.

FIGURE 3.2: Estimates of the mean and between- and within-participant standard deviation (SD) from variance components models fitted to subsets of MBW data according to different validity categories. Points indicate posterior medians and bars indicate 95% credible intervals.



Between-participant variation differed across the criteria, however across the MBW outcomes the pattern was broadly similar. Applying either ‘Stable washin’, ‘Stable washout’, ‘Flow’, or ‘All criteria satisfied’ typically resulted in lower between-participant variation across the three MBW outcomes, in contrast to the ‘Technical’, ‘Regular switchover’, ‘Sighs/pants’, and ‘Ventilation’ criteria.

Within-participant variation did not appear to meaningfully differ across MBW outcomes, with the exception of lower values in ‘All criteria satisfied’ for all MBW outcomes and ‘Ventilation’ for LCI. It seems that any apparent reduction in the within-participant variation

for ‘All criteria satisfied’ was due to the substantial reduction in sample size and repeated within-participant measurements, but the same cannot be said of the ‘Ventilation’ criterion.

Applying either the ‘Technical’ criterion alone, or including the ‘Ventilation’ criterion, appears to result in a large sample of tests that have reasonable MBW outcomes. To a lesser extent, the ‘Regular switchover’ and ‘Sighs/pants’ criteria also produce reasonable MBW outcomes but with comparatively smaller sample sizes. Importantly, deciding whether to apply either the basic ‘Technical’ criterion alone or including the ‘Ventilation’ criterion involves a trade-off between the larger sample size resulting from the former and the lower within-participant variation from the latter.

3.4 Discussion

This chapter examined the influence of variations in ideal breathing patterns on MBW outcomes. We defined inclusion rules, based on a reduced set of checklist items proposed by two previously published studies, to form broader validity criteria. We have endeavoured to make these validity criteria less ambiguous, more reproducible, and more automated by assigning them concrete quantitative decision rules.

Our results suggest that although it is of interest to control the breathing pattern as much as possible during MBW tests, most deviations from an ideal breathing pattern, as defined by the validity criteria, do not substantially affect the overall MBW outcomes. We found, in broad terms, that estimates of the population mean and between- and within-participant variation of MBW outcomes are broadly similar under a range of validity criteria, with some exceptions. When all criteria were satisfied the resulting tests tended to have a lower mean CEV and FRC, and a higher mean LCI, which may be an indication that the individuals within this category represented ‘special cases’, for example children with smaller lung volumes or greater small airway involvement, and that by implementing the most strict validity criteria we may be discarding important information and biasing the MBW outcomes. Alternatively, when all criteria were not satisfied, it may be that tests with errors or breathing artefacts were not removed appropriately, resulting in higher mean FRC and hence lower LCI. With these results in mind, in the absence of the availability of manually reviewed tests, it seems valid to focus on using the validity criteria with the most

favourable statistical properties. In our case, we identify at least two plausible criteria: the basic ‘Technical’ validity criterion applied alone, or with the inclusion of the ‘Ventilation’ criterion.

This study is preliminary, and as such has some limitations. For example, we did not compare the various validity criteria against a ‘gold standard’. This would be necessary to establish which of the criteria is clearly optimal. Currently, the accepted gold standard requires complete manual review by a trained respiratory scientist to establish stable breathing pattern. At the time of writing this has not been done for the MBW tests included in the analysis reported in this chapter. Additionally, the analysis presented in this chapter was limited only to preschool children from a healthy birth cohort. For the less restrictive validity criteria to be accepted into routine use it would be important to also assess the ability of the MBW outcomes estimated under the various validity criteria to detect early lung disease. Although the AREST CF cohort represents an excellent opportunity to make this assessment, at the time of writing the required data were not available. Furthermore, our validity study was limited to N₂ MBW only. MBW tests performed with other gases such as SF₆ are technically different to N₂ MBW tests and may require a different set of validity criteria. Examining the sensitivity and specificity of various validity criteria both with respect to a ‘gold standard’, ideally using multiple raters to ensure agreement, and the ability to detect of early lung disease, and developing validity criteria for non-N₂ MBW would be natural extensions of the analysis performed in this chapter.

Chapter 4

Modelling MBW and estimating LCI from incomplete tests

Contents

4.1 Background	42
4.2 Publication	43

This chapter details a novel statistical model for MBW data and applies it to the problem of estimating LCI from incomplete tests. Section 4.1 provides some of the motivation for this research which was published in the journal *Statistics in Medicine* and is included in Section 4.2.

4.1 Background

The MBW technique is complicated and time consuming [18]. The test takes several minutes to perform to completion and the associated recalibration of equipment and complete clearance of the patient airways takes longer still. The standard protocol is to calculate LCI based on three MBW test runs taken to completion. Many attempts are typically required to obtain acceptable data, particularly in patients with advanced disease, and therefore often large amounts of data are not used. Any flexibility allowing the test to be stopped earlier than usual would save clinician and patient time and effort. It is therefore of interest to shorten the testing procedure to save resources and effort for both patients and clinicians alike.

There have been a number of studies into shortening the test, with varying success. Yamine et al. [74] systematically assessed ways to shorten LCI measurements in children by examining the predictive values, coefficients of variation, and sensitivity and specificity of LCI values taken where the normalised end-tidal concentration was 15%, 9%, 5%, and 2.5%. They found that the LCI based on the mean of three successful N_2 MBW attempts using the 5% end-tidal threshold, and two successful attempts until 2.5%, had similar or better repeatability and sensitivity/specificity for CF in school-aged children than the standard LCI calculated from three successful attempts until 2.5%. Similarly, Hannon et al. [24] showed that in children and adults with CF, the mean LCI from three SF_6 successful attempts until 5% was as reliable and repeatable as until 2.5%. Similar studies by Ahmad et al. [29, 75], and Singer et al. [76] obtained similar results.

Not directly related to the flexibility of end-test thresholds and shortening of tests, but rather the loosening of the requirement for three complete tests, Robinson et al. [77] showed that, similar to Yamine et al. [74], robust estimates of LCI can be obtained from the first two technically acceptable tests while still being sensitive to abnormal peripheral airway function in CF subjects regardless of FRC repeatability.

These studies represent important contributions to the MBW technique, and bring into question the need to run MBW tests until the end-tidal gas concentration is 2.5% to calculate LCI and associated outcomes such as FRC. However, these empirical studies used relatively simple methods to establish equivalence of shortened-test LCI with complete-data LCI. Additionally, the estimates of LCI based on shortened test data only defined LCI up to the corresponding shortened end-test threshold, which is not ideal in cases where like-for-like comparison of LCI is required (either within individuals, for example in a clinical setting, or across individuals, in a research setting).

4.2 Publication

The included journal article was accepted for publication in February 2018 [78]. In it we first develop a novel statistical method for modelling MBW data, validate its use using a variance components model, and then show how this model can be used to estimate MBW outcomes in the absence of complete test data. This type of problem lends itself naturally to Bayesian methods, where prior information on model parameters can be

incorporated into the estimation procedure, potentially improving estimates over other estimation methods such as maximum likelihood estimation (MLE).

RESEARCH ARTICLE

Bayesian modelling of lung function data from multiple-breath washout tests

Robert K. Mahar^{1,2}  | John B. Carlin^{1,2,3}  | Sarath Ranganathan^{2,4,5}  |
Anne-Louise Ponsonby^{2,6}  | Peter Vuillermin^{6,7,8}  | Damjan Vukcevic^{1,9} 

¹Data Science, Murdoch Children's Research Institute, Parkville, Victoria, Australia

²Department of Paediatrics, Faculty of Medicine, Dentistry and Health Services, University of Melbourne, Parkville, Victoria, Australia

³Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Services, University of Melbourne, Parkville, Victoria, Australia

⁴Infection and Immunity, Murdoch Children's Research Institute, Parkville, Victoria, Australia

⁵Department of Respiratory and Sleep Medicine, Royal Children's Hospital, Parkville, Victoria, Australia

⁶Population Health, Murdoch Children's Research Institute, Parkville, Victoria, Australia

⁷School of Medicine, Faculty of Health, Deakin University, Geelong, Victoria, Australia

⁸Department of Paediatrics, Barwon Health, Geelong, Victoria, Australia

⁹School of Mathematics and Statistics, Faculty of Science, University of Melbourne, Parkville, Victoria, Australia

Correspondence

Damjan Vukcevic, Data Science, Murdoch Children's Research Institute, Flemington Road, Parkville 3052, Victoria, Australia.
Email: damjan.vukcevic@mcri.edu.au

Funding information

National Health and Medical Research Council, Grant/Award Numbers: ID 1009044, ID 1035261 and ID 1110200; Victorian Government's Operational Infrastructure Support Program; Australian Government Research Training Program

Paediatric respiratory researchers have widely adopted the multiple-breath washout (MBW) test because it allows assessment of lung function in unsedated infants and is well suited to longitudinal studies of lung development and disease. However, a substantial proportion of MBW tests in infants fail current acceptability criteria. We hypothesised that a model-based approach to analysing the data, in place of traditional simple empirical summaries, would enable more efficient use of these tests. We therefore developed a novel statistical model for infant MBW data and applied it to 1197 tests from 432 individuals from a large birth cohort study. We focus on Bayesian estimation of the lung clearance index, the most commonly used summary of lung function from MBW tests. Our results show that the model provides an excellent fit to the data and shed further light on statistical properties of the standard empirical approach. Furthermore, the modelling approach enables the lung clearance index to be estimated by using tests with different degrees of completeness, something not possible with the standard approach. Our model therefore allows previously unused data to be used rather than discarded, as well as routine use of shorter tests without significant loss of precision. Beyond our specific application, our work illustrates a number of important aspects of Bayesian modelling in practice, such as the importance of hierarchical specifications to account for repeated measurements and the value of model checking via posterior predictive distributions.

KEYWORDS

incomplete data, lung clearance index, multiple-breath washout, Stan, variance components

1 | INTRODUCTION

Chronic adult diseases, and respiratory diseases in particular, often have their origins in early life, and lung function in the early years is a major determinant of lung function in later life.¹ The human lung undergoes rapid development while in utero, continuing up until late infancy and, in some cases, well beyond that. Adverse lung growth and development during gestation and infancy have been associated with lifelong deficits in lung function and respiratory health. In this context, techniques are required that enable accurate and feasible measurement of lung function across the life course to better monitor disease progression and manage responses to therapeutic interventions.^{2,3}

The inert gas multiple-breath washout (MBW) test is increasingly used to measure lung function in individuals with chronic and serious respiratory diseases such as cystic fibrosis, as well as in cohort studies investigating the early life origins of asthma and chronic obstructive pulmonary disease. Recent clinical trials have also adopted the MBW as a primary tool for monitoring patient responses to therapeutic interventions.⁴ The test begins with a “wash-in” phase: a mouthpiece is fitted to the patient and, after several breaths of room air, the patient starts inhaling an inert tracer gas mixture until the concentration of inert gas is in equilibrium. At this point, the patient is switched back to room air and the “washout” phase of the test commences. The molar mass of the respired gases and airflow are measured at high frequency throughout the washout until the lungs are effectively clear of the tracer gas. Specialised software then performs breath detection and derives various quantities that are used to assess lung function. Crucially, the washout phase of the test must be performed during regular, uninterrupted breathing.

In young children the use of MBW is favoured over more conventional lung function tests such as spirometry, since it requires minimal patient cooperation and coordination.^{5,6} While lung function can be assessed by a number of different summaries of the MBW test, the most commonly used is the lung clearance index (LCI),⁷ which is based on the functional residual capacity (FRC), a widely used measure of lung function in its own right. The widespread adoption of the MBW has culminated in a recent consensus statement that recommends guidelines for MBW analysis and highlights areas that need further research.⁵

The data captured during an MBW test and used to calculate the LCI suggest an underlying functional form, yet to our knowledge, no one has thus far used an explicit statistical model to exploit this. The standard method for calculating LCI instead uses a simple algorithm based on the observed values.⁵ By not modelling the underlying process, the standard method does not in general fully exploit the available information. We propose a relatively simple stochastic model for MBW data that leads to a model-based definition of the LCI. Our approach explicitly models the quantities used to calculate LCI, namely, the cumulative expired volume (CEV) and FRC. Based on these explicit definitions of the estimands for LCI, CEV, and FRC, we develop a Bayesian estimation method.

One immediate and novel application of this model is to estimate LCI and FRC from incomplete or shortened MBW tests. Shortening the test procedure is of practical interest because it is time-consuming and prone to disruption, particularly in patients with advanced disease, in preschool children who have shorter attention spans, and among infants who are being tested during natural sleep. In some cases, breathing irregularities during the wash-in, such as sighs, can be acceptable.⁸ However, if breathing irregularities occur during the washout or the patient becomes unsettled, the test data are typically not used and additional testing must be performed or testing abandoned. Indeed, in most studies conducted among sleeping babies, successful MBW measures have only been obtained in 60% to 70% of tested participants. A number of recent studies have shown that under certain circumstances, MBW tests can be substantially shortened while still providing clinically meaningful results.⁹⁻¹¹ These studies represent important innovations with the MBW technique and cast doubt on whether the test must be complete before a meaningful summary statistic can be derived. However, they do this by defining a different type of LCI, which is based only on earlier time points in the test. In contrast, our approach is to model the available shortened test data, allowing us to meaningfully extrapolate it to directly estimate the usual LCI.

Using data from a large birth cohort study, we assess how our model-based method of estimating LCI, CEV, and FRC compares to the standard method using complete MBW tests. We then assess the performance of our model-based method with shortened MBW tests.

We adopt a Bayesian modelling approach as it allows us to naturally incorporate prior information and account for uncertainty in estimation of our model parameters. By assigning prior distributions to model parameters, we are also able to constrain them to more realistic values, which we show is particularly helpful when analysing incomplete tests. We use a subset of tests taken from normal infants to estimate an appropriate informative prior and validate its use on a complementary dataset.

A number of the practical statistical considerations that went into this analysis warrant further explanation. We therefore elaborate on these and consider their broader relevance.

2 | MODELLING MBW DATA

2.1 | Data description

The Barwon Infant Study (BIS) is a birth cohort study in the Barwon region of south-eastern Australia designed to investigate the impact of early-life exposures on immune, allergic, cardiovascular, respiratory, and neurodevelopmental outcomes (at inception $n = 1074$).¹²

Multiple-breath washout data were collected from infants aged between 4 and 12 weeks between February 2011 and December 2013. Acceptable tests exhibited a stable breathing pattern with no artefacts such as sighs, sucking, snoring, or mask leaks, or any observable software artefacts. We obtained a total of 1197 acceptable MBW tests from 432 infants in this study. Of these infants, 89 (21%) had only a single test performed, while 115 (27%), 98 (23%), 81 (19%), 35 (8%), 13 (3%), and 1 (<1%) participants had 2, 3, 4, 5, 6, and 7 replicate measurements available, respectively. Multiple-breath washout tests were performed in accordance with current guidelines.⁵ Infants were tested during natural sleep. All tests were performed with 4% sulphur hexafluoride (SF_6) by using a mainstream ultrasonic flowmeter (Exhalyzer D, Ecomedics, Duernten, Switzerland). For the purposes of this analysis, MBW data were processed with WBreath (version 3.19, ndd Medizintechnik AG, Zurich, Switzerland). There is limited native support for data processing and export in WBreath,¹³ so where possible we automated these functions via the graphical user interface by using AutoIt, a scripting language.¹⁴

2.2 | MBW model

The data collected as part of a routine MBW test consist of air flow (L/s) and molar mass of the gas mixture (g/mol) measured over a short period of time (approximately 1-3 min) at high frequency (200 Hz); see Figure 1. The wash-in component of the test is not typically of interest and is thus discarded. From the subsequent washout phase of the test, the following quantities are derived (Figure 2): the tracer gas quantity (GAS; the current amount of tracer gas, as a proportion of the amount of tracer gas at the start of the washout), CEV of gas mixture (CEVGM), and the CEV of tracer gas (CEVTG). As is typically done, we derive discrete versions of these quantities by calculating them on a per-breath basis (the quantities are therefore interpretable as representing the state at the end of each breath cycle).^{5,15}

We first introduce notation for modelling the data from a single MBW test. Let $k = \{0, 1, \dots, K\}$ index breaths from the start of the washout phase (indexed by $k = 0$). We denote the GAS, CEVGM, and CEVTG series as $c(k)$, $v(k)$, and $r(k)$, respectively. Figure 2 shows $c(k)$, $v(k)$, and $r(k)$ for 3 randomly selected MBW tests. The LCI as defined in the aforementioned consensus statement, herein referred to as the “standard method,” is denoted here as

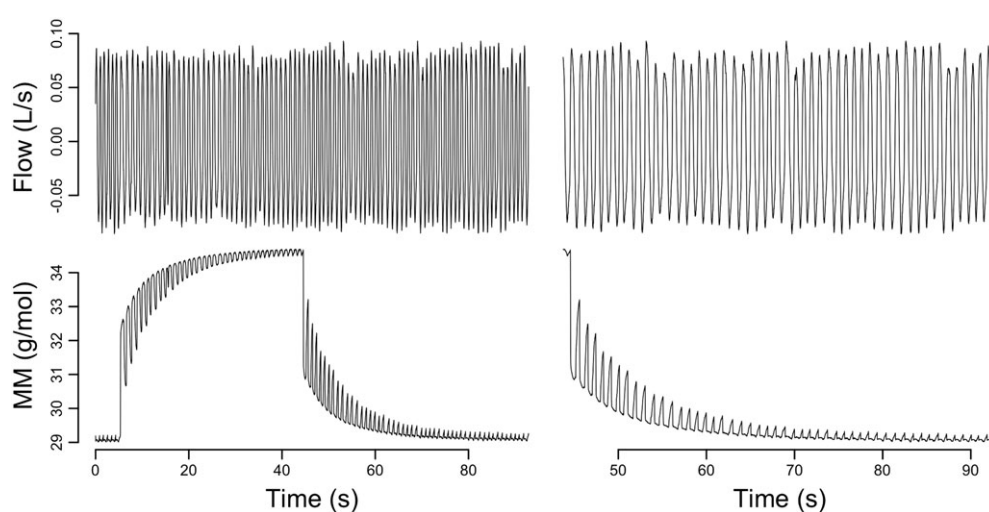


FIGURE 1 Example of raw multiple-breath washout data. Air flow (L/s) and molar mass (g/mol), shown over time (s) for a representative test. The left panels show the data for both the wash-in and washout phases; the right panels show the data for the washout phase only

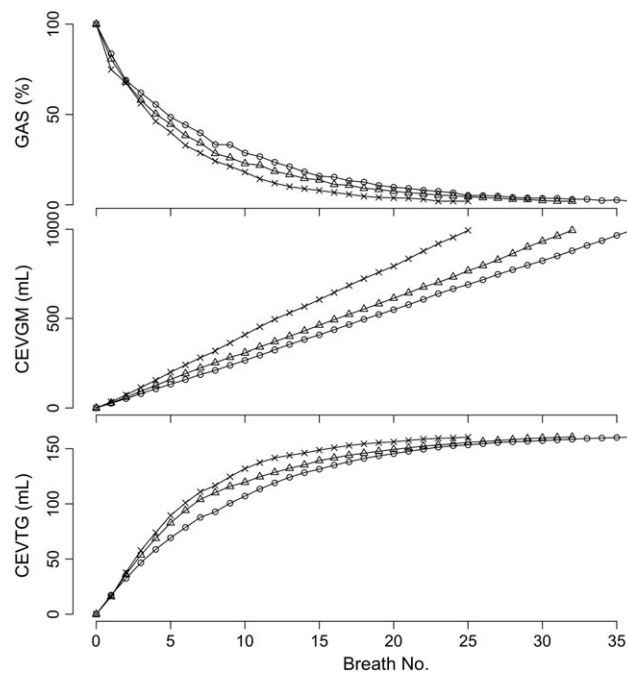


FIGURE 2 Example of derived multiple-breath washout test quantities. These are shown for 3 randomly selected tests. GAS: tracer gas quantity (%), CEVGM: cumulative expired volume of gas mixture (mL), and CEVTG: cumulative expired volume of tracer gas (mL)

$$\text{LCI}^{(s)} = \frac{\text{CEV}^{(s)}}{\text{FRC}^{(s)}} = \frac{v(k^{(40)})}{r(k^{(40)}) / (1 - c(k^{(40)}))}, \quad (1)$$

where the CEV ($\text{CEV}^{(s)}$) denotes the CEVGM measured at breath $k^{(40)}$, the FRC ($\text{FRC}^{(s)}$) denotes an estimate of the volume of a typical expiration (using quantities measured at breath $k^{(40)}$), and the parenthetical s denotes the “standard method.” As recommended in the consensus guidelines,⁵ the $k^{(40)}$ used to calculate the quantity in Equation 1 is the first of 3 consecutive breaths for which $c(k) \leq 1/40$. We define a complete test to be one that allows identification of $k^{(40)}$; ie, it must include data measured up to the third consecutive breath for which $c(k)$ is under $1/40$.

Our approach was to specify statistical models for $c(k)$, $v(k)$, and $r(k)$, with each quantity represented by a smooth function plus an independent random error process.

We introduce notation for vectors of parameters, $\beta = \{\beta_1, \beta_2, \dots, \beta_5\}$ and $\sigma = \{\sigma_c, \sigma_v, \sigma_r\}$, which will be used to characterise models for the 3 quantities of interest. We modelled the tracer gas quantity $c(k)$ by using a 2-phase exponential decay function of k for the (log-scale) mean, which is implied by a theoretical 2-compartment lung model¹⁶:

$$f(k; \beta) = \beta_0 \exp(-\beta_1 k) + (1 - \beta_0) \exp(-\beta_2 k), \quad (2)$$

where $0 < \beta_0 < 1$ is the proportionate weight of the fast decaying component and β_1 and β_2 ($0 < \beta_1 < \beta_2$) denote the fast and slow decay constants, respectively. This model is commonly applied to various physical systems, including the lungs, and at least one recent study has applied the model directly to MBW end-tidal tracer gas concentration in adults, albeit to achieve different aims.¹⁷ Since the data are restricted to be positive, we model c as log-normal with independent and identically distributed errors:

$$c(k) \sim \log N(f(k; \beta), \sigma_c^2). \quad (3)$$

The CEVGM, $v(k)$, was modelled as an increasing linear function of k , with geometric mean

$$g(k; \beta) = \beta_5 k, \quad (4)$$

and the CEVTG, $r(k)$, as an increasing exponential decay function of k , with geometric mean

$$h(k; \beta) = \beta_3(1 - \exp(-\beta_4 k)). \quad (5)$$

Note that the functions v and r are both strictly positive and monotonically increasing, which we capture by modeling their discrete derivatives (denoted here with primes), and do so on a log scale:

$$\ln(g'(k; \beta)) = \ln(\beta_5), \quad (6)$$

$$\ln(h'(k; \beta)) = \ln(\beta_3) + \ln(1 - \exp(-\beta_4)) - \beta_4 k. \quad (7)$$

An added benefit of these transformations is that we can assume independent and identically distributed errors on a natural scale for each quantity, giving us the final model:

$$v'(k) \sim \log N(g'(k; \beta), \sigma_v^2), \quad (8)$$

$$r'(k) \sim \log N(h'(k; \beta), \sigma_r^2). \quad (9)$$

Using this model, we define the end-test breath, which we denote θ , as the real-valued breath index at which the expected value of c is equal to 1/40:

$$f(k = \theta; \beta) = 1/40. \quad (10)$$

We can thus redefine FRC, CEV, and therefore LCI, in terms of the model parameters:

$$\text{CEV}^{(m)} = g(\theta; \beta), \quad (11)$$

$$\text{FRC}^{(m)} = \frac{h(\theta; \beta)}{1 - f(\theta; \beta)}, \quad (12)$$

$$\text{LCI}^{(m)} = \frac{g(\theta; \beta)}{h(\theta; \beta)/(1 - f(\theta; \beta))}, \quad (13)$$

where the parenthetic m denotes the “model-based” method. Note, however, that the standard method defines FRC as a simple summary statistic, by taking the observed CEVTG and dividing it by $1 - c(k)$ at breath $k^{(40)}$, where some gas will remain in the lungs. Using the model-based method, a more natural approach is to use the asymptotic value of $\text{FRC}^{(m)}$, which represents the state of the lungs when completely free of the inert gas:

$$\lim_{\theta \rightarrow \infty} \frac{h(\theta; \beta)}{1 - f(\theta; \beta)} = \beta_3. \quad (14)$$

Thus, we propose a more refined model-based definition, denoted by an asterisk, with $\text{FRC}^{(m^*)} = \beta_3$, giving

$$\text{LCI}^{(m^*)} = \frac{g(\theta; \beta)}{\beta_3}. \quad (15)$$

Given an end-test threshold 1/40 (n.b. the numerator still depends on this threshold), the $\text{LCI}^{(m^*)}$ is therefore wholly defined as a function of the model parameters β .

We fitted the models using a Bayesian approach, assigning either diffuse or informative prior distributions to the parameters. Stan, a probabilistic programming language,¹⁸ was used for parameter estimation (see Appendix A). Diffuse distributions contained only vague information based on the potential scale and theoretical support of a parameter.¹⁹ Therefore, we assumed that $\beta_0 \sim \text{Beta}(2, 2)$, $\beta_1, \beta_2, \beta_4 \sim N(0, 1)$, $\beta_3 \sim N(0, 10^3)$, and $\beta_5 \sim N(0, 10^2)$. These priors have the advantage over wider, less informative priors in that they aid computation and inference while providing only minimal

information. We used half-Cauchy distributions as diffuse priors for the standard deviations, as recommended,^{19,20} such that $\sigma \sim \text{Half-Cauchy}(0, 2.5)$. Informative prior distributions, on the other hand, were intended to characterise the typical distribution of the MBW model parameters and took the form of a multivariate normal distribution such that $\beta \sim \text{MVN}(\mu, \Sigma)$, where μ denotes the mean vector and Σ denotes the variance-covariance matrix. The informative priors were obtained by fitting a 2-level hierarchical model for GAS, CEVGM, and CEVTG to a subset of our data, as outlined in the next section.

We fitted the models with both diffuse and informative priors to individual tests with complete data and calculated the posterior medians of $\text{CEV}^{(m)}$, $\text{FRC}^{(m^*)}$, and $\text{LCI}^{(m^*)}$ in both cases, which we refer to herein as the “model-based” estimates.

2.3 | Obtaining an informative prior via a hierarchical model

The parameters of the informative prior distributions used in this paper were estimated by using hyperparameters obtained by fitting a 2-level hierarchical model for GAS, CEVGM, and CEVTG. This model allows the sharing of information between tests by modelling the lower-level per-test parameters (or random effects) as coming from a common distribution defined by the hyperparameters. This allowed us to summarise and extract the relevant information from a large set of tests into a form suitable for use as a prior distribution. Because our primary interest here was to approximate prior distributions for use with individually modelled tests, we deliberately used a simple 2-level hierarchical model rather than a more complex 3-level model that would account for the fact that we had multiple replicates for each participant. To ensure that we were using data appropriate to this simpler model specification, we reduced the data by taking only the first replicate test for each patient, resulting in 414 tests. We adopted a simple holdout/validation approach whereby we fitted this 2-level model to one half, or 212, of the tests in this dataset. The posterior medians of the hyperparameter distributions were extracted for use as parameters of the informative prior distributions on the validation data.

We define hyperparameters β with lower-level parameters $\mathbf{w}_i = \{w_{1i}, w_{2i}, \dots, w_{6i}\}$, a combined vector of random effects for test $i = \{1, \dots, 212\}$. We specify our hierarchical model as follows. First, the underlying functional forms for each series were

$$\begin{aligned} f(k; \beta, \mathbf{w}_i) &= (\beta_0 + w_{0i}) \exp((- \beta_1 + w_{1i})k) + (1 - (\beta_0 + w_{0i})) \exp((- \beta_2 + w_{2i})k) \\ g(k; \beta, \mathbf{w}_i) &= (\beta_5 + w_{5i})k \\ h(k; \beta, \mathbf{w}_i) &= (\beta_3 + w_{3i})(1 - \exp((- \beta_4 + w_{4i})k)) \end{aligned}$$

The distribution of our observations for a test i can now be summarised as follows:

$$\begin{aligned} c_i(k) &\sim \text{logN}(f(k; \beta, \mathbf{w}_i), \sigma_c^2) \\ v_i'(k) &\sim \text{logN}(g'(k; \beta, \mathbf{w}_i), \sigma_v^2) \\ r_i'(k) &\sim \text{logN}(h'(k; \beta, \mathbf{w}_i), \sigma_r^2) \end{aligned}$$

We specify a multivariate normal distribution for the random effects: $\mathbf{w}_i \sim \text{MVN}(0, \Sigma)$, where Σ is the variance-covariance matrix of the random effects $\mathbf{w}_i$ and has a correlation matrix $\mathbf{P}$. As is common practice in Stan, we define a prior on the correlation matrix, $\mathbf{L} \sim \text{LKJ}(2.0)$, where $\mathbf{L}$ is the Cholesky decomposition of $\mathbf{P}$ and $\mathbf{w}_i = \text{diag}(\sigma_w) \mathbf{L} \mathbf{Z}$, where $\mathbf{Z}$ is a matrix of unit normal random variables corresponding to the individual random effects; this prior and parameterisation implies that the off-diagonals of the correlation matrix are near zero and is consistent with a multivariate noncentred parameterisation on the random effects $\mathbf{w}_i$.^{19,21} All hyperparameters contained in β were specified with diffuse priors as outlined in the previous section. Note that we assume that the standard deviations are fixed across tests and $\sigma \sim \text{Half-Cauchy}(0, 2.5)$.

2.4 | Results

We fitted the hierarchical MBW model to 209 of the 212 MBW tests in the holdout subset. Three of the tests were excluded as the data appeared highly irregular. Parameter convergence was sufficient, with $\hat{R}$ values close to 1, and effective sample sizes ranged across the hyperparameters from 102 to 4000 (median: 452). A summary of the posterior distributions used to create informative priors is shown in Table 1, along with the relevant correlation matrix.

TABLE 1 Summary of the hierarchical model posterior

Parameter	Posterior Summary		Correlations				
	Median	SD	β_1	β_2	β_3	β_4	β_5
β_0	0.68 (0.65, 0.7)	0.15 (0.14, 0.17)	25 (11, 36)	83 (75, 89)	2 (−12, 17)	−15 (−28, −2)	−32 (−44, −18)
β_1	0.129 (0.126, 0.13)	0.023 (0.021, 0.025)	-	25 (8, 40)	−11 (−25, 4)	91 (88, 94)	56 (46, 65)
β_2	0.53 (0.49, 0.58)	0.2 (0.16, 0.23)	-	-	−8 (−25, 8)	−10 (−26, 7)	−17 (−32, −2)
β_3	124.3 (121.85, 126.97)	18.27 (16.5, 20.34)	-	-	-	−14 (−27, 0)	53 (42, 62)
β_4	0.137 (0.134, 0.14)	0.024 (0.022, 0.026)	-	-	-	-	68 (60, 75)
β_5	29.96 (29.09, 30.85)	6.38 (5.83, 7.04)	-	-	-	-	-

Posterior medians and 95% credible intervals of the hyperparameters in the model. “SD” refers to the standard deviation of the random effects associated with each hyperparameter. For clarity, the correlation matrix shows the correlations multiplied by 100 (eg, 0.24 is shown as 24).

We also fitted the individual MBW model using diffuse priors to each of the 1197 individual tests using diffuse priors and to each of 212 individual tests in the validation subset using the informative priors. Parameter convergence was sufficient in most cases, with $\hat{R}$ values close to 1. Although for a small number of tests some parameters exhibited $\hat{R} > 1.1$, we judged that this was unlikely to materially affect the overall results. Effective sample sizes across all tests were judged as sufficient with a median of 1141 (interquartile range: 841, 1378) and 1207 (interquartile range: 900, 1418) for the diffuse and informative priors, respectively.

The model fit for a typical individual's MBW data with diffuse prior information is shown in Figure 3, along with 10 simulated datasets drawn from the posterior predictive distribution for each quantity. The model appears to describe the data well on the transformed scale (left column), and on the original scale (right column) (for detailed description of model fit, see section 5). Although we only show a single test in Figure 3, we observed similar results for other tests.

Upon fitting the above models, we examined the residual errors for the presence of autocorrelation via visual inspection of the residual partial autocorrelation function of 100 randomly selected tests (data not shown). Additionally, we fitted the models described in Equations 3, 8, and 9 to the same 100 individual tests, modifying the error term to have an autoregressive structure with lag 1 (AR1). Overall, there was insufficient evidence to suggest the presence of a clear autocorrelation structure or differences in the estimates of MBW outcomes when using either independent or AR1 errors. We therefore retained the assumption of independent errors for each model.

3 | ASSESSING AGREEMENT AND VARIANCE

3.1 | Model

We assessed agreement between the standard and model-based methods by using a Bayesian variance components model. Since the MBW tests in our study were performed multiple times on the same participant, we needed a model flexible enough to account for replicate measurements. We used a variation of a model proposed by Carstensen et al,²² which accounts for variation between participants and also variation due to replicate measurements. Whereas Carstensen et al²² model the between-participant variation using fixed effects, in the Bayesian setting, and with a large number of participants, we found that these were more naturally modelled as random effects.

We used the above model solely for comparing the standard method and the model-based method with diffuse prior, to examine the agreement between the methods without influence from informative prior information. Using the additional information will simply tend to shrink the model-based MBW outcomes toward the prior mean, so we judged that a comparison with the informative prior was unnecessary. Unlike standard agreement analysis, our goal here is not to establish strict equivalence or interchangeability between the 2 methods, but rather to examine how they differ and whether we are satisfied that any differences are justified. The use of a proper agreement model for this purpose allows us to properly control for the replicate structure within our data.

Denoting the method of measurement by m , where $m = 1$ denotes the standard method, $m = 2$ denotes the model-based method using diffuse priors, the study participants by p ($p = 1, \dots, 432$), and the r 'th replicate of method m on participant p as y_{mpr} , the agreement model was defined as

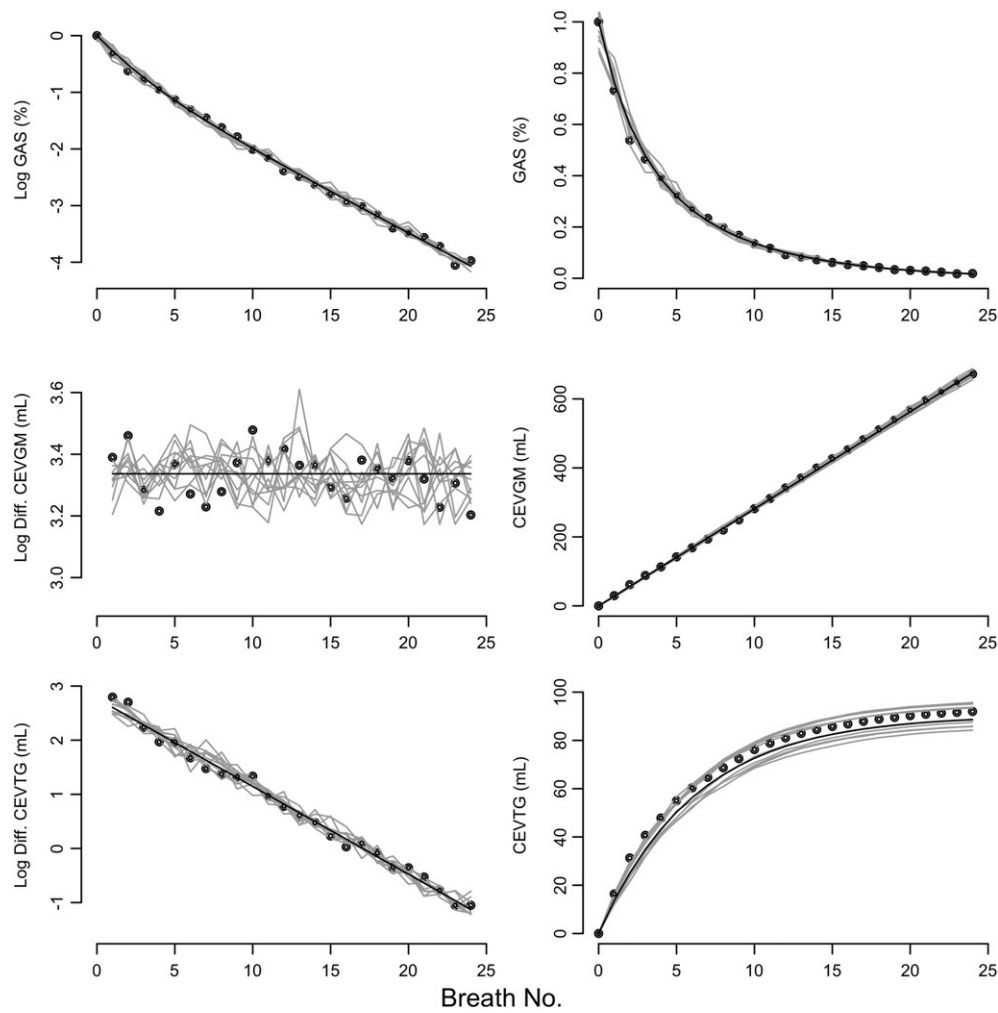


FIGURE 3 Example of a model fit. Plots showing, for a typical test and informative priors, the observed data (points), model-based mean prediction (black line), and a sample of posterior predictive draws (grey lines). The panels in the left column show the quantities on the transformed scale, those in the right column show them on the original scale

$$\begin{aligned}
 y_{mpr} &= \alpha_m + u_p + a_{pr} + c_{mp} + e_{mpr}, \\
 u_p &\sim N(0, \gamma), \quad a_{pr} \sim N(0, \omega), \quad c_{mp} \sim N(0, \tau_m), \quad e_{mpr} \sim N(0, \sigma_m), \\
 \alpha_m &\sim N(0, 10^4), \quad \gamma, \omega, \tau_m, \sigma_m \sim \text{Half-Cauchy}(0, 2.5),
 \end{aligned} \tag{16}$$

where α_m is the mean effect for method m , the variation between participants is captured by γ , the variation between replicates (nested within participants) is captured by ω , the extra variation between participants attributable to each method is captured by τ_m , and the within-participant variation for a specific method m (ie, across replicates) is captured by the residual standard deviation σ_m . Note that the inclusion of the a_{pr} terms makes this correspond to the “linked replicates” (referred to as “nonexchangeable replicates” henceforth) model of Carstensen et al.²² As we are comparing only 2 methods, the different τ parameters become indistinguishable, so we set $\tau_1 = \tau_2 = \tau$; furthermore, the size of this component was estimated to be negligible in our data.

We encountered numerical issues fitting the full model that could not be resolved through either model reparameterisation or fine tuning of the sampler parameters. Therefore, although our data follow a nonexchangeable structure (since the input data used by each method are identical, meaning that the output estimates correspond to each other on a per-replicate basis), we assumed exchangeable replicates, dropping the α_{pr} terms (which have the same value across methods, and thus capture the “linking” structure). This allowed us to estimate a separate residual standard deviation for each method, albeit with the variation due to replicates, which would otherwise be captured by ω , being distributed among the estimable variance components γ , τ , and σ_m . However, it was also of interest to understand the size of the

per-replicate variation, so we fitted a second model where we included α_{pr} but assumed a common residual standard deviation for each method; ie, $\sigma_1 = \sigma_2 = \sigma$.

We again took a Bayesian modelling approach, using Stan for estimation (see Appendix A), and assigned diffuse priors on the model parameters^{19,20}: normal distributions for the mean components, and half-Cauchy distributions for the variance components, as shown in 16.

3.2 | Results

The 2 variance components models were fitted to the 1197 MBW tests using Stan. Parameter convergence was sufficient in all cases, with $\hat{R}$ values close to 1. Effective sample sizes were also judged as sufficient, ranging from 70 to 4000 (median: 2929) and 977 to 4000 (median: 483) across all mean and variance components for the exchangeable replicates model and the nonexchangeable replicates model, respectively.

Comparisons of LCI, CEV, and FRC estimated by each method, relative to the standard method, are shown in Figure 4 (in this figure, we also show, for reference only, a comparison with the estimates when using an informative prior, but as explained above we do not discuss these in this section). Estimates of the parameters of the variance component model assuming exchangeable replicates and method-specific residuals are summarised in Table 2.

For CEV and FRC, the model-based estimates were slightly smaller on average (looking at the differences between mean components). This was presumably because of the use of the model-based end-test definition, which, by nature, is typically earlier than when using the standard method. Nevertheless, the resulting estimates of LCI appeared to be similar between both methods.

The residual variance component of the model-based LCI was slightly higher than under the standard method. In contrast, for CEV, it was the other way around, which may be explained by the different end-test definitions (see previous paragraph). However, all of these differences are minor when compared with the overall magnitude of variation.

The method-participant interaction parameter, τ , was relatively small for each outcome, which was not surprising since we were comparing similar methods. Interestingly, for CEV and FRC, the within-participant (residual) variation for each method was about half the size of the between-participant variation γ , but for LCI, these were roughly similar to each other.

Table 3 shows estimates from the variance component model assuming nonexchangeable replicates and a single common residual term. The main purpose of fitting this model was to investigate the extent to which the exchangeability assumption affected our interpretation of the previous model. As the table shows, the mean components remained similar, but a substantial amount of variation was “redirected” to the replicate-participant interaction term ω from the between-participant interaction γ and the residual σ . This suggests that a substantial amount of the residual variation summarised in Table 2 is in fact because of the variation between replicates, meaning that the scope for substantial differences between methods in this regard is actually rather limited.

Taken together, these results suggest that the diffuse-prior model produced estimates of LCI that could be used interchangeably with the standard method. Furthermore, they suggest that the standard method is in fact a good estimator of MBW outcomes using complete data.

4 | ASSESSING SHORTENED TESTS

4.1 | Method

We assessed the accuracy of our model when using shortened test data. For each test, we truncated each of $c(k)$, $v(k)$, and $r(k)$ up to the first k where $c(k)$ fell below some threshold and remained there for 2 subsequent breaths, in the same way that the standard method is used to empirically determine the end-test breath. This method of truncation is similar to that used by other studies that have evaluated the use of “shortened” MBW tests.⁹⁻¹¹ We used 6 different thresholds for the tracer gas quantity: 1/3, 1/4, 1/5, 1/10, 1/20, and 1/30. We fitted the MBW models to the artificially shortened test data to estimate, by posterior medians, the model-based outcomes LCI and FRC by using either the diffuse or informative prior.

We assessed the accuracy of our model, for each of the 6 truncation thresholds, by summarising estimates of LCI and FRC from shortened tests relative to estimates obtained using complete test data (to make them comparable across tests). We summarised these by examining the distribution of the relative prediction errors and the width of the 95% credible intervals (CIs) of the predictions.

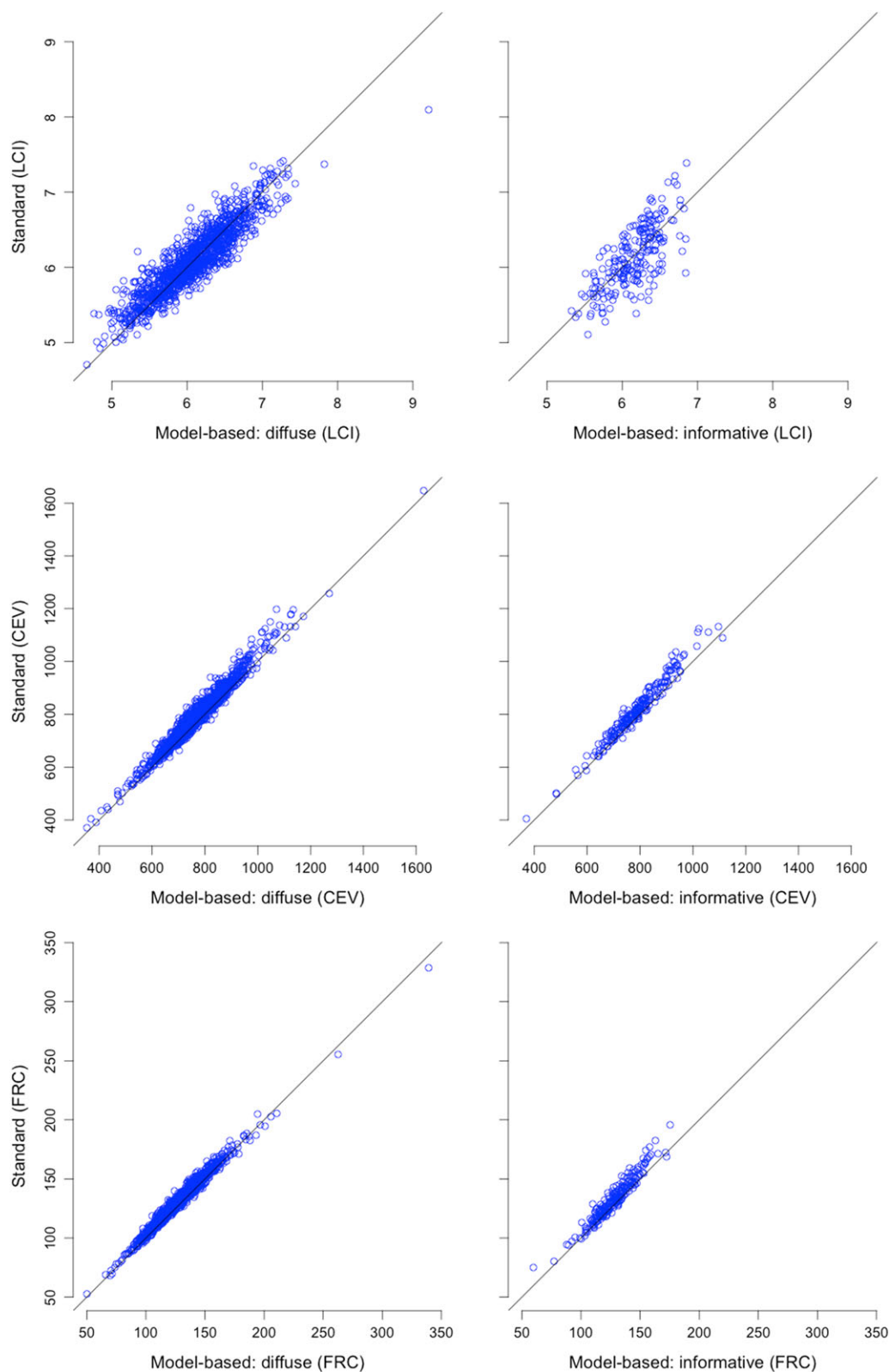


FIGURE 4 Comparison of estimation methods. Scatterplots showing a comparison of multiple-breath washout quantities estimated with complete data using the standard method vs model-based method with diffuse ($n = 1197$) and informative priors ($n = 212$). LCI, lung clearance index; CEV, cumulative expired volume (mL); FRC, functional residual capacity (mL) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Agreement analysis: Exchangeable replicates, method-specific residuals

Parameter	LCI		CEV (mL)		FRC (mL)	
	Median	95% CI	Median	95% CI	Median	95% CI
<i>Mean components</i>						
Standard method (α_1)	6.10	(6.07, 6.14)	796.40	(785.28, 808.21)	131.10	(129.13, 133.15)
Model-based method: diffuse (α_2)	6.08	(6.05, 6.12)	773.28	(761.96, 785.12)	127.90	(125.94, 129.93)
Mean difference ($\alpha_1 - \alpha_2$)	0.02	(-0.01, 0.05)	23.14	(19.07, 27.34)	3.20	(2.51, 3.94)
<i>Variation components</i>						
Between-participant (γ)	0.31	(0.28, 0.34)	110.38	(103.38, 119.63)	20.50	(19.22, 21.96)
Method-participant interaction (τ)	0.01	(0, 0.03)	0.84	(0.03, 2.97)	0.18	(0.01, 0.57)
<i>Residual components</i>						
Standard method (σ_1)	0.35	(0.33, 0.36)	52.58	(50.4, 54.93)	8.70	(8.32, 9.07)
Model-based method: diffuse (σ_2)	0.39	(0.37, 0.41)	48.86	(46.86, 51.05)	9.00	(8.66, 9.41)
Residual ratio (σ_1/σ_2)	0.89	(0.84, 0.95)	1.08	(1.02, 1.14)	1.00	(0.91, 1.02)

Parameter estimates (posterior medians and 95% credible intervals) from the variance components model fitted to the Barwon Infant Study multiple-breath washout outcomes, comparing standard and diffuse prior model-based estimates. Notes: LCI: lung clearance index; CEV: cumulative expired volume; FRC: functional residual capacity; CI: credible interval.

TABLE 3 Agreement analysis: Nonexchangeable replicates, common residuals

Parameter	LCI		CEV (mL)		FRC (mL)	
	Median	95% CI	Median	95% CI	Median	95% CI
<i>Mean components</i>						
Standard method (α_1)	6.10	(6.07, 6.14)	796.44	(785.61, 807.47)	131.00	(129.1, 132.95)
Model-based method: diffuse (α_2)	6.08	(6.05, 6.12)	773.11	(762.43, 783.87)	127.80	(125.88, 129.71)
Mean difference ($\alpha_1 - \alpha_2$)	0.02	(0.01, 0.03)	23.29	(21.88, 24.69)	3.20	(3.02, 3.43)
<i>Variation components</i>						
Between-participant (γ)	0.24	(0.21, 0.28)	104.48	(96.68, 113.02)	19.20	(17.86, 20.87)
Replicate-participant interaction (ω)	0.38	(0.36, 0.4)	55.08	(52.29, 58.07)	9.80	(9.34, 10.37)
Method-participant interaction (τ)	0.03	(0, 0.05)	3.95	(1.11, 5.73)	0.90	(0.68, 1.07)
Common residuals (σ)	0.14	(0.137, 0.15)	15.32	(14.55, 16.14)	2.10	(1.98, 2.18)

Parameter estimates (posterior medians and 95% credible intervals) from the variance component model fitted to the Barwon Infant Study multiple-breath washout outcomes, comparing standard and diffuse prior model-based estimates. Notes: LCI: lung clearance index; CEV: cumulative expired volume; FRC: functional residual capacity; CI: credible interval.

4.2 | Results

We fitted the MBW models to the 6 artificially shortened versions of all 1197 MBW tests by using diffuse priors, and similarly also to the 212 MBW tests in the validation sample by using informative priors.

In total, this resulted in close to 8500 separate analyses. Since it is impractical to assess the convergence of so many tests graphically, for these analyses, convergence was assessed to be sufficient for a single test if all parameters had scale reduction factors ($\hat{R}$) below 1.25. In general, as shown in Table 4, relatively few tests exhibited suboptimal Markov chain Monte Carlo (MCMC) convergence, with the exception of the scenarios that had very little data. In any case, where convergence was not sufficient, it is likely that finer tuning of the MCMC sampler or additional sampling iterations would have resolved the issue. Regardless, dropping the estimates from tests that had convergence issues had negligible impact on these results for any of the diffuse and informative prior information cases, and thus, we did not remove them from the results shown here.

TABLE 4 Summary of MCMC convergence for the analysis of shortened tests

Shortened Test Length	Proportion of Tests With Convergence Issues (%)	
	Diffuse Prior	Informative Prior
1/3	40	18
1/4	8	3
1/5	<1	2
1/10	<1	5
1/20	<1	5
1/30	<1	2
Complete data	<1	3

The proportion of tests that exhibited suboptimal parameter convergence, by having $\hat{R} > 1.25$ for at least one test parameter, in the shortened tests analysis.

Figure 5 shows plots of the median and the 2.5th and 97.5th percentiles of the prediction errors and the width of the 95% CIs, for each of the shortened datasets across either 1197 MBW tests (in the diffuse prior case) or 212 MBW tests (in the informative prior case), as a percentage of the corresponding model-based LCI estimate using complete data. Figure 6 shows a similar summary for estimates of FRC.

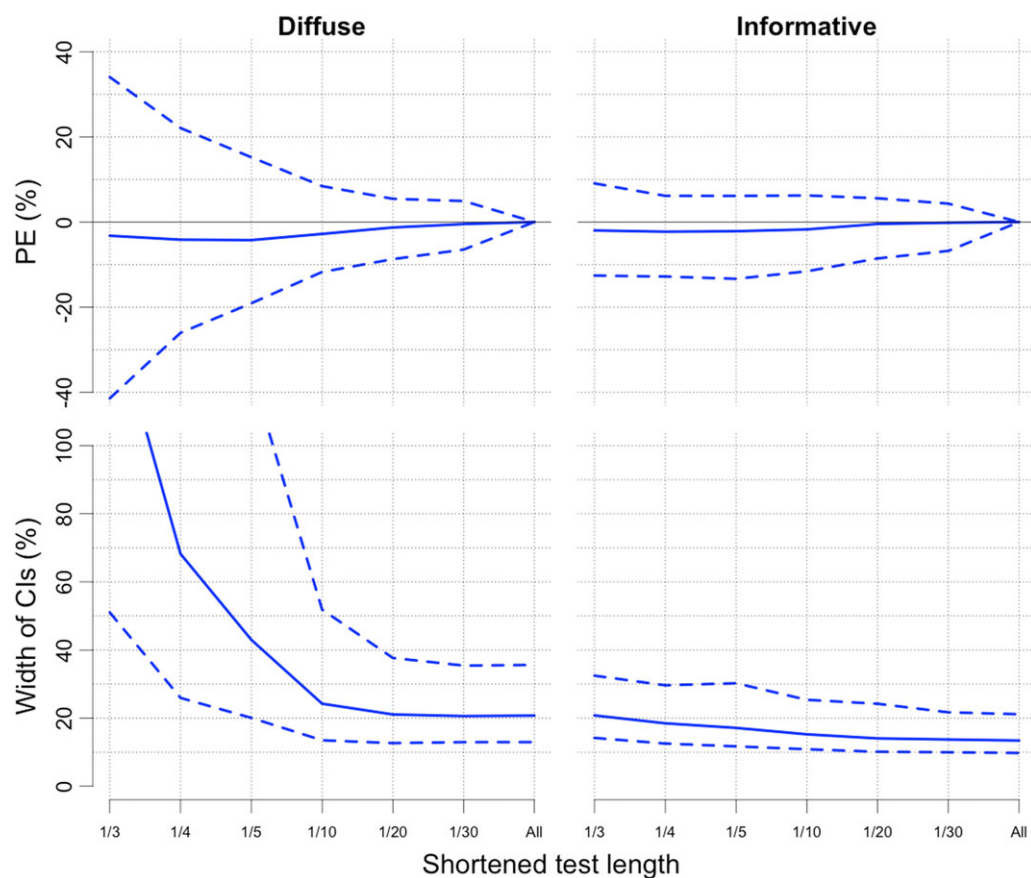


FIGURE 5 Estimation performance for LCI using shortened tests. The distributions of the posterior median prediction errors (PEs) and widths of 95% credible intervals derived from the model-based LCI posteriors, shown for scenarios based on different degrees of test completeness and either the diffuse ($n = 1197$) or informative prior distribution ($n = 212$). Both of the performance measures were first divided by the posterior median using complete test data (on a per-test basis) and are therefore shown as a percentage. Solid blue lines: median of the performance measure across all tests; dashed blue lines: 2.5th and 97.5th percentiles of the performance measure across all tests; shortened test lengths correspond to a tracer gas quantity; “All”: refers to estimates from tests using complete data, included for reference [Colour figure can be viewed at wileyonlinelibrary.com]

For LCI, in both the diffuse and informative prior information cases, the accuracy of the model-based estimate clearly improved with more data. Furthermore, in each shortened test scenario, the median prediction error was less than 10% under either prior. This indicates that the model-based estimates are generally accurate even when the data are highly incomplete. The informative priors helped to moderate the estimates, particularly when the tests were highly incomplete, somewhat reducing the bias that was discernible with the diffuse priors. The range of the prediction errors shrank as we considered longer tests, showing how the estimates become progressively more reliable with additional data. For shortened test thresholds higher than 1/10, the model was clearly more reliable when using informative priors.

As shown in Figure 6, the results were similar for estimation of FRC, although the model-based FRC showed little to no discernible bias in either the diffuse and informative prior cases, and the range of the prediction errors was smaller than that for LCI in both cases, with the exception of tests shortened to tracer gas quantities higher than 1/10 in the diffuse case.

The strong downward bias of LCI for very short tests when using diffuse priors can be understood in the context of the underlying models. The log discrete derivative models, used for CEVGM and CEVTG, are linear and thus can be extrapolated with some accuracy beyond their range. In contrast, the underlying model for θ is a mixture of fast and slow decaying exponential components. With only a limited amount of data available (due to short tests) and the absence of informative priors to constrain the 2 components, the model placed a greater weight on the fast decaying component, which dominates in the earlier stages of the test, resulting in a smaller θ , and thus smaller $CEV^{(m)}$ and $LCI^{(m^*)}$. This was somewhat rectified by using informative priors, which ensured that the model properly acknowledged the existence of a slowly emptying compartment even when there were insufficient data to easily estimate it, resulting in a more accurate estimate of θ , and correspondingly better estimates of $CEV^{(m)}$ and $LCI^{(m^*)}$.

A clear issue with both the LCI and FRC estimates using shortened tests was the degree of uncertainty around each estimate. This is assessed in both Figures 5 and 6 by summarising the proportional width of the 95% CI across the

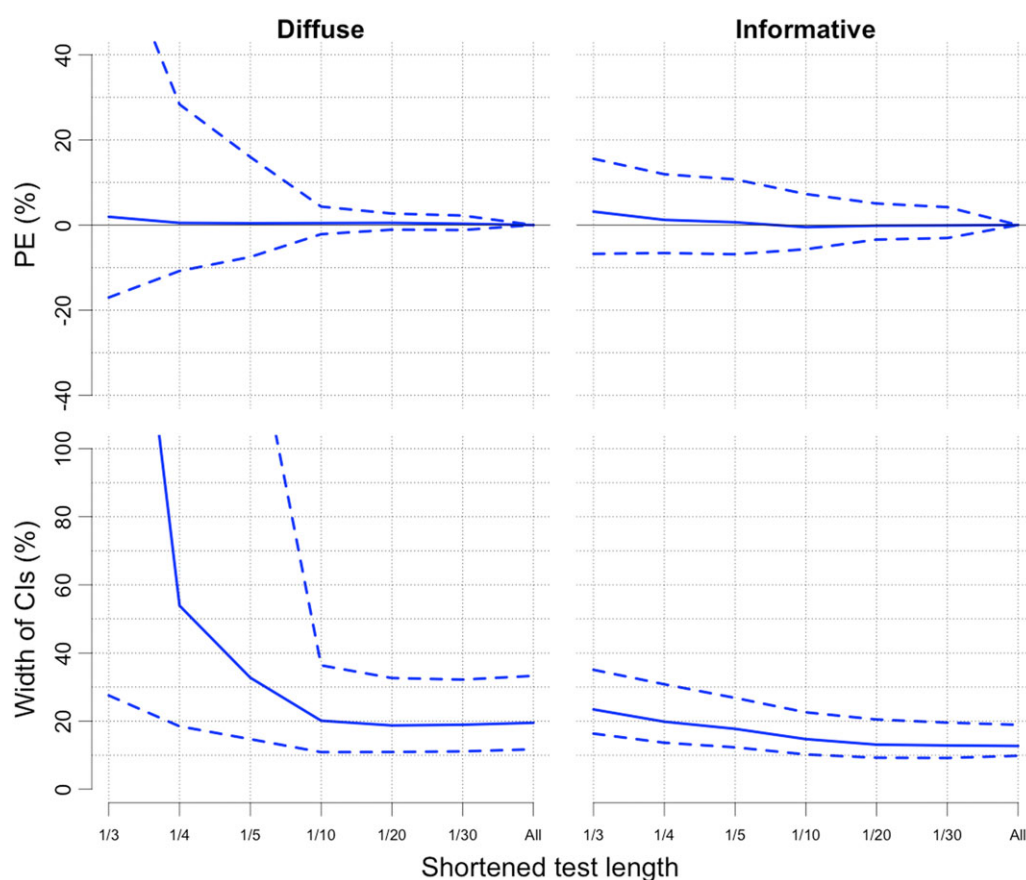


FIGURE 6 Estimation performance for FRC using shortened tests. The same as Figure 5 but now for estimates of functional residual capacity rather than lung clearance index [Colour figure can be viewed at wileyonlinelibrary.com]

estimates for each shortened test scenario. Uncertainty around estimates of both $FRC^{(m')}$ and $LCI^{(m')}$ using data shortened to thresholds higher than 1/10 is high enough to discourage any use of estimates from tests shorter than this level. However, the uncertainty appeared to stabilise at a level similar to the complete data case at thresholds equal to or lower than 1/10, particularly when using informative prior information. The fact that the uncertainty did not reduce beyond this threshold is likely representative of the inherent uncertainty in the underlying LCI definition. Moreover, this suggests that there is little to be gained by performing the test for any longer than up to a tracer gas quantity threshold of around 1/10.

Overall, we were able to obtain reasonable estimates of LCI and FRC even with data shortened to a tracer gas quantity threshold as high as 1/10. On average, a complete test is 30 breaths in length, and a test shortened to the 1/10 tracer gas level is approximately 17 breaths, translating to an average 43% reduction in breaths required. For tests shortened to thresholds higher than this, our estimates became inaccurate. This suggests some portion of the test beyond the start but before the 1/10 threshold contains information necessary to determine LCI at the end of the test. This makes intuitive sense: because the areas of the lung with least ventilation are likely to take longer to empty, the slope of the tracer gas quantity curve is flatter later in the test. Predictions made without any data from this part of the test will typically be underestimates because θ will typically be underestimated. From these results, it appears that, in general, sufficient additional information to reliably estimate LCI becomes available between proportions 1/5 and 1/10 (on average this gap spans approximately 5 breaths).

5 | FURTHER CONSIDERATIONS

5.1 | Modelling cumulative data

The wide variation of the posterior predictive distribution of the CEVTG model on the cumulative scale, as shown in Figure 7, may appear at odds with the observed data. Note, however, that a key feature of the data was the strict monotonicity of the CEVGM and CEVTG series.

We initially specified our model with independent errors directly on the cumulative scale. While this produced posterior median estimates that fit the observed data closely, it also gave rise to posterior predictive distributions that did not reflect the nature of the observed data. In particular, the model violated the monotonicity constraint for both series (see Figure 7). We also found that the posterior distributions of the parameters did not reflect the observed variation across replicates, with CIs for FRC (which is derived from CEVTG) being overly narrow (see Figure 8).

In thinking about alternative models, shifting to the discrete derivative on a log scale seemed a natural choice to preserve monotonicity. When viewed on the cumulative scale, this resulted in series with seemingly overly wide variation. However, upon reflection, we realised that such variation is actually probably indicative of reality and reflects the nature of these data. Namely, CEVTG is a cumulative measure of a decreasing quantity, meaning that the observed

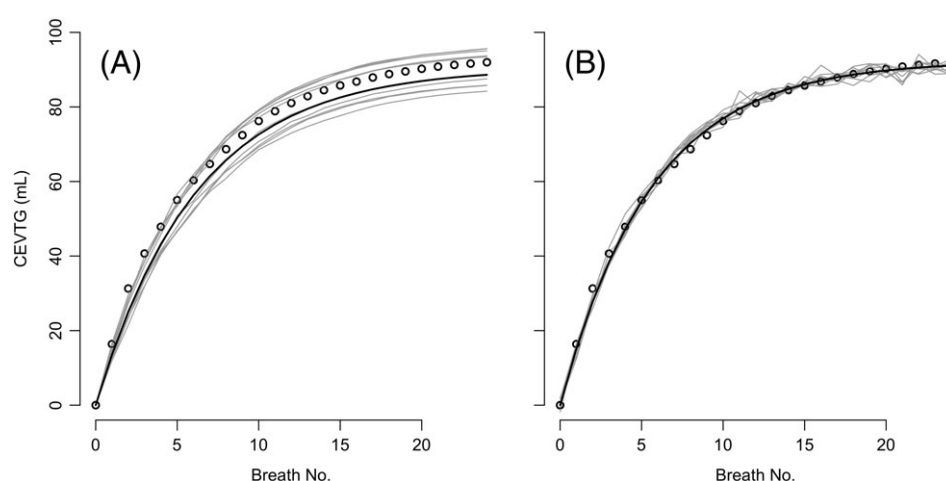


FIGURE 7 Contrasting alternative model fits on the cumulative scale. Plots showing, for a typical test, the observed data (points), model-based median prediction (black lines), and a sample of posterior predictive draws (grey lines) for the fit based on (A) a model for the log discrete derivatives and (B) a model directly for the untransformed cumulative data

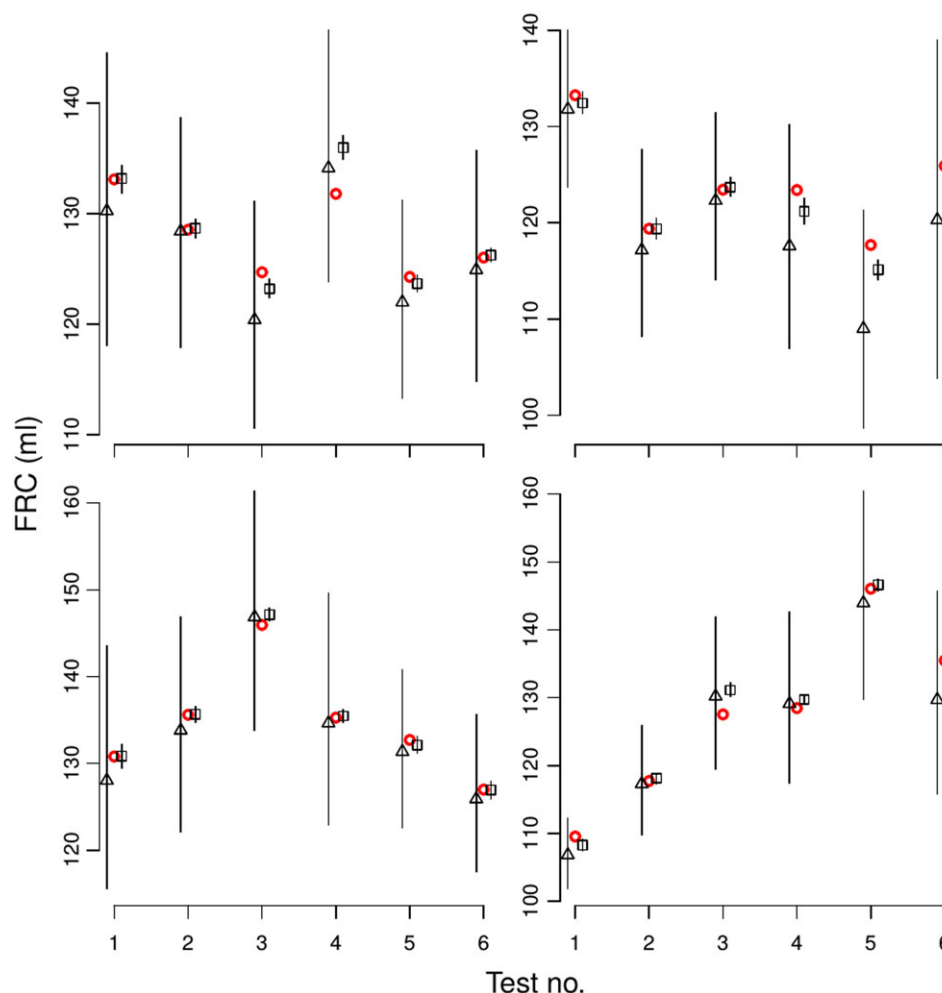


FIGURE 8 Comparison of functional residual capacity (FRC) estimates across replicates. Estimates of FRC using 3 different methods: standard method ($FRC^{(s)}$, red circles), model-based method using discrete derivatives ($FRC^{(m)}$, black triangles), and a similar model but formulated on the cumulative scale (black squares). The model-based estimates were calculated using the diffuse prior and the resulting 95% credible intervals are shown. Notice that intervals from the model we use (discrete derivatives) more accurately portray the observed variation in FRC estimates across replicates [Colour figure can be viewed at wileyonlinelibrary.com]

CEVTG curves are heavily influenced by any noise in the early breaths of the washout. Therefore, the observed asymptote may actually be inaccurate and models should not necessarily try to fit it too closely. Indeed, we found that the posterior variation given by our discrete derivative model matched quite well with the actual variation we observed between replicate tests (see Figure 8).

5.2 | Hyperparameter estimation to obtain informative priors

Our analysis of shortened tests showed the utility of using a parametric model along with genuine prior information with this type of extrapolation problem. Additionally, the complete data analysis of CEVTG curves suggests that using informative prior distributions also shrinks the parameter estimates towards more realistic values. More generally, there is no doubt opportunity to routinely use prior information to aid the analysis of individual “growth/decay-like” data and to estimate unseen quantities of these models where data are limited. This approach should be particularly useful when it is costly or difficult to collect additional data for an individual and one has a solid model of the underlying “growth” or “decay” pattern, as in the case of the MBW test.

We believe that our approach to obtaining informative priors for the MBW model—using the hyperparameters of a hierarchical model—is appropriate and effective for this problem. It builds on the structure of the simpler individual-level model in the right way to allow us to pool information across a dataset. We were able to validate its effectiveness

through a (holdout) validation step to ensure we were not “using the data twice.” Other researchers are now able to use our informative prior directly.

5.3 | Variance components model

We took the approach of examining the properties of our estimation procedures by comparing our model-based results with the traditional empirical approach using a dataset of several hundred healthy children and a variance components model. This was in an applied Bayesian spirit in contrast to a frequentist evaluation of repeated-sampling properties via simulation. This approach also proved useful as it allowed us to better understand the source of the relatively high observed variance in MBW outcomes (which is informative for future analyses with a more epidemiological focus).

When analysing the residual variances of each method, we made the assumption of an exchangeable replicate structure. While it would have been useful and convenient to parse out the variation into individual components to the fullest extent possible, by including an additional interaction term to account for the nonexchangeability of our replicates, we found that it was not possible to do so while achieving sufficient mixing of the posterior chains even after applying common reparameterisations and modifying the operating parameters of the sampler.

Faced with this type of situation, the modeller has 1 of 3 practical choices. They could drop an interaction term and assume exchangeability, allowing estimation of separate residual variance components, with the interpretation that those components contain some variation due to replicates within each participant. They could retain the interaction term for nonexchangeable replicates but assume a common residual variance, or, as we have done, do both and compare the results.

6 | CONCLUSIONS

As far as we know, we are the first to use statistical models to define and then estimate the key parameters of the MBW, in particular LCI and FRC. By modelling the MBW data rather than using simple summaries, we developed a principled way of assessing MBW outcomes. These models are strongly tied to the underlying physiology, and therefore may reveal patient characteristics that are unobservable using simpler methods. Additionally, we applied these models to assess how much test data are required to obtain reasonable estimates of MBW outcomes and how the uncertainty around these estimates changes given different degrees of test completion.

Applying our model to complete MBW tests, we observed posterior predictive distributions that reflected the underlying structure and variation in the observed data. As part of this modelling process, we also examined the variation structure of the MBW parameters in depth using a variance components model. Our complete-data model-based estimates of LCI have good agreement and similar residual variance compared with the same quantities estimated using the standard empirical method. This suggests that the standard method is actually a good estimator of LCI using complete data, something that would have been difficult to establish without implementing a model-based approach. Using informative priors seems to constrain the estimation procedure to produce more realistic values of FRC when faced with noisy data. Our exposition of a model-based LCI predicted from shortened tests provides statistical support to recent studies that have found that an LCI defined using a higher tracer gas quantity threshold can still be clinically meaningful, eg, $k^{(20)}$ and in some cases $k^{(10)}$, instead of $k^{(40)}$.⁹⁻¹¹

Our model can produce reliable estimates of LCI from MBW tests that are, on average, 43% shorter than a complete test (measured by number of breaths). This potentially allows data that would otherwise be unused, due to artefacts or test disruption, to be shortened and subsequently analysed to estimate key quantities of lung function with good accuracy. Given the logistical challenges of conducting MBW testing in unsedated infants during natural sleep, this is an important finding as it may allow the various cohort studies that have undertaken MBW testing to utilise a larger proportion of their MBW data. At a clinical level, this new approach may save both patient and clinician time and effort as potentially fewer tests will be required to achieve the desired number of acceptable LCI measurements. The analyses reported here were limited to a population-derived infant sample. It is not clear how our model would perform with MBW data from participants with lung disease, but it is likely that for these participants there will be an even more pronounced change in the slope in the later part of the tracer gas quantity curve, which could potentially make extrapolation into this area of the test less accurate using only shortened test data with diffuse prior information. A clear direction for future research is therefore to assess these models using MBW tests from both healthy and disease-affected infants.

We note a number of more general aspects of our analyses, in particular, the need to model underlying increments of cumulative data, particularly where there are monotonicity constraints; the utility of informative priors in “growth” or “decay” type problems where a realistic data model exists; and difficulties fitting a Bayesian variance components model with nonexchangeable replicates.

ACKNOWLEDGEMENTS

The authors are grateful to the BIS study participants and their families for their contribution and to Louise King and Hilianty Tardjono for their assistance in obtaining and processing the MBW data. We also thank the BIS investigator team for their collaboration in making this work possible.

This work was supported by the National Health and Medical Research Council: Centre of Research Excellence grant ID 1035261 (Victorian Centre for Biostatistics; ViCBiostat), Senior Research Fellowship ID 1110200 (Anne-Louise Ponsonby), and Project Grant ID 1009044. Robert Mahar was supported by the Australian Government Research Training Program Scholarship. The MBW testing equipment was purchased by the Shane O'Brien Memorial Asthma Foundation. Research at the Murdoch Children's Research Institute is supported by the Victorian Government's Operational Infrastructure Support Program.

ORCID

Robert K. Mahar  <http://orcid.org/0000-0002-3643-923X>

John B. Carlin  <http://orcid.org/0000-0002-2694-9463>

Sarath Ranganathan  <http://orcid.org/0000-0002-6975-0307>

Anne-Louise Ponsonby  <http://orcid.org/0000-0002-6581-3657>

Peter Vuillermin  <http://orcid.org/0000-0002-6580-0346>

Damjan Vukcevic  <http://orcid.org/0000-0001-7780-9586>

REFERENCES

1. Morgan WJ, Stem DA, Sherrill DL, et al. Outcome of asthma and wheezing in the first 6 years of life: follow-up through adolescence. *Am J Respir Crit Care Med*. 2005;172(10):1253-1258.
2. Hülskamp G, Pillow JJ, Stocks J. Lung function testing in acute neonatal respiratory disorders and chronic lung disease of infancy: a review series. *Pediatr Pulmonol*. 2005;40(6):467-470. <https://doi.org/10.1002/ppul.20317>
3. Beydon N, Davis SD, Lombardi E, et al. Pulmonary function testing in preschool children. *Am J Respir Crit Care Med*. 2007;175(12):1304-1345. <https://doi.org/10.1164/rccm.200605-642ST>
4. Saline Hypertonic in Preschoolers with Cystic Fibrosis and lung structure as measured by computed tomography (CT). SHIP-CT study.
5. Robinson PD, Latzin P, Verbanck S, et al. Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *Eur Respir J*. 2013;41(3):507-522. <https://doi.org/10.1183/09031936.00069712>
6. Rosenfeld M, Allen J, Arets BHGM, et al. An official American Thoracic Society workshop report: optimal lung function tests for monitoring cystic fibrosis, bronchopulmonary dysplasia, and recurrent wheezing in children less than 6 years of age. *Ann Am Thorac Soc*. 2013;10(2):S1-S11. <https://doi.org/10.1513/AnnalsATS.201301-017ST>
7. Horsley A. Lung clearance index in the assessment of airways disease. *Respir Med*. 2009;103(6):793-799. <https://doi.org/10.1016/j.rmed.2009.01.025>
8. Vukcevic D, Carlin JB, King L, et al. The influence of sighing respirations on infant lung function measured using multiple breath washout gas mixing techniques. *Physiol Rep*. 2015;3(4):e12347-e12347. <https://doi.org/10.14814/phy2.12347>
9. Singer F, Yamine S, Abbas C, et al. Ways to shorten the lung clearance index measurement II—How long to wash? *Eur Respir J*. 2012;40:3298.
10. Yamine S, Singer F, Abbas C, Roos M, Latzin P. Multiple-breath washout measurements can be significantly shortened in children. *Thorax*. 2013;68(6):586-587. <https://doi.org/10.1136/thoraxjnl-2012-202345>
11. Hannon D, Bradley JM, Bradbury I, Bell N, Elborn JS, O'Neill K. Shortened lung clearance index is a repeatable and sensitive test in children and adults with cystic fibrosis. *BMJ Open Respir Res*. 2014;1(1):e000031.
12. Vuillermin P, Saffery R, Allen KJ, et al. Cohort profile: the Barwon Infant Study. *Int J Epidemiol*. 2015;44(4):1148-1160. <https://doi.org/10.1093/ije/dyv026>

13. Mahar RK, Vukcevic D, King L, et al. Lack of transparency in software used to analyze multiple breath washout data. *Pediatr Pulmonol*. 2016;51(11):1108-1110. <https://doi.org/10.1002/ppul.23420>
14. Bennett J. (2015) AutoIt. AutoIt Consulting
15. Schibler A, Henning R. Measurement of functional residual capacity in rabbits and children using an ultrasonic flow meter. *Pediatr Res*. 2001;49(4):581-588.
16. Bates JHT. *Lung mechanics: an inverse modeling approach*. Leiden: Cambridge University Press; 2009.
17. Gonem S, Scadding A, Soares M, et al. Lung clearance index in adults with non-cystic fibrosis bronchiectasis. *Respir Res*. 2014;15(1):59.
18. Stan Development Team (2016) The Stan C++ Library.
19. Stan Development Team (2016) Stan modeling language: User's guide and reference manual. Version 2.12.0.
20. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal*. 2006;1(3):515-534.
21. Sorensen T, Vasisht S. (2015) Bayesian linear mixed models using Stan: a tutorial for psychologists, linguists, and cognitive scientists. ArXiv Prepr. ArXiv150606201
22. Carstensen B, Simpson J, Gurrin LC. Statistical models for assessing agreement in method comparison studies with replicate measurements. *Int J Biostat*. 2008;4(1).
23. Stan Development Team (2016) RStan: the R interface to Stan.
24. Gelman A, Carlin JB, Stern HS, et al. *Bayesian Data Analysis*. 3rd ed. New York: Chapman and Hall/CRC; 2013.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Mahar RK, Carlin JB, Ranganathan S, Ponsonby A-L, Vuillermin P, Vukcevic D. Bayesian modelling of lung function data from multiple-breath washout tests. *Statistics in Medicine*. 2018;37: 2016–2033. <https://doi.org/10.1002/sim.7650>

APPENDIX A

Model estimation

We took a Bayesian approach to estimating our model parameters for a number of reasons. First, the approach makes it relatively straightforward to define and estimate derived parameters such as the model-based LCI by sampling from the posterior distribution using Stan, a probabilistic programming language.¹⁸ This allows us to quantify the uncertainty around our model-based LCI by using MCMC methods. Second, the Bayesian approach allows us to naturally incorporate prior information in our model parameters. By assigning prior distributions to our model parameters, we are able to constrain them to realistic values, which can be effectively updated with the available data given different levels of test completion. The Stan model code and a description of the data used are provided in the Supporting Information.

We sampled from the posterior distribution of each parameter estimated in this paper by using the Hamiltonian Monte Carlo algorithm implemented within Stan, called from the R software environment.²³ In each case, 4 MCMC chains were run in parallel comprising a discarded “warm-up” phase and a “sampling” phase of 1000 iterations each, with the latter comprising 4000 samples from the posterior distribution being used for parameter estimation.

In line with Gelman et al.²⁴ we judged a sufficient effective sample size to be $5m$, where $m = 2 \times$ number of chains. Thus, with 4 chains, we aimed to have all parameters exhibit an effective sample size of at least 40 to be sufficient for accurate parameter inference. Therefore, where effective sample sizes appeared suboptimal, we state this in the main text.

Convergence of the Markov chains was assessed both graphically, via visual inspection of the MCMC chains, and statistically, by ensuring that the scale reduction factor $\hat{R}$ was near 1 for each parameter.²⁴

Chapter 5

Modelling the spectral densities of tidal flow waveforms

Contents

5.1	Background	64
5.1.1	Aims	68
5.2	Models	68
5.2.1	Time and frequency domains	69
5.2.2	Power-law spectra and scale-invariance	69
5.3	Estimation	71
5.3.1	Time and frequency domains	72
5.3.2	Spectral density estimation	77
5.3.3	Power-law exponent estimation	82
5.4	Simulation	85
5.4.1	Data generation	85
5.4.2	Performance evaluation	88
5.4.3	Results	88
5.5	Discussion	95

With this chapter we change our focus, from combining various forms of the end-tidal MBW test data to estimate commonly accepted mechanistic indices of lung function, such as LCI and FRC, to using the raw, high-frequency tidal flow waveform data to estimate

novel indices of lung function with interpretations closely tied to the complexity of the underlying respiratory system.

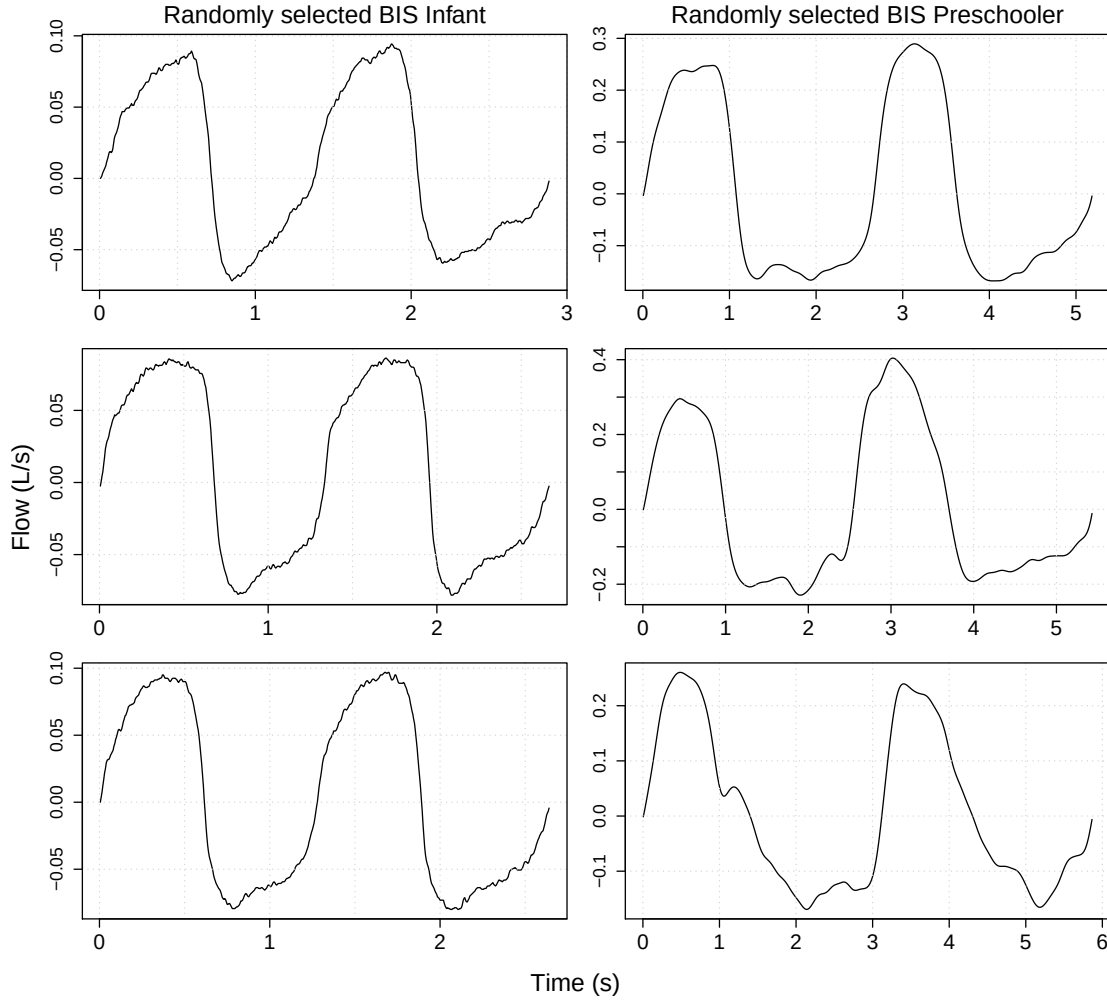
We focus on the estimation of a power-law exponent from time series data with a power-law spectrum. The power-law exponent is often proposed as a measure of ‘organisation’ in complex systems and is encountered throughout the physical sciences. This chapter leads naturally into chapter 6, which explores the power-law exponent estimated from tidal flow waveforms in an epidemiological setting. We describe estimators for the power-law exponent including power spectral density (PSD) estimation methods and time domain methods, propose a minor extension to a commonly used PSD estimator, and examine the statistical properties of the extensions using simulation.

Section 5.1 begins with a summary of the key concepts behind tidal flow waveforms, power-law spectra and power-law exponents, and scale invariance, along with examples from the respiratory medicine literature. An overview of how time series can be represented in the frequency domain, the power spectrum, and scale invariance is provided in Section 5.2. This is followed by a review of common estimators of the power-law exponent, the statistical properties of estimates of spectral density, and a description of our extension to the PSD estimator in Section 5.3. The estimators are then assessed via simulation in Section 5.4, with discussion following in Section 5.5.

5.1 Background

Tidal flow waveforms are an example of easily measured data that are rich with physiological information related to respiratory control and pulmonary mechanical function and can be understood as integrated outputs of highly complex feedback systems of neurorespiratory control [33–35, 39, 57, 79–81] (examples are given in Figure 5.1). This output is determined by a regulated network of physiological components including neural respiratory control centres, central chemoreceptors, higher cortical inputs, and thoraco-pulmonary mechanics [33, 57]. These components may be affected by environmental factors, disease, pharmacological interventions, and physiological development. In a state of health, this system is suggested to operate in a complex dynamic equilibrium [33, 34, 39, 79, 81]. Altered behaviour of any system components due to disease or other factors may disturb this equilibrium and thus modify the tidal flow waveform [33, 34, 39, 79–81].

FIGURE 5.1: Tidal flow waveforms for two breaths of a randomly selected BIS infant (left) and a BIS preschooler (right). Each row corresponds to a different test occasion.



It is common practice to examine the shape of tidal flow waveforms using simple summaries such as tidal volume, respiratory rate, or the ratio of time taken to reach peak expiratory flow to total time of expiration ($t_{\text{TPEF}}/t_{\text{E}}$) [57]. Yet these types of summaries tend to reduce components of the breathing pattern to a single dimension, such as time. Therefore these methods may not sufficiently capture the complexity of the underlying respiratory system. This is because biological systems, such as the neurorespiratory control system, are often highly non-linear, involving multiple inputs and components interacting in complex ways [82–84]. These complex systems cannot be explained satisfactorily by linear models which are typically relatively simple and can be understood by their components alone with no uncharacteristic behaviours [46, 83–85]. The non-linear system cannot be fully understood by modelling its components in isolation as they influence and regulate one another through complex feedback mechanisms. More sophisticated methods

may therefore be required to analyse the complex output of the neurorespiratory control system.

One such approach is to examine the system’s scaling properties [33–35, 39, 46, 83, 86, 87]. This is typically done by estimating the *spectral density* of the system output—a frequency domain representation of a temporal process—and using this new ‘data’ to then estimate a *power-law exponent*. A number of methods exist to estimate a power-law exponent from time series data, which we consider in technical detail in Sections 5.2 and 5.3. We consider estimators that operate in both the frequency and the time domains and show how they are fundamentally related, however we focus primarily PSD estimation.

We know of only two studies that have examined power-law exponents of tidal flow waveforms [33, 35]. In 2001, Frey et al. [33] used the PSD estimation method to examine whether spectral densities of tidal flow waveforms in healthy infants, infants with a history of wheezing disorders, and infants with chronic obstructive lung disorder (CLD) were consistent with a power-law spectrum and whether the spectral properties differed between groups. This work was followed by Habib et al. [35] in 2003, who performed a similar analysis although with very low birthweight (VLBW) preterm babies. Both studies also used a model-free description of the deviation of the tidal flow waveform from a pure sine wave, called the harmonic distortion index [88, 89], but we restrict our focus to the PSD approach.

Frey et al.’s 2001 study [33] focused on assessing the shape of the tidal flow waveform using both the power-law exponent and its association with age, wheezing, chronic lung disease (CLD) and altered acute airway mechanical function. Our focus is on the power-law exponent rather than the harmonic distortion index since we are interested in understanding the shape of the waveform explicitly in terms of its spectral properties. Frey et al. [33] estimated power-law exponents from individual breath cycles from 8–10 second recordings of tidal flow obtained with face-mask and pneumatograph from 10 healthy infants at approximately 1, 6 and 12 months of age, as well as 10 infants with a history of wheezing disorder (mean age: 13.9 ± 1.6 (SD) months) and 10 infants with CLD (mean age: 8.1 ± 7.2 (SD) months). All infants were sedated and in a supine position at the time of the experiment. To test the effect of altered acute airway function, histamine challenges were administered to the 10 healthy infants which reduced airway function—measured via maximal expiratory flow at FRC—by an average of 30%. The power-law exponents were

also compared with $T_{\text{PTEF}}/T_{\text{E}}$ for individual breaths. It was observed that infant tidal flow spectra follow a power-law between 0.13 and 10 Hz, and that the mean values of the power-law exponents were significantly higher in healthy infants than those with CLD or wheezing disorders, but exhibited no difference with respect to age. Further, their estimates of power-law exponents exhibited less breath-to-breath variability than $T_{\text{PTEF}}/T_{\text{E}}$ in general. Importantly, they found negligible difference between power-law exponents before and after histamine challenges, providing valuable evidence that the power-law exponent behaviour is not driven by lung mechanics alone.

Habib et al. [35] followed Frey et al.'s work by comparing power-law exponents derived from 30–60 second tidal flow traces of 16 very low birthweight (VLBW) infants (mean gestation age: 28 ± 2 (SD) weeks) measured simultaneously with a face-mask pneumatograph and respiratory inductance plethysmography (RIP). They observed that, in contrast to the results of Frey et al. [33], values of the power-law exponent from VLBW babies were comparable to older and CLD infants, but significantly lower compared to 1-month-old infants of similar age and size. Importantly, they observed negligible difference between power-law exponent estimates obtained by pneumatograph or RIP.

These results support the notion that health and development may be related to respiratory complexity, as measured by the power-law exponent from tidal flow waveforms, particularly in neonates [90]. Although this work was cited at the time as an important future area of research, particularly as measurements can be taken quickly and non-invasively [48], to our knowledge, similar analyses of tidal flow waveforms using a power-law exponent have not been done outside of the two aforementioned studies other than a handful of historical examples of limited relevance [91].

The research thus far on power-law scaling behaviour of the respiratory system is certainly promising, and much could be done to further understand lung function complexity from routinely collected tidal-breathing data. Of particular interest would be, using the large amount of BIS and AREST CF data, to build upon previous studies that use the PSD method to investigate whether tidal breathing in early childhood exhibits power-law scaling behaviour, and to what extent these change with maturation, disease, or other epidemiological covariates of interest [33, 35]. This would represent the largest applied study of the spectral densities of tidal flow waveforms, that we know of, to date.

However, typically the estimators of power-law exponents are used without due consideration to their statistical properties when applied to data that represent what is actually clinically observed, for example waveforms that are relatively short in length and observed with error. Although a number of studies have examined the statistical properties of the PSD method in detail, they are either for very long series or neglect to take into account the effect of observational error, and only assess power-law exponents with specifically fractal interpretations [46, 47, 92]. Therefore, before using these methods on real-world data, it was necessary to have a detailed understanding of how different power-law estimators perform empirically on tidal-waveform data. We therefore provide a detailed background to the PSD method, before assessing the statistical properties of a number of power-law exponent estimators using simulated data.

5.1.1 Aims

The overarching aim of the research presented in this chapter was to gain a detailed understanding of procedures for estimating a power-law exponent from tidal-breathing data. This aim was broken into the following specific objectives:

- To describe, in detail, the PSD method of estimating a power-law exponent—as used by Eke [46], Frey et al. [33], and Habib et al. [35]—from tidal flow waveforms.
- To specify a number of extensions to the PSD method described by Eke [46].
- To investigate the sampling properties, in particular bias and variation, of the extensions to the PSD method using simulated data.

5.2 Models

This section provides an overview of models relevant to estimating a power-law exponent of a time series in the frequency domain. Section 5.2.1 first describes how a time-series can be represented in the frequency domain and Section 5.2.2 describes the power-law spectrum and its relationship to the concept of scale-invariance.

5.2.1 Time and frequency domains

Fundamental to the theory of spectral analysis is that any signal in the time domain, no matter how complex, can be represented by the sum of cosine functions of varying frequency, amplitude, and phase [93]. The frequencies and amplitudes together define an object called a *spectral density*, which provides the link between the time domain and the frequency domain; we explain this in more detail below. We limit the review to discrete spectra for both brevity and because our focus is on digitally sampled time series data¹.

Let X_t ($t = 1, \dots, N$) be a time series that represents a noiseless signal of interest (in the time domain). Let the function $S(\omega_i)$ be its (discrete) spectral density. The two are mathematically related as follows [93, 95]:

$$X_t = \sum_{i=1}^K S(\omega_i) \cos(\omega_i t + \phi_i). \quad (5.1)$$

Thus our time series is simply the sum of K cosine functions with amplitude $S(\omega_i)$, at frequencies ω_i and with phase ϕ_i ($i = 1, \dots, K$). The phases are distributed randomly in $[0, 2\pi)$, yet are effectively constant for any single realisation of X_t [93]. A process such as this can be observed with perfect accuracy, requiring no estimation procedure, and is therefore not of much relevance in an applied context. Estimating the spectral density of a temporal process from real-world data is considered in Section 5.3.2.

5.2.2 Power-law spectra and scale-invariance

For our purposes, we are primarily interested in a particular functional form of the spectral density known as a *power-law spectrum*. A power-law spectrum is defined as a spectral density, $S(\omega)$, that is an inverse function of the power of a frequency, ω :

$$S(\omega_i) = \alpha / \omega_i^\beta, \quad (5.2)$$

where α is a proportionality factor and β denotes the *power-law exponent* [46, 47]. Note that 5.2 in logarithmic form is expressed as a linear relationship:

$$\log(S(\omega_i)) = \log(\alpha) - \beta \log(\omega_i), \quad (5.3)$$

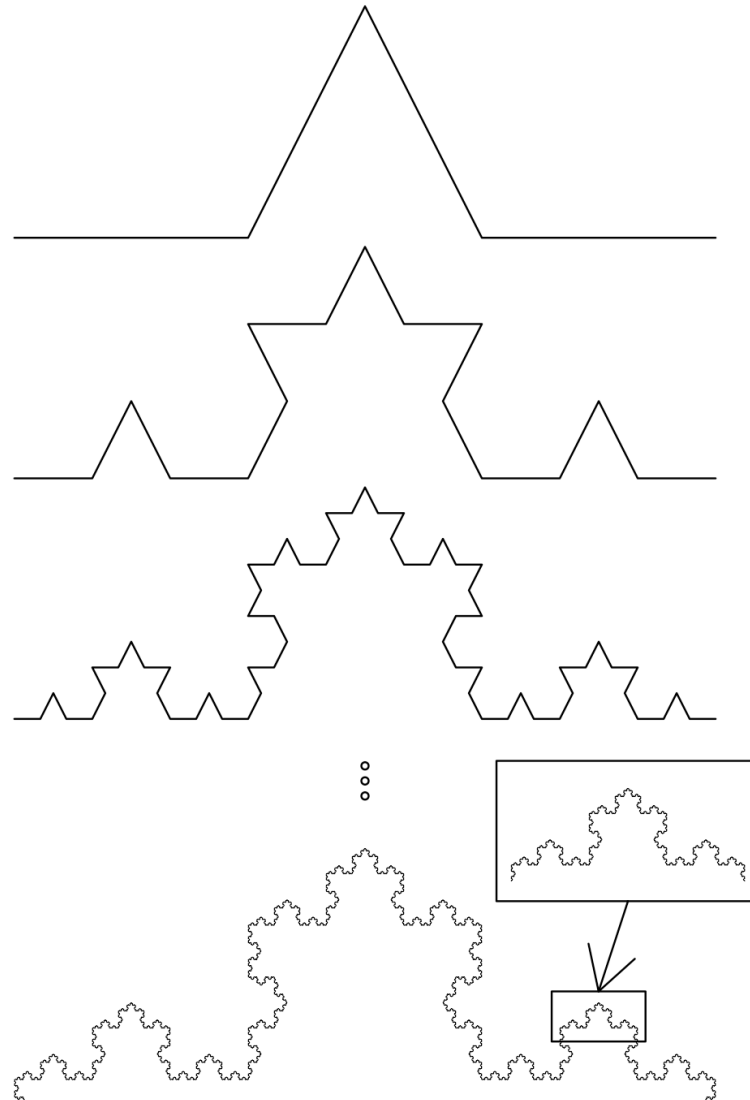
¹More detailed theoretical treatments can be found elsewhere, for example [93, 94], on which this material is based.

Processes that are known to have a power-law spectrum can be defined as *scale invariant*. That is, they lack a characteristic time scale or frequency [39, 84]. The concept of scale invariance is perhaps best demonstrated through geometric objects, many of which are widespread throughout nature [96]. A commonly cited and intriguing example is that provided by Mandelbrot in his seminal paper on fractals in 1967: the length of Britain's coastline [97]. Many of the methods that analyse the scaling behaviour of systems are known collectively as *fractal* methods. A fractal is an object composed of a hierarchy of sub-units that resemble some larger scale structure, such that it is said to be *self-similar* [84, 97], or similarly put, a fractal system can be said to exhibit *scale invariance*. No characteristic scale exists that adequately describes the structure of the system because, at different scale magnitudes, the system's structure appears replicated, with similar detail being observed at every level of magnification.

One well-known artificial geometric example is that of a Koch curve (shown in Figure 5.2), which uses a simple iterative algorithm to produce a complex curve with unusual properties [96]. For example, in contrast to a Euclidean curve, the Koch curve shows similar amount of detail on all length scales, and at different scales similar structure is observed, hence *scale invariance*. An alternative and perhaps more realistic example is the branching structure of an idealised tree. Whether we view the tree as a whole or we view only one of the tree's terminal branches, the branch angles and proportionate diameters remain approximately constant. We can consider this tree example in a more physiological context, for example the pulmonary artery system, which has been demonstrated to branch in a way consistent with a fractal structure for both infants and adults [36–38]. The Koch curve and the pulmonary artery system can be described as, respectively, exact and non-exact fractals [46].

Various methods can be applied to fractal objects, both exact and non-exact, to either directly measure, in the case of the former, or estimate, in the case of the latter, their *fractal dimension* [46]. This is analogous to a power-law exponent of a temporal process [39, 46]. In the following section we consider a number of methods for estimating a power-law exponent from temporal processes.

FIGURE 5.2: The Koch curve: an example of scale invariance. The first three curves show, in order, a graphical depiction of the first three steps of the algorithm used to produce the curve. The final panel shows an example of the curve after many iterations. Note that the curve shows similar detail at different scales, as indicated by the ‘zoomed’ window.



5.3 Estimation

A power-law exponent can be estimated from a sampled temporal process, with a power-law spectrum, in both the time or frequency domain. Methods also exist that operate in the time-frequency domain using, for example, wavelet analysis, however we do not consider

them here [46, 47]. In section 5.3.1 we describe a common approach for estimating the power-law exponent in the frequency domain, using the PSD estimation method, before describing several common time domain approaches and why we consider the PSD method to be a better approach in the context of the analysis of tidal flow waveforms. We describe the statistical properties of the estimator of spectral density used in the PSD method, the *periodogram*, in Section 5.3.2. Section 5.3.3 describes the PSD estimator in more detail, and proposes some minor extensions to the common approach.

5.3.1 Time and frequency domains

The method we focus on to estimate the power-law exponent in the frequency domain is the PSD estimation as described by Eke [46] and used to estimate power-law exponents from tidal flow waveforms by Frey et al. [33] and Habib et al. [35].

PSD uses the discrete Fourier transform of the data to compute a set of periodogram ‘ordinates’, described in detail in the next section but defined here as $I(\omega)$, which estimates the spectral density of a signal at the harmonics of its lowest observable frequency.

A signal with a power-law spectrum can then be modelled as:

$$\log(I(\omega_i)) = \log(\alpha) - \beta \log(\omega_i) + r, \quad (5.4)$$

where r represents some random error (the distribution of which is defined in the next section). An example of how tidal flow waveforms are represented in the power spectrum as periodogram ordinates is given in Figure 5.3.

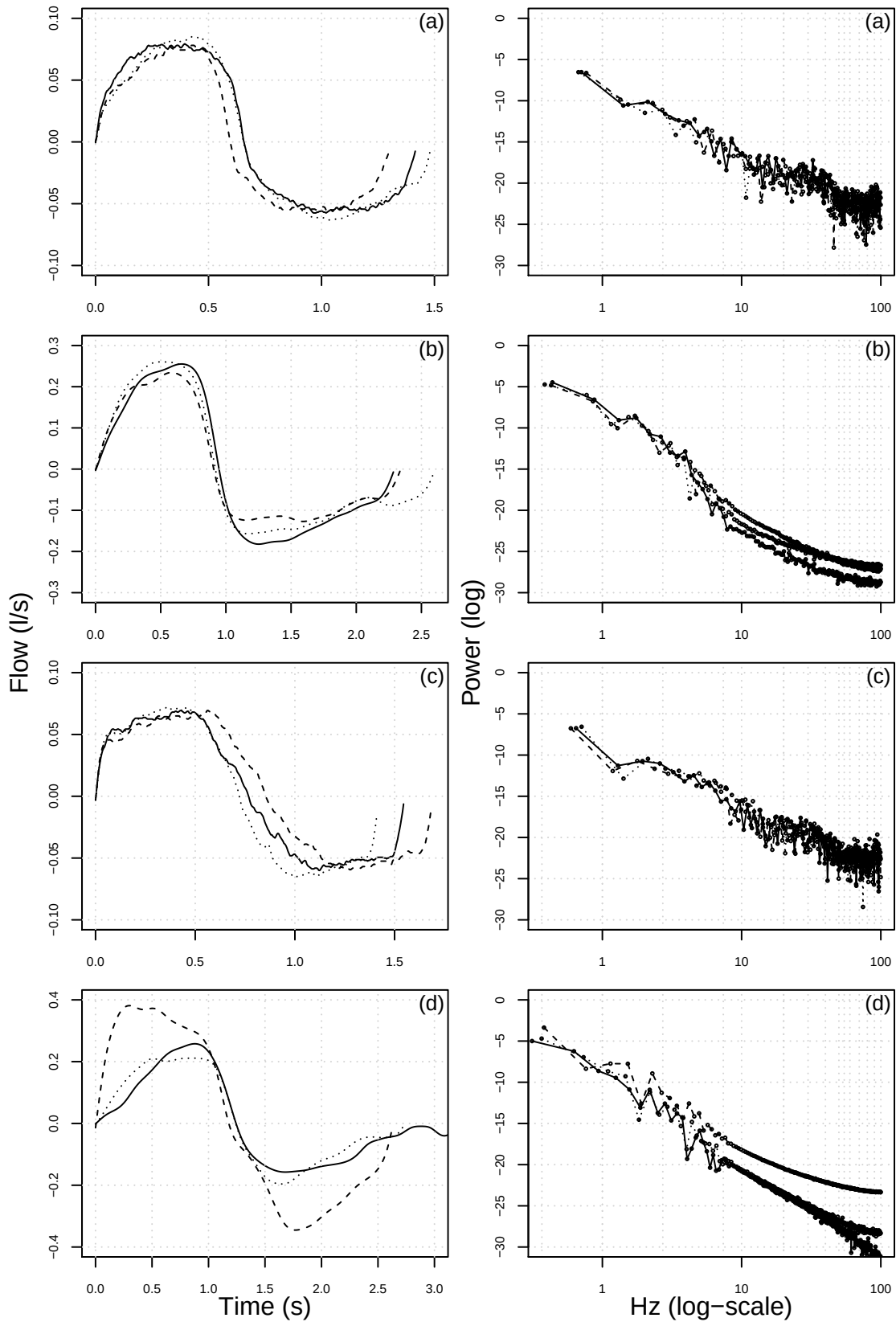
Ordinary least squares is then used to estimate the power-law exponent β by minimising the sum of the squared residuals.

$$\sum_{i=1}^N r_i^2 \quad (5.5)$$

where $r_i = \log(I(\omega_i)) - (\log(\alpha) + \beta \log(\omega_i))$.

Fitting only to periodogram ordinates from lower frequencies appears to be a common approach in the literature [46, 98], where it is suggested that only frequencies up to 1/8 of the maximum frequency (in this case the Nyquist frequency, or one-half of the sampling rate) be used for analysis.

FIGURE 5.3: tidal flow waveforms in the time domain (left-column) and their associated log-log periodograms in the frequency domain (right-column). Shown are three randomly chosen waveforms for a: BIS participant approximately 3 months (a) and 4 years (b); AREST CF participant aged approximately 1 year (c) and 4 years (d).



A number of methods are available to estimate a power-law exponent in the time domain, for example detrended fluctuation analysis (DFA), dispersional analysis, Hurst's rescaled range analysis, and adaptive fractal analysis [46, 99].

These methods all take a similar approach in that they all focus on estimating what is known as a Hurst exponent, $0 < H < 1$, by splitting the data into windows of length n and taking some locally descriptive measure of the data (for example variance) within each window, which we will denote simply as m_n . For a process with a power-law spectrum there will exist a power-law scaling relationship such that:

$$\log(m_n) = \log(p) + H \log(n) + e_n, \quad (5.6)$$

where p is some proportionality factor and $0 < H < 1$ is the Hurst exponent and e represents some random error. Thus the statistical measures of the signal over a large observation window are approximately proportional to the statistical measures of the series over a smaller observation window when the series is rescaled to the smaller window [46, 47, 98, 100]. Like PSD, the Hurst exponent is estimated from 5.6 using OLS, although unlike PSD there is no need to discard data from the smallest window sizes (analogous to higher frequency estimates of spectral densities).

Importantly, as we describe in some more detail below, the Hurst coefficient has been shown to map to the power law exponent β if the process belongs to one of the following distinct classes: fractional Gaussian noise (fGn) or fractional Brownian motion (fBm) [46, 47, 98]. Thus, each method can be understood in context of the system's spectral density. Both fGn and fBm are classes of time series that have a power spectrum. The distinction between them is that fGn are stationary (the mean and variance remain the same over time) while fBm are non-stationary (the mean and variance can change over the length of the observation period). Importantly, an fBm signal can be converted to fGn by first-order differencing and, vice-versa, an fGn signal can be converted to an fBm via cumulative summation.

Generally, which time domain method to use to estimate the Hurst exponent depends on the class of the process that is being analysed, resulting in two class-specific Hurst exponents denoted H_{fGn} and H_{fBm} [46, 98]. For both fGn and fBm, $H = 0.5$ is a special case, corresponding to uncorrelated white noise for the former and a random walk/Brownian motion for the latter. Importantly, the Hurst exponent from either class and the

power spectrum $1/\omega^\beta$ are related, with $H_{fGn} = (\beta + 1)/2$ and $H_{fBm} = (\beta - 1)/2$. Thus $0 < H_{fGn} < 1$ maps to $-1 < \beta < 1$ and $0 < H_{fBm} < 1$ maps to $1 < \beta < 3$ [46, 47]. Special cases exist for β corresponding to Gaussian white noise ($\beta = 0$), and Brownian motion ($\beta = 2$) [46, 47]. The boundary case, where $\beta = 1$, represents the very particular kind of process, whereby the strength of each frequency is proportional to its oscillation and thus the power is distributed evenly and independently across the spectrum [92]. This ‘spectral independence’ suggests under this limiting assumption local perturbations to the underlying system are likely to be adapted to without affecting the global stability of the system [92].

DFA is a popular and relatively simple method for estimating power-law exponents from signals in the time domain [87, 101, 102]. Since its development some 20 years ago, DFA has been adopted by researchers as a tool for analysing all manner of time series; particularly so in respiratory research [39].

As summarised by Thamrin et al. [39], several studies have used DFA to estimate power-law exponents in tidal-breathing measurements or inter-breath interval signals, from pre-natal to elderly stages of life. Govindan et al. [44] used ultrasound to measure breathing, gross-body and extremity movements of 49 fetuses from healthy singleton pregnancies with a gestational age range between 28 to 39 weeks. They found fetal breathing movements were consistent with a power-law process and increased mildly from a white noise process toward the boundary condition $\beta = 1$ with gestational age, which they argue is an indication of the maturation of the fetal lung and the development respiratory aspect of the fetal central nervous system. Cernelc et al. [40] demonstrated that the breath-to-breath time series of tidal volume, oxygen and carbon-dioxide concentrations measured from healthy infants in quiet sleep exhibit power-law behaviour. Importantly, they found that this behaviour exhibited what is known as a crossover, or two different power-law exponents for different scales, in approximately half of their patients, indicating that scale invariance may change over longer time scales. They suggest that as this crossover behaviour was sensitive to gestational age, DFA could be used as a marker of changes in the control of breathing in immature infants. Rostig et al. [41] use DFA to study long-range correlation across a number of respiratory time signals in 29 healthy adult subjects (mean age 25.8) and found that the power spectrum of tidal-volumes went from something similar to white noise when in non-REM sleep to a power-law spectrum in wakefulness and REM sleep. Peng et al. [103] assessed the long-range correlation of inter-breath intervals in two groups

of healthy adult subjects, one group young (mean age 27, range 21–34 years) and one group elderly (mean age 74, range 68–81 years). They found that for measurements from both young and old subjects the power-law exponent decreased from $\beta = 1$ with age, indicating that respiratory system dynamics may become less robust with age. However, the focus on temporal fluctuations of per-breath quantities such as inter-breath intervals or tidal volume is somewhat problematic in early childhood. Time domain methods for estimating a power-law exponent typically require not only many data points—requirements are highly dependent on the method used and strength of the power-law relationship, although general estimates range from needing 256 (although 1,024 is cited as a vast improvement) [92], to perhaps 2,048 [46], and somewhere in-between [47]—but also uninterrupted recordings of the tidal flow waveform to provide reasonable results. Given that it is remarkably difficult to obtain uninterrupted tidal flow recordings of this length in young children, examination of the ‘long-range’ power-law scaling behaviour in this age group has therefore very limited clinical feasibility.

Several studies exist with variable success in ‘stitching’ together non-contiguous recordings of stride length in gait studies, but these methods focus more on recordings that are non-contiguous due to removal of errant measurements of stride, rather than on joining recordings separated by intervals of an arbitrary length, and are therefore not relevant in our case [104–106].

An alternative approach to studying the scale invariant behaviour of the respiratory system is to shift the focus away from per-breath quantities such as inter-breath intervals over long timeframes, and instead focus on very high-frequency measurements over much shorter time scales, for example the tidal flow recording over an individual breath cycle. Although these series themselves are limited in length by the physiology of the individual and the measurement apparatus², the signal is generally uninterrupted and measurement is clinically feasible. The power-law exponent for a large number of individual breaths can be estimated and simply averaged, an approach recommended by Delignieres et al. [92] where series length is an issue. Rather than use DFA or a similar time-based method, a more natural approach to estimating the power-law exponent in the context of short tidal flow waveforms is to use PSD.

²Via the respiratory and sampling rates. For example, a short breath of 1 second in length sampled at 200 Hz provides a signal of length 200. Either an increase in the sampling rate to 400 Hz or the breath length to 2 seconds breath would provide a signal of length 400.

The PSD approach has the benefit that results can be more directly interpreted in terms of the relationship of the breathing frequency and its upper harmonics, and also the sinusoidal shape of the tidal flow waveform (or lack thereof). Thus, it has a direct and intuitive physiological interpretation. Furthermore, a major benefit of using the PSD method is that β can be estimated directly; it is not necessary for the signal to be first categorised into an fGn or fBn class prior to analysis, and in fact it does not even need to belong to either class in order for a power-law to be estimated. This is particularly relevant in the context of the findings by Frey et al. [33] and Habib et al. [35], who find values of the power-law exponent that in general fall outside of the range that would class tidal flow waveforms into either fGn or fBm.

5.3.2 Spectral density estimation

As described in Section 5.2.1, a noiseless signal in the time domain can be represented as the sum of a set of cosine functions. A more useful model assumes X_t is a random variable defined by some error such that:

$$X_t = \sum_{i=1}^K S(\omega_i) \cos(\omega_i t + \phi_i) + e_t, \quad (5.7)$$

where $e_t \sim N(0, \sigma_e^2)$. Given N observations³ $\{x_1, \dots, x_N\}$ we have unknown parameters $K, S(\omega_i), \omega_i, \sigma_e^2$ that need to be estimated. If we assume that the ω_i are known (and thus also K) we can redefine model (5.7)⁴:

$$X_t = \sum_{i=1}^K (A_i \cos(\omega_i t) + B_i \sin(\omega_i t)) + e_t, \quad (5.8)$$

where $A_i = S(\omega_i) \cos(\phi_i)$ and $B_i = -S(\omega_i) \sin(\phi_i)$ [93, 94]. This is now in the form of a multiple linear regression model. We can estimate the regression coefficients by minimising the objective function (5.9):

$$Q = \sum_{i=1}^N (X_t - \sum_{i=1}^K (A_i \cos(\omega_i t) + B_i \sin(\omega_i t)))^2, \quad (5.9)$$

³We restrict N to be an even number to simplify what follows.

⁴Using the identity $\cos(x + y) = \cos(x) \cos(y) - \sin(x) \sin(y)$.

The fundamental frequency of a signal, denoted ω_0 , is the lowest observable frequency and is the reciprocal of the signal length, so $\omega_0 = 1/N$. The Nyquist frequency is the highest observable frequency, and is half the sampling rate. A harmonic frequency, denoted ω_p , is an integer multiple, up to the Nyquist frequency, of the fundamental frequency, so $\omega_i = 2\pi p_i/N$, $p_1, \dots, p_K$ are integers, and $0 \leq p_i \leq (N/2)$. It is shown in Priestley [93] that for harmonic frequencies:

$$\begin{aligned}\hat{A}_i &= \frac{2}{N} \sum_{t=1}^N X_t \cos \omega_i t, \\ \hat{B}_i &= \frac{2}{N} \sum_{t=1}^N X_t \sin \omega_i t,\end{aligned}\tag{5.10}$$

and from this it is shown:

$$\begin{aligned}E(\hat{A}_i) &= A_i, \\ E(\hat{B}_i) &= B_i, \\ \text{var}(\hat{A}_i) &= \text{var}(\hat{B}_i) = 2\sigma_e^2/N.\end{aligned}\tag{5.11}$$

Here, $\hat{A}_i$ and $\hat{B}_i$ denote least squares estimates from (5.9). The results in (5.11) hold approximately when the frequencies are known but not of harmonic form. Thus, if we knew our time signal was an exact mixture of a specific set of frequencies ω_i , we could reduce the variance on our spectral estimates by simply observing the process longer (i.e. increasing N).

However, it is unlikely that the frequencies ω_i that comprise a real world signal are known, and thus OLS estimation is unlikely to have practical relevance beyond simply describing the estimation procedure. When the ω_i are unknown, the typical approach to spectral estimation is to construct a set of the periodogram ordinates.

The periodogram, as defined by Priestley [93], is a function over $\omega \in [0, 2\pi]$ that takes the form

$$I_N(\omega) = A(\omega)^2 + B(\omega)^2,\tag{5.12}$$

which comprises the sum of the squared *cosine* and *sine transforms* of the data,

$$\begin{aligned}A(\omega) &= \sqrt{\frac{2}{N}} \sum_{t=1}^N X_t \cos \omega t, \\ B(\omega) &= \sqrt{\frac{2}{N}} \sum_{t=1}^N X_t \sin \omega t.\end{aligned}\tag{5.13}$$

Note that both the cosine and sine transforms are functions of the random process X_t and are therefore random variables. They can also be equivalently expressed as the squared absolute value of the discrete Fourier transform of the data:

$$I_N(\omega) = \frac{2}{N} \left| \sum_{t=1}^N X_t e^{-i\omega t} \right|^2. \quad (5.14)$$

The intuition behind the periodogram is given in Priestley and summarised as follows [93]. Where ω are unknown, if we take a guess as to what they are, and that guess is far from any frequency that comprises the data, then we are effectively estimating non-existent parameters and thus $I_N(\omega)$ will be close to zero. The more accurately we guess the value of ω , the further from 0 the periodogram ordinates will be. To more easily distinguish between the small and large values of the periodogram, a normalising factor is typically applied which, in the case of Priestley is $N/2$, hence the appearance of the $\sqrt{2/N}$ in (5.13).

In order to preserve independence among ordinates, $I_N(\omega)$ is evaluated over the harmonic frequencies described earlier, which we now denote ω_p , which leads to the periodogram as it is encountered in practice [93],

$$I_p = I_N(\omega_p). \quad (5.15)$$

In general, the number of harmonic frequencies ω_p (and therefore number of periodogram ordinates) available is dependent on both the signal length and the sampling rate. Because harmonics are simply integer multiples of the fundamental frequency—the reciprocal of the signal length—with a fixed sampling rate the number of harmonics varies with signal length.

Where the frequencies of model (5.7) correspond exactly to the harmonic frequencies, ω_p , we can use (5.10) and (5.13) to show the relationship between the least squares estimates of A_i and B_i and the cosine and sine transforms:

$$\begin{aligned} A(\omega_p) &= \sqrt{\frac{N}{2}} \times \frac{2}{N} \sum_{t=1}^N X_t \cos \omega_p t = \sqrt{\frac{N}{2}} \hat{A}_i, \\ B(\omega_p) &= \sqrt{\frac{N}{2}} \times \frac{2}{N} \sum_{t=1}^N X_t \sin \omega_p t = \sqrt{\frac{N}{2}} \hat{B}_i. \end{aligned} \quad (5.16)$$

With this we can then express the expectation of the periodogram as follows:

$$\begin{aligned} E(I_p) &= E(A(\omega_p)^2) + E(B(\omega_p)^2) \\ &= \frac{N}{2}(E(\hat{A}_i^2) + E(\hat{B}_i^2)). \end{aligned} \quad (5.17)$$

We can then follow Priestley [93] and re-express the expectation in (5.17) in terms of the original model parameters in (5.7). Using (5.11),

$$\begin{aligned} E(\hat{A}_i^2) &= E(\hat{A}_i)^2 + \text{var}(\hat{A}_i) \\ &= A_i^2 + 2\sigma_\epsilon^2/N, \end{aligned} \quad (5.18)$$

with a similar result for $E(\hat{B}_i^2)$, we therefore have

$$\begin{aligned} E(I_p) &= E(A(\omega_p)^2) + E(B(\omega_p)^2) \\ &= \frac{N}{2}(E(\hat{A}_i^2) + E(\hat{B}_i^2)) \\ &= \frac{N}{2}(E(\hat{A}_i)^2 + \text{var}(\hat{A}_i) + E(\hat{B}_i)^2 + \text{var}(\hat{B}_i)) \\ &= \frac{N}{2}(A_i^2 + B_i^2 + 4\sigma_\epsilon^2/N) \\ &= \frac{N}{2}S(\omega_p)^2 + 2\sigma_\epsilon^2. \end{aligned} \quad (5.19)$$

Thus, the periodogram has a close connection to the least squares estimates of the spectral density. This result is exact if our signal has a purely discrete spectrum with frequencies of the form ω_p [93]. But, as mentioned earlier, this is not a common real-world occurrence, although if measurement noise is low then we might expect it to be an adequate approximation.

Having described the expectation of the periodogram under restrictive assumptions, that is, when $\omega_i = \omega_p$, we now turn our attention to its full sampling distribution when $\omega_i \neq \omega_p$.

$$X_t = e_t, \quad e_t \sim \mathcal{N}(0, \sigma_X^2) \quad (5.20)$$

The variance of the cosine transform, where $\omega_p \neq 0$ and $p \neq N/2$, is shown to be:

$$\begin{aligned}
 \text{var}\left(A(\omega_p)\right) &= \text{var}\left(\sqrt{\frac{2}{N}} \sum_{t=1}^N X_t \cos \omega_p t\right) \\
 &= \frac{2}{N} \text{var}\left(\sum_{t=1}^N X_t \cos \omega_p t\right) \\
 &= \frac{2\sigma_X^2}{N} \left(\sum_{t=1}^N \cos^2 \omega_p t\right) \\
 &= \sigma_X^2,
 \end{aligned} \tag{5.21}$$

because:

$$\text{var}\left(\sum_{t=1}^N X_t \cos \omega_p t\right) = \sigma_x^2 \sum_{t=1}^N \cos^2 \omega_p t = \sigma_X^2 \frac{N}{2}. \tag{5.22}$$

The result is similar for the sine transform, thus:

$$\begin{aligned}
 A(\omega_p) &\sim \mathcal{N}(0, \sigma_X^2), \\
 B(\omega_p) &\sim \mathcal{N}(0, \sigma_X^2).
 \end{aligned} \tag{5.23}$$

Priestley [93] also briefly shows that:

$$\text{cov}(A(\omega_p), B(\omega_q)) = \frac{2\sigma_X^2}{N} \left(\sum_{t=1}^N \cos \omega_p t \sin \omega_q t\right) = 0 \tag{5.24}$$

for all p and q , and similarly,

$$\text{cov}(A(\omega_p), A(\omega_q)) = \text{cov}(B(\omega_p), B(\omega_q)) = 0, \tag{5.25}$$

which implies that the sine and cosine transforms are mutually independent, given that they are both functions of X_t . Thus, each periodogram ordinate is the sum of two independent, squared, Gaussian random variables,

$$I_p \sim \sigma_X^2 (\mathcal{N}(0, 1)^2 + \mathcal{N}(0, 1)^2). \tag{5.26}$$

which implies that the ordinate is distributed according to a scaled chi-squared distribution with 2 degrees of freedom (χ_2^2) such that:

$$I_p \sim \sigma_X^2 \chi_2^2. \tag{5.27}$$

Note that for the process defined by (5.20), $S(\omega) = \sigma_X^2$. This result is shown similarly in Shumway and Stoffer [94] and Timmer and König [95], who further show that, in general, for any normal linear process, not just for stationary Gaussian processes as defined by (5.20):

$$I_p \sim S(\omega_p) \chi_2^2. \quad (5.28)$$

Because $E(\chi_\nu^2) = \nu$ and $\text{var}(\chi_\nu^2) = 2\nu$, it follows that $E(I_p) = 2S(\omega_p)$ and $\text{var}(I_p) = 4S(\omega_p)^2$. Thus, although periodogram ordinates are proportionally unbiased, they have large variances and are not consistent (since the variances do not depend on N) [93, 95, 107]. It is suggested that this makes the raw periodogram near unusable for most real-world applications [93].

Common approaches to reducing the variance of periodogram ordinates are either smoothing of the periodogram ordinates using some form of moving average or filter, or splitting the original time series into equal length subsets and averaging the periodogram estimates from each subset (otherwise known as ‘windowing’). However, neither of these methods is entirely satisfactory if we are interested in estimating the power-law exponent as the former results in correlated estimates of spectral density, while the latter may both result in highly variable estimates of the power-law exponent and also fail to capture valuable low frequency information, particularly with short signals [108].

5.3.3 Power-law exponent estimation

We now define two estimators of the power-law exponent using periodogram ordinates. The first, OLS, is perhaps the most commonly used linear estimation procedure, both for estimating the power-law exponent and in general, and thus requires minimal explanation. The second is a novel approach based on an explicit model of the error distribution described in (5.31), fitted using maximum likelihood estimation (MLE). We also describe two additional estimators that implement these methods but use only low-frequency periodogram ordinates.

OLS is used to estimate our power-law exponent by minimising the sum of the squared residuals:

$$\sum_{i=1}^N r_i^2, \quad (5.29)$$

where,

$$r_i = \log(I(\omega_i)) - (\log(\alpha) + \beta \log(\omega_i)). \quad (5.30)$$

We implement OLS using the `lm` function in the R programming environment.

One potential issue with the periodogram method that we identified relates to the error distribution of the periodogram ordinates. As described in Fan & Yao [107] and Zhao-Guo & Hannan [109], the implications of the distributional result in (5.28) is that the error term in (5.4) has a probability density function proportional to

$$f_r(x) = \exp(-\exp(x) + x). \quad (5.31)$$

To see how this may affect the estimates of the parameter β , we must refer to the standard assumptions of the Gauss-Markov theorem which states that OLS is the best linear unbiased estimator (BLUE) where:

1. The model parameters α and β are linear.
2. That the expected value of the residuals is zero: $E(r_i) = 0$
3. That the conditional variance remains constant for all i : $\text{var}(r_i) = \sigma^2$
4. That the conditional covariance of errors is zero: $\text{cov}(r_i, r_j) = 0, i \neq j$
5. That X is uncorrelated with the error term: $\text{cov}(X_i, r_j) = 0$

We can be confident that the Gauss-Markov assumptions 1, 4, and 5 are satisfied. The heteroskedasticity present in the raw periodogram of a power-law process is controlled for using the logarithmic transformation, thus satisfying assumption 3. However, it is clear that assumption 2, the zero conditional mean is not satisfied. This is however, not a major concern for estimating the power-law exponent β , which should remain unbiased, although this will not be the case for the intercept term α .

Yet even after log transforming the periodogram, the error distribution will remain non-normal. According to the Gauss-Markov theorem there is no requirement for the normality of errors for OLS estimators to be unbiased. The only real requirement for normality is either when the model will be used for prediction [70] or to ensure that the maximum-likelihood estimate of the parameters will be the same as that of OLS. It is suggested,

however, that OLS estimation in the presence of either heavy tailed or highly skewed error distributions, while BLUE, may not be the most efficient [110]. Thus OLS may not be the optimal approach for estimating β , particularly for shorter signals with fewer periodogram ordinates to use in modelling. We could not identify any studies that have assessed the impact of the non-normality of the periodogram error distribution on subsequent estimates of the power-law exponent.

The known distribution of the log periodogram (5.31) allows us to construct the following log-likelihood where r_i is defined by (5.30):

$$\ell(\alpha, \beta) = \sum_{i=1}^N -\exp(r_i) + r_i. \quad (5.32)$$

This likelihood function is maximised numerically for α and β using the `optim` function in the R programming environment. Herein we will refer to this estimator as the log chi-squared error estimator (and denote it LCE).

As shown in (5.19), the expected value of the periodogram is altered by observational noise in the original signal, which produces what is termed a ‘noise floor’ under which higher frequency ordinates are masked [93]. While it is likely that either dropping the high-frequency estimates from the periodogram, as recommended by Eke [46, 98], and/or only using estimates at frequencies with physiological relevance, as seemingly done by Frey et al. [33] and Habib et al. [35], may mitigate this effect, such a cut-off may discard potentially useful data and thus affect the subsequent estimates of β .

Therefore, in addition to the OLS and LCE estimators, we also follow the general approach of Frey et al. [33] and Habib et al. [35], that is, to remove high-frequency periodogram ordinates. While neither study was explicit in what cut-off was used, the general recommendation we followed was to remove all ordinates corresponding to frequencies over 1/8 of the Nyquist frequency [46]. In our case the Nyquist frequency was 100 Hz (half of the sampling rate, 200 Hz), so we used a cut-off of 12.5 Hz.

We hypothesised that this general recommendation may not be sufficient to achieve unbiased estimates of the power-law exponent, given the signal-to-noise ratio (SNR) of our tidal flow measurements. We therefore also applied a more extreme cut-off of 1/20 of the Nyquist frequency, in this case 5 Hz.

We indicate these threshold-based estimators using the subscripts ‘mid’ and ‘low’ to indicate the use of ordinates from mid- and low-range frequencies only. Overall, this gave us four additional estimators, which we refer to as OLS_{mid} , LCE_{mid} , OLS_{low} and LCE_{low} .

That the periodogram ordinates are highly variable is perhaps not a major issue in the context of estimating a power-law exponent, because we are not interested in any particular estimate of spectral density, but rather the OLS estimates of the power-law function using the spectral density estimates as the input ‘data’. This is somewhat analogous to the smoothing and filtering approaches described in the previous section. Although with shorter signals the high variability of periodogram ordinates means that the resulting power-law exponent estimates will also be highly variable, we take the approach recommended by Delignieres et al. [92] and used by Frey et al. [33] and Habib et al. [35], that is, to average the estimates from a number of individual shorter signals.

Note that this averaging approach requires an assumption about the statistical properties of the shape of the tidal flow waveforms, in particular that the shape of each tidal flow waveform for a given test is independently and identically distributed over replicate breaths. We note that this assumption is supported by Figure 5.1 which shows broadly similar shapes of not only the two tidal flow waveforms presented within each test, but also across individual test occasions for each individual.

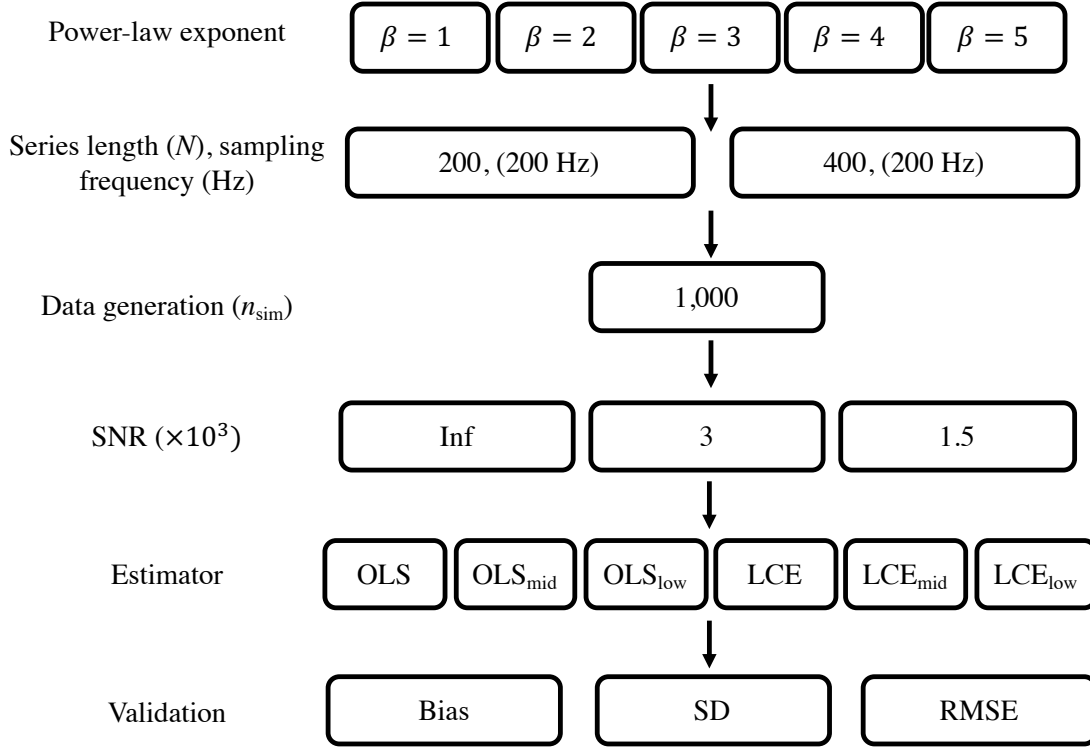
5.4 Simulation

Having defined estimators of the power-law exponent, this section examines their statistical properties using simulated data with various properties similar to a tidal flow waveform. A schematic of the simulation is provided in Figure 5.4. The details of the simulation scheme are provided in the following subsections.

5.4.1 Data generation

Timmer & König [95] described a straightforward algorithm to simulate periodic time series that incorporates spectral variation and results in periodogram ordinates distributed as

FIGURE 5.4: Scheme for simulating a time series with an underlying power-law spectrum using the Timmer & König algorithm [95]. Hz: hertz; SNR: signal-to-noise ratio; OLS: ordinary least squares; LCE: log chi-squared estimator; RMSE: root mean squared error; SD: standard deviation.



described in (5.28). The simulation model,

$$X_t = \sum_{i=1}^N S(\omega_i) \cos(\omega_i t - \phi(\omega_i)), \quad (5.33)$$

describes the time domain data $X_1, \dots, X_t$ as a mixture of periodic frequencies of various phases, $\phi(\omega_i)$, and amplitudes, $S(\omega_i)$. The ω_i here are defined as harmonic frequencies $i = (1, \dots, K)$.

This algorithm builds on a previously described method that assumed a deterministic amplitude and only randomised the phase [111] to simulate galactic X-ray emissions, and has been used in many studies in various fields as described in Emmanoulopoulos et al. [112]. The algorithm, in broad terms, is implemented as follows⁵:

⁵An R-language implementation of this algorithm is given in the Appendix D.

- Define the underlying power-law spectrum by β as in (5.2).
- For each harmonic frequency ω_p , draw two numbers from a standard normal distribution and multiply them by $\sqrt{S(\omega_i)}$. For each frequency ω_p , the two numbers form the real and imaginary part of a Fourier transform of the data, denoted $f_{\text{sim}}(\omega_p)$.
- Use $f_{\text{sim}}(\omega_i)$ to reconstruct the signal in the time domain, $X_{\text{sim}}(t)$, using the inverse Fourier transform.

This algorithm is convenient as by randomising the sine and cosine transforms (in the form of the Fourier transform) it avoids the need to define an explicit distribution for the phase. But the phase is randomised and a consequence is that the data will typically not appear to resemble a tidal flow waveform in the time-domain exactly. However, this is not an issue for assessing the performance of our power-law estimators as the phase information is not captured in the periodogram outside of the sampling distribution of the ordinates. Thus the spectral estimates will be consistent with those of a tidal flow waveform as we show in Figure 5.5.

In addition to uncertainty in phase and amplitude, a measured signal will likely also incorporate some degree of observational error as described in 5.7. We therefore add an additional step to the Timmer & König algorithm that incorporates additive error e_t to the time signal,

$$X_t = \sum_{i=1}^N S(\omega_i) \cos(\omega_i t - \phi(\omega_i)) + e_t, \quad (5.34)$$

where $e_t \sim N(0, \sigma_e^2)$. We tie this error to the strength of the signal using the signal-to-noise ratio (SNR), which can be defined as:

$$\text{SNR} = \frac{\text{var}(X_t)}{\sigma_e^2} \quad (5.35)$$

Where X_t is simply the noiseless signal as defined in 5.33.

As described in Figure 5.4 we assume $\beta = (1, \dots, 5)$ to cover a range of different power-law exponents we might expect in our data. We also use a range of different additive noise scenarios. It is difficult to find exact estimates of the signal-to-noise ratio of commonly used ultrasonic transducers used for recording tidal flow waveforms, however Singer et al. [113] note that it is somewhere between 3,000–3,500. We therefore set $\text{SNR} = (\text{Inf}, 3000, 1500)$ —where $\text{SNR} = \text{Inf}$ is simply a noiseless signal, $\text{SNR} = 3,000$ represents a typical amount

of noise encountered using ultrasonic technology, and $\text{SNR} = 1,500$ represents double the typical noise levels.

Commercially available ultrasonic flowmeters used for tidal-breathing tests, for example the Exhalyzer D (Exhalyzer D, Ecomedics, Duernten, Switzerland), typically take samples at a rate of 200 Hz; we reflected this in our simulations. To understand the effect of estimating β from different tidal flow waveform lengths, and therefore the number of periodogram ordinates, we simulated waveforms of both 200 samples (1 second) and 400 samples (2 seconds) in length, which broadly reflects respiratory rates observed in early childhood [114].

5.4.2 Performance evaluation

For each scenario we performed the simulation 1,000 times (i.e. $n_{\text{sim}} = 1,000$). For each simulation, β was estimated using the estimators previously described. The performance of each estimator was evaluated using bias, empirical standard deviation, and root mean square error (RMSE). Bias was defined as the difference between the mean estimate and the target parameter, $\delta = E(\hat{\beta}) - \beta$, with standard deviation defined as the square root of the variance of the estimates, $\sqrt{\text{var}(\hat{\beta})}$, and RMSE, a combined measure of bias and variance, defined as $\sqrt{\delta^2 + \text{var}(\hat{\beta})}$. We assessed the uncertainty around the bias using the Monte Carlo standard error of the bias ($\text{MCSE}_{\text{bias}}$), as defined in Morris et al. [115]:

$$\text{MCSE}_{\text{bias}} = \sqrt{\frac{\text{var}(\hat{\beta})}{n_{\text{sim}}}} \quad (5.36)$$

5.4.3 Results

We performed the simulations as described in the previous section. Examples of simulated waveforms of 1 second length for each combination of β and SNR are provided in Figure 5.5. A graphical summary of the periodogram applied to simulated data for each combination of β and SNR is shown in Figure 5.6. This figure demonstrates clearly the properties of the periodogram described in Section 5.3: the substantial variation of the periodogram ordinates around the underlying spectral density, the asymmetrical non-normal distribution (as described by equation 5.31), and the masking of the higher-frequency ordinates by observational errors, particularly for high values of β .

Figure 5.7 provides the results of the simulation study. For demonstrative purposes, Figure 5.8 shows the fitted OLS estimates of the power-law exponent from a single simulated signal under each condition as described in Figure 5.4. What is clear from these results is that, with the exception of trivial examples where there is no signal noise, none of the estimators we have proposed perform well under each combination of signal length, power-law exponent, and SNR. The highest empirical standard deviation of any estimator was for OLS_{low} at a value of approximately 1. Using a value of 1 as the numerator in (5.36), with $n_{\text{sim}} = 1,000$ as the denominator, resulted in a ‘worst-case’ Monte Carlo standard error of the bias of approximately 0.03, which we considered acceptable in the context of possible values of power-law exponents in general.

The accuracy of each estimator is highly dependent on β as well as signal length and noise. Where there is no signal noise, all estimators are unbiased, but this is not a likely scenario encountered in practice. We can see from Figure 5.7, where the signal contains noise, as β increases the need to discard higher frequency ordinates also increases in order to preserve accuracy. For example, where $\beta = 3$ estimators that use all ordinates become severely biased relative to those that discard high frequencies. This result reflects common recommendations to restrict the use of ordinates below 1/8th of the Nyquist frequency when using the PSD approach [46]. However, we can see that even this restriction may not be conservative enough. For values of $\beta > 4$, where there is signal noise, the only estimator with acceptable accuracy appears to be those that ignore ordinates above 1/20th of the Nyquist frequency and, even then, accuracy appears to be mildly affected by signal noise.

Perhaps counter-intuitively, increasing the signal length in the presence of signal noise actually *increases* bias. To see why, note that increasing the signal length lowers the fundamental frequency, and for a fixed sampling rate, increases the number of harmonics available to estimate the spectral density. Where a noise floor is clearly present it is then populated with additional noisy periodogram ordinates. Because none of the estimators differentiate between the importance of high and low frequency ordinates, a disproportionate amount of the estimation is based on ordinates that simply reflect the degree of noise in the signal. This disproportion increases with increasing signal length.

All estimators exhibit sampling variation under the stated conditions that is likely to prohibit their usefulness when applied to a single breath signal in isolation, apart from the unrealistic example of where there is no signal noise and none of the higher frequencies

are discarded. This sampling variation would be reduced by increasing the number of noise-unaffected periodogram ordinates that are then input into the estimation procedure. However this is not possible. The number of ordinates is a function of signal length—which is dependent on underlying physiology—and sampling rate—which is dependent on the signal measurement apparatus and cannot be increased at whim. Although some signals will contain more ordinates than others, the number cannot be increased naturally by increasing the signal length or the sampling rate. This may not represent a substantial problem in practice, however, as estimation will proceed for a large number of breaths per individual, the precision of the individual level estimate can be improved in proportion to the number of breaths by averaging breath-specific estimates—in effect obtaining more data from similar breaths rather than lengthening the signal. This is the approach used by Frey et al. [33] and Habib et al. [35]. Given that our interest is in accurately estimating an individual’s underlying respiratory complexity, via the power-law exponent, rather than that of a single breath, we should focus on using the estimator that is least biased in all cases. Both OLS and LCE estimators perform similarly in terms of RMSE within their sub-categories, but the OLS estimator has better properties. Although LCE estimators appear to be marginally more efficient than OLS, particularly in the presence of signal noise, this tends to come at the expense of accuracy. Under the conditions we examine here, the OLS_{low} estimator is arguably the least biased of all estimators and most suitable for practical estimation of the power-law exponent.

FIGURE 5.5: Simulated time signals of 1 Hz length with an underlying power-law spectrum for various β with (A) no additive noise, (B) SNR = 1,500, and (C) SNR = 3,000. SNR: signal-to-noise ratio.

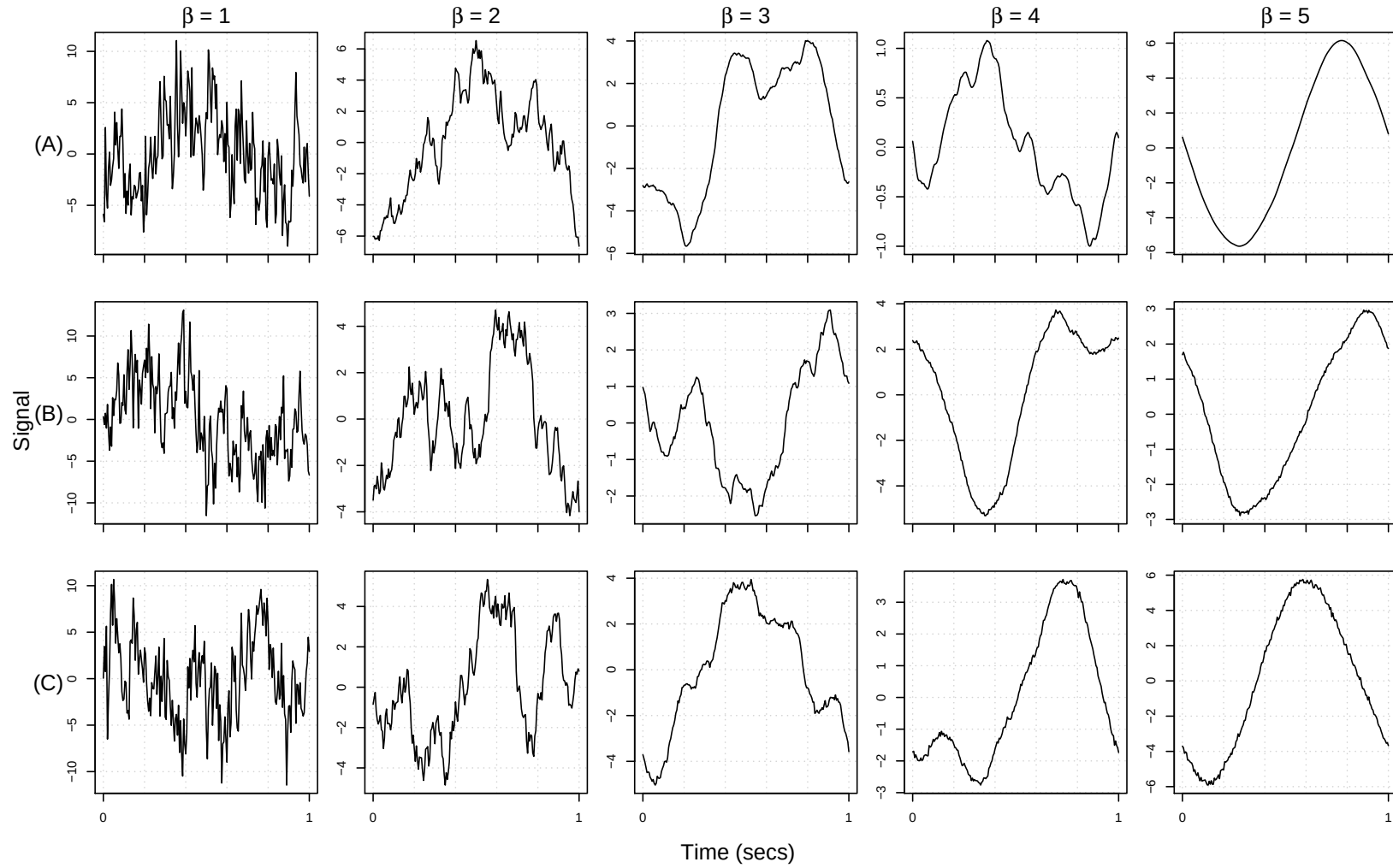


FIGURE 5.6: Periodogram applied to simulated time signals with an underlying power-law spectrum ($n = 1,000$) for various β with (A) no additive noise, (B) $\text{SNR} = 3,000$, and (C) $\text{SNR} = 1,500$. Black points are the values of a typical spectrum power used for the simulation (in this case the exact discrete power-law spectrum). Black and grey lines shows the mean and 2.5% and 97.5% quantiles for the periodogram estimates of spectral density. SNR: signal-to-noise ratio.

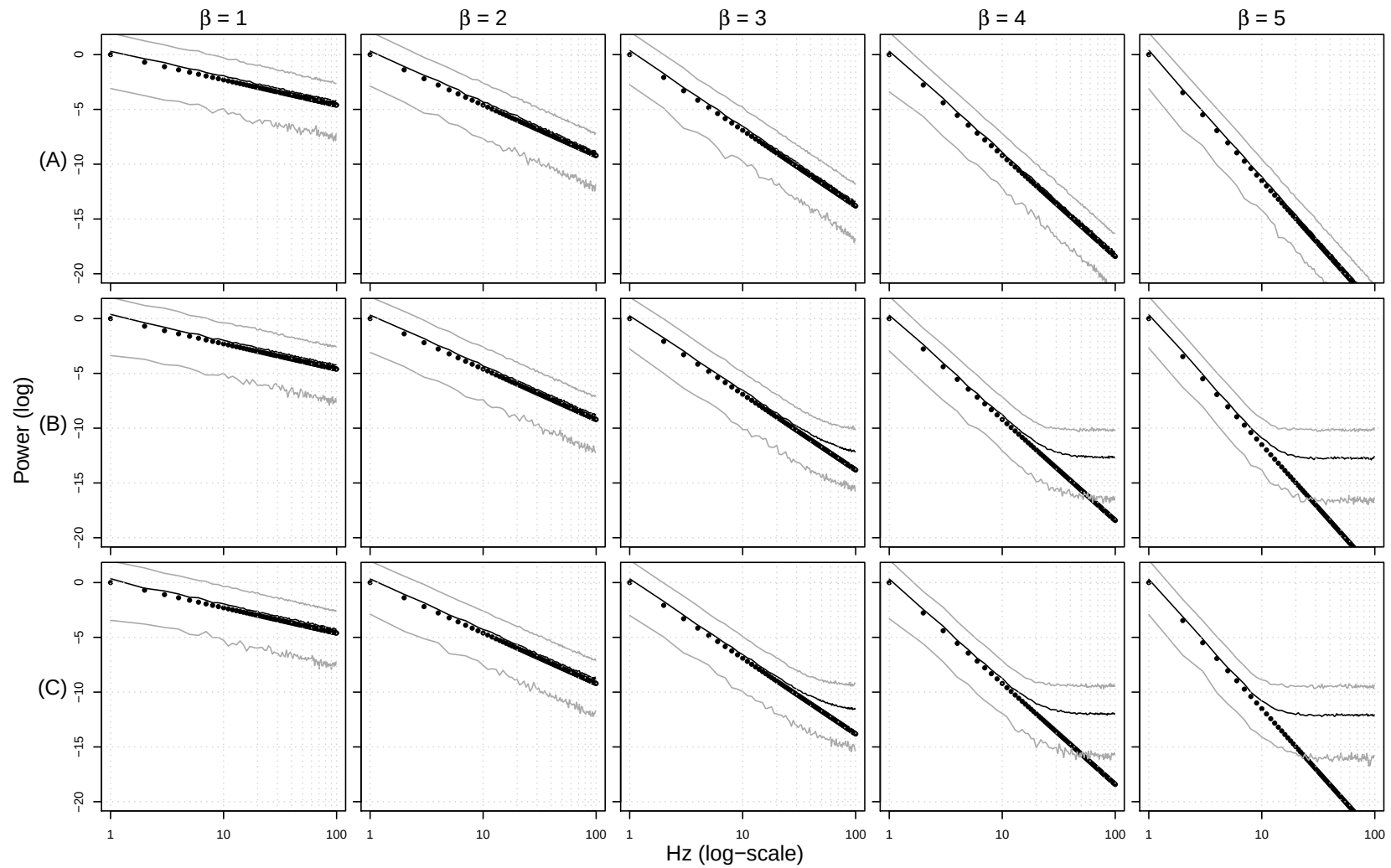


FIGURE 5.7: Results from estimation of power-law exponent from simulated data. Estimation is evaluated using bias and root mean squared error (RMSE). Methods assessed are ordinary least squares (OLS), log chi-square error estimation (LCE) and variants thereof which use only frequencies below 12.5 Hz and 5 Hz (denoted by subscript ‘mid’ and ‘low’, respectively). ω_0 : fundamental frequency; SD: standard deviation; SNR: signal-to-noise ratio ($\times 10^3$).

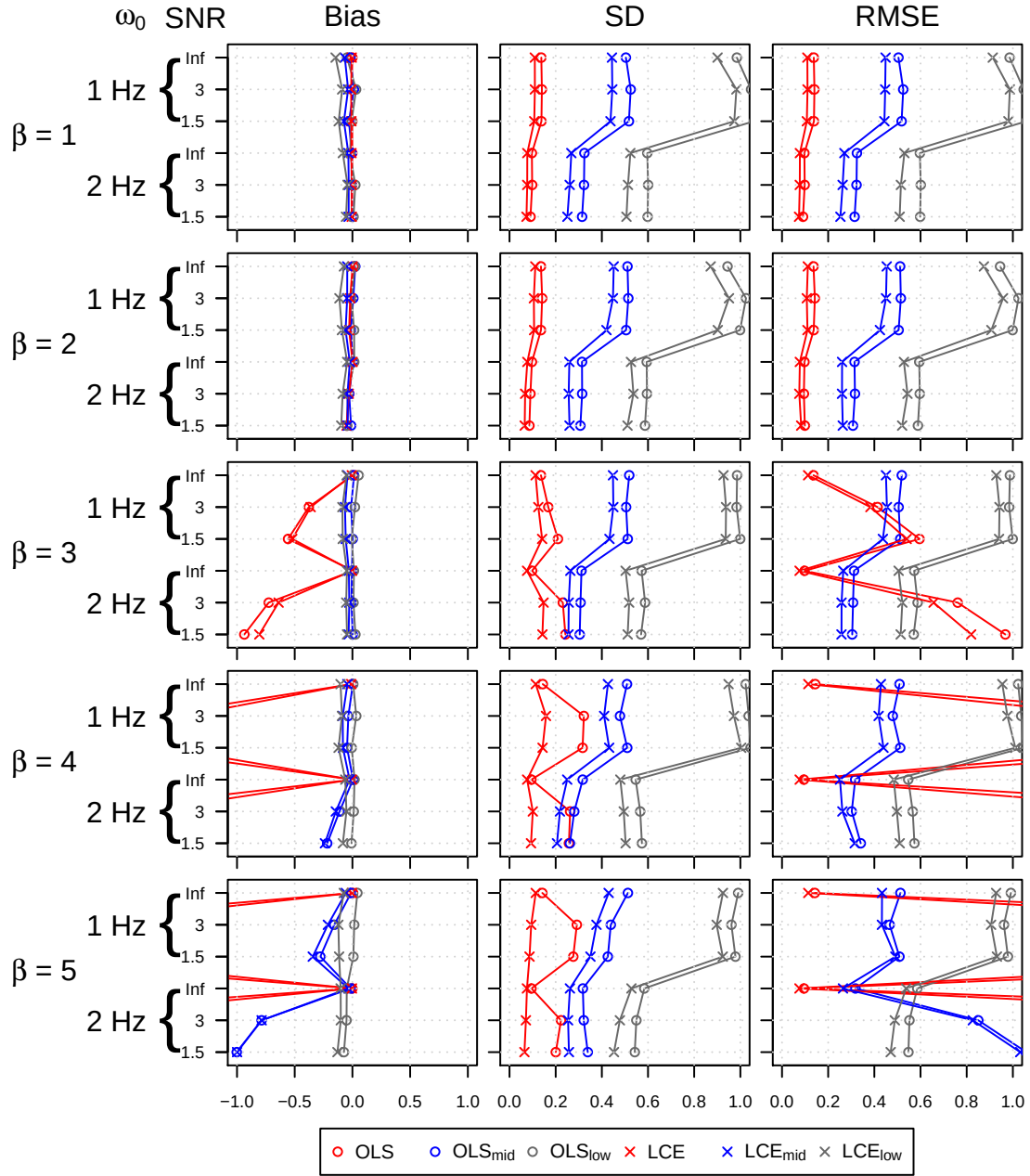
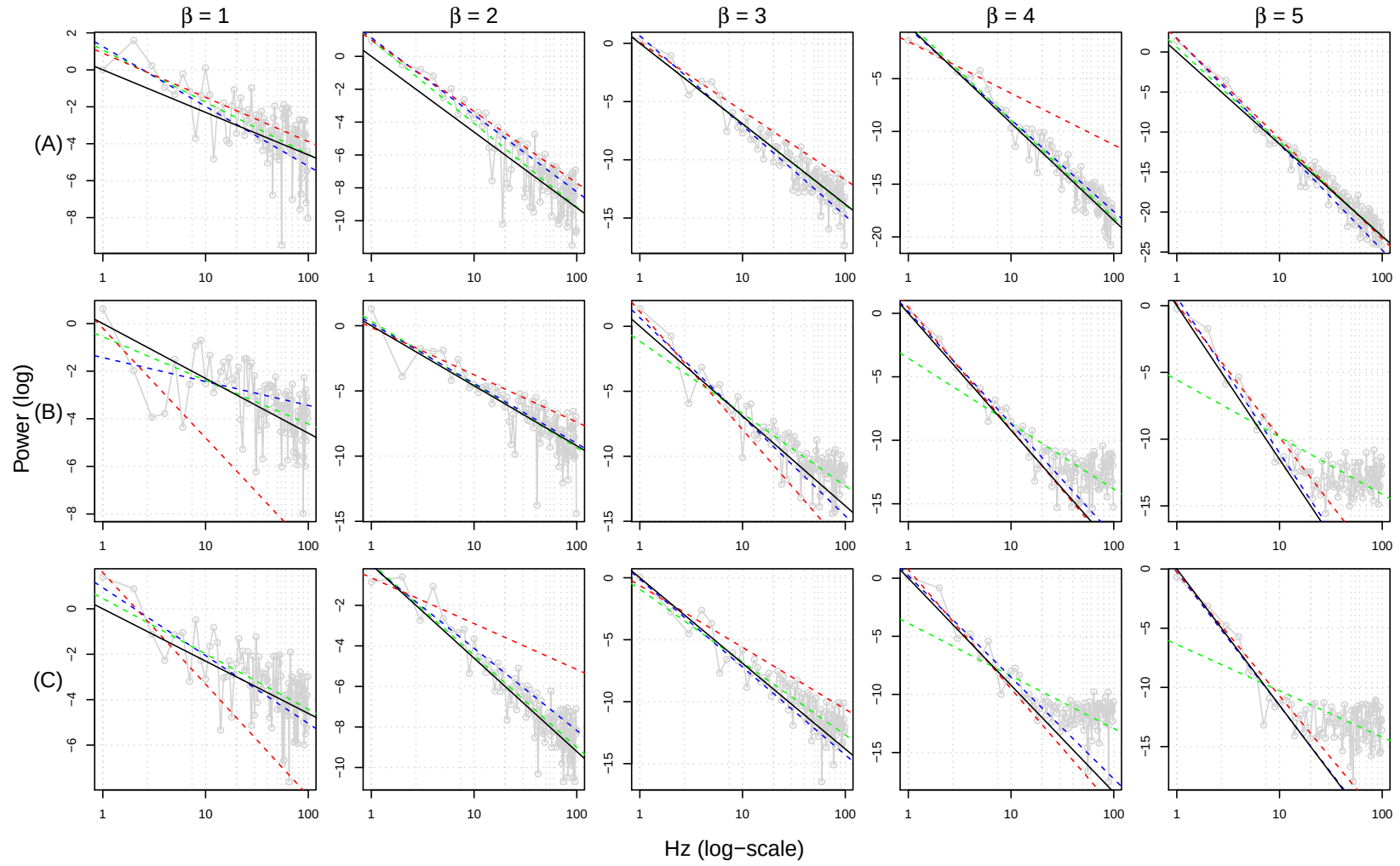


FIGURE 5.8: Fitted ordinary least squares (OLS). Shown are the periodogram ordinates (grey), the underlying power-law function (black), and fitted OLS function using frequencies up to 100 Hz (green), 12.5 Hz (blue), and 5 Hz (red).



5.5 Discussion

This chapter has introduced the concepts of scale invariance and estimation of a power-law exponent from the outputs of complex systems, specifically tidal flow waveforms measured during tidal-breathing. We reviewed various estimation methods for power-law exponents with a focus on the PSD methods used by Frey et al. [33] and Habib et al. [35]. We then examined the statistical properties of various estimators using simulated data that have similar properties to tidal flow waveforms.

We examined the statistical properties of not only OLS estimators of the power-law exponent from periodogram ordinates but, as far as we know uniquely in the literature, also novel estimators that properly account for the non-normal distribution of the periodogram ordinates. We made this investigation in the context of short time signals in combination with observational error. This investigation was necessary to ensure that, when examining scale invariance in BIS and AREST CF tidal flow waveforms, we were using an estimator with reasonable statistical properties. Although other studies have made similar investigations [46, 47, 92], they have neither opted for a thorough explanation of the statistical properties of the PSD method nor examined the performance of their methods in a context similar to that presented here.

We have shown that the sampling properties of PSD estimators are highly sensitive to the value of the underlying power-law exponent, signal length and signal noise. Where there is even a small amount of observational error (relative to the underlying signal), bias is substantially increased. However, we have shown that dropping high-frequency ordinates from the estimation procedure mostly mitigates this bias when using OLS, albeit at a cost to efficiency. While dropping high frequencies is not a new idea [46], we believe that the recommendation to drop frequencies above $1/8$ of the Nyquist frequency may be too conservative, particularly if the signal-to-noise ratio of a signal is known. Furthermore, we have shown that the asymmetry of the conditional error distribution of the periodogram ordinates does not substantially affect the proper estimation of the power-law exponent using OLS. We recommend that in practice, the power-law exponent should be estimated using the OLS_{low} estimator—which has the lowest bias in general—averaged across a sufficiently large number of individual breaths to increase precision.

Outside of Frey et al. [33] and Habib et al. [35], studies of scale invariance behaviour of the respiratory system to date are limited to using end-tidal quantities and time-domain methods such as DFA. Although analysing tidal flow waveforms using the ‘periodogram method’ may not capture some of the long-range correlation properties that may be evident in long-series of end-tidal quantities, it may contain complementary information about respiratory complexity, and we believe that it may be worthwhile to study both in tandem. Given that the method using tidal flow waveforms is vastly more feasible in a clinical setting there may be some merit in understanding whether complexity at the ‘micro’ level, in our case in the tidal flow waveform, is in any way associated with complexity at the ‘macro’ level, for example in end-tidal quantities collected over a much longer timeframe.

This study has focused specifically on power-law estimation from data similar to tidal flow waveforms, which are very short and have a high signal-to-noise ratio. Even in this context, we found that performance of our estimators were sensitive to very minor observational error in the original signal. As far as we know, observational error is not typically considered when applying PSD methods, apart from the general recommendation to ignore frequencies over $1/8$ of the Nyquist frequency, however as we have shown doing so, even in the context of very low observational error, may result in severely biased power-law exponent estimates, even after discarding high-frequencies. It would be illustrative and valuable for future research to investigate the sampling properties of PSD estimators across a wider range of signal lengths and observational noise.

Chapter 6

Epidemiology of tidal flow power-law spectra

Contents

6.1	Background	98
6.1.1	Aims	99
6.1.2	Limitations	99
6.2	Methods	102
6.2.1	Data	102
6.2.2	Models	110
6.2.3	Estimation	112
6.3	Results	114
6.3.1	Power-law spectra	114
6.3.2	Model of age and cystic fibrosis	115
6.3.3	Model of BIS power-law exponents and covariates	120
6.3.4	MBW and non-MBW agreement analysis	130
6.4	Discussion	130

In this chapter we focus on the application of estimators of the power-law exponent, which were described in technical detail in Chapter 5, to tidal flow waveform data collected as part of the AREST CF and BIS studies, with the aim of exploring the potential value of this modelling concept for understanding the epidemiology of lung diseases of various types.

We begin with a brief background in Section 6.1, noting that many of the relevant technical details have been covered in Chapter 5. In Section 6.2 we follow with a description of the data and models used in the analysis. Results and discussion are provided in Sections 6.3 and 6.4.

6.1 Background

Traditional summaries of lung function taken from tidal flow recordings, for example the ratio of time until peak tidal expiratory flow (T_{PTEF}) to expiratory time (T_{E}) ($T_{\text{PTEF}}/T_{\text{E}}$), are perhaps overly simple in that they tend to focus on a single dimension, such as time, in an effort to gain insight into what is ultimately a highly complex system [33, 35, 116]. Frey et al. [33] and Habib et al. [35] proposed more sophisticated analytical methods to characterise the shape of the entire tidal flow waveform, via its spectral properties, to provide a deeper understanding of the underlying physiological complexity responsible for respiration [33, 35]. These studies provide some guidance on the development and implementation of such methods, the relationship of the power-law exponent to covariates such as age, disease, and birthweight, and the novel insights that may be gained by their practical, routine use. A detailed theoretical background to the power-law exponent and its estimation from tidal flow data, as well as the related concepts of scale invariance and complexity was provided, in Chapter 5. We focus here on the application of the method to clinical data similar to those of Frey et al. [33] and Habib et al. [35]. While various procedures have been used to estimate power-law exponents from long recordings of per-breath quantities such as inter-breath intervals or tidal volume [39–41, 43, 44], Frey et al. [33] and Habib et al. [35] were the first to adapt such methods to individual breath cycles.

Despite being earmarked as a potentially fruitful area of research [33, 35, 48], we have not identified any work extending the results of Frey et al. [33] and Habib et al. [35]. Given their findings, that tidal flow recordings are a fairly ubiquitous form of lung function measurement, and that the power-law method uniquely characterises the entire complex tidal flow waveform, there exists a clear opportunity to revisit this novel approach to the understanding of lung function. In doing so we may gain a better insight into the impact of disease and early life exposures on the developing respiratory system. This may enable, for example, better disease monitoring and intervention, particularly because tidal flow recordings are relatively simple to obtain, requiring a minimum of equipment and patient

discomfort compared to, for example, more complicated washout and spirometric systems. Encouraged by positive personal correspondence with various experts in respiratory medicine and lung function testing, and with access to a large amount of longitudinally collected tidal flow recordings from the AREST CF and BIS studies, we identified this gap in the literature as an opportunity to explore the potential value of this modelling concept.

6.1.1 Aims

The overarching aim of the research presented in this chapter was, similar to Frey et al. [33] and Habib et al. [35], to estimate a power-law exponent from tidal flow waveforms collected in early childhood and to explore the potential value of this modelling concept for understanding the epidemiology of lung diseases of various types. In contrast to the two aforementioned studies however, we have relatively long tidal flow recordings from a very large number of study participants and use an estimator that explicitly accounts for noise to reduce bias in the estimation procedure, as described in Chapter 5.

The overall aim was broken into the following three specific objectives:

1. To explore whether the power-law spectra of tidal flow waveforms differ between children with and without cystic fibrosis, and whether any such difference changes with age, using tidal flow recordings collected during MBW testing from both the AREST CF and BIS studies.
2. To explore whether differences between the power-law spectra of tidal flow waveforms of infants and preschool aged children vary with other covariates, using tidal flow recordings collected during MBW testing from the BIS study.
3. To examine differences in power-law exponents estimated from SF₆ MBW and non-MBW tidal flow recordings from essentially healthy children using data collected during the BIS study.

6.1.2 Limitations

Importantly this analysis was limited by three key factors, namely that:

- There are major differences in the MBW data collection procedure both between and within the AREST CF and BIS studies.
- Covariates of epidemiological interest were available only for the BIS cohort.
- The number of follow-ups differed substantially between the two studies, with only two time points available for BIS participants: a baseline observation at around 1 month of age and a follow-up observation around 4 years of age.

As described in Chapter 2, for both the AREST CF and BIS studies, two different MBW procedures were used depending on whether the child was aged above or below 3 years, namely SF₆ MBW for children under 3 years of age and N₂ MBW for children over 3 years of age.

Not only was the washout gas different in each case, but each procedure had its own specific software and hardware for performing and processing the collected data. For example, SF₆ MBW was performed with a mainstream ultrasonic flowmeter via a facemask interface, with SF₆ being calculated from changes in molar mass measured by the mainstream ultrasonic flowmeter, and tests were analysed and exported using WBreath (version 3.19, ndd Medizintechnik AG, Zurich, Switzerland), whereas N₂ was performed using a mainstream ultrasonic flowmeter via a mouthpiece interface with a sidestream O₂ gas analyser and tests were analysed and exported using Spiroware (Spiroware 3.1.6, Ecomedics AG, Duernten, Switzerland). It has been established that various lung function indices differ between software and procedure [28, 58]. This is likely due to differences in the analytical procedure and software used to analyse the trace of the molar mass of respired air.

It has also been established that various lung function indices derived from tidal flow recordings differ according to whether they are collected via MBW or while simply respiring room air [27]. Furthermore, changes to breathing pattern upon exposure to 100% O₂ have been documented for both infants and older children [117, 118]. Yet whether differences in the washin gas used—for example, SF₆ is significantly heavier than the O₂ used in N₂ MBW (146.06 g/mol vs 32 g/mol, respectively)—or use of a more elaborate apparatus than a simple room air recording affect tidal flow breathing pattern at the spectral level is unknown. We do know that for data exported using Spiroware (N₂ MBW) some degree of smoothing or filtering is applied, which obfuscates the periodogram at higher frequencies (see Figure 5.3 for an example). We assume that this smoothing does not

effect spectral estimation of lower-frequency harmonics, and therefore using the OLS_{low} estimation procedure will provide accurate results, but we cannot confirm this from the software documentation [49].

Habib et al. [35] found only a small difference between power-law exponents estimated from tidal flow data obtained using pneumatograph and respiratory inductance plethysmography (RIP), suggesting that very different methods for measuring tidal flow waveforms may not have a significant impact on the spectral properties of the waveforms. The N_2 MBW, SF_6 MBW, and non-MBW tidal-breathing recordings all measure tidal flow using an ultrasonic transducer, which leads us to expect that the spectral properties of the waveforms will also be similar. As described in Chapter 2, non-MBW tidal flow data were available for a large number of BIS infants who also performed the MBW. With these data we can improve our understanding of the differences between power-law exponents estimated from data collected by different measurement devices, thus addressing the third objective stated in Section 6.1.1.

Additionally, tests under 3 years of age were performed under sedation for AREST CF participants and during natural sleep for BIS participants. For children over 3 years, in both AREST CF and BIS cohorts, MBW was performed while sitting upright and distracted using a children's film or similar. Unfortunately, we do not have data to examine the effects of sedation, sleep, or body position on the spectral characteristics of the breathing pattern. Nevertheless, it is important to recall from Chapter 5 that Frey et al. [33] did not establish a meaningful difference between power-law exponents estimated in healthy infants before and after histamine challenge. While a histamine challenge may not have direct parallels with sedation or body position, Frey et al.'s [33] result suggests that the spectral densities of tidal flow waveforms, and therefore the power-law exponent, is determined by respiratory control systems rather than alterations to airway mechanics. We might reasonably assume that body position affects airway mechanics more than it affects respiratory control, and therefore has little effect on tidal flow power spectra, but sedation is known to affect the control of breathing and therefore its effect on tidal flow spectra cannot be ignored.

The lack of covariates of epidemiological interest in AREST CF compared to BIS is due to differences in the purposes of each study. AREST CF is an early surveillance program that aims to assess, treat, and monitor cystic fibrosis in young children. The type of covariates

collected during testing, for example genetic information and mode of presentation, are more related to these aims rather than broader epidemiological investigation. BIS, on the other hand, specifically aimed to examine the effect of early-life epidemiological exposures on various health outcomes, and therefore a range of covariates of epidemiological interest were available for analysis (see Table 6.2). Furthermore, having only two time points available for each BIS participant presented an analytical challenge because of the difficulties in both estimating an age effect for each individual while also allowing for residual variation over time.

Because of these limitations, the analyses performed herein are firmly intended to be descriptive and exploratory only, that is, we intend to generate hypotheses rather than establish firm associations or causal effects.

6.2 Methods

6.2.1 Data

This study used both MBW and non-MBW tidal flow waveform data from AREST CF and BIS (see Chapter 2.2 and 2.3 for description of data collection and management procedures) to estimate a power-law exponent for every available breathing cycle in both studies.

An automated breath detection algorithm—described to be particularly robust in the presence of severely disturbed respiratory signals or irregular breathing patterns [57, 61]—was implemented to identify each individual breathing cycle (see Section 2.4 for details). The first and last detected breaths of each test were removed as they were often incomplete.

Figure 6.1 provides an example of the individual breath cycles detected by the algorithm. As we can see, when the breathing pattern was stable, as it appears to be in panels (A), (C), and (D) of Figure 6.1, the breath cycles were easily identified by the algorithm. When the breathing pattern was less stable, for example in panel (B) of Figure 6.1, the breathing cycles are still well-identified by the algorithm, but a number of breath cycles could be described as ‘abnormal’—for example, where flow stalls due to breath holding.

We chose to include the ‘abnormal’ breaths in the analysis due to the sheer volume of individual breath cycles (making it difficult to filter them out reliably) and the fact that

abnormal breaths may contain relevant information about respiratory complexity. For example, we assessed 2,745 individual MBW tests which comprised tidal flow recordings totalling over 200,000 breath cycles. It was clearly infeasible to review the ‘normality’ of each breath cycle. We therefore placed our trust in the algorithm to identify the majority of breath cycles correctly, which it seemed to upon reviewing a random subset of 10 tests for each study (AREST/BIS) and age group (< 3 years/ ≥ 3 years). Although implementing the algorithm in this way will result in a number of ‘abnormal’ breaths included in the analysis, this is not necessarily undesirable. We are interested in the spectral content of tidal flow waveforms without restricting ourselves only to stable breathing patterns.

Each participant typically had several tests performed at each follow-up occasion, which we refer to here as ‘replicate tests’, and each replicate test comprised many individual breath cycles. A power-law exponent was estimated from each breath cycle—identified using the aforementioned algorithm—using the OLS_{low} procedure as described in Chapter 5 (see panel (individual breaths) in Figure 6.2). Recall that, as described in Chapter 5, the OLS_{low} estimates of a power-law exponent from a single breath cycle were shown to have a substantial amount of irreducible error due to the properties of spectral density estimation, yet are unbiased. We therefore pooled all of the breath cycles across replicates for each individual at each age, and then averaged each such pool of breath cycles for each individual at each age (we show these averages joined over time in Figure 6.2). This approach discards information about the variation between replicate tests for individuals.

Means and standard deviations of the respiration rate (number of breaths per minute), number of breathing cycles per test, number of replicate tests per individual test occasion, and number of individuals by follow-up occasion and study are given in Table 6.1. The follow-up categories are defined by centred one-year intervals with closed lower bounds except for the youngest and oldest follow-up groups which were defined as aged less than 9 months and 5.5 years or older, respectively. A summary of the covariates of BIS MBW participants is provided in Table 6.2. For well populated groups gender was approximately in balance and age was approximately centred around the nominal follow-up time. We note that, in general, younger children had a larger number of breaths per test, due to faster respiratory rates. Although AREST CF participants overall and BIS infant participants had, in general, an average of at least 3 replicate tests on each follow-up occasion, the older BIS participants only had approximately 2 replicate tests on average.

FIGURE 6.1: Individual breath cycles identified within a randomly selected tidal flow waveform from AREST CF SF₆ (A) and N₂ MBW (B), and BIS SF₆ (C) and N₂ MBW (D). The left column shows an overlay of the individual breath cycles detected from the corresponding waveforms in the right column.

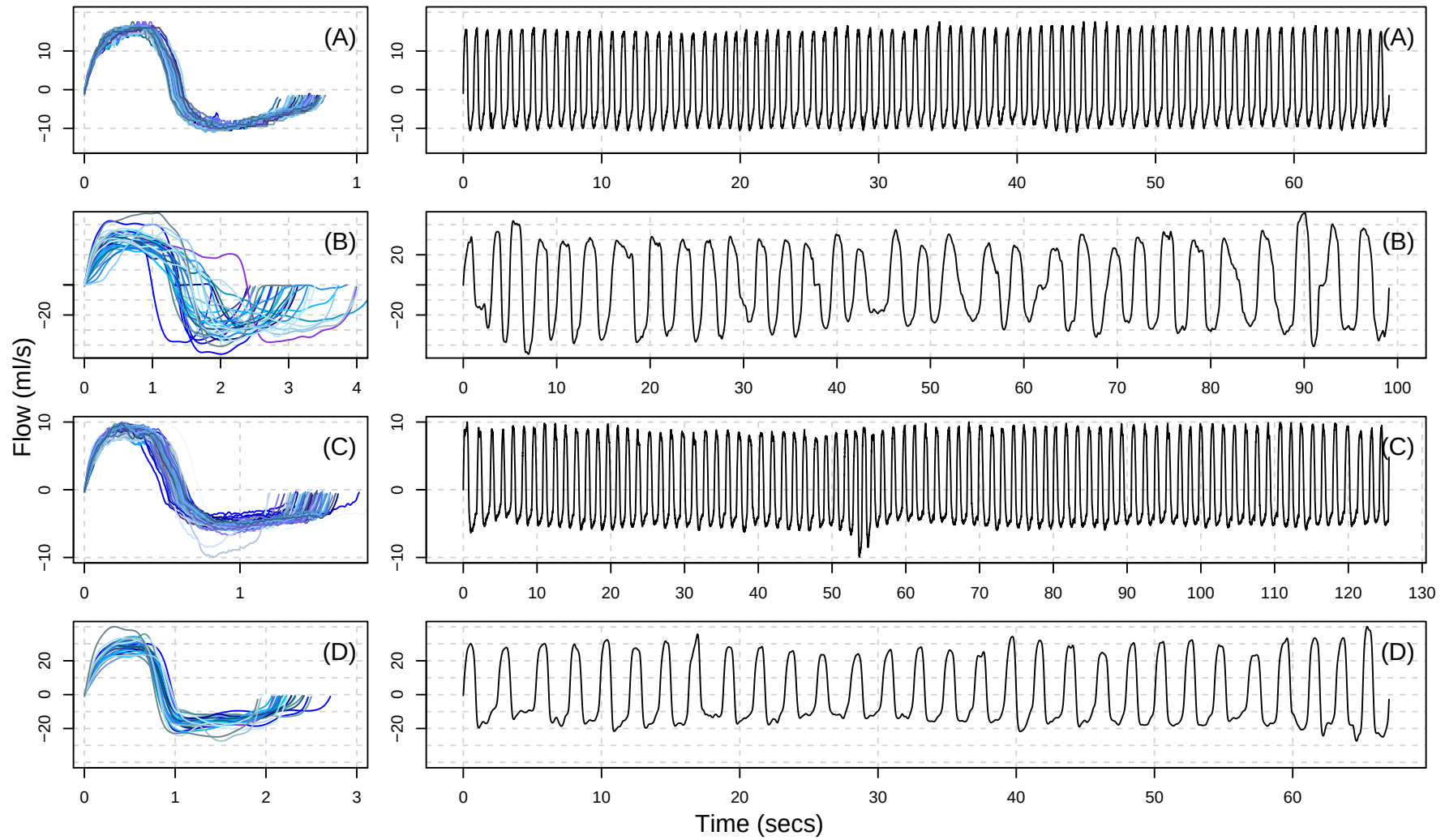


TABLE 6.1: Physiological characteristics of study participants. Continuous measures are summarised as means with standard deviations in parentheses. Repl.: replicate; RR: respiratory rate (number of breaths per minute).

Study	Follow-up	No. participants	Male (%)	Age (yrs)	No. repl. tests	No. breaths p/repl. test	RR
AREST-CF	12 w	45	55.6	0.3 (0.1)	3.1 (1.1)	82.9 (16.4)	34.6 (9.3)
	1 yr	47	46.8	1.1 (0.1)	3 (1.1)	84.2 (13.1)	60.8 (15)
	2 yr	40	60.0	2 (0.1)	3.1 (0.8)	75.8 (23.6)	50 (14.9)
	3 yr	5	40.0	3.2 (0.2)	2.6 (0.5)	49.5 (14.3)	30.3 (14.1)
	4 yr	16	56.2	4 (0.1)	3 (1.6)	42.4 (9.1)	26.9 (13.3)
	5 yr	17	58.8	5.1 (0.2)	3.4 (1.5)	41.6 (9.9)	30.2 (17.1)
	6+ yr	19	52.6	6 (0.1)	3.5 (1.8)	46.5 (10.3)	27.2 (14.2)
BIS	4 w	532	52.6	0.1 (0)	3.2 (1.7)	93.1 (16)	51.2 (13.8)
	4 yr	217	55.3	4.2 (0.1)	1.9 (0.9)	32 (5.1)	23.7 (10.2)
	5 yr	30	50.0	4.8 (0.3)	1.8 (0.8)	31.9 (5.6)	22.9 (11.7)
	6+ yr	2	0.0	5.5 (0)	1 (0)	32 (1.4)	22.1 (8.9)

TABLE 6.2: Characteristics of BIS MBW participants by follow-up period. Infant follow-up relates to 4 week follow-up, preschool follow-up relates to follow-up at 4 years and older (see Table 6.1). pT: preconception; T1: 1st trimester; T2: 2nd trimester; T3: 3rd trimester.

Characteristics	Infant follow-up	Preschool follow-up
n	435	249
Gas used (%)		
N ₂	0 (0.0)	249 (100.0)
SF ₆	435 (100.0)	0 (0.0)
Gender (%)		
Female	206 (47.4)	114 (45.8)
Male	229 (52.6)	135 (54.2)
Gestational age (%)		
32 to <37 weeks	19 (4.4)	7 (2.8)
37 to 42 weeks	416 (95.6)	242 (97.2)
>42 weeks	0 (0.0)	0 (0.0)
Birthweight (g) (%)		
1 to 1.995 kg	1 (0.2)	1 (0.4)
2 to 2.995 kg	57 (13.1)	22 (8.8)
3 to 3.995 kg	299 (68.7)	179 (71.9)
4 to 4.995 kg	77 (17.7)	46 (18.5)
5 to 5.995 kg	1 (0.2)	1 (0.4)
Mode of delivery (%)		
Vaginal	302 (69.4)	167 (67.1)
Emergency	76 (17.5)	46 (18.5)
Elective	57 (13.1)	36 (14.5)
Plurality (%)		
Singleton	431 (99.1)	249 (100.0)
Twin	4 (0.9)	0 (0.0)
Delivering hospital (%)		
Geelong Hospital	311 (71.5)	163 (65.5)
St John of God	124 (28.5)	86 (34.5)
Mother's age (%)		
Less than 20 yrs	6 (1.4)	1 (0.4)
20 to 29 yrs	151 (34.7)	79 (31.7)
30 to 39 yrs	261 (60.0)	161 (64.7)
40 to 49 yrs	17 (3.9)	8 (3.2)
Father's age (%)		
Less than 20 yrs	1 (0.2)	1 (0.4)
20 to 29 yrs	115 (27.4)	56 (23.2)
30 to 39 yrs	255 (60.7)	148 (61.4)
40 to 49 yrs	47 (11.2)	34 (14.1)
More than 50 yrs	2 (0.5)	2 (0.8)

Continued on next page

TABLE 6.2 (continued).

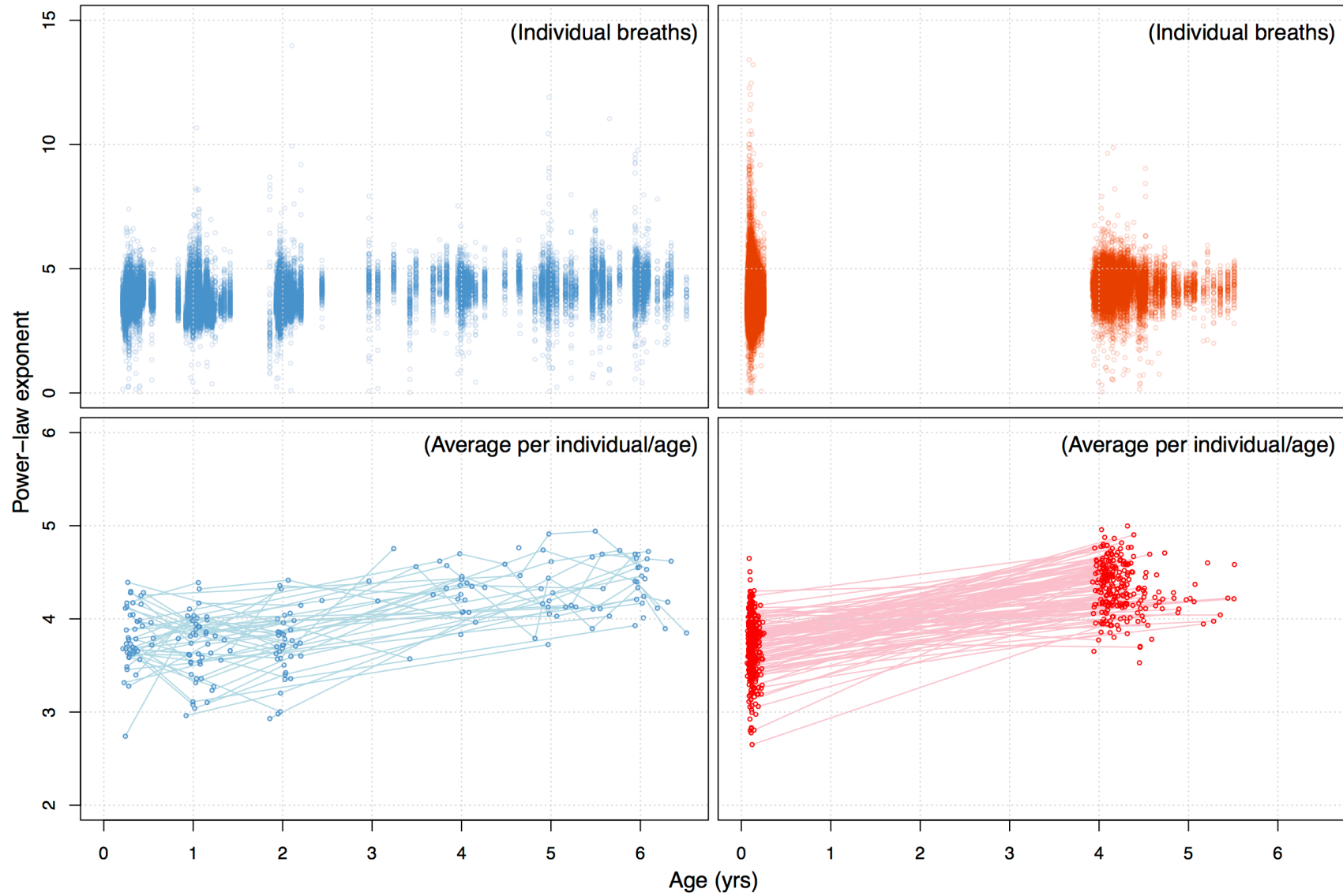
Characteristics	Infant follow-up	Preschool follow-up
Mother's highest education (%)		
Don't know	0 (0.0)	0 (0.0)
Declined to answer	0 (0.0)	0 (0.0)
Less than year 10	5 (1.2)	3 (1.2)
Year 10 or equivalent	19 (4.4)	11 (4.5)
Year 12 or equivalent	66 (15.2)	24 (9.7)
Trade/Cert/Dip	121 (27.9)	59 (23.9)
Bachelor degree	150 (34.6)	99 (40.1)
Postgraduate degree	73 (16.8)	51 (20.6)
Other	0 (0.0)	0 (0.0)
Father's highest education (%)		
Don't know	0 (0.0)	0 (0.0)
Declined to answer	0 (0.0)	0 (0.0)
Less than year 10	9 (2.1)	7 (2.9)
Year 10 or equivalent	25 (5.9)	16 (6.6)
Year 12 or equivalent	74 (17.5)	40 (16.5)
Trade/Cert/Dip	173 (40.9)	77 (31.7)
Bachelor degree	102 (24.1)	75 (30.9)
Postgraduate degree	40 (9.5)	28 (11.5)
Other	0 (0.0)	0 (0.0)
Number of siblings (%)		
No siblings	179 (41.1)	99 (39.8)
One sibling	167 (38.4)	93 (37.3)
Two siblings	68 (15.6)	41 (16.5)
Three or more siblings	21 (4.8)	16 (6.4)
Household yearly income (%)		
Unsure or declined to answer	10 (2.3)	6 (2.4)
Less than \$25k	11 (2.5)	8 (3.2)
\$25,000 to \$49,999	40 (9.2)	17 (6.8)
\$50,000 to \$74,999	78 (17.9)	38 (15.3)
\$75,000 to \$99,999	100 (23.0)	67 (26.9)
\$100,000 to \$149,999	149 (34.3)	85 (34.1)
More than \$150,000	47 (10.8)	28 (11.2)
Any maternal cigarette smoking in previous 12 months (%)		
No	363 (85.0)	223 (89.9)
Yes	64 (15.0)	25 (10.1)
Extent of maternal cigarette smoking (%)		
None	363 (85.2)	223 (89.9)
Light pT	21 (4.9)	9 (3.6)
Moderate pT	4 (0.9)	1 (0.4)
Light pT or T1	10 (2.3)	5 (2.0)
Mod pT or T1	5 (1.2)	1 (0.4)
Heavy pT,T1,T2,T3	23 (5.4)	9 (3.6)

Continued on next page

TABLE 6.2 (continued).

Characteristics	Infant follow-up	Preschool follow-up
Exposure to passive smoke in previous 12 months (%)		
No	327 (80.7)	202 (85.6)
Yes	78 (19.3)	34 (14.4)
Household pets (%)		
No	101 (23.3)	75 (30.2)
Yes	332 (76.7)	173 (69.8)
Household livestock (%)		
No	399 (92.8)	224 (91.4)
Yes	31 (7.2)	21 (8.6)
First degree family history of asthma (%)		
No	197 (46.4)	114 (46.5)
Yes	228 (53.6)	131 (53.5)
First degree family history of eczema (%)		
No	224 (52.8)	128 (52.7)
Yes	200 (47.2)	115 (47.3)
First degree family history of hayfever (%)		
No	152 (36.0)	75 (30.4)
Yes	270 (64.0)	172 (69.6)

FIGURE 6.2: Power-law exponents by age from AREST CF (blue) and BIS (red) data. Shown are individual breaths and means per individual and age combination.



6.2.2 Models

Model of age and cystic fibrosis

To address the first objectives stated in Section 6.1.1, we modelled the relationship between the power-law exponent and age linearly for both children with and without cystic fibrosis, allowing for some variation in the intercept and slope of the linear relationship across individuals.

For i indexing the sequence of follow ups attended an individual, and j indexing the individual, we modelled the ‘observed’ (estimated by the averaging process described in Section 6.2.1) power-law exponent from MBW tests only, denoted y_{ij} , as a function of age in years, denoted age_{ij} , and cystic fibrosis status, denoted cf_j , where $\text{cf}_j = (0, 1)$ according to whether an individual did or did not have cystic fibrosis, respectively.

$$y_{ij} = (\beta_0 + \theta_0 \text{cf}_j + b_{0j}) + (\beta_1 + \theta_1 \text{cf}_j + b_{1j}) \text{age}_{ij} + e_{ij} \quad (6.1)$$

In the model described by (6.1), β_0 represents the intercept and modelled value of the BIS individuals at birth, β_1 the linear effect of age, and θ_0 and θ_1 model the differences in the baseline mean and linear effect of age for individuals with cystic fibrosis. The b_{0j} and b_{1j} are random intercept and slope effects for the j -th individual, respectively, and e_{ij} is the residual error term with $e_{ij} \sim N(0, w_{ij}\sigma_e^2)$. The $w_{ij} = \bar{n}/n_{ij}$ is a normalised weighting factor to account for the different number of breaths n_{ij} used to produce the averages y_{ij} for each individual at each follow-up visit. To capture any correlation in the individual-level intercepts and slopes we specify the random effects b_{0j} and b_{1j} to follow a bivariate normal distribution:

$$\begin{pmatrix} b_{0j} \\ b_{1j} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{b0}^2 & \rho\sigma_{b0}\sigma_{b1} \\ \rho\sigma_{b0}\sigma_{b1} & \sigma_{b1}^2 \end{pmatrix} \right], \quad (6.2)$$

where σ_{b0} and σ_{b1} denote the standard deviations of the random effects with correlation parameter ρ . As usual, we assume that ρ is independent of the error term distribution.

Model of BIS power-law exponents and covariates

To address the second objective stated in Section 6.1.1, we modelled the power-law exponent ‘observed’ at follow-up (estimated by the averaging process described in Section 6.2.1) from MBW tests only, denoted y_{1j} , conditional on the observed power-law exponent at baseline, denoted y_{0j} , as a function of a covariate with K categories with a reference category captured by the intercept and the remaining $K - 1$ levels captured by x_{jk} .

$$y_{1j} = \beta_0 + \sum_{k=1}^{K-1} \theta_k x_{jk} + (\beta_1 + \sum_{k=1}^{K-1} \delta_k x_{jk}) y_{0j} + e_j. \quad (6.3)$$

In the model described by (6.3), β_0 represents the intercept and mean value of the reference category, β_1 the linear effect of the observed baseline power-law exponent. For the remaining $K - 1$ categories, θ_k models the mean changes in intercept and δ_k models the mean changes in the linear effect of observed baseline power-law exponent. The residual error term is normally distributed $e_j \sim N(0, w_{1j} \sigma_e^2)$, where $w_{1j} = \bar{n}/n_{1j}$ is a normalised weighting factor to account for the different number of breaths n_{1j} used to produce the averages y_{1j} for each individual at the preschool follow-up visit. This analysis is analogous to a traditional analysis of covariance (ANCOVA), although we have extended the standard ANCOVA model to include an interaction component.

Model of agreement

To address our third objective as stated in Section 6.1.1—to examine differences between power-law exponents estimated from either SF₆ MBW or non-MBW tidal flow recordings—we used a variance components model similar to that used in Chapter 4 and also Carstensen et al. [119, 120], herein referred to as the ‘agreement model’:

$$\begin{aligned} y_{mp} &= \alpha_m + u_p + e_{mp}, \\ u_p &\sim N(0, \gamma^2), \quad e_{mp} \sim N(0, \sigma_m^2) \end{aligned} \quad (6.4)$$

where y_{mp} denotes the the ‘observed’ (estimated by the averaging process described in Section 6.2.1) power-law exponent for method m and study participant p at the initial BIS follow-up, α_m is the mean effect for method m , the variation between participants is captured by γ , and the residual variation is captured by σ_m .

The parameters of (6.4) were used to calculate 95% limits of agreement (LoA) [121–123]:

$$\text{LoA} = (\alpha_1 - \alpha_2) \pm 2\sqrt{\sigma_1^2 + \sigma_2^2}. \quad (6.5)$$

The LoA and mean difference were superimposed on a Bland-Altman plot [121, 122]. This plots the difference between two methods on a vertical axis and the mean of the two methods on the horizontal axis, allowing us to clearly assess agreement between the two methods and whether systemic differences in the methods exist in relation to the magnitude of the measurements.

6.2.3 Estimation

We took a Bayesian approach to estimating the parameters of the models described by (6.1) and (6.2), and (6.4). The motivation for the Bayesian approach is that relatively complex data and correlation structures—such as unbalanced longitudinal structures and correlated random effects—could be easily estimated in Stan, and parameter inference could be summarised intuitively in terms of probabilities. The models described by (6.3) were relatively simple and were estimated using weighted least squares implemented by the `lm()` function in the R software environment.

The Bayesian approach required that we assign prior probability distributions to our model parameters. We assigned diffuse prior distributions, which contain only vague information based on the potential scale and theoretical support of a parameter, as often recommended [69]. For the model described by (6.1) we assigned prior distributions $\beta_0, \beta_1, \theta_k, \delta_k \sim \text{N}(0, 10)$, $L_b \sim \text{LKJ}(1)$, where L_b is the Cholesky decomposition of the 2×2 correlation matrix containing ρ and LKJ is a useful prior distribution described by Lewandowski et al. [124] and the Stan Development Team [69], and $\sigma_{b0}, \sigma_{b1}, \sigma_e \sim \text{half-Cauchy}(0, 2.5)$. These types of priors are commonly recommended and found in practice [69, 70, 125]. For the model described by (6.4), we assigned similarly diffuse priors: $\alpha_m \sim \text{N}(0, 10)$ and $\gamma, \tau, \sigma_m \sim \text{half-Cauchy}(0, 2.5)$.

We sampled from the joint posterior distribution of the parameters using Stan [69], called from the R software environment using RStan [126] (see Appendix A). In each case, a number of MCMC chains were run in parallel comprising a discarded ‘warm-up’ phase and a ‘sampling’ phase, the latter being used for parameter inference.

The number of MCMC chains, iterations, and the thinning factor used differed for each model depending on how well the parameter chains converged in preliminary analyses. Convergence of the chains was assessed both graphically, via visual inspection of the chains, and statistically, by ensuring that the scale reduction factor $\hat{R}$ was between 1 and 1.1 for each parameter [73]. We ran 8 parameter chains for the model described in (6.1) for 2,000 warm-up iterations followed by 2,000 sampling iterations, which were then thinned by a factor of 4. In contrast, there was little autocorrelation in the parameter chains for the models described by (6.4), we therefore ran 4 chains for 1,000 warm-up iterations followed by 1,000 iterations with no thinning. In all cases the resulting posterior distribution contained 4,000 samples for each parameter.

To guide Stan’s Hamiltonian Monte Carlo (HMC) sampler away from areas of high posterior curvature, which are difficult to sample from, random components were given a unit normal prior distribution that was then rescaled by the relevant variance component; for example, with a convenience parameter $\gamma^* \sim N(0, 1)$ and $\gamma = \gamma^* \sigma_\gamma^2 \implies \gamma \sim N(0, \sigma_\gamma^2)$, as described in [127]. This rescaling of the random components is often necessary to achieve parameter convergence using Stan’s HMC sampler. Note that the LKJ prior specification corresponds to a multivariate non-centred parameterisation [127]. Posterior medians and 95% credible intervals (CI) were chosen to describe the posterior distributions. All other data management and analyses were performed using R [72]. Where non-Bayesian analyses were performed, parameters were summarised by estimated means and confidence intervals (also denoted CI).

Sensitivity analysis

As mentioned in Section 6.1.2 a maximum of only two time points were available for any BIS participant. A random effects model for the BIS data alone was thus inappropriate as the individual effects and residuals could not be properly estimated because the intercept, slope and residual variance components would ‘compete’ with each other and the prior distribution. Our approach was, as (6.1) describes, to combine the AREST CF and BIS data and specify a common intercept, slope, and residual standard deviation for both study data. In principle, this approach allows information about the residual variation to be ‘shared’ between studies. It was of interest to understand how sensitive the estimates would be to a prior on the residual standard deviation that was more informative

than the half-Cauchy prior, with a probability mass centred around some value that was either considerably less than or greater than what was expected given the scale of the data. Understanding the sensitivity of the estimates to substantially different prior distributions provided a gauge as to whether the estimation procedure, particularly of the main effects, was being unduly influenced by the prior choice, and thus provided a way to characterise the strength of the estimation procedure given the data limitations. We therefore estimated the model described by (6.1) and (6.3) using three prior scenarios: $\sigma_e \sim \text{half-Cauchy}(0, 2.5)$, $\sigma_e \sim \text{logN}(-4, 0.1)$, and $\sigma_e \sim \text{logN}(1.5, 0.5)$ with the latter two distributions, described as sensitivity prior 1 and sensitivity prior 2, having means of approximately 0.018 and 5.1 and standard deviations of 0.002 and 2.7, respectively. The parameters of the latter two prior distributions were chosen to ensure that, given the scale of the data, the prior probability mass on the residual error was concentrated approximately close to zero, in the case of sensitivity prior 1, or distributed more broadly, from zero to some moderately large number, in the case of the sensitivity prior 2. These two sensitivity priors therefore provided us with two scenarios (out of many alternatives). Sensitivity prior 1 represents a rather extreme scenario where the prior assumption is that the residuals are almost certainly very small, and sensitivity prior 2 represents an alternative where the prior assumption is that the residuals are more likely to be quite large.

Posterior predictive distributions (see Appendix A for a general description) for each scenario were produced to assess whether the data generated from the estimated model were representative of the observed data and thus the appropriateness of the model.

6.3 Results

6.3.1 Power-law spectra

Figure 6.3 provides an example of individual breath cycles, their periodograms, and their power-law exponents estimated using the OLS_{low} , OLS_{mid} and OLS_{high} estimators that were described in Chapter 5. In this example, the spectral density appears to be well-modelled by a power-law, after accounting for the noise floor that appears evident for most of the estimated spectral densities. This effect was observed in several similar randomly selected subsets, and reflects results reported by Frey et al. [33] and Habib et al. [35], thus

supporting the notion that the spectral densities of tidal flow waveforms tend to follow a power-law.

6.3.2 Model of age and cystic fibrosis

The linear random effects model, described by (6.1) and (6.2), was fitted to power-law exponents (estimated by the averaging process described in Section 6.2.1) from the MBW tidal flow recordings of AREST CF and BIS study participants. The posterior distributions of the model parameters fitted using either AREST CF data only, BIS data only, or by combining both datasets, are summarised in Table 6.3. A single sample of the posterior predictive distribution generated by each model described in Table 6.3 is shown in Figures 6.4 and 6.5.

Parameter convergence was sufficient for all model parameters, with $\hat{R}$ values mostly between 1 and 1.03 with none larger than 1.13, and adequate effective sample sizes given the number of chains used. Upon modifying the sampler's target acceptance rate from 0.8 to 0.99 any divergent transitions, which are the result of numerically exploring extreme posterior curvature and are known to lead to biased inference, were avoided in all cases with the exception of that using BIS data only and diffuse priors (where there were 10 divergent transitions, which we judged did not materially affect the overall results).

The main difference between the models estimated using either the AREST CF data only or the BIS data only was in the estimated residual standard deviations of 0.29 (CI: 0.25, 0.33) and 0.05 (CI: 0.04, 0.06), respectively. It is unlikely we can infer a great deal from the BIS data only results. Considering for the BIS data there was only a maximum of two time points per individual—a pattern of data that could be consistent with either minimal intercept and slope variation and large random error, or large intercept and slope variation and minimal random error—without using an strongly informative prior, we could not produce predictable estimates.

By using the combined data to estimate the model, in which both the AREST CF and BIS groups shared common variance components, we obtained an estimate of the residual standard deviation that was somewhere between that estimated from either AREST CF data alone or BIS data alone. Although such an estimate has its deficiencies, note that when only the BIS data were used, the posterior predictive distributions did not reflect the

uncertainty around the data in ages 3 and above where, on average, individuals had many fewer breaths available to estimate power-law exponents with than younger children (Figures 6.4 and 6.5). The model using the combined data appeared to allow the uncertainty due to the number of observed breath cycles to be better demonstrated in the posterior predictive distributions for the BIS data.

Importantly, our sensitivity analysis shows that our main model estimates $\beta_0, \beta_1, \theta_0$, and θ_1 were similar when using either the AREST CF data only, the BIS data only, or the combined data. Further, the main model estimates remained similar across the different prior scenarios. The variance components did not exhibit similar stability across the different scenarios. For example, in the case of AREST CF data only, sensitivity prior 1 appeared to halve the residual standard deviation with a corresponding increase in the intercept and slope standard deviation whereas sensitivity prior 2 did not affect estimates in any meaningful way, although this result was not unexpected given the nature of the priors. For BIS data only, the sensitivity priors did not have much affect, which given the unidentifiability of the variance components was also not surprising. The parameters of the model are fully identified when using the AREST CF data only, and if we take the corresponding results as a reference set, we can see that the variance components when using the combined data are not substantially different, with the exception of the the residual variation which appears to be some sort of compromise between the estimate using either AREST CF only or BIS data only. Thus we can be reasonably confident in the estimates of the main model parameters, and possibly also in the estimates of the intercept and slope standard deviations, particularly when using diffuse priors, however only in the case of AREST CF data only can we have confidence in the estimates of the residual standard deviation.

The estimated intercept term suggests that at birth children either with or without cystic fibrosis have quite complex tidal flow waveforms which are decidedly non-sinusoidal in shape (see Figure 5.5 for simulated examples). The strong positive estimated age effect suggests that as the respiratory system develops with age, the tidal flow waveforms become less complex and more sinusoidal in shape. For children with cystic fibrosis this loss of complexity is slightly more moderate. The fairly broad distribution of individuals around the intercept and slope suggests that there is substantial population variation in the development of tidal flow waveforms in both infancy and early childhood. Further, the residual standard deviation estimated from AREST CF data only, which provides

a good indication of the the remaining within-subject variation after accounting for all other effects, was approximately equivalent to the standard deviation across individuals, suggesting that the power-law exponent is highly variable, similar to other summaries of lung function (see Chapter 3 for similar results for MBW outcomes).

We note that although our results suggest, similar to Frey et al. [33] and Habib et al. [35], that infant tidal flow spectra can be well-modelled by a power-law, there are noticeable differences. In our case the estimated intercept, 3.7 (CI: 3.67, 3.73), was substantially lower than that of Frey et al. [33], who estimated the mean for healthy 1 month old infants at 4.24 (CI: 4.14, 4.34). Frey et al. [33] also did not find any significant change in the power-law exponents of healthy infants between 1 month and 12 months whereas we have estimated a strong age effect. However, we suspect that by modelling the linear effect of age across the range of 1 month to 6 years that the our estimate is driven by data from older children (see Figure 6.2). Furthermore, where Frey et al. [33] estimated lower mean values of the power-law exponent in infants and children with CLD or wheezing disorders, we did not observe a similar effect for children with cystic fibrosis or any covariate for that matter. Oddly, our mean result for BIS infants most closely reflected those of Habib et al. [35] for VLBW babies (approximately 1.2 kg at birth) i.e. 3.73 (CI: 3.62, 3.85) or 3.51 (CI: 3.46, 3.66) depending on method used, however we cannot compare their results to similarly low birthweight BIS babies because we did not have enough data to provide meaningful inference for this category. Note that the results presented from Frey et al. [33] and Habib et al. [35] are estimated means and confidence intervals, as opposed to summaries of the posterior distribution.

Apparent differences between studies could be due to a number of reasons but they are likely due to either study design, for example quite different testing procedures and measurement devices, or estimation procedures, for example differences in spectral density or power-law estimation. We also cannot discount the potential confounding effect of the differences in study design between the AREST CF and BIS studies, and between the age groups within each study, as described in detail in the section on limitations.

FIGURE 6.3: Power-law exponents estimated from individual breath cycles for three randomly selected breaths by study and gas: (A) AREST CF/SF₆, (B) AREST CF/N₂, (C) BIS/SF₆, and (D) BIS/N₂. Shown, side-by-side for each group, are the tidal flow waveforms (y-axis: Flow (l/s)) and periodogram (y-axis: Power (log)) with fitted power-law exponent functions and corresponding power-law exponents: OLS_{low} (red), OLS_{mid} (blue), and OLS_{high} (green). Note the ‘smoothing’ of the higher frequencies for the N₂ MBW data (this is done automatically by the software with no explanation, see Section 2.4).

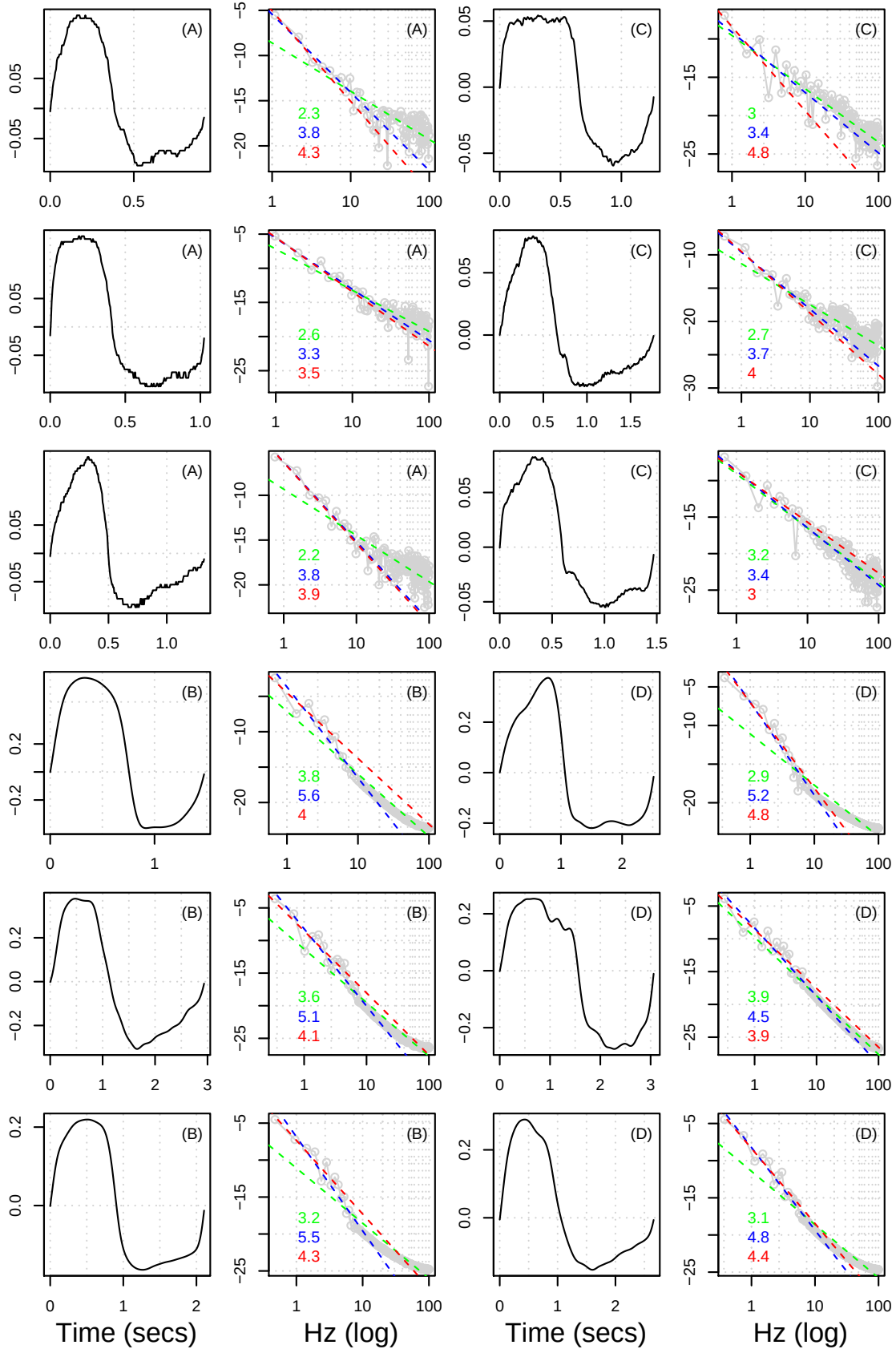


TABLE 6.3: Summary of posterior distributions and convergence diagnostics. Sensitivity analysis.

Model	Parameter	Diffuse priors		Sensitivity prior 1		Sensitivity prior 2	
		Median (95% CI)	$\hat{R}$	Median (95% CI)	$\hat{R}$	Median (95% CI)	$\hat{R}$
AREST CF only	Intercept (β_0)	3.64 (3.55, 3.73)	1.00	3.63 (3.53, 3.73)	1.01	3.64 (3.54, 3.73)	1.00
	Age (β_1)	0.12 (0.08, 0.15)	1.00	0.1 (0.06, 0.15)	1.01	0.12 (0.08, 0.15)	1.00
	Intercept SD (σ_{b0})	0.3 (0.22, 0.39)	1.00	0.39 (0.33, 0.48)	1.00	0.29 (0.2, 0.38)	1.00
	Age SD (σ_{b1})	0.06 (0.02, 0.11)	1.01	0.15 (0.12, 0.19)	1.00	0.06 (0.02, 0.1)	1.00
	Correlation (ρ)	-0.76 (-0.96, -0.26)	1.00	-0.67 (-0.8, -0.49)	1.00	-0.76 (-0.96, -0.21)	1.00
	Residual SD (σ_e)	0.29 (0.25, 0.33)	1.00	0.14 (0.13, 0.15)	1.01	0.3 (0.27, 0.35)	1.00
BIS only	Intercept (β_0)	3.69 (3.67, 3.72)	1.10	3.69 (3.66, 3.72)	1.07	3.69 (3.66, 3.72)	1.04
	Age (β_1)	0.15 (0.14, 0.17)	1.04	0.15 (0.14, 0.16)	1.03	0.16 (0.15, 0.17)	1.01
	Intercept SD (σ_{b0})	0.3 (0.28, 0.32)	1.03	0.3 (0.28, 0.32)	1.07	0.29 (0.27, 0.32)	1.03
	Age SD (σ_{b1})	0.09 (0.07, 0.1)	1.01	0.09 (0.08, 0.1)	1.03	0.08 (0.07, 0.09)	1.01
	Correlation (ρ)	-0.73 (-0.8, -0.63)	1.02	-0.72 (-0.78, -0.63)	1.03	-0.77 (-0.84, -0.67)	1.01
	Residual SD (σ_e)	0.03 (0.02, 0.04)	1.08	0.02 (0.02, 0.02)	1.01	0.05 (0.04, 0.06)	1.02
Both studies	Intercept (β_0)	3.7 (3.67, 3.73)	1.00	3.7 (3.67, 3.73)	1.01	3.7 (3.67, 3.73)	1.00
	Age (β_1)	0.16 (0.14, 0.17)	1.00	0.16 (0.14, 0.17)	1.00	0.16 (0.14, 0.17)	1.00
	CF (θ_0)	-0.07 (-0.15, 0.01)	1.00	-0.07 (-0.15, 0.01)	1.00	-0.07 (-0.15, 0.01)	1.00
	CF:Age (θ_1)	-0.04 (-0.07, -0.01)	1.00	-0.05 (-0.08, -0.02)	1.00	-0.04 (-0.07, -0.01)	1.00
	Intercept SD (σ_{b0})	0.27 (0.25, 0.3)	1.00	0.29 (0.27, 0.31)	1.00	0.27 (0.25, 0.29)	1.00
	Age SD (σ_{b1})	0.07 (0.05, 0.09)	1.00	0.09 (0.08, 0.1)	1.01	0.07 (0.05, 0.09)	1.00
	Correlation (ρ)	-0.75 (-0.86, -0.63)	1.00	-0.72 (-0.8, -0.62)	1.00	-0.76 (-0.87, -0.63)	1.00
	Residual SD (σ_e)	0.17 (0.15, 0.18)	1.00	0.13 (0.12, 0.13)	1.00	0.17 (0.16, 0.19)	1.00

6.3.3 Model of BIS power-law exponents and covariates

The model of follow-up power-law exponents between covariate categories, conditional on observed baseline power-law exponents, described by (6.3), was fitted to power-law exponents estimated from the MBW tidal flow recordings of BIS study participants. Figure 6.6 plots the observations by categories for each covariate, along with their associated model fits for (6.3) assuming either a common slope parameter or a linear covariate/slope interaction. Detailed summaries of the model fits are located in Appendix F. We assessed whether the inclusion of each covariate independently of one another explained more variation than the baseline model without any additional covariates by an analysis of variance implemented using R's `add1` function (see Appendix Table F.20 for detailed results).

There was little evidence suggesting any major differences between covariate categories. The models that included interaction terms appeared to be overfitting some of the categories, and thus this model specification was perhaps not an appropriate one. There is no evidence to suggest that the addition of any covariates to the model explains more variation than the model without covariates. While it may be tempting to conclude from Appendix Table F.20 that the model including the number of siblings is an improvement, if we look at the fits for each category there is no clear dose-response relationship and it seems likely that the category of three or more siblings is different simply due to chance. Furthermore, to perform this analysis we required complete cases, which reduced our data to 105 observations. This meant that often only a handful of observations, if any, were available for many of the categories within covariates. It is unlikely that if differences do exist, we could estimate them precisely enough using these data.

FIGURE 6.4: AREST CF data and posterior predictive distributions (PPD) from models using either BIS data only, or the combined AREST/BIS data, and the three different prior scenarios—either the diffuse or the two ‘sensitivity’ priors on the residual standard deviation—are used.

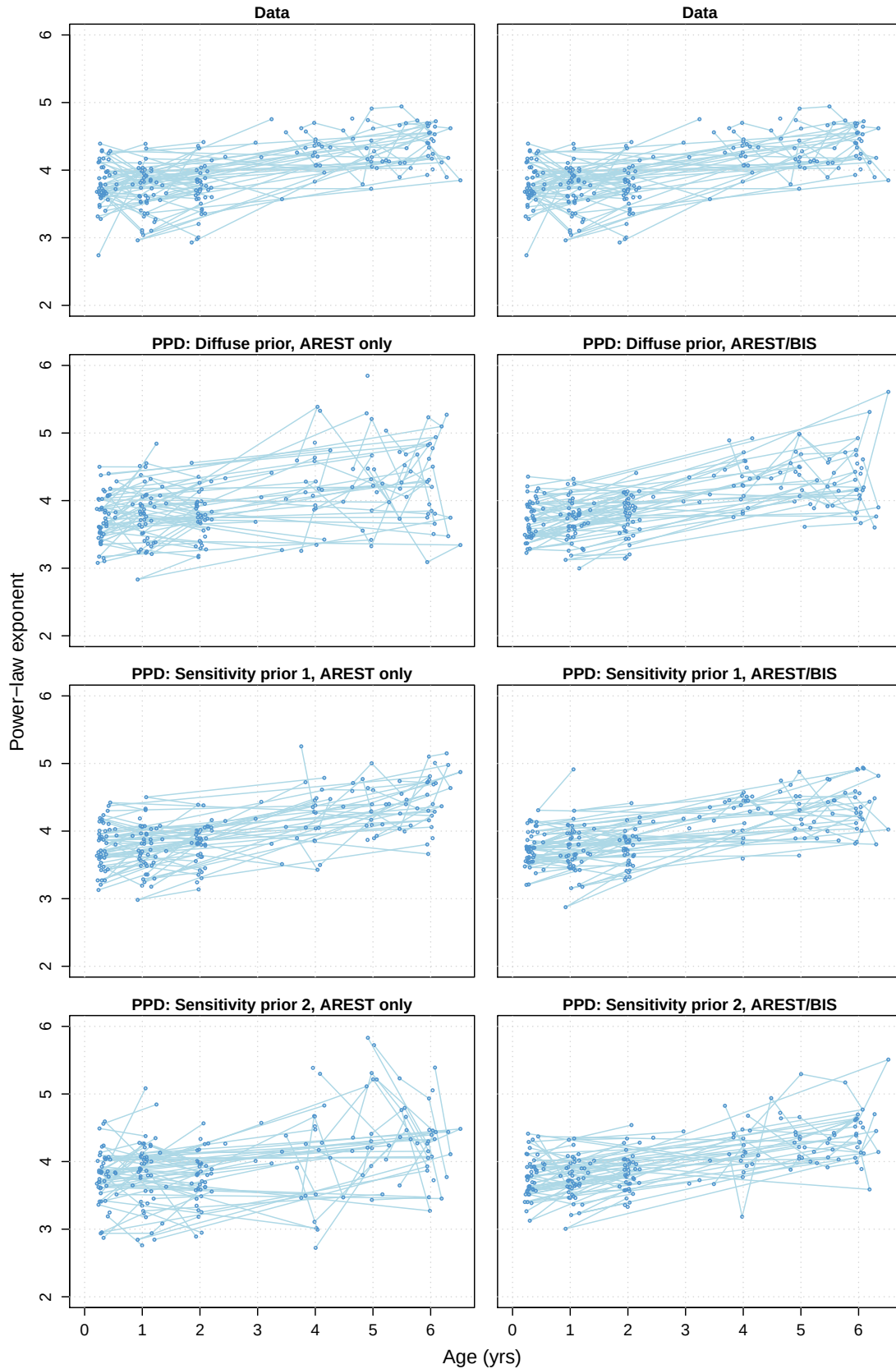


FIGURE 6.5: BIS data and posterior predictive distributions (PPD) from models using either BIS data only, or the combined AREST/BIS data, and the three different prior scenarios—either the diffuse or the two ‘sensitivity’ priors on the residual standard deviation—are used.

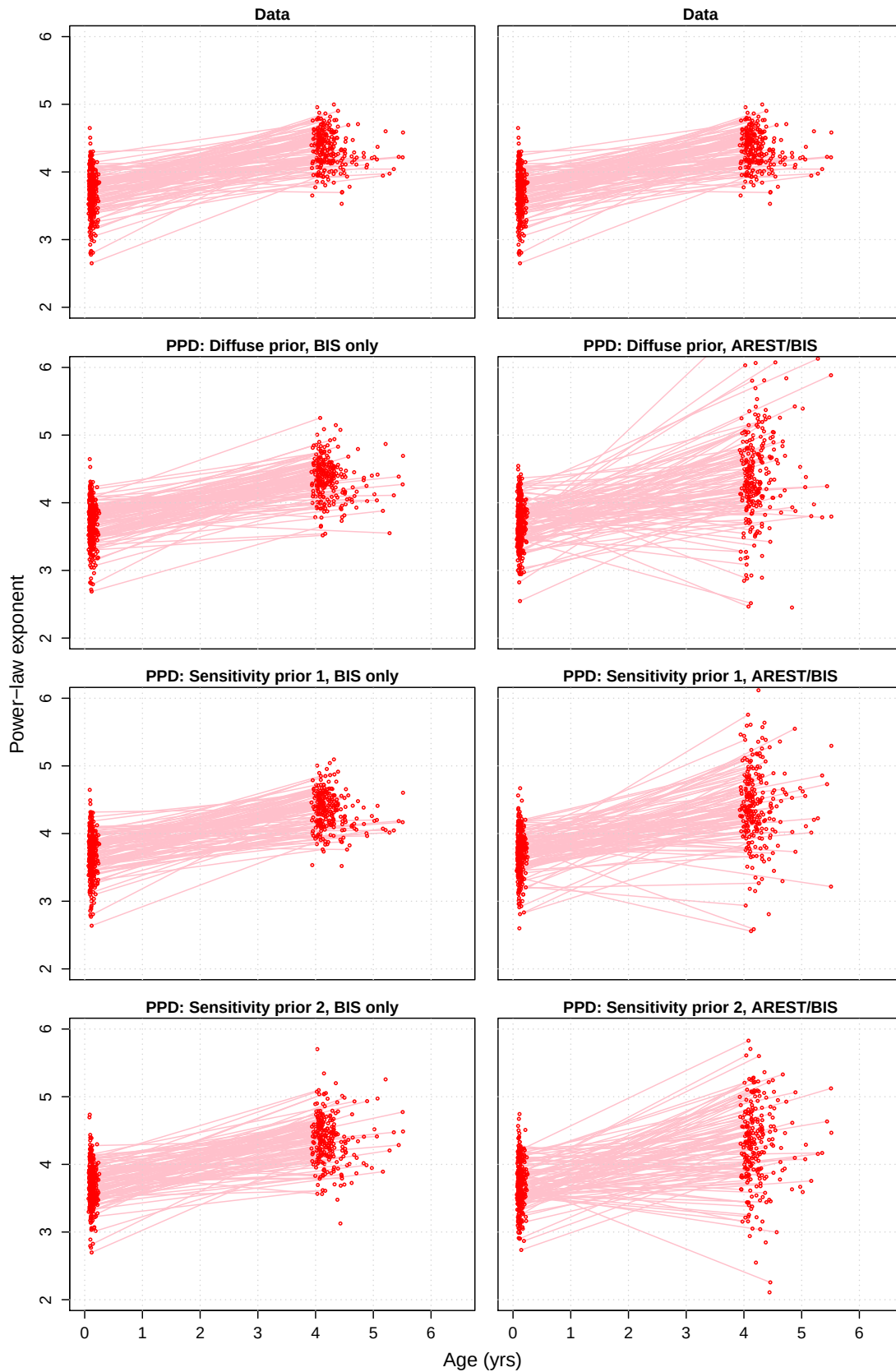


FIGURE 6.6: Scatterplot of power-law exponents (PLE) estimated from BIS participants at baseline (infancy) and follow-up (preschool) ($n = 105$). Model fits shown as solid lines with shaded 95% confidence intervals. Left columns correspond to models with a common slope parameter, right columns correspond to models assuming a slope/covariate interaction.

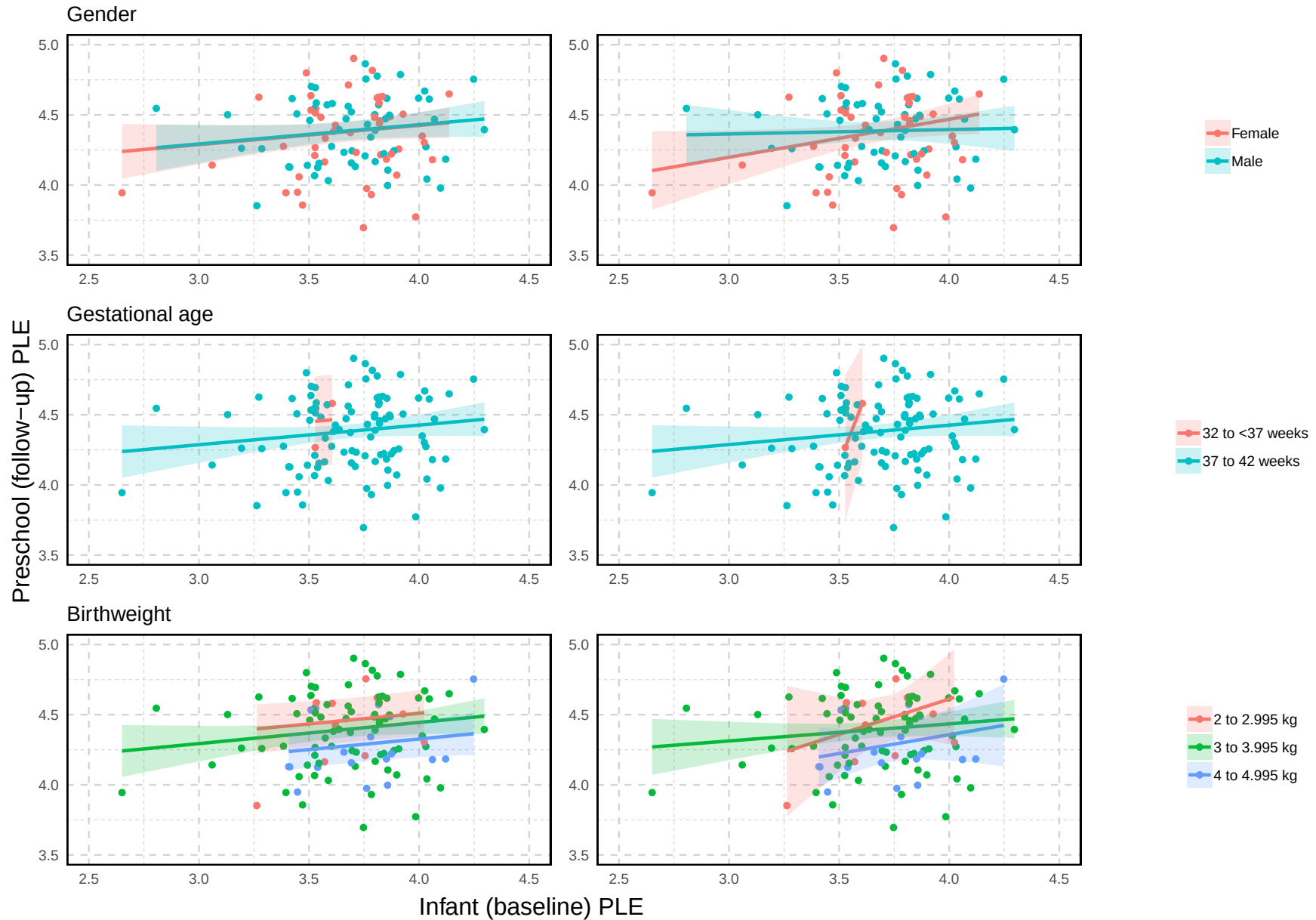


FIGURE 6.6: (continued).

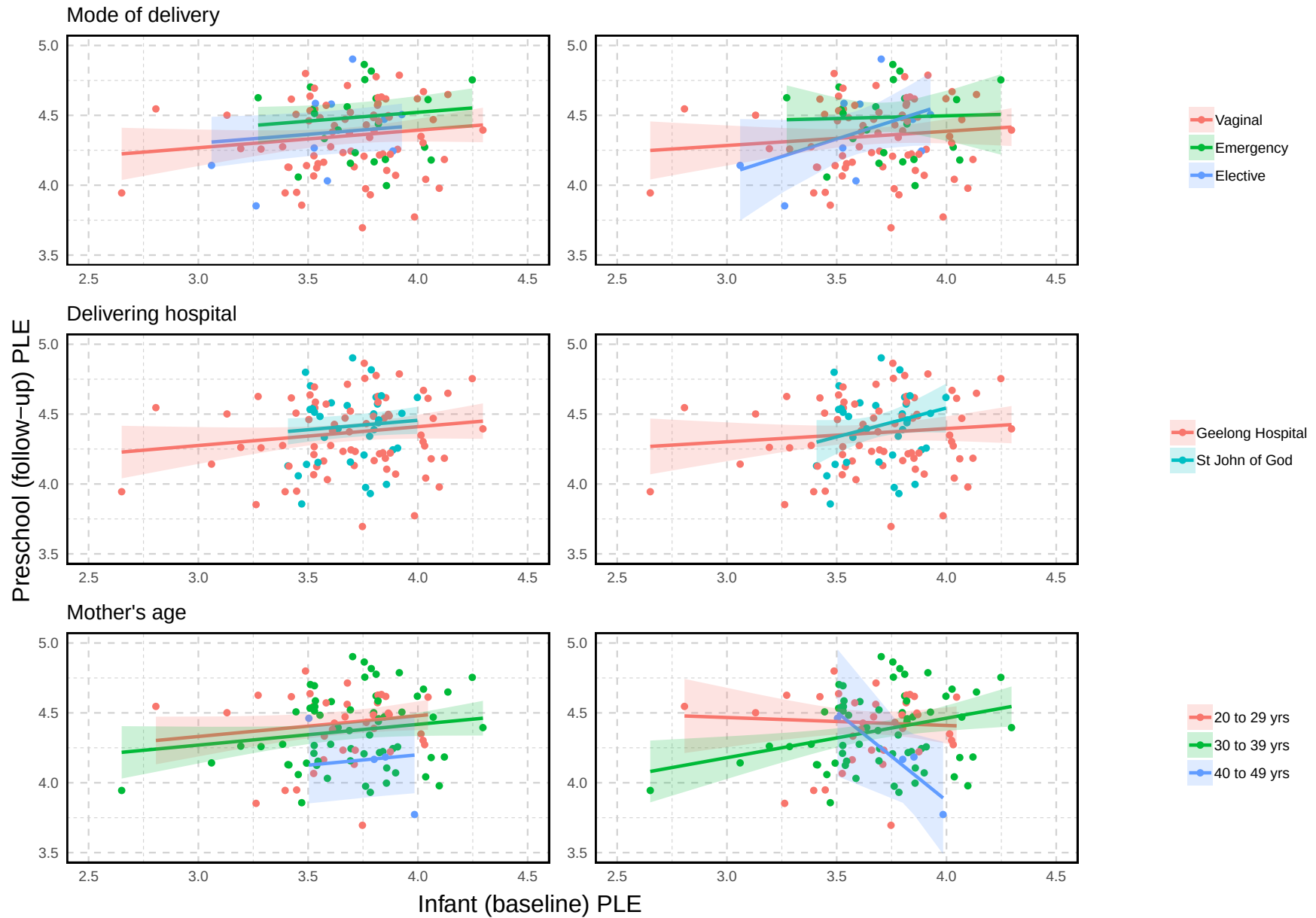


FIGURE 6.6: (continued).

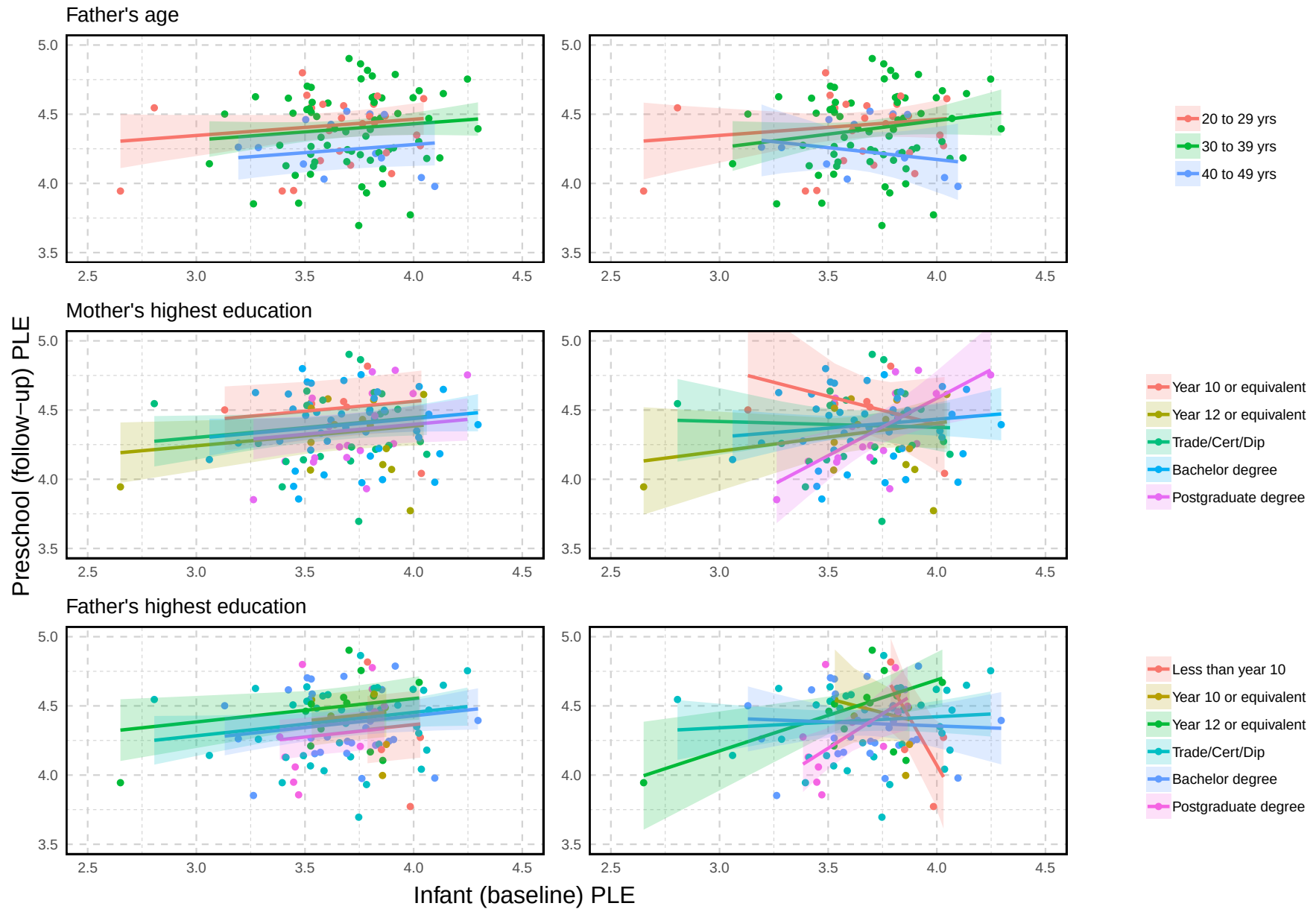


FIGURE 6.6: (continued).

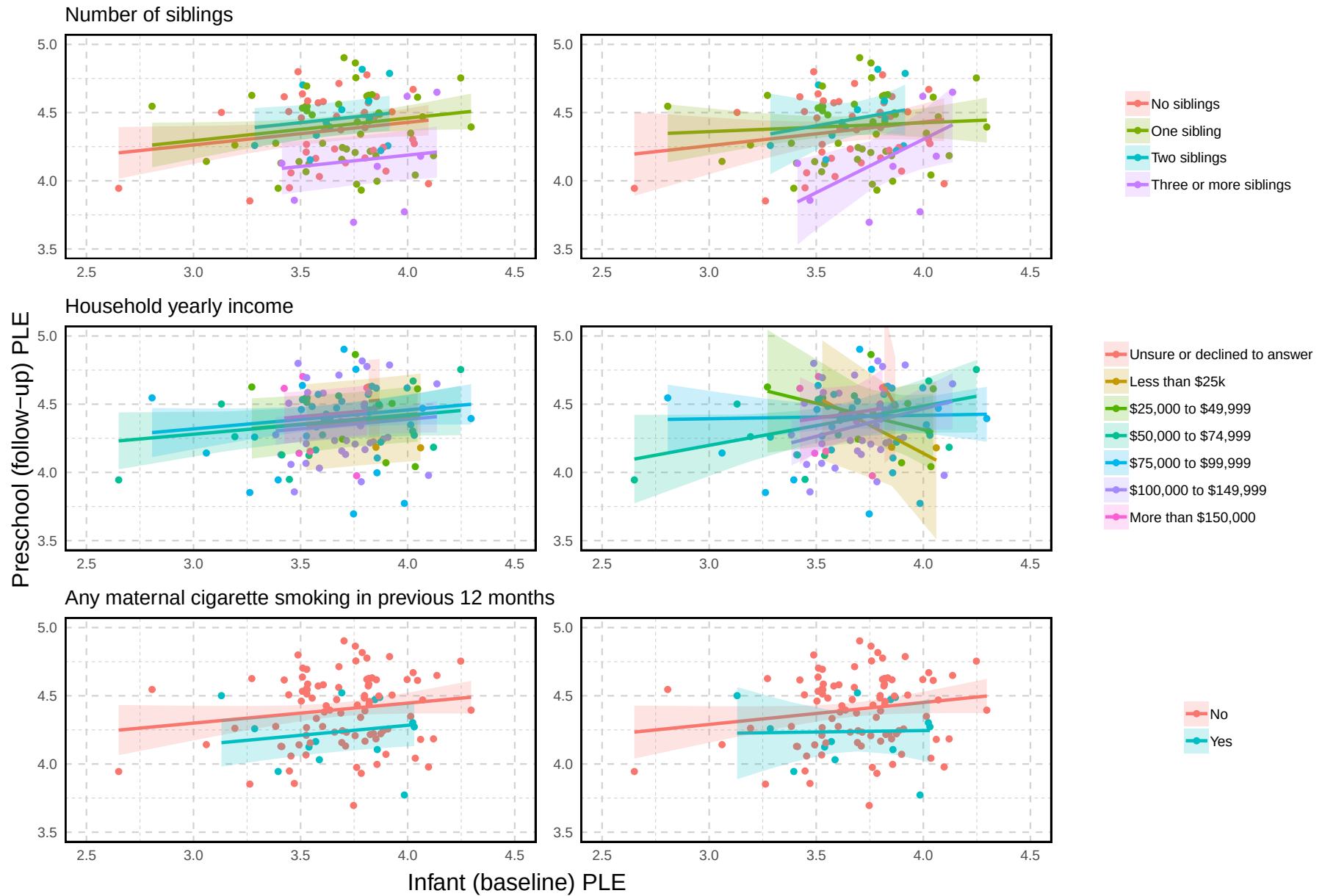


FIGURE 6.6: (continued).

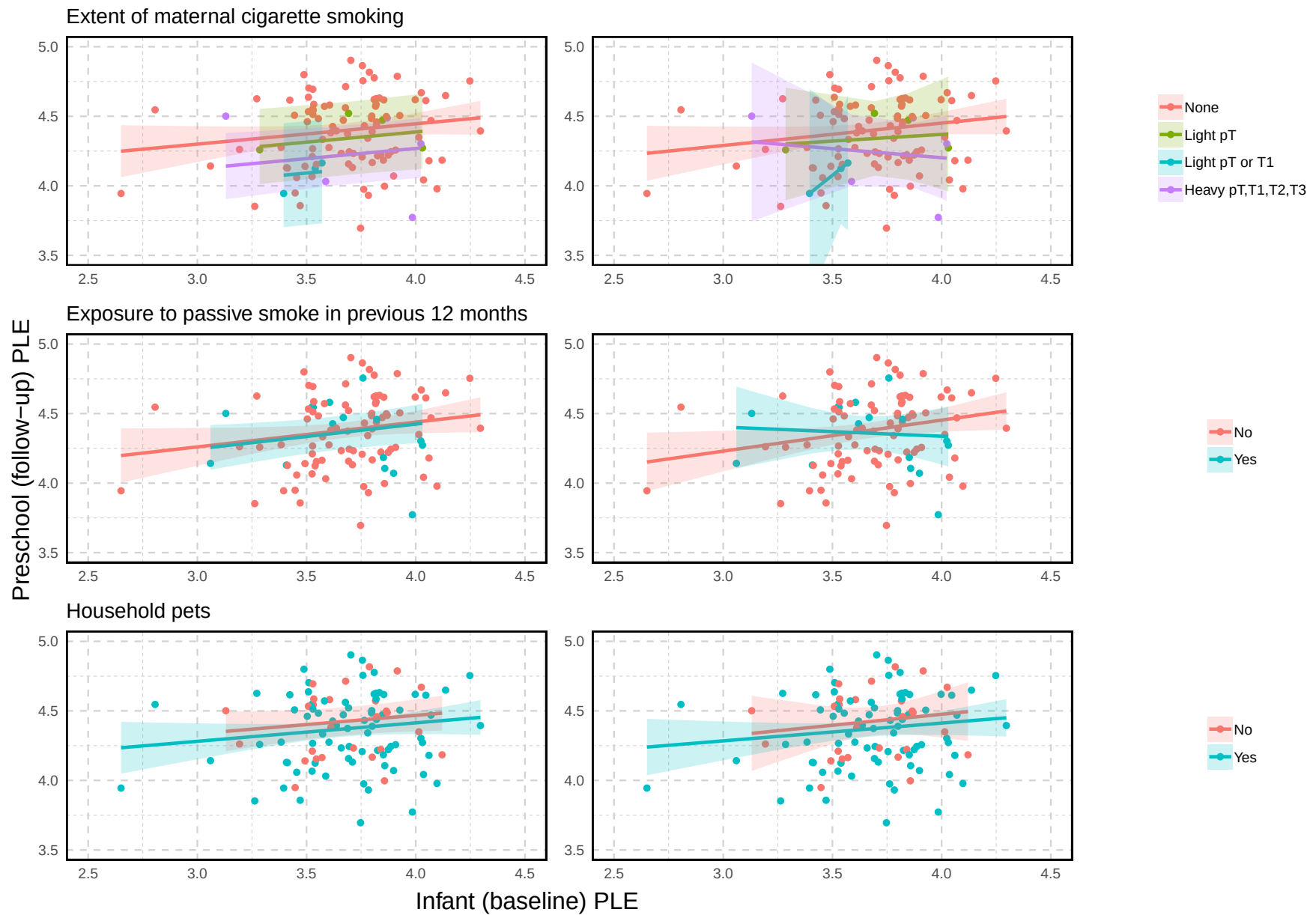


FIGURE 6.6: (continued).

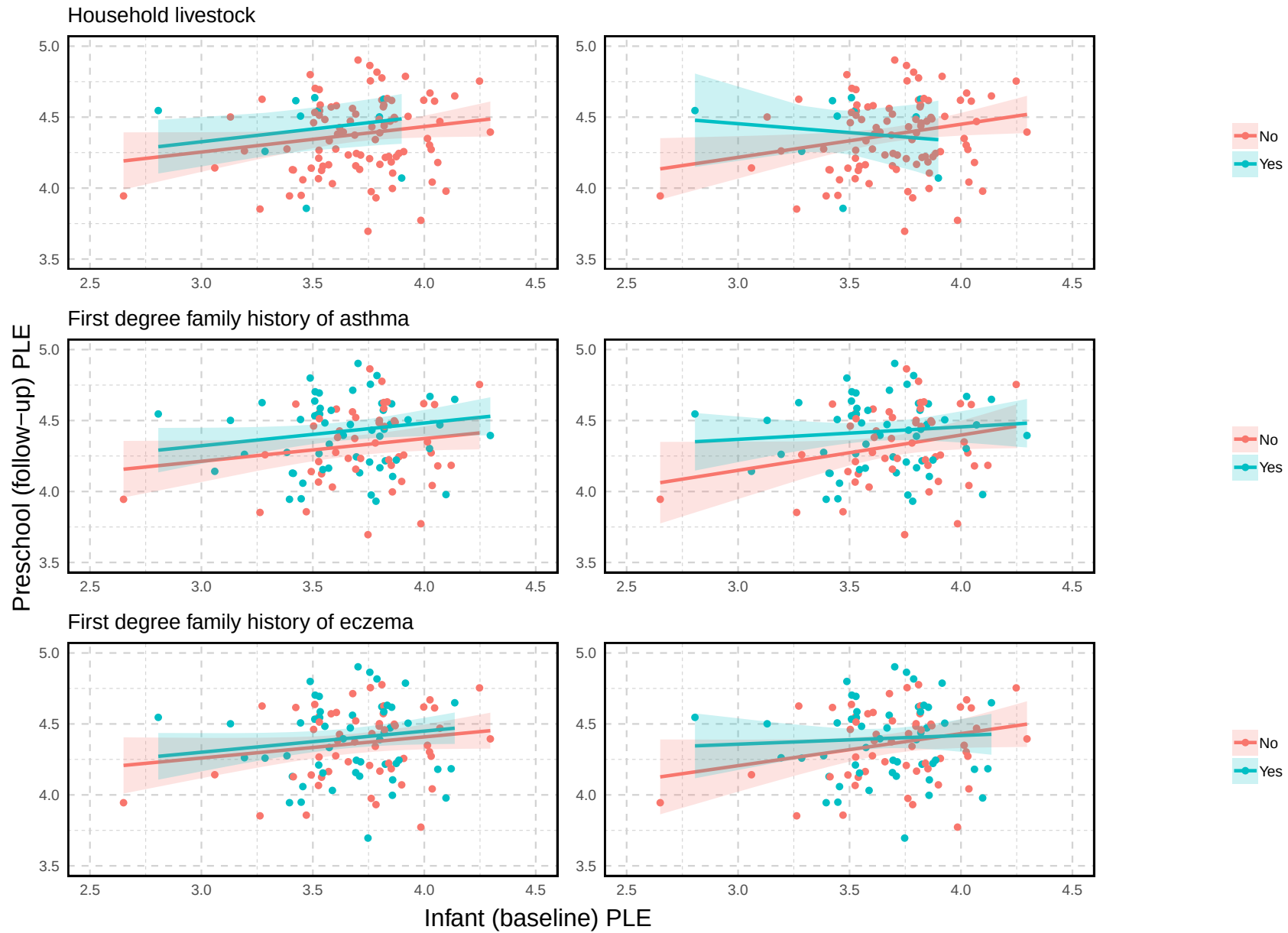
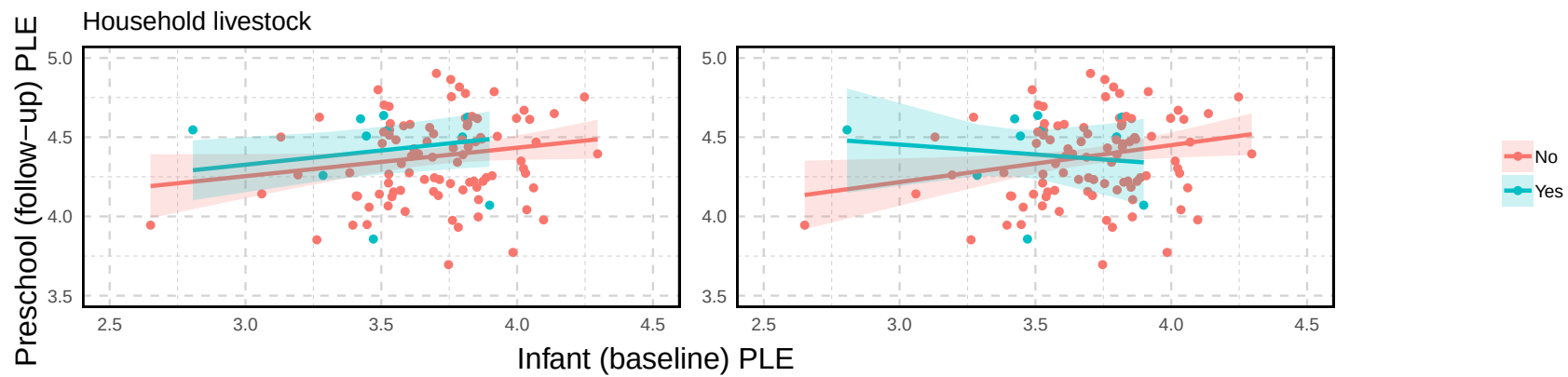


FIGURE 6.6: (continued).



6.3.4 MBW and non-MBW agreement analysis

The agreement model described by (6.4) was fitted to power-law exponents estimated from the breath cycles of 302 individuals from the BIS infant cohort who successfully completed both SF₆ MBW and non-MBW tidal flow recordings.

Table 6.4 summarises the posterior distributions and convergence diagnostics for the mean, variance, and agreement components from the fitted agreement model. Parameter convergence was sufficient for all model parameters, with $\hat{R}$ values very close to one, adequate effective sample sizes given the number of chains used (not shown), and no divergent transitions.

The estimated bias is discernably positive and, due to non-overlapping credible intervals, the residual standard deviations for each method are also discernably different. Yet each of the bias and standard deviations are relatively small in magnitude, which suggests that the two methods are providing measurably similar outcomes.

Figure 6.7 shows the differences, for individuals, between power-law exponents estimated using either SF₆ MBW or non-MBW tidal flow recordings. There appears to be no systematic relationship in the differences between each method with magnitude of the estimate. We have included upper and lower LoA, of which the interval is quite narrow at approximately $\pm 10\%$ of the mean values. Typically, we can confidently use two methods interchangeably if there is minimal bias, no systematic relationship in the differences, and the LoA satisfy some predetermined acceptability criterion. In the absence of an established reference, we cannot say whether these LoA are acceptable or not, but we note that an ‘agreement interval’ of $\pm 10\%$ would be considered reasonable for other measures of lung function such as LCI [77].

Our estimates of model (6.4) suggest that the SF₆ MBW and non-MBW tidal flow recordings produce quite similar power-law exponents at an individual level, even though the overall bias is not zero.

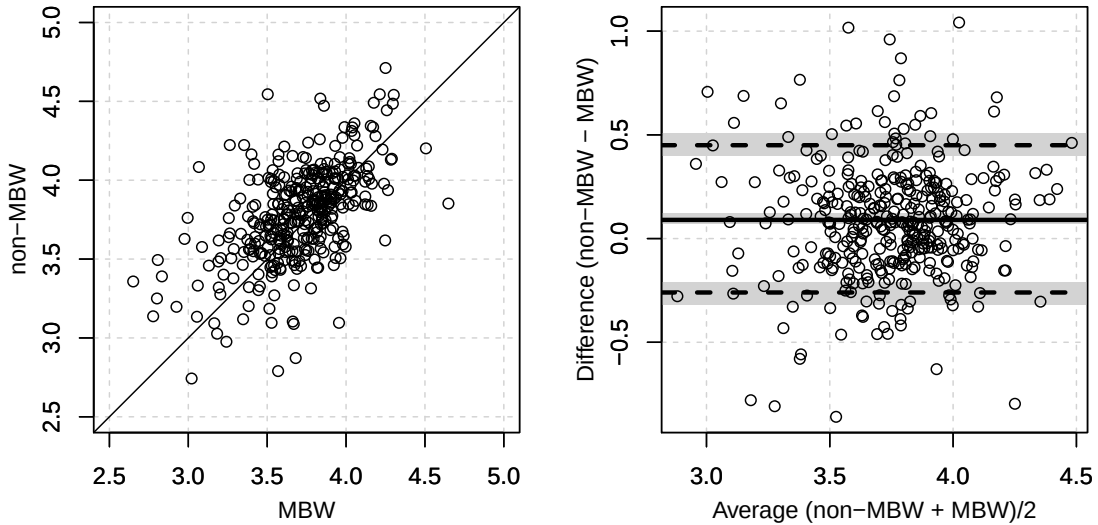
6.4 Discussion

This chapter focused on the application of estimators of the power-law exponent, as described in theoretical detail in Chapter 5, to MBW tidal flow waveforms collected as part

TABLE 6.4: Parameter estimates (posterior medians and 95% credible intervals) from the agreement model fitted to power-law exponents estimated using SF₆ MBW tidal flow (MBW) and non-MBW tidal flow (non-MBW) recordings from the BIS infant cohort ($n = 302$). CI: credible interval

Parameter	Median	95% CI	$\hat{R}$
Mean components			
Non-MBW (α_1)	3.81	(3.78, 3.84)	1.00
MBW (α_2)	3.72	(3.69, 3.75)	1.00
Variance components			
Between-participant (γ)	0.24	(0.21, 0.26)	1.01
Non-MBW SD (σ_1)	0.13	(0.11, 0.14)	1.01
MBW SD (σ_2)	0.22	(0.17, 0.27)	1.02
Agreement components			
Bias ($\alpha_1 - \alpha_2$)	0.09	(0.07, 0.12)	1.00
LoA (upper)	0.45	(0.4, 0.51)	1.01
LoA (lower)	-0.26	(-0.32, -0.21)	1.01

FIGURE 6.7: Agreement plots comparing power-law exponents estimated using SF₆ MBW tidal flow (SF₆ MBW) and non-MBW tidal flow (non-MBW) recordings from the BIS infant cohort ($n = 302$). The left panel shows a scatterplot of SF₆ MBW vs non-MBW power-law exponents. The right panel shows a Bland-Altman plot (comparing differences vs averages) for SF₆ MBW vs non-MBW power-law exponents, with the estimated mean difference (solid line), 95% limits of agreement (dashed lines), and 95% credible intervals (shaded areas).



of AREST CF and BIS, with the aim of exploring the potential value of these modelling concepts for understanding the epidemiology of lung diseases of various types.

We observed a number of interesting results from our statistical models. First, age has a strong positive effect on power-law exponent and therefore complexity of the tidal flow waveform. Frey et al. [33] did not observe a similar effect for healthy infants under 12 months, but it is possible that the strength of our estimated age effect is driven by children aged older than 12 months. Second, at birth the complexity of tidal flow waveforms was

on average similar for children with and without cystic fibrosis but later in childhood children with cystic fibrosis had slightly more complex tidal flow waveforms. This pattern of developing complexity was not observed by Frey et al. [33] for infants with CLD and history of wheezing or Habib et al. [35] for VLBW babies. Third and finally, for the BIS children, after conditioning on observed power-law exponents at baseline, none of the covariates appeared to be associated with the power-law exponent at preschool age. However, many of the categories within covariates are quite sparsely populated which made estimation difficult and in some cases required the categories to be omitted from the analysis.

When we compared the power-law exponents estimated from SF₆ and non-MBW tidal flow recordings from infants without cystic fibrosis, we did not find a meaningful difference between power-law exponents estimated from the same individuals but from different measurement methods. This result is in contrast to the findings of Banton et al. [27], who found significant differences between tidal breathing indices derived from either SF₆ MBW or non-MBW tidal flow recordings. It is possible that the power-law exponent may be more robust to changes in measurement systems compared to existing tidal breathing indices, particularly considering the results of Habib et al. [35] who found similarly small differences between power-law exponents estimated using tidal flow recordings from either a pneumatograph or RIP.

It is insightful to consider these results in the context of scale-invariance and system complexity. As we described in Chapter 5, different values of the power-law exponent have particular interpretations in terms of scale-invariance. A value of $\beta = 1$ represents a special condition at the boundary between fractional Gaussian noise and fractional Brownian motion where the strength of each frequency is exactly proportional to its oscillation, a form of complexity that is likely to be globally robust to local perturbations. Values that differ from 1 in either direction represent a departure from this special condition and a breakdown of the robustness that comes with this special property of complexity. Although we estimated the power-law exponent of some individual breath cycles as near 1, when we average over breath cycles, we do not observe values of β anywhere near 1, but rather in between 3 and 5. The power-law exponent appears to increase with age quite substantially, but this increase is not as rapid for children with cystic fibrosis. How might we interpret this in the context of system complexity and robustness?

Frey et al. [33] and Habib et al. [35] reason that if the respiratory system is driven by a pure sine wave with linear inputs, the output would be a pure sine wave, representing very little complexity. The power-law exponent could therefore be interpreted as an index of system complexity even outside of the typical range of $\beta \in [0, 3]$. In the context of the respiratory system, this complexity may be due to either system non-linearities (i.e. a non-sinusoidal respiratory oscillator), non-linear inputs to the oscillator via feedback mechanisms, or a combination of both [33]. Our results suggest that the tidal flow of either developing respiratory systems, for example those of young infants, or respiratory systems under stress, for example those of children with cystic fibrosis, produce somewhat *more* complex outputs than the respiratory systems of otherwise older or healthier individuals. This contrasts with what is described in the literature on scale-invariance and complexity, which describes complexity as something associated with robustness or health [39, 82, 84]. However, this discussion is typically in reference to scale-invariance in long time-series—examples such as inter-breath intervals are discussed in Chapter 5—rather than something so closely tied to the operating physiology as tidal breath cycles which, because humans tend to balance an inspiration immediately with an expiration and vice-versa, will almost always contain a strong sinusoidal component reflecting this balance that dominates any higher frequency oscillations. On one hand, while we might see a purely sinusoidal waveform as something undesirable and representative of an underlying pathology, for example an impediment to the complex feedback system that underlies neurorespiratory control, on the other hand a waveform that is too complex with many local extrema may represent a system that is not balancing its complex inputs well, potentially due to some underlying systemic disease, to produce a smooth and efficient breathing pattern. Therefore in terms of systems where we expect a strong sinusoidal component, such as tidal breath cycles, a robust respiratory system might be somewhere in between, where local extrema are minimised but the system output is not purely sinusoidal.

The power-law exponent appears to be a straightforward and powerful approach to characterising the shape and complexity of entire tidal flow waveform with interpretations related to system complexity. We have observed that the shape of the tidal flow waveform is affected by age and cystic fibrosis. However, the results presented in this chapter indicate that respiratory complexity may not be associated with any of the covariates we have included, although we note that our analysis used a relatively small sample size and with more precision some effects may be better identified.

Our analysis uses combined data that was collected for purposes not directly related to our research and the results presented here should be interpreted with this in mind. For example, the effect of various aspects of the study designs cannot be discounted, including the use of different flow and gas measurement systems, software, sedation and sleep states, and body positioning. Having data corresponding to a maximum of two follow-ups for BIS also made it difficult to properly identify and estimate residual standard deviation for our linear random effects model, although we note that for the other model parameters, particularly the main effects, the estimates were surprisingly similar across a range of scenarios. Furthermore, there may be value in exploring other models that can account for differences in residual variation between age groups, which we have not investigated. While extending the results presented here goes beyond the scope of this exploratory analysis, we note however that a third, primary school wave of BIS has been funded, and will include further MBW testing. Tidal flow recordings from this third wave could be used to augment the current BIS data and allow the model using combined data to be extended and fully estimated.

We have only analysed the effect of covariates measured at baseline in children without cystic fibrosis, and in some cases the categories in these covariates are sparsely populated. Furthermore, our analysis did not include data on the severity of cystic fibrosis and therefore we cannot establish whether there is dose-response relationship between cystic fibrosis severity and respiratory complexity. It also is of interest to understand how the power-law exponent might predict the occurrence of respiratory diseases in later childhood. If further analyses are to be undertaken to establish the value of the power-law exponent as an epidemiological tool they may benefit from the inclusion of variables related to both disease severity and future health.

Chapter 7

Discussion

Contents

7.1 Summary of research	135
7.1.1 Data collection procedures	136
7.1.2 Effect of breathing pattern on MBW outcomes	136
7.1.3 Modelling MBW data	139
7.1.4 Modelling the spectral densities of tidal flow waveforms	142
7.1.5 Epidemiology of tidal flow power-law spectra	144
7.2 Conclusion	145

This concluding chapter summarises the key findings arising from research presented in this thesis. Section 7.1 provides a broad summary of the findings arising from the research contained within each chapter, along with reflections on the statistical challenges faced when addressing the research objectives, and highlights potential avenues for further investigation. Section 7.2 provides a concluding statement.

7.1 Summary of research

The research presented in this thesis aimed to develop and apply novel statistical approaches to modelling lung function data. We summarise Chapters 2 to 6 in turn in the following sections.

7.1.1 Data collection procedures

Prior to undertaking the methodological components of this thesis, a major hurdle to overcome was that of understanding the data, how it was collected, and how it was handled by the collecting devices and software. This required a solid understanding of the technical specifications and design of a number of complicated measurement devices in addition to various aspects of respiratory physiology. Much of this work is documented in Chapter 2, but related themes are discussed in one way or another throughout the thesis. While at face value this work may not appear to fit with the broader statistical themes of this thesis, it was in fact fundamentally important to the subsequent modelling.

The quality of any statistical model is dependent on the quality of the data. In this sense, ensuring that we had a firm grasp of the limitations of the data collection procedures, prior to modelling, was fundamental to our subsequent methodological work and analyses. MBW procedures vary substantially, with many ‘moving parts’ that need to be considered in order to fully understand the raw data and the summary outcomes they produce. After working with the commercial software for some time, discussing with collaborators about data collection and management, and attempting to automate our workflow, it became clear that we would have to appropriately document the flow of data from the collection device to the exported dataset in order to identify where we ‘lost control’ of the data workflow so that we could try to regain it. The need to document the software workflow motivated our publication ‘Lack of transparency in software used to analyze multiple breath washout data’ in *Pediatric Pulmonology* [49]. Unfortunately, several ‘black boxes’ were identified in the commercial software which we were unable to address without access to the underlying algorithms. The research represents a valuable contribution to the lung function testing literature in that it brought the issues with data workflow into the realm of the known, highlighting the details of the issues that remain unknown.

Similar and related concerns have been published by many other research groups around the world [11, 19, 28, 58, 128–131].

7.1.2 Effect of breathing pattern on MBW outcomes

The analysis of the effect of variations in breathing patterns on MBW outcomes had several motivations: previous work by Vukcevic et al. [32] that showed that the occurrence of a

sigh may not necessarily require MBW tests to be discarded; a recent study by Jensen et al. [45] who proposed a checklist to assess MBW tests for breathing pattern stability, which served as a good starting point; the desire for the process of assessing MBW data to be more automated; and, at a more practical level, the need to obtain a working dataset for subsequent analyses in this thesis in a timely manner. The intention was to save the time and effort of clinicians, researchers, and participants involved in MBW testing, and also ensure that the principles of scientific transparency and reproducibility were upheld through unambiguous and automatable descriptions of our methods. By focusing on improving the lung function data workflow, this component of the thesis therefore echoed the sentiment of the initial work as described in the previous section.

We addressed our aims by first defining a set of rules, using Jensen et al.'s checklist [45] and the MBW consensus statement [11] as a guide, to develop a set of broad validity criteria. We endeavoured to make these rules as unambiguous as possible so that they could be implemented directly as software, rather than manual assessment. We used all N₂ MBW tests collected from the BIS preschool group to assess the extent that different validity criteria identified meaningfully different subsets of tests using a variance components model was fitted to the repeated measures MBW outcomes from tests deemed successful by each validity criterion.

We found that most of the deviations from ideal breathing pattern, according to our unambiguous validity criteria, do not affect overall MBW outcomes substantially. Further, we show that the criterion corresponding to tests that were technically sufficient—which essentially ensured that the measured end-tidal N₂ concentration started and ended at the appropriate levels and decayed consistently throughout the washout—provided a combination of MBW outcomes with reasonable statistical properties as well as the largest sample size. This is in contrast to the current recommendations that broadly suggest that the breathing pattern must satisfy a number of strict criteria in order to be considered a successful MBW. Indeed we observed that the MBW outcomes from tests that satisfied our strictest validity criterion were perhaps indicative of being special cases in that the characteristics of the participants' physiology might be conducive to producing short, clean tests and potentially unrepresentative MBW outcomes.

This analysis was not without its limitations. This included the lack of a 'gold standard' for defining test validity to compare against, the lack of avenues to codify subjective inclusion

rules, and the focus on N₂ MBW tests performed on otherwise healthy preschool children. The accepted gold standard is currently a complete manual review by a trained MBW technician. Ideally we would examine the sensitivity and specificity of our validity criteria against the gold standard but at the time of writing such gold standard assessments were not available. In the absence of the gold standard, we compared the statistical properties of each validity criterion to one another, and used known representative values of the MBW outcomes to assess the broad utility of each criterion. A number of potentially important inclusion rules are also still firmly in the domain of subjective manual assessment by the MBW technician and were not defined for our analysis. For example, whether on close inspection minor variations in the N₂ concentration could be attributable to a system leak currently involves fairly subjective consideration of the potential pathologies of various MBW signals over a subset of the test. Codifying such rules using basic logical statements, as we have done for our inclusion rules, represents a major challenge and indeed may not be completely possible. The focus on N₂ MBW from preschool children from a healthy birth cohort was because the rules proposed by Jensen et al. [45] were specific to N₂ MBW and because the only tests that required assessment for the purposes of our research were from the preschool wave of BIS. Other variants of MBW tests collect different data and tend to contain less useful information about breathing pattern as compared to N₂ MBW. For example, end-tidal CO₂ concentration is not typically measured in SF₆ MBW tests. Any adaptation of our methods to non-N₂ MBW would require substantial modification to the current set of inclusion rules.

Addressing these limitations represents a clear opportunity for extending this research. For example, the only constraint to obtaining gold standard assessment of the MBW is economic, that is to say, the time and resources required to have multiple highly trained technicians manually assess a large number (perhaps thousands) of MBW tests. Alternatively, our validity criteria could be applied by other research groups who may have already collected and assessed a similarly large number of N₂ MBW data for quality. Furthermore, rather than struggling to codify the various subjectively assessed inclusion rules as a collection of logical clauses, there may be value in leveraging statistical learning algorithms to train classification models to identify tests that satisfy such rules. Indeed, with a large amount of data for training a predictive model, it is conceivable that restrictive rules-based approaches such as ours could be superseded by an ‘artificial intelligence’ approach that can identify currently unknown features of the data that are highly predictive

of the quality of an MBW test.

7.1.3 Modelling MBW data

The development of novel statistical approaches to modelling MBW data was primarily motivated by the fact that the MBW test is difficult to perform and prone to disruption or breathing artefacts which typically result in the data from many tests being unused. The cost is not only economic, in terms of researcher and patient time and resources, but also ethical, in that to obtain the required number of complete tests the patient potentially spends more time under experimentation than may be absolutely necessary. Indeed, successful MBW measurements are only obtained in around 60% to 70% of sleeping babies, with this proportion decreasing rapidly with age in early childhood as participants are no longer tested during sleep and are easily distracted. The consensus statement implicitly highlights this issue with reference to a number of important future research directions, of which we have addressed two [11]. The first was, in the words of the authors, the use of ‘interpolation’ methods to determine exact end-of-test for LCI, and the second was the flexibility of MBW end-test thresholds and the shortening of tests. By developing and testing our novel statistical model on a large amount of MBW data we addressed both issues.

The modelling process involved correctly identifying and adapting a set of linear and non-linear functions that could appropriately capture the underlying growth and decay structure of the MBW data, summarised at the end of each breath, for individual tests. In the model of the gas decay curve we were guided by the theoretical two-compartment model of the lung, which can be defined by a bi-exponential decay function. The remaining models for the cumulative expired volume of the gas mixture (CEVGM) and cumulative expired volume of the tracer gas (CEVTG) at first glance appeared to be, respectively, relatively simple linear growth and increasing exponential decay models. However we found that while the mean estimates from these models fitted the individual test data very well, the posterior predictive distributions both violated a monotonicity constraint—due to the cumulative nature of the data—and were not representative of the observed within-subject variation across tests. We rectified this by modelling the original model’s discrete derivatives on a log scale. Satisfyingly, the resulting posterior predictive distributions were far better representations of the observed data.

The importance of respecting monotonicity constraints and the power of posterior predictive distributions as a modelling tool was a key statistical insight gained from this research. Graphical examination of the posterior predictive distribution for the monotonic data allowed us to clearly see the deficiencies in our approach and validate our revision in the context of repeated measurements. The fitting of non-linear functions to growth or decay data is now commonplace and relatively straightforward, and we assume in many cases these functions are applied to data with similar monotonicity constraints. If one simply examines the model fit using classical methods such as residual analysis, as we showed, one may be led to thinking that the model fit is satisfactory, indeed it may appear excellent. However, in our case, posterior predictive checking showed the implausibility of the predicted data as it clearly violated the monotonicity constraint and also led to severe overfitting of the observed data when considered in context of within-subject variation.

With these models defined we were able to formulate fully parameterised model-based alternatives to the standard empirical summaries of lung function, namely LCI and FRC, which were not dependent on the discretely observed data at the end of the nearest breath corresponding to an end-of-test threshold, but rather on an underlying set of continuous functions. Indeed, our model-based FRC was shown to be estimable independently of the end-of-test threshold by taking its asymptotic value. We fitted these models using Bayesian methods, with two sets of prior probability distributions on the model parameters: one set of *diffuse* priors, containing very little information about the model parameters, and one set of *informative* priors, intended to characterise the typical distribution of the model parameters including an explicit covariance structure jointly linking all three data models.

We found that the diffuse prior model produced estimates of LCI that could be used interchangeably with the standard method, suggesting that the standard method is in fact a good estimator of MBW outcomes using complete data. Thus if we consider our model-based approach to be an example of an ‘interpolation’ method, as described in the consensus statement, we can conclude that interpolation between end-of-breath quantities provides similar point estimates to the standard empirical methods when complete data are available. However, in contrast to the standard method, our model-based approach allows the uncertainty of the estimates to be directly quantified and therefore represents a valuable innovation.

A real strength of our model was in cases where complete data were *not available*, such

as when the test ends prior to the gas concentration reaching an end-of-test threshold. This is because our model allowed us to extrapolate the quantities of interest—in this case the model-based LCI or FRC—using the available data. We improved upon these extrapolations by using informative prior distributions on the model parameters. We found that our model produced, on average, reasonable estimates of LCI and FRC once the tracer gas reaches 1/10 of the starting concentration, representing an average test shortening of 43%. Tests any shorter than this did not produce reliable estimates.

This result is encouraging, particularly in the context of similar, yet simpler, studies of shortened tests which found that tests which used an end-of-test threshold of 1/20 of the initial gas concentration led to moderately reliable estimates. That shorter tests produced unreliable estimates even when using an informative prior distribution suggests that it is unlikely that the test can be shortened to thresholds greater than 1/10, irrespective of which estimation procedure is used.

Allowing the use of shorter tests would mean that more data can be used to accurately estimate key quantities of lung function, an important finding given the challenges of obtaining complete tests from children. This applies both to data already collected as part of various cohort studies globally, and at a clinical level where clinicians may be able to use partially completed tests to achieve the desired number of lung function measurements. Our analysis was limited to infants from a population-derived birth cohort. An examination of the performance of our methods using a comparison group with lung disease would have been ideal. We did plan to include data from the AREST CF cohort in the analysis as a comparison group but, at the time of publication of the results, the required data were not available for use, however we consider this an opportunity for future research. For example, it is of interest to know how accurately we can estimate lung function from incomplete tests in infants with lung disease, who typically have slower gas clearance profiles, hence substantially longer tests, and may therefore benefit the most from the use of an informative prior model. While we would like to see this method refined and tested further we admit that its implementation may be relatively complex, particularly to MBW technicians with minimal statistical experience. This hurdle could be easily overcome, however, by ensuring that the method is properly abstracted and integrated into the commercial software, removing the burden of model fitting from the technician so that they can focus solely on the estimates and their uncertainty. We note that previous methods

to empirically shorten the tests to 1/20 have already been integrated into WBreath at the time of writing.

7.1.4 Modelling the spectral densities of tidal flow waveforms

tidal flow waveforms are an example of easily measured data that are rich with physiological information related to respiratory control and pulmonary mechanical function. They can be understood as integrated outputs of highly complex feedback systems of neurorespiratory control [33–35, 39, 57, 79–81]. Our research into the modelling of the spectral densities of tidal flow waveforms was motivated by our desire to process these signals using all of the sampled data to provide a summary of respiratory complexity to complement the established MBW outcomes and simpler summary measures.

We focused on the estimation of the power-law exponent from the spectral density of tidal flow waveforms. Power-law exponents are often proposed as a measure of ‘organisation’ in complex systems and are encountered throughout the physical sciences. The spectral density describes, in the frequency domain, the sinusoidal components that comprise the complete waveform in the time domain. If the spectral density follows what is known as a power-law distribution, then this density can be modelled (in the log scale) using a simple linear function. In practice, the spectral density cannot be observed directly but must be estimated. Estimation is commonly done using a discrete Fourier transform to construct a periodogram. The periodogram has its own theoretical sampling distribution which we hypothesised may affect the estimation of the power-law exponent.

Although the power-law exponent of tidal flow waveforms had been estimated in practice by Frey et al. [33] and Habib et al. [35] using these methods, not much information was available concerning the statistical properties of the power-law exponent estimator. While some studies have looked at the sampling properties of commonly used power-law exponent estimators, including the PSD method, they typically did so either for long series, without accounting for observational error, or using power-law exponents with specific fractal interpretations (i.e. taking values between 0 and 3). We therefore aimed to gain a detailed understanding of the sampling distribution of the power spectral density (PSD) estimator of the power-law exponent both theoretically, from the principles of spectral estimation, and empirically, via simulation. With this knowledge, we proposed a novel estimator for power-law exponents that explicitly accounted for the non-normal

distribution of periodogram ordinates, in addition to tailoring the existing standard OLS estimator to account for noise encountered in tidal flow recordings.

We showed that the sampling properties of PSD estimators are highly sensitive to the value of the underlying power-law exponent, signal length, and even small amounts of noise, but robust to the non-normal sampling distribution of the periodogram.

We concluded that the standard approach for mitigating the biasing effect of noise on the power-law exponent—dropping components of the spectral density for frequencies above $1/8$ of the Nyquist frequency—was not conservative enough. Unbiased estimation of power-law exponents of value greater than 3 from data similar to tidal flow waveforms was only possible by discarding a greater proportion of the spectral density, specifically for frequencies above $1/20$ of the Nyquist frequency. While the more conservative approach improved the precision of the estimator, it reduced the accuracy considerably. Our recommendation is therefore to improve precision by taking the average of multiple estimates of the power-law exponent, taken over multiple breaths from the same individual. However, this implies that the PSD method is unlikely to be of much use for analysing individual breaths.

Counter-intuitively, longer breaths containing more samples actually increased the estimation bias in the presence of noise. This is because increasing signal length lowers the fundamental frequency which, for a fixed sampling rate, increases the number of harmonics available for estimating the spectral density. These extra harmonics will be of higher frequency and be disproportionately affected by sampling noise, as compared with the lower frequencies that are less noisy. In such situations the noise will have a large influence on the estimates and it is therefore necessary to account for the relative importance of high and low frequency periodogram ordinates.

Furthermore, when explicitly accounting for the non-normal distribution of the spectral density estimates using the LCE estimator, we found only a minor increase in estimation efficiency compared to using OLS, but at the cost of increased bias. Because we can control for sampling variation by simply averaging many breaths, it is optimal to focus on the least biased estimator, OLS, and therefore we do not recommend using the LCE estimator.

This research allowed us to properly apply these methods to epidemiological data while explicitly accounting for bias and variance of the estimators. It also serves as a detailed

reference for practitioners interested in applying similar methods to data. It is possible that a more elaborate power-law exponent estimator, which adapts to signal noise of different strengths rather than discarding all frequencies above a hard threshold, could further improve upon our proposed estimators. We consider this to be a potentially important avenue for extending this research.

7.1.5 Epidemiology of tidal flow power-law spectra

Equipped with a detailed understanding of the sampling distribution of the PSD approach to power-law exponent estimation, we turned our attention to exploring whether the power-law spectra of tidal flow waveforms were affected by age and other covariates.

Specifically we estimated power-law exponents from tidal flow waveforms recorded from 2,745 MBW tests comprising over 200,000 breath cycles, and then used a linear model to explore the relationship between the exponents and age. We also allowed for individual-specific variation in both parts of the model. The effect of the covariates was measured by quantifying their impact on the power-law exponents observed at follow-up conditioned on baseline observations.

We found that, on average, the tidal flow waveforms of newborns were quite complex with distinctly non-sinusoidal forms. This complexity is reduced significantly with age, although for children with cystic fibrosis this reduction is less strong than children without cystic fibrosis. There was little evidence to suggest any real association between covariates and power-law spectra, and therefore the shapes, of tidal flow waveforms. Further, the between-individual variation around the main effects and the residual variation was substantial in all cases, suggesting that observed power-law exponents and hence tidal flow waveform complexity varied substantially between individuals and within individuals over time. We note this is the case for many summaries of lung function.

We performed an additional analysis to compare differences in power-law exponents estimated from SF₆ and non-MBW tidal flow waveforms. We found that there was no meaningful mean difference between power-law exponents estimated from the same individuals but from the different measurement methods. Furthermore the limits of agreement were within bounds that would be acceptable for similar measures of lung function. This work can be considered as a major extension to the work of Frey et al. [33] and Habib et

al. [35], who undertook a similar analysis, albeit with far less data, a reduced scope, and without explicitly controlling for signal noise in the estimation procedure. Furthermore we are the first to perform this analysis with data from children with cystic fibrosis. Our analysis, while rich with data and using estimation methods validated in Chapter 5, was decidedly exploratory and intended to understand the value of the power-law exponent for epidemiological modelling rather than establish firm causal associations. A number of limitations of our study design limit the extent to which we can control for any confounding by, for example, differences in sedation and measurement devices. We performed our analysis using the data available to us at the time, which included data from Melbourne's AREST CF participants and the infant and preschool waves of BIS, however we only had covariates of epidemiological interest for BIS. Furthermore the usefulness of our linear random effects model results was limited when applied to the BIS data because it has only been measured, at the time of writing, at two time points. In particular, we found we could not predictably estimate residual variation when using BIS data because it has only been collected for two time points at the time of writing. We note that an there is an opportunity to augment the BIS data with tidal flow recordings from future follow-ups to overcome some of these limitations.

The value of the power-law exponent for understanding lung function in either a clinical or predictive setting remains to be seen. Further work using a more tailored study design and variables related to future states of health may be required to better understand the utility of the power-law exponent model for understanding the epidemiology of lung disease and development.

7.2 Conclusion

This thesis makes important contributions to respiratory medicine and applied statistics. Building on an already growing body of research, we have provided an in-depth examination of the effect of both software algorithms and breathing pattern variations on MBW outcomes. This information will allow respiratory scientists to better adapt and refine the MBW method and therefore gain a deeper understanding of lung function. Innovatively, we have shown that by modelling the MBW data and using Bayesian estimation methods we are able to use data that would otherwise have been discarded. This has the potential to reduce both the economic and ethical cost of MBW testing. This modelling process

also provided us with valuable insight into the proper modelling of data with monotonicity constraints, a consideration that we do not often encounter in applied settings.

These results are all the more important in the context of the burgeoning global uptake of the MBW as an important tool for understanding lung function. We have developed and tested various estimators of the power-law exponent to both improve upon the statistical properties of existing estimation procedures and also describe the major limitations of the method. In doing so we have provided a valuable statistical perspective on the method which allowed us to confidently apply it to a large amount of data to describe the relationship of the complex tidal flow waveform, age, disease, and other covariates. This work will serve as both a valuable reference for applied scientists using similar methods and as a useful starting point for further methodological refinements.

We hope that the findings contained within this thesis will encourage and facilitate more sophisticated statistical analysis of lung function data. Through this we may begin to reveal novel characteristics of the underlying physiology to the ultimate benefit of people living with the consequences of lung diseases.

Bibliography

1. Carraro, S, Scheltema, N, Bont, L, *et al.* Early-life origins of chronic respiratory diseases: understanding and promoting healthy ageing. *European Respiratory Journal* 44: 1682–1696 (2014).
2. Dezateux, C & Stocks, J. Lung development and early origins of childhood respiratory illness. *British Medical Bulletin* 53: 40–57 (1997).
3. Mullassery, D & Smith, NP. Lung development. *Seminars in Pediatric Surgery* 24: 152–155 (2015).
4. Joshi, S & Kotecha, S. Lung growth and development. *Early Human Development* 83: 789–794 (2007).
5. Shi, W, Bellusci, S & Warburton, D. Lung development and adult lung diseases. *CHEST* 132: 651–656 (2007).
6. Stick, S. The contribution of airway development to paediatric and adult lung disease. *Thorax* 55: 587–594 (2000).
7. Stocks, J, Hislop, A & Sonnappa, S. Early lung development: lifelong effect on respiratory health and disease. *The Lancet Respiratory Medicine* 1: 728–742 (2013).
8. Beydon, N, Davis, SD, Lombardi, E, *et al.* An official American Thoracic Society/European Respiratory Society statement: pulmonary function testing in preschool children. *American Journal of Respiratory and Critical Care Medicine* 175: 1304–1345 (2007).
9. Rosenfeld, M, Allen, J, Arets, BHGM, *et al.* An official American Thoracic Society workshop report: optimal lung function tests for monitoring cystic fibrosis, bronchopulmonary dysplasia, and recurrent wheezing in children less than 6 years of age. *Annals of the American Thoracic Society* 10: S1–S11 (2013).

10. Thamrin, C, Hardaker, K & Robinson, PD. Multiple breath washout: from renaissance to enlightenment? *Pediatric Pulmonology* 51: 447–449 (2015).
11. Robinson, PD, Latzin, P, Verbanck, S, *et al.* Consensus statement for inert gas washout measurement using multiple- and single- breath tests. *European Respiratory Journal* 41: 507–522 (2013).
12. Schulzke, SM & Frey, U. Consensus statement on inert gas washout measurement: at the threshold of clinical use. *European Respiratory Journal* 41: 500–502 (2013).
13. Horsley, AR, Gustafsson, PM, Macleod, KA, *et al.* Lung clearance index is a sensitive, repeatable and practical measure of airways disease in adults with cystic fibrosis. *Thorax* 63: 135–140 (2008).
14. Fuchs, SI & Gappa, M. Lung clearance index: clinical and research applications in children. *Paediatric Respiratory Reviews* 12: 264–270 (2011).
15. Horsley, A. Lung clearance index in the assessment of airways disease. *Respiratory Medicine* 103: 793–799 (2009).
16. Fowler, WS, Cornish Jr, ER & Kety, SS. Lung function studies. VIII. Analysis of alveolar ventilation by pulmonary N₂ clearance curves. *Journal of Clinical Investigation* 31: 40–50 (1952).
17. Fuchs, SI, Eder, J, Ellemunter, H, *et al.* Lung clearance index: normal values, repeatability, and reproducibility in healthy children and adolescents. *Pediatric Pulmonology* 44: 1180–1185 (2009).
18. Horsley, A & Wild, JM. Ventilation heterogeneity and the benefits and challenges of multiple breath washout testing in patients with cystic fibrosis. *Paediatric Respiratory Reviews* 16: 15–18 (2015).
19. Anagnostopoulou, P, Yammine, S, Schmidt, A, *et al.* False normal lung clearance index in infants with cystic fibrosis due to software algorithms. *Pediatric Pulmonology* 50: 970–977 (2015).
20. Verbanck, S, Paiva, M, Paeps, E, *et al.* Lung clearance index in adult cystic fibrosis patients: the role of convection-dependent lung units. *European Respiratory Journal* 42: 380–388 (2013).
21. Fuchs, SI, Ellemunter, H, Eder, J, *et al.* Feasibility and variability of measuring the lung clearance index in a multi-center setting. *Pediatric Pulmonology* 47: 649–657 (2012).

22. Singer, F, Houlitz, B, Latzin, P, *et al.* A realistic validation study of a new nitrogen multiple-breath washout system. *PLoS ONE* 7: e36083 (2012).
23. Yammine, S, Singer, F, Abbas, C, *et al.* Ways to shorten the lung clearance index measurement I—are three measurements needed? in *European Respiratory Society Annual Congress 2012* 40: (2012), 4604.
24. Hannon, D, Bradley, JM, Bradbury, I, *et al.* Shortened lung clearance index is a repeatable and sensitive test in children and adults with cystic fibrosis. *BMJ Open Respiratory Research* 1: e000031 (2014).
25. Egger, B, Jost, K, Anagnostopoulou, P, *et al.* Lung clearance index and moment ratios at different cut-off values in infant multiple-breath washout measurements: moment ratios at different cut-off values. *Pediatric Pulmonology* 51: 1373–1381 (2016).
26. Anagnostopoulou, P, Egger, B, Lurà, M, *et al.* Multiple breath washout analysis in infants: quality assessment and recommendations for improvement. *Physiological Measurement* 37: L1–L15 (2016).
27. Banton, GL, Hall, GL, Tan, M, *et al.* Multiple breath washout cannot be used for tidal breath parameter analysis in infants. *Pediatric Pulmonology* 51: 531–540 (2016).
28. Summermatter, S, Singer, F, Latzin, P, *et al.* Impact of software settings on multiple-breath washout outcomes. *PLOS ONE* 10: e0132250 (2015).
29. Ahmad, F, Irving, S, Alton, E, *et al.* Multiple breath washouts in children can be shortened without compromising quality. *European Respiratory Journal* 46: 1814–1816 (2015).
30. Yammine, S, Lenherr, N, Nyilas, S, *et al.* Using the same cut-off for sulfur hexafluoride and nitrogen multiple-breath washout may not be appropriate. *Journal of Applied Physiology* 119: 1510–1512 (2015).
31. Mahar, R. *Model-based estimation of the lung clearance index from complete and incomplete multiple breath washout data.* MBIostat Workplace Project Portfolio. (University of Queensland, Australia, 2014).
32. Vukcevic, D, Carlin, JB, King, L, *et al.* The influence of sighing respirations on infant lung function measured using multiple breath washout gas mixing techniques. *Physiological Reports* 3: e12347 (2015).

33. Frey, U, Silverman, M & Suki, B. Analysis of the harmonic content of the tidal flow waveforms in infants. *Journal of Applied Physiology* 91: 1687–1693 (2001).
34. Baldwin, DN, Pillow, JJ, Stocks, J, *et al.* Lung-function tests in neonates and infants with chronic lung disease: tidal breathing and respiratory control. *Pediatric Pulmonology* 41: 391–419 (2006).
35. Habib, RH, Pyon, KH, Courtney, SE, *et al.* Spectral characteristics of airway opening and chest wall tidal flows in spontaneously breathing preterm infants. *Journal of Applied Physiology* 94: 1933–1940 (2003).
36. Frey, U, Hislop, A & Silverman, M. Branching properties of the pulmonary arterial tree during pre- and postnatal development. *Respiratory Physiology and Neurobiology* 139: 179–189 (2004).
37. Huang, W, Yen, RT, McLaurine, M, *et al.* Morphometry of the human pulmonary vasculature. *Journal of Applied Physiology* 81: 2123–2133 (1996).
38. Kitaoka, H, Takaki, R & Suki, B. A three-dimensional model of the human airway tree. *Journal of Applied Physiology* 87: 2207–2217 (1999).
39. Thamrin, C, Stern, G & Frey, U. Fractals for physicians. *Paediatric Respiratory Reviews* 11: 123–131 (2010).
40. Cernelc, M, Suki, B, Reinmann, B, *et al.* Correlation properties of tidal volume and end-tidal O₂ and CO₂ concentrations in healthy infants. *Journal of Applied Physiology* 92: 1817–1827 (2002).
41. Rostig, S, Kantelhardt, JW, Penzel, T, *et al.* Nonrandom variability of respiration during sleep in healthy humans. *Sleep* 28: 441–417 (2005).
42. Peng, L, Ye, L, Guo, X, *et al.* Evaluation of formocresol versus ferric sulphate primary molar pulpotomy: a systematic review and meta-analysis. *International Endodontic Journal* 40: 751–757 (2007).
43. Govindan, R, Wilson, J, Preißl, H, *et al.* Detrended fluctuation analysis of short datasets: an application to fetal cardiac data. *Physica D* 226: 23–31 (2007).
44. Govindan, R, Wilson, J, Murphy, P, *et al.* Scaling analysis of paces of fetal breathing, gross-body and extremity movements. *Physica A* 386: 231–239 (2007).
45. Jensen, R, Stanojevic, S, Klingel, M, *et al.* A systematic approach to multiple breath nitrogen washout test quality. *PLOS ONE* 11: e0157523 (2016).

46. Eke, A, Herman, P, Kocsis, L, *et al.* Fractal characterization of complexity in temporal physiological signals. *Physiological Measurement* 23: R1–R38 (2002).
47. Schaefer, A, Brach, JS, Perera, S, *et al.* A comparative analysis of spectral exponent estimation techniques for $1/f\beta$ processes with applications to the analysis of stride interval time series. *Journal of Neuroscience Methods* 222: 118–130 (2014).
48. Stocks, J & Lum, S. Applications and future directions of infant pulmonary function testing. *Progress in Respiratory Research* 33: 78–91 (2005).
49. Mahar, RK, Vukcevic, D, King, L, *et al.* Lack of transparency in software used to analyze multiple breath washout data. *Pediatric Pulmonology* 51: 1108–1110 (2016).
50. Gustafsson, P, Aurora, P & Lindblad, A. Evaluation of ventilation maldistribution as an early indicator of lung disease in children with cystic fibrosis. *European Respiratory Journal* 22: 972–979 (2003).
51. Pillow, JJ, Ljungberg, H, Hülskamp, G, *et al.* Functional residual capacity measurements in healthy infants: ultrasonic flow meter versus a mass spectrometer. *European Respiratory Journal* 23: 763–768 (2004).
52. Fuchs, SI, Buess, C, Lum, S, *et al.* Multiple breath washout with a sidestream ultrasonic flow sensor and mass spectrometry: a comparative study. *Pediatric Pulmonology* 41: 1218–1225 (2006).
53. Buess, C, Burger, R & Guggenbühl, W. *Ultrasonic respiration analysis* in *Annual International Conference of the IEEE Engineering and Medicine and Biology Society* 13: (1991), 1597–1598.
54. Blumenfeld, W, Wilson, PD & Turney, S. A mathematical model for the ultrasonic measurement of respiratory flow. *Medical and Biological Engineering* 12: 621–625 (1974).
55. Ndd Medizintechnik AG. *Multiple breath washout FAQ: application note*. Duern-ten, Switzerland, 2013. <http://www.nddmed.com/inc/data/support/AppNote-EOPLAB-MBW-FAQ-V03R.pdf>.
56. Gustafsson, PM, Robinson, PD, Lindblad, A, *et al.* Novel methodology to perform sulfur hexafluoride (SF₆)-based multiple-breath wash-in and washout in infants using current commercially available equipment. *Journal of Applied Physiology* 121: 1087–1097 (2016).

57. Bates, JHT, Schmalisch, G, Filbrun, D, *et al.* Tidal breath analysis for infant pulmonary function testing. *European Respiratory Journal* 16: 1180–1192 (2000).
58. Ebdon, A-M, Skoric, B, Robertson, C, *et al.* *Comparing analysis software versions in infant multiple breath washout* in *European Respiratory Society Annual Congress* 42: (2013), 1099.
59. Vuillermin, P, Saffery, R, Allen, KJ, *et al.* Cohort profile: the Barwon Infant Study. *International Journal of Epidemiology* 44: 1148–1160 (2015).
60. Bennett, J & AutoIt Team. *AutoIt, Version 3.3.14.0* 2015.
61. Schmidt, M, Foitzik, B, Wauer, RR, *et al.* Comparative investigations of algorithms for the detection of breaths in newborns with disturbed respiratory signals. *Computers and Biomedical Research* 31: 413–425 (1998).
62. Ndd Medizintechnik AG. *WBreath Application Help, Version 3.19.6.0* Duernten, Switzerland, 2007.
63. Harnoncourt, K & Buess, C. *Method for non-cooperative lung function diagnosis using ultrasound. U.S Patent 7635339 B2.* 2009. <http://www.google.com/patents/US7635339http://patentimages.storage.googleapis.com/pdfs/US7635339.pdf>.
64. Latzin, P, Sauteur, L, Thamrin, C, *et al.* Optimized temperature and deadspace correction improve analysis of multiple breath washout measurements by ultrasonic flowmeter in infants. *Pediatric Pulmonology* 42: 888–97 (2007).
65. Schibler, A & Henning, R. Measurement of functional residual capacity in rabbits and children using an ultrasonic flow meter. *Pediatric Research* 49: 581–8 (2001).
66. Singer, F, Kieninger, E, Abbas, C, *et al.* Practicability of nitrogen multiple-breath washout measurements in a pediatric cystic fibrosis outpatient setting. *Pediatric Pulmonology* 48: 739–746 (2013).
67. Wilson, AJ, Franks, CI & Freeston, IL. Algorithms for the detection of breaths from respiratory waveform recordings of infants. *Medical and Biological Engineering and Computing* 20: 286–292 (1982).
68. Centers for Disease Control. *CDC growth charts* 2016. https://www.cdc.gov/growthcharts/cdc_charts.htm.

69. Stan Development Team. *Stan Modeling Language: User's Guide and Reference Manual. Version 2.12.0*. 2016.
70. Gelman, A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1: 515–533 (2006).
71. Stan Development Team. *The Stan C++ Library, Version 2.10.0* 2016. <http://mc-stan.org>.
72. R Core Team. *R: a language and environment for statistical computing* Vienna, Austria, 2016.
73. Gelman, A, Carlin, JB, Stern, HS, *et al.* *Bayesian Data Analysis* 3rd ed. (Chapman & Hall/CRC, Boca Raton, Florida, 2014).
74. Yammine, S, Singer, F, Abbas, C, *et al.* Multiple-breath washout measurements can be significantly shortened in children. *Thorax* 68: 586–587 (2013).
75. Ahmad, F, Irving, S, Bush, A, *et al.* Multiple breath washouts in children can be significantly shortened without compromising measurement quality. *Thorax* 69: A166–A167 (2014).
76. Singer, F, Yammine, S, Abbas, C, *et al.* Ways to shorten the lung clearance index measurement II—how long to wash? in *European Respiratory Society Annual Congress 2012* 40: (2012), 3298.
77. Robinson, PD, Stocks, J, Aurora, P, *et al.* Abbreviated multi-breath washout for calculation of lung clearance index. *Pediatric Pulmonology* 48: 336–343 (2013).
78. Mahar, RK, Carlin, JB, Ranganathan, S, *et al.* Bayesian modelling of lung function data from multiple-breath washout tests. *Statistics in Medicine* 37: 2016–2033 (2018).
79. Baldwin, D. *Impact of maturation and disease on respiratory system dynamics in infants. Phd Thesis*. (University of Western Australia, Australia, 2008).
80. Fauroux, B & Lofaso, F. Control of breathing. *European Respiratory Monograph* 47: 240–250 (2010).
81. Frey, U, Maksym, G & Suki, B. Temporal complexity in clinical manifestations of lung disease. *Journal of Applied Physiology* 110: 1723–1731 (2011).
82. Pearce, N & Merletti, F. Complexity, simplicity, and epidemiology. *International Journal of Epidemiology* 35: 515–519 (2006).

83. Rickles, D, Hawe, P & Shiell, A. A simple guide to chaos and complexity. *Journal of Epidemiology and Community Health* 61: 933–937 (2007).
84. Goldberger, AL. Non-linear dynamics for clinicians: chaos theory, fractals, and complexity at the bedside. *The Lancet* 347: 1312–1314 (1996).
85. Glenny, RW, Robertson, HT, Yamashiro, S, *et al.* Application of fractal analysis to physiology. *Journal of Applied Physiology* 70: 2351–2367 (1991).
86. Thamrin, C & Frey, U. Complexity and respiratory growth: a developing story. *Journal of Applied Physiology* 106: 753–754 (2009).
87. Thamrin, C & Stern, G. New methods: what do they tell us? Fluctuation analysis of lung function. *European Respiratory Monograph* 47: 310–324 (2010).
88. Zhang, Q, Suki, B & Lutchen, KR. Harmonic distortion from nonlinear systems with broadband inputs: applications to lung mechanics. *Annals of Biomedical Engineering* 23: 672–681 (1995).
89. Suki, B, Hantos, Z, Daróczy, B, *et al.* Nonlinearity and harmonic distortion of dog lungs measured by low-frequency forced oscillations. *Journal of Applied Physiology* 71: 69–75 (1991).
90. Thamrin, C, Stern, G, Strippoli, M-PF, *et al.* Fluctuation analysis of lung function as a predictor of long-term response to β 2-agonists. *European Respiratory Journal* 33: 486–493 (2009).
91. McCall, CB, Hyatt, RE, Noble, FW, *et al.* Harmonic content of certain respiratory flow phenomena of normal individuals. *Journal of Applied Physiology* 10: 215–218 (1957).
92. Delignieres, D, Ramdani, S, Lemoine, L, *et al.* Fractal analyses for ‘short’ time series: a re-assessment of classical methods. *Journal of Mathematical Psychology* 50: 525–544 (2006).
93. Priestley, M. *Estimation in the frequency domain in Spectral Analysis and Time Series* 389–449 (Academic Press, London, 1981).
94. Shumway, RH & Stoffer, DS. *Time Series Analysis and Its Applications with R Examples* 2nd ed. (Springer, New York, 2006).
95. Timmer, J & König, M. On generating power law noise. *Astronomy and Astrophysics* 300: 707–710 (1995).

96. Voss, RF. *Fractals in nature: from characterization to simulation* in *The Science of Fractal Images* 21–70 (Springer, New York, 1988).
97. Mandelbrot, B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science* 156: 636–638 (1967).
98. Eke, A, Hermán, P, Bassingthwaite, J, *et al.* Physiological time series: distinguishing fractal noises from motions. *Pflügers Archiv—European Journal of Physiology* 439: 403–415 (2000).
99. Riley, MA, Bonnette, S, Kuznetsov, N, *et al.* A tutorial introduction to adaptive fractal analysis. *Frontiers in Physiology* 3: 1–10 (2012).
100. Eke, A. Pitfalls in fractal time series analysis: fMRI BOLD as an exemplary case. *Frontiers in Physiology* 3: 1–24 (2012).
101. Peng, C, Buldyrev, SV, Havlin, S, *et al.* Mosaic organization of DNA nucleotides. *Physical Review E* 49: 1685–1689 (1994).
102. Peng, C, Havlin, S, Stanley, HE, *et al.* Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos* 5: 82–87 (1995).
103. Peng, C-K, Mietus, JE, Liu, Y, *et al.* Quantifying fractal dynamics of human respiration: age and gender effects. *Annals of Biomedical Engineering* 30: 683–692 (2002).
104. Kirchner, M, Schubert, P, Liebherr, M, *et al.* Detrended fluctuation analysis and adaptive fractal analysis of stride time data in parkinson’s disease: stitching together short gait trials. *PLOS ONE* 9: e85787 (2014).
105. Chen, Z, Ivanov, PC, Hu, K, *et al.* Effect of nonstationarities on detrended fluctuation analysis. *Physical Review E* 65: 041107 (2002).
106. Hausdorff, JM. Gait dynamics in Parkinson’s disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling. *Chaos* 19: 026113 (2009).
107. Fan, J & Yao, Q. *Spectral density estimation and its applications* in *Nonlinear Time Series: Nonparametric and Parametric Methods* 275–310 (Springer-Verlag, New York, 2003).
108. Papadakis, I & Lawrence, A. Improved methods for power spectrum modelling of red noise. *Monthly Notices of the Royal Astronomical Society* 624: 612–624 (1993).

109. Zhao-Gua, C & Hannan, E. The distribution of periodogram ordinates. *Journal of Time Series Analysis* 1: 73–82 (1980).
110. Ricci, L, Verardi, V & Vermandele, C. A highly efficient regression estimator for skewed and/or heavy-tailed distributed errors. *Working Papers from the European Stability Mechanism* 19: 1–8 (2016).
111. Done, C, Madejski, GM, Mushotzky, RF, *et al.* The X-ray variability of NGC 6814: power spectrum. *The Astrophysical Journal* 400: 138–152 (1992).
112. Emmanoulopoulos, D, McHardy, IM & Papadakis, IE. Generating artificial light curves: revisited and updated. *Monthly Notices of the Royal Astronomical Society* 433: 907–927 (2013).
113. Singer, F, Stern, G, Thamrin, C, *et al.* Tidal volume single breath washout of two tracer gases—a practical and promising lung function test. *PLoS ONE* 6: e17588 (2011).
114. Fleming, S, Thompson, M, Stevens, R, *et al.* Normal ranges of heart rate and respiratory rate in children from birth to 18 years of age: a systematic review of observational studies. *The Lancet* 377: 1011–1018 (2011).
115. Morris, TP, White, IR & Crowther, MJ. Using simulation studies to evaluate statistical methods. *Statistics in Medicine* 38: 2074–2102 (2019).
116. Ranganathan, S, Goetz, I, Hoo, A-F, *et al.* Assessment of tidal breathing parameters in infants with cystic fibrosis. *European Respiratory Journal* 22: 761–766 (2003).
117. Gustafsson, PM, Bengtsson, L, Lindblad, A, *et al.* The effect of inert gas choice on multiple breath washout in healthy infants: differences in lung function outcomes and breathing pattern. *Journal of Applied Physiology* 123: 1545–1554 (2017).
118. Moore, C, Horton, E, Benseler, N, *et al.* The effects of 100% oxygen on breathing pattern are not limited to young children. *European Respiratory Journal* 51: 1702055 (2018).
119. Carstensen, B. Comparing and predicting between several methods of measurement. *Biostatistics* 5: 399–413 (2004).
120. Carstensen, B. Comparing methods of measurement: extending the LoA by regression. *Statistics in Medicine* 29: 401–410 (2009).

121. Bland, JM & Altman, DG. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research* 8: 135–160 (1999).
122. Bland, J & Altman, DG. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet* 327: 307–310 (1986).
123. Carstensen, B, Simpson, J & Gurrin, LC. Statistical models for assessing agreement in method comparison studies with replicate measurements. *The International Journal of Biostatistics* 4: Article 16 (2008).
124. Lewandowski, D, Kurowicka, D & Joe, H. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100: 1989–2001 (2009).
125. Sorensen, T & Vasishth, S. Bayesian linear mixed models using Stan: a tutorial for psychologists, linguists, and cognitive scientists. *arXiv:1506.06201* (2015).
126. Stan Development Team. *RStan: the R interface to Stan, Version 2.10.0*. 2015.
127. *Stan Modeling Language: User's Guide and Reference Manual. Version 2.12.0*. 2016.
128. Poncin, W, Singer, F, Aubriot, AS, *et al.* Agreement between multiple-breath nitrogen washout systems in children and adults. *Journal of Cystic Fibrosis* 16: 258–266 (2017).
129. Kentgens, A-C, Guidi, M, Korten, I, *et al.* Infant multiple breath washout using a new commercially available device: Ready to replace the previous setup? *Pediatric Pulmonology* 53: 628–635 (2018).
130. Kentgens, A-C, Anagnostopoulou, P, Singer, F, *et al.* Unexpected results in infant multiple breath washout: is the truth hidden in the washin? *Pediatric Pulmonology* 52: 717–719 (2017).
131. Tonga, KO, Robinson, PD, Farah, CS, *et al.* In vitro and in vivo functional residual capacity comparisons between multiple-breath nitrogen washout devices. *ERJ Open Research* 3: 00011–2017 (2017).
132. Robert, C & Casella, G. A short history of Markov chain Monte Carlo: subjective recollections from incomplete data. *Statistical Science* 26: 102–115 (2011).
133. Neal, RM. *MCMC using Hamiltonian dynamics* in *Handbook of Markov Chain Monte Carlo* 2nd ed., 113–162 (Chapman & Hall/CRC, Boca Raton, Florida, 2011).

134. Monnahan, CC, Thorson, JT & Branch, TA. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution* 8: 339–348 (2017).
135. Betancourt, M & Girolami, M. Hamiltonian Monte Carlo for hierarchical models. *arXiv.1312.0906* (2015).
136. Gelman, A & Shirley, K. *Inference from simulations and monitoring convergence* in *Handbook of Markov Chain Monte Carlo* 1st ed., 163–174 (Chapman & Hall/CRC, Boca Raton, Florida, 2011).
137. Gelman, A, Lee, D & Guo, J. Stan: a probabilistic programming language for bayesian inference and optimization. *Journal of Educational and Behavioral Statistics* 40: 530–543 (2015).
138. Carpenter, B, Gelman, A, Hoffman, M, *et al.* Stan: a probabilistic programming language. *Journal of Statistical Software* 20: 1–37 (2016).
139. Lunn, DJ, Thomas, A, Best, N, *et al.* WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 10: 325–337 (2000).
140. Plummer, M. *JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling* in *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (Vienna, Austria, 2003).
141. RStudio Team. *RStudio: integrated development environment for R, Version 1.1.383*. Vienna, Austria, 2016.

Appendix A

Computational methods

Computational Bayesian methods were used for most of the statistical analysis performed throughout this thesis. In large part this was driven by the relatively complex structure of our MBW data. For example, our data was typically collected over time and with repeated measurements on individuals at each time point. We found our proposed models for this data were most naturally handled in a Bayesian setting.

This chapter provides a brief overview of Bayesian methods, models and implementations used in chapters 3, 4, and 6. It serves as both a reference as well as a foundation for many of the analyses found in the main thesis body and the other technical appendices.

Section A.1 outlines the principles of Bayesian inference. Section A.2 describes common computational methods used to perform modern Bayesian analyses and section A.3 describes their practical implementation. Section A.3.2 describes the tools used in the production of this thesis to support reproducible and efficient statistical analysis.

A.1 Bayesian data analysis

Although the principles of Bayesian methods have been detailed elsewhere, it is useful to outline them here before discussing our motivation for taking this approach and the different computational methods that were used to perform the analyses.

This section relies heavily on the material contained in Gelman et al. [73] and the Stan Development Team [69]. More detailed aspects of each analysis performed in this thesis are considered within each relevant chapter.

A.1.1 Bayesian principles

Bayesian inference begins with a full probability model of the problem at hand. This model takes the form of a joint probability distribution of the observable and unobservable quantities that are related to the problem. We adopt Gelman et al.'s [73] notation where $p(\cdot | \cdot)$ denotes the conditional probability distribution and $p(\cdot)$ similarly denotes a marginal distribution for both (continuous) probability density functions and (discrete) probability mass functions. In the simplest terms, for an unobserved parameter (or set of parameters) θ , and observed data y , we can express the joint probability distribution as the product of the ‘prior’ distribution of the parameter $p(\theta)$ and the likelihood $p(y | \theta)$:

$$p(\theta, y) = p(y | \theta)p(\theta) \quad (\text{A.1})$$

With the joint distribution defined, we can now condition this probability model on observed data to form a ‘posterior’ probability distribution of the unobservable quantities of the problem, given the evidence in the form of the observable data. Using the same notation, the posterior is as follows, using Bayes’ Rule:

$$p(\theta | y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y | \theta)p(\theta)}{\int p(y | \theta)p(\theta)d\theta} \quad (\text{A.2})$$

Note that $p(y) = \sum_{\theta} p(y | \theta)p(\theta)$ if θ is discrete. The factor $p(y)$ simply expresses that we are summing over all possible values of θ . Omitting this allows us to restate the posterior distribution in an unnormalised form, which is often more convenient:

$$p(\theta | y) \propto p(y | \theta)p(\theta). \quad (\text{A.3})$$

Bayesian analysis typically proceeds by carefully considering the problem, exploring data, reviewing previous knowledge, and obtaining input from domain experts. A full probability

model is then developed and estimated. The posterior predictive distribution is then assessed in context of the observable data, the substantive conclusions, and the sensitivity of the model fit to the assumptions underlying the probability model used for the analysis.

A.1.2 Motivation

Our main motivation in taking a Bayesian approach to modelling reflects many of the strengths of the approach as cited by Gelman et al. [73], namely that:

- information can be easily combined from multiple sources,
- randomness can be incorporated at multiple levels which lends itself well to the analysis of repeated measures, a key feature of much of our data,
- it allows the sensitivity of inferences to be assessed based on different prior assumptions and probability models, and
- inferences are based upon probability distributions, often summarised by a credible interval, rather than point estimates (with credible intervals taking on an intuitive probabilistic interpretation).

A.2 Markov chain Monte Carlo methods

Obtaining the exact posterior distribution analytically is often not realistic outside relatively simple problems. Analytical solutions are often non-existent, or algebraically infeasible, when working with real-world models. In these cases numerical methods such as Markov chain Monte Carlo (MCMC) can be used to obtain approximate posterior distributions. Although the first MCMC methods were developed in the 1950s, it was not until the early 1990s that they were routinely used in applied contexts [132]. The following section describes some of these methods and builds toward a description of the Hamiltonian Monte Carlo (HMC) method, a recently developed algorithm which is used throughout this thesis. This section relies heavily on the material presented in Gelman et al. [73].

Commonly used and relatively simple MCMC algorithms include Gibbs sampling and the Metropolis algorithm. These algorithms have been modified or extended, for example the

Metropolis-Hastings algorithm and HMC [133] With the advent of specialised software to implement these methods and their variants, discussed in Section A.3, it is not essential to have a complete understanding of the operating characteristics of the algorithms when working with simple models. However, real world models are often complex and may only be estimated properly using MCMC after appropriate model transformations or modifications to the MCMC operating parameters. Therefore, at a minimum, some understanding of the algorithms and their implementation is required for practical applications. We therefore briefly describe Gibbs sampling and the Metropolis algorithm before providing an overview to the HMC method as they appear in Gelman et al. [73]).

A.2.1 Gibbs sampling

Gibbs sampling is perhaps the most conceptually simple of the MCMC algorithms. Rather than sampling directly from the posterior distribution, it samples from the conditional posterior distributions iteratively. Suppose we can partition the parameter vector θ into a collection of d subvectors, $\theta = (\theta_1, \dots, \theta_d)$. Gibbs sampling samples each subvector conditional on the others.

Specifically, each subvector θ_j^t at iteration t with $j \in [1, d]$ is updated conditional on the most recent iteration of all subvectors complementary to it. Thus, for a subvector θ_j with complement θ_{-j} :

$$p(\theta_j^t \mid \theta_{-j}^{t-1}, y)$$

For a two parameter model, choosing an arbitrary initial starting point to sample from $p(\theta_1 \mid \theta_2^0, y)$, Gibbs sampler would then sample from $p(\theta_2^t \mid \theta_1^{t-1}, y)$, followed by $p(\theta_1^{t-1} \mid \theta_2^t, y)$, and so on.

An important requirement for Gibbs sampling is the ability to sample from each of these conditional distributions. If this is not possible, other types of MCMC need to be considered.

A.2.2 Metropolis algorithm

The Metropolis algorithm is the most basic form of the general Metropolis-Hastings algorithm and is more general than Gibbs sampling [73].

The Metropolis algorithm begins with a starting distribution $p_0(\theta)$ from which θ^0 is sampled and satisfies $p(\theta^0 | y) > 0$. For time $t = (1, 2, \dots)$ the algorithm proceeds by sampling a proposal θ^* from a proposal distribution, $J_t(\theta^* | \theta^{t-1})$. The next value θ^t is determined by taking the ratio of the two densities

$$r = \frac{p(\theta^* | y)}{p(\theta^{t-1} | y)} \quad (\text{A.4})$$

and setting:

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise} \end{cases}$$

This acceptance rule allows the algorithm to converge toward the stationary target distribution, and explore it once there, by favouring proposals θ^* with greater probability than the current value of θ . In its specific form, the Metropolis algorithm requires that the proposal distribution is symmetric, that is, that $J_t(\theta_a | \theta_b) = J_t(\theta_b | \theta_a)$ for all θ_a , θ_b , and t . In the more general Metropolis-Hastings algorithm, the proposal transitions are not required to be symmetric.

A.2.3 Hamiltonian Monte Carlo (HMC)

HMC is a generalisation of the Metropolis algorithm that mitigates some of the latter's relative inefficiencies due to its random walk behaviour by incorporating ideas from physics [73, 133–135].

The idea behind HMC is to treat the parameter vector θ_j as if it were a coordinate vector for a particle in space with an associated momentum vector $\phi \sim \text{MVN}(0, M)$, where M denotes a covariance matrix, which is usually diagonal. HMC proceeds by simultaneously updating θ and ϕ using a discrete approximation of physical Hamiltonian dynamics which describe the position and momentum of the particle in continuous time [73]. The technical

details of this approximation can be found elsewhere including Gelman et al. [73], Neal [133], Monnahan et al. [134], and Betancourt [135]. The algorithm proceeds in a similar fashion to Metropolis, although we now compute:

$$r = \frac{p(\theta^* | y)p(\phi^*)}{p(\theta^{t-1} | y)p(\phi^{t-1})}, \quad (\text{A.5})$$

and again setting:

$$\theta^t = \begin{cases} \theta^* & \text{with probability } \min(r, 1) \\ \theta^{t-1} & \text{otherwise} \end{cases}$$

It should be clear that there are similarities between the HMC and Metropolis algorithms. The main difference is in how the jumping distribution is determined. In HMC's case, using Hamiltonian dynamics to determine the jumping distribution allows the target distribution to be explored more rapidly than either Gibbs or Metropolis sampling schemes [73].

A.2.4 Performing MCMC and diagnostics

There are a number of best practices to follow when performing MCMC and conducting inference on the resulting posterior distributions [69, 73, 136]. These practices aim to ensure that the MCMC chains have converged to the target distribution of interest and that the effective number of samples is sufficient to perform valid inference. In this thesis we have adhered to these best practices, where possible, by taking the following approach to performing MCMC:

- *Running multiple chains.* Multiple chains are required to assess whether convergence has been achieved. A single chain may be stable and appear to be exploring the parameter space well in isolation, but may be doing so far from the actual target distribution. Using multiple chains makes it more likely that, if we observe that all chains have converged, we are exploring the target distribution properly. Our approach was always to use four chains and, typically, run the simulation for 2,000 samples.

- *Setting overdispersed initial values.* Setting overdispersed initial values of the MCMC chains allows each chain to approach the target distribution from a different region of parameter space. If all parameters converge from various initial values, we can be more confident that stationary and mixing chains have converged to the target distribution. In some cases, for example hierarchical models with constrained variables, it was necessary to set initial values to some approximate summary of the target distribution.
- *Discarding initial samples.* It is common practice to discard initial MCMC samples, typically known as ‘warm-up’ or ‘burn-in’ iterations, to remove the influence of the initial values. We tended to adopt the conservative approach and discarded the first half of the samples from each chain [69, 73]. If running four chains for 2,000 samples, this results in a total of 4,000 samples from the posterior distribution.
- *Assessing mixing and stationarity.* We can assess convergence by observing the mixing and stationarity characteristics of the MCMC chains. Briefly, this is done by splitting each chain (after discarding the warm-up samples) at the halfway point and checking that the half-chains have mixed. If the chains are mixing well and stationary, each half-chain should traverse the same distribution. This approach is formalised by a ‘scale reduction factor’, $\hat{R}$, which should be between 1 and 1.1 for all model parameters (see Gelman et al. for derivation [73]).
- *Ensuring sufficient effective sample size.* Even though we may have 4,000 samples from the posterior distribution, due to the iterative nature of MCMC, samples are often heavily autocorrelated, reducing sampling efficiency. We therefore conduct inference assuming a sufficient ‘effective’ sample size. We took the approach of Gelman et al. [73], who state that an effective sample size should be at least ten times the number of chains used. Therefore with four chains a minimum acceptable effective sample size is 40.
- *Performing graphical diagnostics.* Graphical diagnostics were the main method for debugging and refining our models. The main types of diagnostic plots used were ‘trace’ plots, which show the value of each MCMC chain across its iterations, and ‘pairs’ plots, which show the bivariate distributions of selected parameters in matrix form. Beyond debugging, graphical checking is useful to assess convergence and mixing of the chains. Using summary methods, such as the $\hat{R}$, is far from foolproof.

In some cases $\hat{R}$ can be close to one, yet the MCMC simulations exhibits pathologies, such as chains becoming ‘stuck’ in areas of the distribution, that may indicate a problem with the model.

- *Minimising divergent iterations.* Divergent transition are specific to HMC sampling. A divergent transition occurs when the HMC sampler encounters a region of extreme curvature in the target distribution which it cannot adequately explore. This type of extreme curvature is common in hierarchical settings, where complex posterior geometries are common. It is desirable to eliminate divergent transitions where possible otherwise proper inference may be compromised [69]. The two most common ways of eliminating divergent transitions are reparameterisation and modification of the sampler operating parameters (see Gelman et al. [73] and the Stan Development Team [69] for details).

A.2.5 Model checking

Assessing the performance of our MCMC chains, as described in the previous section, only confirms that our simulation has converged to the target distribution as determined by the model. However, if the model is poor the inferences may not make much sense. It is therefore necessary to validate the model fit in the context of the observed data.

A common approach to assess the fit of Bayesian models is posterior predictive checking. This involves generating replicated data under the specified model and priors, known as the posterior predictive distribution, and comparing it with the observed data. Generating this replicated data is accomplished by simulating from the joint posterior predictive distribution which, continuing with notation from earlier sections, is defined as:

$$p(y^{rep} | y) = \int p(y^{rep} | \theta) p(\theta | y) d\theta, \quad (\text{A.6})$$

where y^{rep} is the replicated data.

A.3 Computational tools

A.3.1 Stan

Rather than implement an MCMC sampler from scratch, it is common in applied problems to use existing tools to perform analyses. These tools are often in the form of domain-specific programming languages that allow high-level control of some form of sampler. There are a number of these tools available for use. We performed our Bayesian analyses using Stan [71, 137].

In the words of its developers, Stan is a domain-specific ‘probabilistic programming language’ for specifying statistical models [138]. Specifically, Stan allows full Bayesian inference to be conducted for continuous variable models using the HMC method. It is similar to other programming languages such as BUGS (Bayesian Inference using Gibbs Sampling) [139] and JAGS (Just Another Gibbs Sampler) [140].

Stan is relatively new and was developed at Columbia University. The first version was released in 2012. At the time of writing it has reached version 2.16. Compared to BUGS or JAGS it is unique in that it has highly developed interfaces for most common statistical languages and packages, and can be run in Windows, Mac OS X, or Linux operating systems. It is completely free, open-source, and has an active development and user community with, at the time of writing, an online development repository (<https://github.com/stan-dev/stan>) and active user forum (<http://mc-stan.org/users/>).

In practice, a Stan program (of which a simple example is shown in Listing A.1) is compiled to a binary for efficient computation which then performs HMC using the no-U-turn sampler [73, 137] to simulate samples from the posterior distribution defined by the Stan program and the data. These posterior samples are returned, allowing them to be summarised to assess convergence and perform inference. Usefully, posterior predictive distributions can be automatically simulated from within a Stan program, making model validation relatively straightforward.

A Stan program is written in blocks named as follows: data, transformed data, parameter, transformed parameter, model, and generated quantities. Within these blocks are variable type declarations and statements corresponding to their usage. The `model` block is used to define components of the probability model, commonly using the distributional

LISTING A.1: Stan code corresponding to a simple linear regression model

```
// Define data.
data {
  vector<lower=0>[20] x; // Dependent variable.
  vector<lower=0>[20] y; // Independent variable.
}

// Define model parameters.
parameters {
  real alpha; // Slope.
  real beta; // Intercept.
  real sigma; // Standard deviation.
}

// Define model likelihood and priors.
model {
  // Create local variable to contain mean function.
  real mu;
  mu = alpha * x + beta;

  // Specify priors
  alpha ~ normal(0, 100);
  beta ~ normal(0, 100);
  sigma ~ normal(0, 100);

  // Specify the likelihood.
  y ~ normal(mu, sigma);
}

// Generate posterior predictive distribution.
generated quantities {
  real y_rep[N];
  y_rep = normal_rng(mu, sigma);
}
```

shorthand ‘ $\sim$ ’. Variable constraints are added when typing variables and the program is read imperatively (such that it is executed in order). The Stan program is compiled into a C++ program which is then re-compiled as a platform-specific executable (e.g. for Linux). This can either be run directly or via one of the provided interfaces (e.g. RStan [126]).

To illustrate, consider a simple example of a Stan program shown in Listing A.1. This is for a simple linear regression of y_i on x_i with $i = (1, \dots, 20)$ and diffuse normal prior distributions,

$$y_i = \alpha x + \beta + e_i, \quad e_i \sim N(0, \sigma^2).$$

Note that this Stan program also produces the posterior predictive distribution in the `generated quantities` block.

To the user, the main practical difference between Stan and BUGS or JAGS is its relative flexibility, and its use of HMC rather than Gibbs sampling or Metropolis-Hastings.

Compared to these other methods, HMC has been shown to be more robust, flexible, and efficient, although the difference is less apparent when working with conjugate models and posteriors with low correlation [137, 138].

A.3.2 Ancillary computational resources

Modern statistical analysis—and not necessarily just Bayesian data analysis—often requires the use of multiple computational resources beyond that of simply a desktop statistics package. These include, but are not limited to, underlying operating systems, statistical programming environments, version control systems, and parallel computation. Many of the analyses in this thesis required the analysis of a large number of data-rich MBW tests, and this guided our choice of what computational resources to use.

A.3.2.1 The R programming environment

We decided early on to use the R programming environment as the workhorse statistical software (i.e. for data management and cleaning, exploratory data analysis, scripting, simple statistical analysis, and producing graphics) (<https://www.r-project.org>) [72]. The choice was driven by the fact that R is free and open-source, capable of handling relatively large datasets, extensively documented, and has a large and supportive online community. It is fairly widely accepted as one of the most commonly used programming environments in data science, and as such has a large number of up-to-date, and in many cases state-of-the-art, statistical software libraries. While other software environments, such as Stata (<https://www.stata.com>), SAS (<https://www.sas.com/>), or Python (<https://www.python.org>), fulfil at least some of these criteria they often lack in at least one. For example, Stata and SAS are neither free nor open-source. Python, on the other hand, is free, relatively easy to use and well-documented, and has a large online community, however does not have the same degree of statistical sophistication compared to R. Helpfully, a free integrated development environment (IDE) exists for R, called RStudio [141], which was used to write and edit the majority of the software used for this thesis.

A.3.2.2 Version control system

In line with Dr Vukcevic's (PhD co-supervisor) approach, and the data scientific community in general, a version control system for all software and scripts was implemented using Git and the collaborative remote version control services Bitbucket and GitHub (<https://bitbucket.org>, <https://github.com>).

Not only did using this approach ensure that our work was securely backed-up at all times at a remote site, it also made it straightforward to return to previous versions of code and, perhaps above all, work collaboratively with others in a convenient way. Further, this approach was very useful when operating on remote servers or between machines, as we could use it to easily and reliably synchronise code between machines.

A.3.2.3 Murdoch Children's Research Institute (MCRI) High Performance Cluster

Although the Bayesian approach has many strengths, it is computationally intensive. This is particularly so when working with either complex or multi-level models, or performing large amounts of sequential analyses. Therefore, where possible, much of the analysis was performed using MCRI's High Performance Cluster (HPC).

A HPC is effectively a large collection of computers with a central interface to use them in tandem. Some large and complex problems can be simplified into a large number of smaller single problems, which can be solved individually in parallel by sending each smaller problem to be solved on a separate processing unit. This approach is typically known as parallel computing, and while it can be performed in some circumstances on a smaller scale, for example using the `mcpParallel` R package to distribute jobs on a single, Unix-based personal computer with more than two processing units, its real strength is when a large number of dedicated processing units are available in a HPC.

I had access to MCRI's HPC, which had 8 compute nodes each with 32 processing units comprising 256 gigabytes of RAM. In practice, parallelisable tasks were placed in a queue using a script known as a portable batch system (`.pbs`) file. This script defined several parameters that would be used by the processing units when performing the task, for example, the maximum amount of RAM to be used and a maximum allowable job time, in

addition to the software environment used to perform the task (for example, R), where to find the input data, where to save the resulting output, and where to keep the task logs.

The HPC was essential to successfully performing some of the analyses, particularly those in Chapter 4, where using the cluster often brought the total computing time from weeks down to minutes.

A.3.2.4 Unix-based computing

On the advice of Dr Vukcevic, computing was done on a Unix-based operating system. Initially using Linux (Ubuntu distribution), albeit running a Windows 7 virtual machine for access to Microsoft products, before eventually to Mac OS X (now macOS) to have access to many of the powerful features of a Unix machine, for example the command line interface allowing text-based access to the operating system via the Unix-shell, while being able to use Microsoft products on the same machine.

The command line interface proved to be invaluable to our data analysis workflow. For example, command line input could be used interactively to manage our Git repositories, automate mundane tasks such as renaming or shifting large amounts of files, and perform tasks that are simple when done programmatically but messy when using a graphical user interface (GUI), for example, converting images between formats. Yet another benefit was that it facilitated direct access to MCRI's Unix servers, for example when running R scripts on the HPC, without resorting to a more cumbersome GUI. This was particularly useful when working on complex parallel analyses remotely via a virtual private network.

Appendix B

Stan model of MBW outcomes and breathing patterns

The variance components model described by Equation 3.1 in Chapter 3 was fit using the Stan model described in Listing B.1.

This model requires the use of Stan through one of its interfaces, e.g. `RStan`, `PyStan`, `CmdStan`. We used `RStan` (version 2.12.1) implemented in `R` (version 3.3.1). More information is available at <http://mc-stan.org/>.

Data for the individual should be a list containing the following elements:

- `N`: an integer count of the number data objects corresponding to an observed lung function quantity (i.e. LCI, FRC, or CEV) for an included test,
e.g. `int 200`
- `N_item`: an integer count of the number of ‘items’ or participants,
e.g. `int 50`
- `y`: a real-valued vector containing the observed lung function quantity (i.e. LCI, FRC, or CEV) for an included test,
e.g. `num [1:200] 6.7 6.9 7.4 5.8 ...`
- `item`: an integer valued vector containing the index number of each participant,
e.g. `num [1:200] 1 1 1 2 2 2 3 3 3 ...`

LISTING B.1: Stan code for variance components model described by (3.1)

```
data {  
  int<lower = 1>      N;  
  int<lower = 1>  N_item;  
  real              y[N];  
  int<lower = 1> item[N];  
}  
  
parameters {  
  real<lower = 0>      theta;  
  real<lower = 0> sigma_between;  
  real<lower = 0>      mu[N_item];  
  real<lower = 0>      sigma_within;  
}  
  
model {  
  mu ~ normal(theta, sigma_between);  
  for (i in 1:N){  
    y[i] ~ normal(mu[item[i]], sigma_within);  
  }  
}
```

Appendix C

Stan models of MBW data and agreement

This appendix contains supplementary material to Chapter 4 as it appears in:

Mahar, RK, Carlin, JB, Ranganathan, S, Ponsonby, AL, Vuillermin, P, Vukcevic, D. Bayesian modelling of repeated measures lung function data from multiple-breath washout tests. *Statistics in Medicine* 37: 2016–2033 (2018).

Supplementary material for “Bayesian modelling of lung function data from multiple-breath washout tests”

We include Stan code for four models included in the main manuscript:

- Individual MBW models:
 - Using the diffuse prior
(`mbw_individual_diffuse.stan`)
 - Using the informative prior
(`mbw_individual_informative.stan`)
- Linear mixed-effects (LME) models:
 - Exchangeable replicates and method-specific residuals
(`lme_exchangeable_separate.stan`)
 - Linked replicates and common residuals
(`lme_linked_common.stan`)

These models require the use of Stan through one of its interfaces, e.g. `RStan`, `PyStan`, `CmdStan`. We used `RStan` (version 2.12.1) implemented in R (version 3.3.1). More information is available at <http://mc-stan.org/>.

Data for the individual MBW models (`mbw_individual_diffuse.stan`, `mbw_individual_informative.stan`) should be a list containing the following elements (note that the model using diffuse priors does not require the elements `mu` and `Sigma`):

- `M`: an integer count of the number of washout breaths measured in this test,
e.g. `int 30`
- `k`: an integer-valued vector containing the index number of each breath (counting from the start of the washout),
e.g. `num [1:30] 0 1 2 3 4 5 6 7 8 9 ...`
- `gas`: a real-valued vector containing the tracer gas quantity (GAS) at the end of the breath indexed by `k`,
e.g. `num [1:30] 1 0.701 0.501 0.404 0.390 ...`

- **cevgm**: a real-valued vector containing the observed cumulative expired volume of gas mixture (CEVGM) at the end of the breath indexed by *k*,
e.g. `num [1:30] 0 36.2 72.0 107.0 141.4 ...`
- **cetgv**: a real-valued vector containing the observed cumulative expired volume of tracer gas (CETGV) at the end of the breath indexed by *k*,
e.g. `num [1:30] 0 22.1 36.4 46.9 55.0 ...`
- **mu**: a real-valued vector containing 6 elements corresponding to informative prior means for, in order, $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$,
e.g. `num [1:6] 0.67 0.13 0.14 ...`
- **Sigma**: an real-valued matrix containing the elements of the informative cross-variance matrix corresponding to the values in **mu**.

e.g. `num [1:6, 1:6]`

	<code>Sigma[,1]</code>	<code>Sigma[,2]</code>	<code>Sigma [,3]</code>	<code>...</code>
<code>Sigma[1,]</code>	0.024	0.001	0.027	<code>...</code>
<code>Sigma[2,]</code>	0.001	0.001	0.001	<code>...</code>
<code>Sigma[3,]</code>	0.027	0.001	0.048	<code>...</code>
<code>...</code>	<code>...</code>	<code>...</code>	<code>...</code>	<code>...</code>

Data for the LME models (`lme_exchangeable_separate.stan`, `lme_linked_common.stan`) should be a list containing the following elements (note that the LME model using exchangeable replicates does not require the vector `repl_ind`):

- **n_item**: an integer count of the number of items (participants) in the analysis,
e.g. `int 10`
- **n_all**: an integer count of the total amount of observations,
e.g. `int 100`
- **n_meth**: an integer count of the total number of methods to be compared,
e.g. `int 3`
- **y**: a real-valued vector containing the observations to compare,
e.g. `num [1:100] 111.1 100.1 120.2 110.8 180.9 ...`

- `item`: an integer-valued vector specifying the item (participant) index for each observation,
e.g. `int [1:100] 1 1 1 1 1 1 2 2 2 2 ...`
- `meth`: an integer-valued vector specifying the method index for each observation,
e.g. `int [1:100] 1 1 1 1 1 1 1 1 1 1 ...`
- `repl_ind`: an integer-valued vector specifying the replicate index for each observation (i.e. a separate index for each actual test conducted),
e.g. `int [1:100] 1 2 3 4 5 6 7 8 9 10 ...`

LISTING C.1: Stan code corresponding to `mbw_individual_diffuse.stan`

```

functions {
  /**
   * Evaluate the tracer gas quantity at a particular time point.
   * This is a two-component exponential decay model.
   *
   * @param t      Time point at which to evaluate the gas quantity
   * @param beta0  Mixture weight parameter
   * @param beta   Exponents for each component
   *
   * @return Gas quantity (between 0 and 1) at the given time point.
   */
  real gas_curve(real t, real beta0, vector beta) {
    return beta0 * exp(-beta[1] * t) + (1 - beta0) * exp(-beta[2] * t);
  }

  /**
   * Determine if the tracer gas quantity crosses a given threshold between
   * two time points, t1 and t2.
   *
   * @param t1      First time point (left end of desired time interval)
   * @param t2      Second time point (right end of desired time interval)
   * @param threshold Threshold for the gas quantity
   * @param beta0   Mixture weight parameter
   * @param beta    Exponents for the mixture components
   *
   * @return Boolean value: true if the threshold is crossed, false otherwise
   */
  int gas_passes_threshold(real t1, real t2, real threshold, real beta0,
                          vector beta) {
    return gas_curve(t1, beta0, beta) > threshold &&
           gas_curve(t2, beta0, beta) <= threshold;
  }
}

```

```

data {
  int          M ; // number of observations
  real         gas[M]; // tracer gas quantity (GAS)
  real         cevgm[M]; // cumulative expired volume of gas mixture (CEVGM)
  real         cevtg[M]; // cumulative expired volume of tracer gas (CEVTG)
  int          k[M]; // breath index
}

transformed data {
  int          nmax          ;
  real         cevgm_tf[M - 1];
  real         cevtg_tf[M - 1];

  // Resolution for line search used to determine the real-valued stopping
  // 'breath'.
  nmax = 100;

  // Transform cumulative data to log increment scale.
  for (m in 1:(M - 1)) {
    cevgm_tf[m] = log(cevgm[m + 1] - cevgm[m]);
    cevtg_tf[m] = log(cevtg[m + 1] - cevtg[m]);
  }
}

parameters {
  // GAS curve.
  real<lower=0, upper=1> beta0;
  positive_ordered[2]    beta; // a.k.a 'beta1' and 'beta2' (log scale)
  real<lower=0>          error_gas;

  // CEVGM curve.
  real<lower=0>          beta5;
  real<lower=0>          error_cevgm;

  // CEVTG curve.
  real<lower=0>          beta3;
  real<lower=0>          beta4;
  real<lower=0>          error_cevtg;
}

transformed parameters {
  // Stopping breaths.
  real<lower=0>          m_stop_40;
  real<lower=0>          m_stop_40_cont;

  // MBW quantities using standard stopping breath.
  real<lower=0>          lci40;

```

```

real<lower=0> cev40;

// MBW quantities using model-based stopping breath.
real<lower=0> lci40_cont;
real<lower=0> cev40_cont;

// Define MBW quantities from model parameters.
// Use a threshold of 1/40 (= 0.025) for the gas concentration.
for (m in 2:100) {
  if (gas_passes_threshold(m - 1, m, 0.025, beta0, beta)) {
    // Use the standard stopping breath, k(40).
    m_stop_40 = m;
    cev40 = beta5 * m_stop_40;
    lci40 = cev40 / beta3; // n.b. FRC = beta3

    // Use the model-based stopping 'breath', theta.
    // This stopping breath is real-valued rather than integer-valued.
    // It is determined by a simple line search.
    for (i in ((m - 1) * nmax):(m * nmax)) {
      if (gas_passes_threshold(1.0 / nmax * (i - 1),
                              1.0 / nmax * i, 0.025, beta0, beta)) {
        m_stop_40_cont = 1.0 / nmax * i;
        cev40_cont = beta5 * m_stop_40_cont;
        lci40_cont = cev40_cont / beta3; // n.b. FRC = beta3
      }
    }
  }
}

}

model {
  // Declare local variables.
  // These are used to store the mean and sd of the 3 curves, see below.
  vector[M]      gas_mean;
  vector[M - 1]  cevgm_mean;
  vector[M - 1]  cevtg_mean;
  vector[M]      gas_sd;
  vector[M - 1]  cevgm_sd;
  vector[M - 1]  cevtg_sd;

  // Priors (diffuse).

  beta0 ~ beta(2, 2); // a.k.a. beta0
  beta[1] ~ normal(0, 1);
  beta[2] ~ normal(0, 1);

  beta3 ~ normal(0, 1000);

```

```
beta4 ~ normal(0, 1);

beta5 ~ normal(0, 100);

error_gas ~ cauchy(0, 2.5);
error_cevgm ~ cauchy(0, 2.5);
error_cevtg ~ cauchy(0, 2.5);

// Calculate the mean and sd of the 3 curves.
for (m in 1:M) {
  gas_mean[m] = gas_curve(k[m], beta0, beta);
  gas_sd[m] = error_gas;
}
for (m in 1:(M - 1)) {
  cevgm_mean[m] = log(beta5);
  cevtg_mean[m] = log(beta3) + log(1 - exp(-beta4)) - (beta4 * (k[m]));

  cevgm_sd[m] = error_cevgm;
  cevtg_sd[m] = error_cevtg;
}

// Define the probability models for the 3 curves.
gas ~ lognormal(log(gas_mean), gas_sd);
cevgm_tf ~ normal(cevgm_mean, cevgm_sd);
cevtg_tf ~ normal(cevtg_mean, cevtg_sd);
}
```

LISTING C.2: Stan code corresponding to `mbw_individual_informative.stan`

```

functions {
  /**
   * Evaluate the tracer gas quantity at a particular time point.
   * This is a two-component exponential decay model.
   *
   * @param t      Time point at which to evaluate the gas quantity
   * @param beta0  Mixture weight parameter (a.k.a. beta0)
   * @param beta   Exponents for each component
   *
   * @return Gas quantity (between 0 and 1) at the given time point.
   */
  real gas_curve(real t, real beta0, vector beta) {
    return beta0 * exp(-beta[1] * t) + (1 - beta0) * exp(-beta[2] * t);
  }

  /**
   * Determine if the tracer gas quantity crosses a given threshold between
   * two time points, t1 and t2.
   *
   * @param t1      First time point (left end of desired time interval)
   * @param t2      Second time point (right end of desired time interval)
   * @param threshold Threshold for the gas quantity
   * @param beta0   Mixture weight parameter (a.k.a. beta0)
   * @param beta    Exponents for the mixture components
   *
   * @return Boolean value: true if the threshold is crossed, false otherwise
   */
  int gas_passes_threshold(real t1, real t2, real threshold, real beta0,
                          vector beta) {
    return gas_curve(t1, beta0, beta) > threshold &&
           gas_curve(t2, beta0, beta) <= threshold;
  }
}

data {
  int      M ; // number of observations
  real     gas[M]; // tracer gas quantity (GAS)
  real     cevgm[M]; // cumulative expired volume of gas mixture (CEVGM)
  real     cevtg[M]; // cumulative expired volume of tracer gas (CEVTG)
  int      k[M]; // breath index
  vector[6] mu; // informative prior means
  matrix[6, 6] Sigma; // informative cross-covariance matrix
}

transformed data {
  int      nmax      ;

```

```

real  cevgm_tf[M - 1];
real  cevtg_tf[M - 1];

// Resolution for line search used to determine the real-valued stopping
// 'breath'.
nmax = 100;

// Transform cumulative data to log increment scale.
for (m in 1:(M - 1)) {
  cevgm_tf[m] = log(cevgm[m + 1] - cevgm[m]);
  cevtg_tf[m] = log(cevtg[m + 1] - cevtg[m]);
}
}

parameters {
  // GAS curve.
  real<lower=0, upper=1>  beta0;
  positive_ordered[2]     beta; // a.k.a 'beta1' and 'beta2'
  real<lower=0>           error_gas;

  // CEVGM curve.
  real<lower=0>           beta5;
  real<lower=0>  error_cevgm;

  // CEVTG curve.
  real<lower=0>           beta3;
  real<lower=0>           beta4;
  real<lower=0>  error_cevtg;
}

transformed parameters {

  // Vector to keep different parameters in when specifying multi_normal prior.
  vector[6] zeta;

  // Stopping breaths.
  real<lower=0>  m_stop_40;
  real<lower=0>  m_stop_40_cont;

  // MBW quantities using standard stopping breath.
  real<lower=0>  lci40;
  real<lower=0>  cev40;

  // MBW quantities using model-based stopping breath.
  real<lower=0>  lci40_cont;
  real<lower=0>  cev40_cont;

  // Store parameters in single vector.

```

```

zeta[1] = beta0;
zeta[2] = beta[1];
zeta[3] = beta[2];
zeta[4] = beta3;
zeta[5] = beta4;
zeta[6] = beta5;

// Define MBW quantities from model parameters.
// Use a threshold of 1/40 (= 0.025) for the gas concentration.
for (m in 2:100) {
  if (gas_passes_threshold(m - 1, m, 0.025, beta0, beta)) {
    // Use the standard stopping breath, k(40).
    m_stop_40 = m;
    cev40 = beta5 * m_stop_40;
    lci40 = cev40 / beta3; // n.b. FRC = beta3

    // Use the model-based stopping 'breath', theta.
    // This stopping breath is real-valued rather than integer-valued.
    // It is determined by a simple line search.
    for (i in ((m - 1) * nmax):(m * nmax)) {
      if (gas_passes_threshold(1.0 / nmax * (i - 1),
                              1.0 / nmax * i,      0.025, beta0, beta)) {
        m_stop_40_cont = 1.0 / nmax * i;
        cev40_cont = beta5 * m_stop_40_cont;
        lci40_cont = cev40_cont / beta3; // n.b. FRC = beta3
      }
    }
  }
}

model {
  // Declare local variables.
  // These are used to store the mean and sd of the 3 curves, see below.
  vector[M]      gas_mean;
  vector[M - 1]  cevgm_mean;
  vector[M - 1]  cevtg_mean;
  vector[M]      gas_sd;
  vector[M - 1]  cevgm_sd;
  vector[M - 1]  cevtg_sd;

  // Informative multivariate normal prior.

  zeta ~ multi_normal(mu, Sigma);
  error_gas ~ cauchy(0, 2.5);
  error_cevtg ~ cauchy(0, 2.5);
  error_cevgm ~ cauchy(0, 2.5);

```

```
// Calculate the mean and sd of the 3 curves.
for (m in 1:M) {
  gas_mean[m] = gas_curve(k[m], beta0, beta);
  gas_sd[m]   = error_gas;
}

for (m in 1:(M - 1)) {
  cevgm_mean[m] = log(beta5);
  cevtg_mean[m] = log(beta3) + log(1 - exp(-beta4)) - (beta4 * (k[m]));

  cevgm_sd[m] = error_cevgm;
  cevtg_sd[m] = error_cevtg;
}

// Define the probability models for the 3 curves.
gas      ~ lognormal(log(gas_mean), gas_sd);
cevgm_tf ~ normal( cevgm_mean , cevgm_sd);
cevtg_tf ~ normal( cevtg_mean , cevtg_sd);
}
```

LISTING C.3: Stan code corresponding to `lme_exchangeable_seperate.stan`

```

data {
  int<lower=0>      n_all ; // number of observations
  int<lower=0>      n_item; // number of participants (items)
  int<lower=0>      n_meth; // number of methods
  real             y[n_all]; // outcome variable
  int<lower=1>      item[n_all]; // item index
  int<lower=1>      meth[n_all]; // method index
}

transformed data {
  int<lower=1>      n_factor;
  n_factor = n_all / n_meth;
}

parameters {
  // Parameters for each method.
  vector[n_meth]    alpha; // mean effect
  vector[n_meth]    sigma; // sd

  // Random effects due to item.
  vector[n_item]    u_raw; // per-item random effect
  real<lower=0>      gamma; // sd

  // Random effects due to interaction between items and methods.
  vector[n_item]    c_raw[n_meth]; // per-combination random effect
  real<lower=0>      tau           ; // sd
}

model {
  vector[n_item]    u           ; // per-item random effect
  vector[n_item]    c[n_meth]; // per-combination random effect

  // Prior for fixed components.
  alpha ~ normal(0, 10000);

  // Priors for variance components.
  gamma ~ cauchy(0, 2.5);
  sigma ~ cauchy(0, 2.5);
  tau   ~ cauchy(0, 2.5);

  // Distributions of random effects.
  u_raw ~ normal(0, 1);
  u = gamma * u_raw;

  for (i in 1:n_meth) {
    c_raw[i] ~ normal(0, 1);
  }
}

```

```
    c[i] = tau * c_raw[i];
  }

  // Distribution of the outcome variable.
  for (n in 1:n_all) {
    y[n] ~ normal(alpha[meth[n]] +
                  u[item[n]] +
                  c[meth[n], item[n]],
                  sigma[meth[n]]);
  }
}

generated quantities {
  real<lower=0> sigma_ratio[1];
  real alpha_diff[1];

  sigma_ratio[1] <- sigma[1]/sigma[2];
  alpha_diff[1]  <- alpha[1]-alpha[2];
}
```

LISTING C.4: Stan code corresponding to `lme_linked_common.stan`

```

data {
  int<lower=0>      n_all ; // number of observations
  int<lower=0>      n_item; // number of participants (items)
  int<lower=0>      n_meth; // number of methods
  real             y[n_all]; // outcome variable
  int<lower=1>      item[n_all]; // item index
  int<lower=1>      meth[n_all]; // method index
  int<lower=1>      repl_ind[n_all]; // item-replicate index
}

transformed data {
  int<lower=1>      n_factor;
  n_factor = n_all / n_meth;
}

parameters {
  // Parameters for each method.
  vector[n_meth]    alpha; // mean effect
  real              sigma; // sd

  // Random effects due to item.
  vector[n_item]    u_raw; // per-item random effect
  real<lower=0>      gamma; // sd

  // Random effects due to interaction between items and methods.
  vector[n_item]    c_raw[n_meth]; // per-combination random effect
  real<lower=0>      tau           ; // sd

  // Random effects due to interaction between items and replicates.
  vector[n_factor]  a_raw; // item-replicate interaction effect
  real<lower=0>      omega; // sd
}

model {
  vector[n_item]    u           ; // per-item random effect
  vector[n_item]    c[n_meth]; // per-combination random effect
  vector[n_factor]  a           ; // item-replicate interaction effect

  // Prior for fixed components.
  alpha ~ normal(0, 10000);

  // Priors for variance components.
  gamma ~ cauchy(0, 2.5);
  sigma ~ cauchy(0, 2.5);
  tau   ~ cauchy(0, 2.5);
  omega ~ cauchy(0, 2.5);

```

```
// Distributions of random effects.
u_raw ~ normal(0, 1);
u = gamma * u_raw;

for (i in 1:n_meth) {
  c_raw[i] ~ normal(0, 1);
  c[i] = tau * c_raw[i];
}

for (i in 1:n_factor) {
  a_raw[i] ~ normal(0, 1);
  a[i] = omega * a_raw[i];
}

// Distribution of the outcome variable.
for (n in 1:n_all) {
  y[n] ~ normal(alpha[meth[n]] +
                u[item[n]] +
                a[repl_ind[n]] +
                c[meth[n], item[n]],
                sigma);
}

}

generated quantities {
  real alpha_diff[1];
  alpha_diff[1] <- alpha[1]-alpha[2];
}
```

Appendix D

Power law simulation in R

The following R functions were used to simulate the tidal flow waveform signals with an underlying power spectrum in Chapter 5.

LISTING D.1: R functions to simulate a noisy signal with a known power law spectrum

```
make_f0 <- function(time)
# Specifies fundamental frequency.
#
# Args:
#   time: Scalar. Time of sequence in Hz.
#
# Returns:
#   Fundamental frequency (in Hz) corresponding to time.
{
  1/time
}

make_harmonics <- function(time, acq)
# Calculates harmonic frequencies given time and sampling frequency.
#
# Args:
#   time: Scalar. Time of sequence in Hz.
#   acq: Scalar. Sampling (acquisition) frequency in Hz.
# Returns:
#   Vector containing harmonic frequencies up to the Nyquist frequency
#   (i.e. acq/2)
{
  f0 <- make_f0(time)
  harmonics <- seq((acq/2)/f0) * f0
}
```

```

make_power_spectrum <- function(beta, time, acq)
# Produces a power spectrum for a given power exponent.
#
# Args:
#   beta: Scalar. Defines the underlying power-law exponent.
#   time: Scalar. Time of sequence in Hz.
#   acq:  Scalar. Sampling frequency in Hz.
# Returns:
#   List containing two elements:
#   power: Vector: Spectral density.
#   freq:  Vector: Frequency harmonics in Hz.
{
  harmonics      <- make_harmonics(time, acq)
  log_harmonics <- log(harmonics)
  log_spec      <- 0 - beta * log(harmonics)
  spec          <- exp(log_spec)
  return(list("power" = spec, "freq" = harmonics))
}

simulate_tk <- function(beta, time, snr = NULL)
# Creates a random time signal with an underlying power-law distribution and noise.
#
# Args:
#   beta: Scalar. Underlying power-law exponent.
#   time: Scalar. Length of signal in Hz.
#   snr:  Scalar. Signal-to-noise ratio. Default = NULL.
#
# Returns:
#   Vector. Signal with observational noise corresponding to signal-to-noise ratio.
{
  spec <- make_power_spectrum(beta, time, acq)
  strength <- spec[["power"]]
  freq     <- spec[["freq"]]
  real <- sqrt(strength) * rnorm(length(strength), 0, 1)
  imag <- sqrt(strength) * rnorm(length(strength), 0, 1)
  z     <- complex(real = real, imaginary = imag)
  z     <- c(complex(real = 0, imaginary = 0), z, rev(Conj(z)))
  ft    <- ts(Re(fft(z, inverse = T)), frequency = 200)
  noise <- rnorm(ft)
  k     <- sqrt(var(ft)/(snr*var(noise)))
  if(snr == NULL)
  {
    ft <- ft
  } else {
    ft <- ft + k*noise
  }
}

```

Appendix E

Stan models for random effects and paired comparisons.

The random effects model described by equation (6.1), (6.2) in Chapter 6 was fit using the Stan model described in Listing E.1. The paired comparison described by equation (6.4) in Chapter 6 was fit using the Stan model described in Listing E.2.

These models require the use of Stan through one of its interfaces, e.g. `RStan`, `PyStan`, `CmdStan`. We used `RStan` (version 2.12.1) implemented in `R` (version 3.3.1). More information is available at <http://mc-stan.org/>.

Data for the model described in Listing E.1 should be a list containing the following elements:

- **Nobs**: an integer count of the number observed power-law exponents
e.g. `int 900`
- **Npreds**: an integer count of the number main effects being estimated,
e.g. `int 4`
- **Nggroups**: an integer count of the number of ‘items’ or participants,
e.g. `int 600`
- **y**: a real-valued vector containing the power-law exponents,
e.g. `num [1:900] 3.7 3.9 4.4 2.8 ...`

- **x**: an R model matrix,

		(Intercept)	age	cf	age:cf
	1	1	0.09	0	0
e.g. num [1:900, 1:4]	2	1	0.08	0	0
	3	1	4.1	1	4.1
	...	...	...	...	...

- **timevar**: a real-value vector containing age in years (corresponding to the ‘age’ vector in the model matrix,
e.g. num [1:900] 0.09 0.08 4.1 ...
- **weight**: a real-value vector containing a normalised weighting factor,
e.g. num [1:900] 1.03 2.2 0.52 ...
- **weighted**: an integer indicating whether to apply the weighting: 0 = No, 1 = Yes,
e.g. num 1
- **sensitivity**: an integer indicating which prior to put on the residual standard deviation: 0 = diffuse, 1 = sensitivity prior 1, 2 = sensitivity prior 2,
e.g. num 1
- **group**: an integer-valued vector containing the index number of each participant,
e.g. num [1:900] 1 2 2 3 ...

Data for the model described in Listing E.2 should be a list containing the following elements:

- **n_all**: an integer count of the number data objects corresponding to an observed lung function quantity (i.e. LCI, FRC, or CEV) for an included test,
e.g. int 350
- **n_item**: an integer count of the number of ‘items’ or participants,
e.g. int 700
- **n_meth**: an integer count of the number of method being compared (in this case two),
e.g. int 2

- **y**: a real-valued vector containing the observed power-law exponent,
e.g. `num [1:700] 3.7 3.9 3.4 3.8 ...`
- **item**: an integer valued vector containing the index number of each participant,
e.g. `num [1:200] 1 1 2 2 3 3 ...`
- **meth**: an integer valued vector containing the index number of each method,
e.g. `num [1:200] 1 2 1 2 1 2 ...`
- **weight**: a real-value vector containing a normalised weighting factor,
e.g. `num [1:200] 1.9 1.04 2.1 2.2 ...`

LISTING E.1: Stan code for linear random effects model described by (6.1) and (6.2)

```

data {
  int<lower = 0>      Nobs;
  int<lower = 0>      Npreds;
  int<lower = 0>      Nggroups;
  vector[Nobs]       y;
  matrix[Nobs, Npreds] x;
  vector[Nobs]       timevar;
  vector[Nobs]       weight;
  int<lower = 0>      weighted;
  int<lower = 0>      sensitivity;
  int<lower = 1, upper = Nggroups> group[Nobs];
}

transformed data {
}

parameters {
  vector[Npreds]      beta;
  real<lower=0>        sigma_e;

  vector<lower = 0>[2]  sigma_u;
  cholesky_factor_corr[2] L_u;
  matrix[2, Nggroups]  z_u;
}

transformed parameters {
  vector[Nobs] yhat;
  matrix[2, Nggroups] u;

  u = diag_pre_multiply(sigma_u, L_u) * z_u;

  for(i in 1:Nobs)
    yhat[i] = x[i] * beta + u[1, group[i]] + u[2, group[i]] * timevar[i];
}

model {
  L_u ~ lkj_corr_cholesky(2);
  to_vector(z_u) ~ normal(0, 1);
}

```

```
sigma_u ~ cauchy(0, 2.5);
if(sensitivity == 0) {
  sigma_e ~ cauchy(0, 2.5);
} else if(sensitivity == 1) {
  sigma_e ~ lognormal(-4, 0.1);
} else if(sensitivity == 2) {
  sigma_e ~ lognormal(1.5, 0.5);
}

beta ~ normal(0, 10);
for(i in 1:Nobs){
  y[i] ~ normal(yhat[i], sigma_e * weight[i]);
}

}

generated quantities {

  real y_rep[Nobs];
  corr_matrix[2] rho_u;
  real a[Nobs];

  rho_u = multiply_lower_tri_self_transpose(L_u);

  for(i in 1:Nobs){
    a[i] = x[i] * beta + u[1, group[i]] + u[2, group[i]] * timevar[i];
    y_rep[i] = normal_rng(a[i], sigma_e * weight[i]);
  }

}
```

LISTING E.2: Stan code for paired comparison model described by (6.4)

```

data {
  int<lower=0>          n_all;
  int<lower=0>          n_item;
  int<lower=0>          n_meth;
  real                 y[n_all]; // outcome variable
  int<lower=1>          item[n_all]; // item index
  int<lower=1>          meth[n_all]; // method index
  real<lower = 0>       weight[n_all];

}

transformed data {

}

parameters {

  // Parameters for each method.
  vector<lower = 0>[n_meth] alpha; // mean effect
  vector<lower = 0>[n_meth] sigma; // sd

  // Random effects due to item.
  vector[n_item] u_raw; // per-item random effect
  real<lower = 0> gamma; // sd

}

transformed parameters {

}

model {

  vector[n_item] u; // per-item random effect

  // Priors for variance components.
  alpha ~ normal(0, 10);
  gamma ~ cauchy(0, 2.5);
  sigma ~ cauchy(0, 2.5);

  // Distributions of random effects.
  u_raw ~ normal(0, 1);

  // Transform the random effect parameters.
  u = gamma * u_raw; // ==> u ~ normal(0, gamma)

```

```
// Distribution of the outcome variable.
for (n in 1:n_all) {
  y[n] ~ normal(alpha[meth[n]] +
                u[item[n]],
                sigma[meth[n]] * weight[n]);
}

}

generated quantities {

  real          bias;
  real          loa[2];

  bias = alpha[1] - alpha[2];
  loa[1] = bias + 2 * sqrt(sigma[1] ^ 2 + sigma[1] ^ 2);
  loa[2] = bias - 2 * sqrt(sigma[1] ^ 2 + sigma[1] ^ 2);

}
```

Appendix F

Regression and analysis of variance results from Chapter 6

Included here in Tables F.1 to F.19 are the detailed regression results corresponding to the model described by (6.3). Table F.20 shows the detailed results from the analysis of variance described in Section 6.3.

TABLE F.1: Regression results of BIS power-law exponents and gender.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Male)	3.881*** (3.239, 4.523)	4.271*** (3.416, 5.125)
Infant (baseline) PLE	0.137 (−0.037, 0.311)	0.031 (−0.201, 0.263)
Female	−0.005 (−0.103, 0.092)	−0.887 (−2.172, 0.399)
Female [†]		0.240 (−0.109, 0.589)
Observations	106	106
R ²	0.023	0.040
Adjusted R ²	0.004	0.012
Residual Std. Error	1.992 (df = 103)	1.984 (df = 102)
F Statistic	1.200 (df = 2; 103)	1.411 (df = 3; 102)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.2: Regression results of BIS power-law exponents and gestational age.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (37 to 42 weeks)	3.867*** (3.225, 4.508)	3.873*** (3.231, 4.515)
Infant (baseline) PLE	0.140 (−0.034, 0.314)	0.138 (−0.036, 0.313)
32 to <37 weeks	0.093 (−0.228, 0.415)	−13.810 (−43.740, 16.120)
32 to <37 weeks [†]		3.889 (−4.481, 12.260)
Observations	106	106
R ²	0.026	0.034
Adjusted R ²	0.007	0.005
Residual Std. Error	1.989 (df = 103)	1.991 (df = 102)
F Statistic	1.359 (df = 2; 103)	1.181 (df = 3; 102)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.3: Regression results of BIS power-law exponents and birthweight.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (3 to 3.995 kg)	3.841*** (3.204, 4.479)	3.948*** (3.255, 4.641)
Infant (baseline) PLE	0.151* (−0.023, 0.325)	0.122 (−0.068, 0.311)
4 to 4.995 kg	−0.118* (−0.251, 0.015)	−0.663 (−2.727, 1.402)
2 to 2.995 kg	0.065 (−0.100, 0.230)	−1.353 (−5.012, 2.307)
4 to 4.995 kg [†]		0.146 (−0.406, 0.699)
2 to 2.995 kg [†]		0.382 (−0.603, 1.368)
Observations	105	105
R ²	0.061	0.068
Adjusted R ²	0.033	0.021
Residual Std. Error	1.972 (df = 101)	1.984 (df = 99)
F Statistic	2.176* (df = 3; 101)	1.450 (df = 5; 99)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.4: Regression results of BIS power-law exponents and mode of delivery.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Vaginal)	3.891*** (3.256, 4.526)	3.979*** (3.268, 4.691)
Infant (baseline) PLE	0.126 (−0.047, 0.298)	0.102 (−0.092, 0.295)
Emergency	0.128** (0.009, 0.247)	0.362 (−1.615, 2.340)
Elective	0.033 (−0.131, 0.197)	−1.404 (−3.725, 0.917)
Emergency [†]		−0.063 (−0.595, 0.469)
Elective [†]		0.400 (−0.244, 1.044)
Observations	106	106
R ²	0.064	0.078
Adjusted R ²	0.036	0.032
Residual Std. Error	1.959 (df = 102)	1.963 (df = 100)
F Statistic	2.308* (df = 3; 102)	1.703 (df = 5; 100)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.5: Regression results of BIS power-law exponents and delivering hospital.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Geelong Hospital)	3.872*** (3.233, 4.511)	4.018*** (3.337, 4.700)
Infant (baseline) PLE	0.135 (−0.039, 0.308)	0.095 (−0.091, 0.280)
St John of God	0.046 (−0.054, 0.146)	−1.128 (−3.056, 0.800)
St John of God [†]		0.319 (−0.204, 0.842)
Observations	106	106
R ²	0.030	0.044
Adjusted R ²	0.011	0.015
Residual Std. Error	1.984 (df = 103)	1.980 (df = 102)
F Statistic	1.605 (df = 2; 103)	1.550 (df = 3; 102)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.6: Regression results of BIS power-law exponents and mother's age.

	<i>Dependent variable:</i>	
	No interactions	Interactions
Intercept (30 to 39 yrs)	3.824*** (3.191, 4.457)	3.332*** (2.572, 4.092)
Infant (baseline) PLE	0.148* (−0.023, 0.320)	0.283*** (0.076, 0.489)
20 to 29 yrs	0.061 (−0.040, 0.163)	1.306* (−0.008, 2.619)
40 to 49 yrs	−0.218 (−0.492, 0.056)	5.607** (0.330, 10.880)
20 to 29 yrs [†]		−0.339* (−0.697, 0.018)
40 to 49 yrs [†]		−1.549** (−2.949, −0.150)
Observations	106	106
R ²	0.063	0.128
Adjusted R ²	0.035	0.085
Residual Std. Error	1.960 (df = 102)	1.909 (df = 100)
F Statistic	2.276* (df = 3; 102)	2.948** (df = 5; 100)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.7: Regression results of BIS power-law exponents and father's age.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (30 to 39 yrs)	3.962*** (3.319, 4.605)	3.671*** (2.711, 4.632)
Infant (baseline) PLE	0.117 (−0.055, 0.290)	0.196 (−0.063, 0.454)
20 to 29 yrs	0.033 (−0.079, 0.144)	0.327 (−1.031, 1.684)
40 to 49 yrs	−0.150* (−0.302, 0.002)	1.187 (−0.866, 3.240)
20 to 29 yrs [†]		−0.079 (−0.448, 0.289)
40 to 49 yrs [†]		−0.367 (−0.929, 0.195)
Observations	104	104
R ²	0.064	0.080
Adjusted R ²	0.036	0.033
Residual Std. Error	1.942 (df = 100)	1.945 (df = 98)
F Statistic	2.290* (df = 3; 100)	1.697 (df = 5; 98)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.8: Regression results of BIS power-law exponents and mother's highest education.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Bachelor degree)	3.868*** (3.215, 4.520)	3.919*** (2.889, 4.949)
Infant (baseline) PLE	0.143 (−0.034, 0.319)	0.129 (−0.151, 0.408)
Trade/Cert/Dip	0.006 (−0.125, 0.138)	0.626 (−0.975, 2.227)
Postgraduate degree	−0.042 (−0.184, 0.101)	−2.648** (−5.009, −0.287)
Year 12 or equivalent	−0.054 (−0.200, 0.092)	−0.324 (−2.009, 1.362)
Year 10 or equivalent	0.125 (−0.095, 0.344)	2.099 (−1.220, 5.417)
Trade/Cert/Dip [†]		−0.171 (−0.609, 0.267)
Postgraduate degree [†]		0.700** (0.066, 1.334)
Year 12 or equivalent [†]		0.074 (−0.385, 0.533)
Year 10 or equivalent [†]		−0.534 (−1.430, 0.362)
Observations	105	105
R ²	0.048	0.126
Adjusted R ²	0.0004	0.043
Residual Std. Error	2.000 (df = 99)	1.956 (df = 95)
F Statistic	1.007 (df = 5; 99)	1.522 (df = 9; 95)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.9: Regression results of BIS power-law exponents and father's highest education.

	<i>Dependent variable:</i>	
	Preschool (follow-up)	PLE
	No interactions	Interactions
Intercept (Trade/Cert/Dip)	3.777*** (3.109, 4.446)	4.103*** (3.238, 4.967)
Infant (baseline) PLE	0.169* (−0.013, 0.351)	0.080 (−0.156, 0.316)
Bachelor degree	−0.025 (−0.150, 0.101)	0.484 (−1.197, 2.165)
Year 12 or equivalent	0.100 (−0.058, 0.258)	−1.467* (−3.092, 0.158)
Postgraduate degree	−0.093 (−0.253, 0.066)	−3.383** (−6.078, −0.688)
Year 10 or equivalent	0.023 (−0.172, 0.217)	1.888 (−3.083, 6.858)
Less than year 10	−0.088 (−0.346, 0.169)	10.900** (2.271, 19.530)
Bachelor degree [†]		−0.138 (−0.595, 0.319)
Year 12 or equivalent [†]		0.433* (−0.014, 0.880)
Postgraduate degree [†]		0.913** (0.166, 1.660)
Year 10 or equivalent [†]		−0.490 (−1.806, 0.825)
Less than year 10 [†]		−2.813** (−5.027, −0.599)
Observations	103	103
R ²	0.075	0.225
Adjusted R ²	0.017	0.131
Residual Std. Error	1.982 (df = 96)	1.863 (df = 91)
F Statistic	1.297 (df = 6; 96)	2.404** (df = 11; 91)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.10: Regression results of BIS power-law exponents and number of siblings.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (One sibling)	3.798*** (3.170, 4.425)	4.165*** (3.313, 5.017)
Infant (baseline) PLE	0.165* (−0.005, 0.336)	0.065 (−0.167, 0.297)
No siblings	−0.031 (−0.138, 0.075)	−0.430 (−1.804, 0.943)
Two siblings	0.049 (−0.092, 0.190)	−0.739 (−3.262, 1.784)
Three or more siblings	−0.272*** (−0.463, −0.080)	−2.986** (−5.585, −0.387)
No siblings [†]		0.109 (−0.266, 0.484)
Two siblings [†]		0.214 (−0.468, 0.896)
Three or more siblings [†]		0.716** (0.033, 1.399)
Observations	106	106
R ²	0.105	0.143
Adjusted R ²	0.070	0.082
Residual Std. Error	1.925 (df = 101)	1.912 (df = 98)
F Statistic	2.963** (df = 4; 101)	2.340** (df = 7; 98)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.11: Regression results of BIS power-law exponents and household yearly income.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (\$100,000 to \$149,999)	3.833*** (3.150, 4.515)	2.846*** (1.222, 4.470)
Infant (baseline) PLE	0.139 (−0.043, 0.322)	0.405* (−0.032, 0.842)
\$75,000 to \$99,999	0.069 (−0.053, 0.190)	1.471 (−0.448, 3.389)
\$50,000 to \$74,999	0.030 (−0.124, 0.183)	0.487 (−1.517, 2.492)
More than \$150,000	0.089 (−0.111, 0.289)	0.729 (−4.113, 5.571)
\$25,000 to \$49,999	0.034 (−0.181, 0.248)	3.016* (−0.320, 6.352)
Less than \$25k	0.007 (−0.314, 0.328)	4.603 (−1.139, 10.350)
Unsure or declined to answer	0.159 (−0.151, 0.469)	12.240 (−37.310, 61.790)
\$75,000 to \$99,999 [†]		−0.380 (−0.898, 0.139)
\$50,000 to \$74,999 [†]		−0.117 (−0.664, 0.431)
More than \$150,000 [†]		−0.170 (−1.500, 1.159)
\$25,000 to \$49,999 [†]		−0.792* (−1.675, 0.090)
Less than \$25k [†]		−1.233 (−2.769, 0.304)
Unsure or declined to answer [†]		−3.144 (−16.000, 9.710)
Observations	106	106
R ²	0.045	0.108
Adjusted R ²	−0.024	−0.018
Residual Std. Error	2.019 (df = 98)	2.014 (df = 92)
F Statistic	0.655 (df = 7; 98)	0.857 (df = 13; 92)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.12: Regression results of BIS power-law exponents and any maternal cigarette smoking in previous 12 months.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (No)	3.860*** (3.229, 4.490)	3.807*** (3.140, 4.475)
Infant (baseline) PLE	0.147* (-0.025, 0.318)	0.161* (-0.020, 0.342)
Yes	-0.163** (-0.315, -0.010)	0.352 (-1.743, 2.446)
Yes [†]		-0.139 (-0.706, 0.427)
Observations	105	105
R ²	0.064	0.067
Adjusted R ²	0.046	0.039
Residual Std. Error	1.953 (df = 102)	1.960 (df = 101)
F Statistic	3.510** (df = 2; 102)	2.400* (df = 3; 101)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.13: Regression results of BIS power-law exponents and extent of maternal cigarette smoking.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (None)	3.862*** (3.222, 4.501)	3.807*** (3.128, 4.487)
Infant (baseline) PLE	0.146 (−0.028, 0.320)	0.161* (−0.024, 0.345)
Heavy pT,T1,T2,T3	−0.177 (−0.392, 0.037)	0.919 (−2.292, 4.129)
Light pT	−0.059 (−0.323, 0.205)	0.177 (−2.916, 3.269)
Light pT or T1	−0.281 (−0.654, 0.092)	−4.103 (−23.000, 14.800)
Heavy pT,T1,T2,T3 [†]		−0.292 (−1.145, 0.561)
Light pT [†]		−0.064 (−0.908, 0.779)
Light pT or T1 [†]		1.088 (−4.288, 6.464)
Observations	104	104
R ²	0.071	0.077
Adjusted R ²	0.033	0.009
Residual Std. Error	1.970 (df = 99)	1.994 (df = 96)
F Statistic	1.881 (df = 4; 99)	1.138 (df = 7; 96)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.14: Regression results of BIS power-law exponents and exposure to passive smoke in previous 12 months.

	<i>Dependent variable:</i>	
	No interactions	Interactions
Intercept (No)	3.726*** (3.062, 4.389)	3.560*** (2.840, 4.281)
Infant (baseline) PLE	0.178* (−0.002, 0.358)	0.223** (0.028, 0.419)
Yes	−0.014 (−0.146, 0.117)	1.042 (−0.771, 2.855)
Yes [†]		−0.289 (−0.785, 0.206)
Observations	100	100
R ²	0.038	0.051
Adjusted R ²	0.019	0.022
Residual Std. Error	2.002 (df = 97)	1.999 (df = 96)
F Statistic	1.935 (df = 2; 97)	1.731 (df = 3; 96)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.15: Regression results of BIS power-law exponents and household pets.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Yes)	3.883*** (3.245, 4.522)	3.901*** (3.201, 4.601)
Infant (baseline) PLE	0.133 (−0.041, 0.306)	0.128 (−0.063, 0.318)
No	0.054 (−0.061, 0.169)	−0.056 (−1.815, 1.703)
No [†]		0.030 (−0.445, 0.505)
Observations	106	106
R ²	0.031	0.031
Adjusted R ²	0.012	0.002
Residual Std. Error	1.984 (df = 103)	1.993 (df = 102)
F Statistic	1.630 (df = 2; 103)	1.081 (df = 3; 102)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.16: Regression results of BIS power-law exponents and household livestock.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (No)	3.716*** (3.034, 4.398)	3.517*** (2.781, 4.254)
Infant (baseline) PLE	0.179* (−0.005, 0.364)	0.233** (0.034, 0.432)
Yes	0.072 (−0.093, 0.238)	1.313 (−0.472, 3.097)
Yes [†]		−0.359 (−0.872, 0.155)
Observations	104	104
R ²	0.036	0.053
Adjusted R ²	0.017	0.025
Residual Std. Error	1.972 (df = 101)	1.964 (df = 100)
F Statistic	1.865 (df = 2; 101)	1.878 (df = 3; 100)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.17: Regression results of BIS power-law exponents and first degree family history of asthma.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Yes)	3.842*** (3.214, 4.470)	4.104*** (3.259, 4.950)
Infant (baseline) PLE	0.160* (-0.012, 0.332)	0.088 (-0.145, 0.320)
No	-0.109** (-0.206, -0.013)	-0.701 (-1.980, 0.577)
No [†]		0.161 (-0.186, 0.508)
Observations	105	105
R ²	0.065	0.073
Adjusted R ²	0.047	0.045
Residual Std. Error	1.934 (df = 102)	1.936 (df = 101)
F Statistic	3.547** (df = 2; 102)	2.637* (df = 3; 101)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.18: Regression results of BIS power-law exponents and first degree family history of eczema.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (No)	3.813*** (3.163, 4.463)	3.528*** (2.638, 4.418)
Infant (baseline) PLE	0.149* (−0.026, 0.324)	0.226* (−0.014, 0.466)
Yes	0.041 (−0.057, 0.139)	0.646 (−0.646, 1.937)
Yes [†]		−0.165 (−0.515, 0.186)
Observations	105	105
R ²	0.031	0.039
Adjusted R ²	0.012	0.010
Residual Std. Error	1.988 (df = 102)	1.989 (df = 101)
F Statistic	1.622 (df = 2; 102)	1.362 (df = 3; 101)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.19: Regression results of BIS power-law exponents and first degree family history of hayfever.

	<i>Dependent variable:</i>	
	Preschool (follow-up) PLE	
	No interactions	Interactions
Intercept (Yes)	3.823*** (3.178, 4.468)	3.543*** (2.802, 4.283)
Infant (baseline) PLE	0.158* (−0.019, 0.335)	0.236** (0.032, 0.439)
No	−0.057 (−0.159, 0.045)	1.085 (−0.426, 2.595)
No [†]		−0.307 (−0.713, 0.099)
Observations	106	106
R ²	0.034	0.054
Adjusted R ²	0.015	0.027
Residual Std. Error	1.981 (df = 103)	1.969 (df = 102)
F Statistic	1.807 (df = 2; 103)	1.953 (df = 3; 102)

Note:

*p<0.1; **p<0.05; ***p<0.01

[†]denotes effect of interaction with baseline PLE.

Estimated means and 95% confidence intervals shown.

PLE: power-law exponent.

TABLE F.20: Analysis of variance of BIS power-law exponents and covariates

	Df	Sum of Sq	RSS	F value	Pr(>F)
No baseline covariate			6.50		
Gender	1	0.08	6.42	1.23	0.2691
BIS Gestational age	1	0.01	6.49	0.13	0.7156
Birthweight	3	0.11	6.38	0.61	0.6096
Mode of delivery	2	0.18	6.32	1.46	0.2379
Delivering hospital	1	0.06	6.43	1.03	0.3128
Mother's age	2	0.19	6.31	1.56	0.2160
Father's age	2	0.11	6.39	0.85	0.4296
Mother's highest education	5	0.17	6.32	0.55	0.7362
Father's highest education	5	0.16	6.33	0.51	0.7649
Number of siblings	3	0.54	5.96	3.09	0.0303
Household yearly income	6	0.23	6.27	0.60	0.7314
Any maternal cigarette smoking in previous 12 months	1	0.21	6.29	3.48	0.0650
Extent of maternal cigarette smoking	4	0.38	6.12	1.56	0.1904
Exposure to passive smoke in previous 12 months	1	0.00	6.50	0.00	0.9709
Household pets	1	0.02	6.48	0.26	0.6141
Household livestock	1	0.00	6.50	0.00	0.9874
First degree family history of asthma	1	0.09	6.41	1.47	0.2280
First degree family history of eczema	1	0.03	6.47	0.41	0.5221
First degree family history of hayfever	1	0.03	6.46	0.53	0.4700

RSS: residual sum of squares.

Df: degrees of freedom.

Sum of sq: sum of squares.