University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Gasparini, L;Shepherd, DA;Bavin, EL;Eadie, P;Reilly, S;Morgan, AT;Wake, M

Title:
Using machine-learning methods to identify early-life predictors of 11-year language outcome

Date:
2023-08-01

Citation:
Gasparini, L., Shepherd, D. A., Bavin, E. L., Eadie, P., Reilly, S., Morgan, A. T. & Wake, M. (2023). Using machine-learning methods to identify early-life predictors of 11-year language outcome. Journal of Child Psychology and Psychiatry and Allied Disciplines, 64 (8), pp.1242-1252. https://doi.org/10.1111/jcpp.13733.

Persistent Link:
https://hdl.handle.net/11343/337542

License:
CC BY

Using machine-learning methods to identify early-life predictors of 11-year language outcome

Loretta Gasparini[1,2], Daisy A. Shepherd[1,2], Edith L. Bavin[1,3], Patricia Eadie[4], Sheena Reilly[1,2,5], Angela T. Morgan[1,6,7]*, Melissa Wake[1,2,8]*

[1] Murdoch Children's Research Institute; [2] Department of Paediatrics, The University of Melbourne; [3] School of Psychology and Public Health, La Trobe University; [4] Melbourne Graduate School of Education, The University of Melbourne; [5] Menzies Health Institute Queensland, Griffith University; [6] Department of Audiology and Speech Pathology, The University of Melbourne; [7] Royal Children's Hospital Melbourne; [8] Liggins Institute, The University of Auckland

* Joint senior authors

Correspondence concerning this article should be addressed to Loretta Gasparini, Royal Children's Hospital, 50 Flemington Rd, Parkville VIC 3052 AUSTRALIA. Email: loretta.gasparini@mcri.edu.au

© 2022. This manuscript is made available under the CC-BY license https://creativecommons.org/licenses/by/4.0/

Author Note

Citation:

This paper was published on 7th December 2022 in *Journal of Child Psychology & Psychiatry*. Please cite as: Gasparini, L., Shepherd, D. A., Bavin, E. L., Eadie, P., Reilly, S., Morgan, A. T., & Wake, M. (2023). Using machine-learning methods to identify early-life predictors of 11-year language outcome. *Journal of Child Psychology & Psychiatry, 64*(8), 1242-1252. https://doi.org/10.1111/jcpp.13733

Correction:

In this article published in *Journal of Child Psychology & Psychiatry* on 7 December, 2022, the variable *peds6_w3* was incorrectly attributed as "Do you have any concerns about how your child uses his/her arms and legs?", when it is in fact "Do you have any concerns about how your child behaves?". Each correction in the current manuscript and the Supporting information is accompanied by a Corrections statement.

Authors:

Loretta Gasparini: https://orcid.org/0000-0002-1561-5572; @lottiegasp

Daisy A. Shepherd: https://orcid.org/0000-0001-8540-0473; @_DaisyShepherd

Edith L. Bavin: https://orcid.org/0000-0002-7580-149X

Patricia Eadie: https://orcid.org/0000-0003-4722-8013; @paeadie

Sheena Reilly: https://orcid.org/0000-0001-6506-4767; @sreilly_reilly

Angela T. Morgan: https://orcid.org/0000-0003-1147-7405

Melissa Wake: https://orcid.org/0000-0001-7501-9257; @Melissa_Wake

Registry:

This paper was first registered with Open Science Framework (OSF) on 02/02/2022 (available at https://osf.io/fpdzk/)

Supporting information:

All Supporting information is available at https://osf.io/fpdzk/

CRediT author statement:

LG: Conceptualisation, Methodology, Software, Formal analysis, Writing – Original Draft, Review & Editing

DS: Conceptualisation, Methodology, Writing – Review & Editing, Supervision

EB: Investigation, Resources, Writing – Review & Editing

PE: Investigation, Resources, Writing – Review & Editing, Project administration

SR: Investigation, Resources, Writing – Review & Editing

AM: Conceptualisation, Methodology, Writing – Review & Editing, Supervision

MW: Conceptualisation, Methodology, Investigation, Resources, Writing – Review & Editing, Supervision

Conflicts of interest:

The authors have no conflicts of interest to disclose

Funding

ELVS was supported by the Australian National Health and Medical Research Council (NHMRC; grant numbers 237106, 436958, 1041947) and small grants from the Murdoch Children's Research Institute and La Trobe University. Research at the Murdoch Children's

Research Institute is supported by the Victorian Government's Operational Infrastructure Support Program. LG was supported by an Australian Government Research Training Program (RTP) Scholarship and MCRI PhD Top Up Scholarship. SR is supported by an NHMRC Practitioner Fellowship 491210), MW is supported by a Senior Research Fellowship 1046518 and Principal Research Fellowship 1160906. The funding organizations are independent of all researchers.

Acknowledgements

We thank the participating children and parents, the Victorian Maternal and Child Health nurses who supported recruitment, Bibi Gerner, Eileen Cini, Gabor Mihala and other ELVS researchers. For insightful contributions thank you to Varnika Aggarwal, Vânia de Aguiar, Simranpreet Kaur, Evan Kidd, Katie McBain, Caroline Rowland, Ashleigh Shipton, MCRI's Speech & Language Group, The University of Melbourne Languages and Linguistics seminar group and two anonymous reviewers.

Abstract

Background: Language is foundational for neurodevelopment and quality of life but an estimated 10% of children have a language disorder at age 5. Many children shift between classifications of typical and low language if assessed at multiple times in the early years, making it difficult to identify which children will have persisting difficulties and benefit most from support. This study aims to identify a parsimonious set of preschool indicators that predict language outcomes in late childhood, using data from the population-based Early Language in Victoria Study (n=839). **Methods:** Parents completed surveys about their children at ages 8, 12, 24 and 36 months. At 11 years, children were assessed using the Clinical Evaluation of Language Fundamentals 4th Edition (CELF-4). We used random forests to identify which of the 1990 parent-reported questions best predict children's 11-year language outcome (CELF-4 score ≤81 representing low language) and used SuperLearner to estimate the accuracy of the constrained sets of questions. **Results:** At 24 months, 7 predictors relating to vocabulary, symbolic play, pragmatics and behavior yielded 73% sensitivity (95% CI: [57,85]) and 77% specificity (95% CI: [74,80]) for predicting low language at 11 years [Corrections made on 5 May 2023, after first online publication: In the preceding sentence 'motor skills' has been corrected to 'behavior' in this version]. At 36 months, 7 predictors relating to morphosyntax, vocabulary, parent-child interactions and parental stress yielded 75% sensitivity (95% CI: [58,88]) and 85% specificity (95% CI: [81,87]). Measures at 8 and 12 months yielded unsatisfactory accuracy. **Conclusion:** We identified two short sets of questions that predict language outcomes at age 11 with fair accuracy. Future research should seek to replicate results in a separate cohort.

Keywords: language development, language disorders, sensitivity and specificity, longitudinal studies, machine learning

Abbreviations: ATP: Australian Temperament Project; AUC: Area Under the receiving operating characteristic Curve; BITS: Brigance Parent-Child Interactions Scale; CDI-III: MacArthur-Bates Communicative Development Inventories Third Edition; CDI-WG: MacArthur-Bates Communicative Development Inventories (Words and Gestures); CDI-WS: MacArthur-Bates Communicative Development Inventories (Words and Sentences); CELF-4: Clinical Evaluation of Language Fundamentals, Fourth Edition, Australian version; CI: confidence interval; CSBS: Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist; ELVS: Early Language in Victoria Study; Kessler-K6: The Kessler Psychological Distress Scale (K6 version); LSAC: Growing Up in Australia: The Longitudinal Study of Australian Children; OSF: Open Science Framework; PEDS: Parents' Evaluation of Developmental Status; SES: socioeconomic status; SD: standard deviation

Using machine-learning methods to identify early-life predictors of 11-year language outcome

Introduction

Lifelong language skills are foundational for good socio-behavioral, academic, employment and quality of life outcomes (Conti-Ramsden & Durkin, 2012; Eadie et al., 2018, 2021; Yew & O'Kearney, 2013). Language Disorder is defined as language difficulties impacting everyday functioning (Bishop et al., 2017) which affects an estimated 10% of children aged 5 (Norbury et al., 2016). Two-thirds of these cases cannot be attributed to any specific condition or environmental factor (Norbury et al., 2016). Ideally, we could identify children across the whole population who will have persisting language difficulties from an early age, enabling us to provide targeted support while avoiding over-servicing. This goal is complicated however, by children shifting between classifications of typical and low language if assessed at multiple times in early childhood (approximately 6% of all children if assessed between 4-11 years, McKean, Wraith, et al., 2017), which yields a high error rate in identifying the estimated 10% of all children who will have persisting difficulties and most benefit from intervention. To best utilize the limited available resources to support children who will have persisting difficulties, we need early measures that accurately predict Language Disorder in later childhood.

Failing to punctually identify children at risk for persisting Language Disorder also hinders early intervention research. Currently, it is difficult to test the effectiveness of early language interventions because many recruited children catch up spontaneously, which drowns out the interventions' impact for the children who would otherwise have persisting difficulties (Wake et al., 2011). Accurately identifying Language Disorder in the preschool

years will enable more precise recruitment into intervention programs, helping us to provide children with the language support they need to thrive.

Known predictors of poor language outcome observable before 3 years of age include: delayed gesture production, limited receptive and/or expressive vocabulary size, impaired syntactic development including absence of two-word combinations, family history of Language Disorder, low socioeconomic status (SES) and poorer quality and quantity of communicative interactions (Sansavini et al., 2021). Whilst there is growing consensus on predictors for language outcomes, the accuracy of language screening tools remains inconclusive as replications are sparse, most studies have assessed accuracy over short timeframes (1-2 years and within the preschool years) and their accuracy tends to attenuate with age (Wallace et al., 2015).

Prioritizing only items that uniquely contribute to predicting language outcome could lead to accurate, time-efficient instruments suited to public health use. This approach draws upon other health areas that have identified short sets of questions that clinicians can readily collect via patient report or history (e.g., for 5-year mortality in adults, Ganna & Ingelsson, 2015). Rudolph and Leonard (2016) asked parents retrospectively when their child was aged 4-7 years at what age their child started combining words. This question alone yielded 55% sensitivity and 93% specificity for classifying Language Disorder cases (at ages 4-7). Sensitivity improved (71%) but specificity decreased (84%) when they added family history of a Speech/Language Disorder and maternal education to the predictive model. Another study found that measures of vocabulary size, grammatical complexity, and semantic network structure (semantic relatedness of vocabulary items the child is reported to know) collected between 16 and 30 months yielded >90% accuracy for classifying children with language or reading disorder (based on parental, school record or clinician report between 4-7 years) (Borovsky et al., 2021). Vocabulary size and semantic network structure measures however,

require prolonged assessments such as administration of the entire MacArthur-Bates Communicative Development Inventories (Words and Sentences, CDI-WS) vocabulary checklist (680 items, Fenson et al., 2006) which limits the feasibility of implementing these measures at a population-based level. By contrast, measures of grammatical complexity can be quick to administer and would be less burdensome for parents to complete if found to have adequate accuracy alone.

Another challenge in accurately predicting persisting Language Disorder is that salient predictors vary with age (Bishop et al., 2017). For a given set of predictors, we would want to identify the optimal age of data collection for best predictive accuracy. For instance, there is little value in collecting measures of morphosyntax at ages when they show floor effects across the population. Furthermore, as children encounter various health and education services in the early years, we would want various sets of predictors available at different child ages.

The Early Language in Victoria Study (ELVS) is a population-based cohort study tracking language and communication from infancy to adolescence (Reilly et al., 2018). One study reporting on the ELVS data identified seven questions on the 12-month parent survey that accounted for school-entry language outcome with “fair” accuracy (Area Under the receiving operating characteristic Curve, AUC, of .73, McKean et al., 2016). These items related to gesture, vocabulary, family history of language/literacy difficulties, maternal education, SES and parent-child interactions. Later ELVS studies dichotomized the expressive vocabulary CDI-WS score at 2 years at a standardized score below the 10th percentile and found it did not reliably predict 7- (McKean, Reilly, et al., 2017) or 11-year (Eadie et al., 2021) language. ELVS collected early morphosyntax measures at 2- and 3-years of age via the CDI-WS and MacArthur-Bates Communicative Development Inventory -

Third Edition (CDI-III, Dale, 2007), but these morphosyntax items are yet to be analyzed for their predictive accuracy.

The first aim of this study is to estimate the individual importance of a range of parent-reported items (individual questions) collected in infancy for predicting low language outcome in late childhood (11 years). Note that 'importance' is used as a technical term here, see *Aim 1: Individual variable importance* for how it is defined and calculated. The second aim is to use the results from the first aim along with knowledge of previous literature on language development, to identify a parsimonious set of parent-reported items that could be used at various infant and toddler ages to accurately predict low language outcome at 11 years (see *Outcome* for how this is operationalized).

Methods

We preregistered the statistical analysis methods on Open Science Framework (OSF) on 02/02/2022 at https://osf.io/fpdzk/ prior to data access by the first author, who conducted the analysis. We published and timestamped protocol amendments and all Supporting information on the OSF repository. Ethical approval was obtained from the Royal Children's Hospital (23018, 27078, 82096) and La Trobe University Human Ethics Committee (03-32). Parents provided written, informed consent.

Sample

The Early Language in Victoria Study (ELVS) is a prospective longitudinal cohort study which recruited a community cohort of 1910 infants aged 7.5-10 months in 6 metropolitan areas of Melbourne, Victoria. Metropolitan areas were selected to represent all SES strata. Participants were recruited via the Maternal and Child Health Service (82% uptake rate), hearing screening visits and press advertising. The participants have been assessed in 11 waves from ages 8 months to 13 years (Reilly et al., 2018).

Participants were excluded from joining ELVS if at recruitment (when they were 8 months of age) the Maternal and Child Health Nurse identified them as having a disability likely to interfere with their language development, including developmental delay, cerebral palsy, a syndrome, intellectual or physical disability, cleft lip or palate, vision or hearing impairment. Children were also excluded if their parents did not understand English sufficiently to complete questionnaires designed for grade 6 level. As we are interested in predicting low language outcome for any child, we included participants with any co-occurring conditions that did not preclude them from participating in ELVS (e.g., Autism, speech disorder). The primary analysis included all ELVS participants with data available at wave 10 (11 years old, n=839, 44.6% of baseline participants).

Predictors

At waves 1-4 (ages 8, 12, 24, 36 months), parents of ELVS participants completed surveys relating to their child, the family and home environment. The predictor measures in this study comprise all multiple-choice or numerical items collected in these waves. Appendix S1 summarizes all constructs and instruments administered and describes our data cleaning and exclusion practices. Our analysis included 1990 variables (182 at 8 months, 644 at 12 months, 916 at 24 months, 248 at 36 months).

Outcome

The 11-year language outcome was measured using the Clinical Evaluation of Language Fundamentals, Fourth Edition, Australian version (CELF-4, Semel et al., 2006), a standardized tool that yields a core language score with a mean of 100 and standard deviation (SD) of 15. The CELF-4 is used as a diagnostic tool for Language Disorder and a core score below 77.5 was found to have 100% sensitivity and 89% specificity for diagnosing Language Disorder (Pearson Education, 2008). Because information on the everyday functional impact

of children's language abilities is also needed for a diagnosis of Language Disorder (Bishop et al., 2016), we consider a low CELF-4 score as a potential indicator of Language Disorder. We operationalized the dichotomous outcome as "low" (≥1.5SD below the mean) and "typical" (<1.5SD below the mean) language. We selected this cut-off as a balance between maximizing diagnostic accuracy (according to Pearson Education, 2008) and ensuring the low language group had sufficient participants for the models to successfully run.

Statistical methods

We conducted statistical analyses in RStudio (R Core Team, 2020; RStudio Team, 2020) and code is available on the OSF repository. We provide more details of the statistical methods in Appendices S2 (random forests) and S3 (SuperLearner).

Aim 1: Individual variable importance

We ran random forests, a tree-based machine-learning method, separately by data collection wave to estimate the importance of all 1990 variables. Of the 642 variables with missing values, we excluded variables with >50% missingness (172 variables). For the remainder, we imputed missing values using the rfImpute command in the *randomForest* package (Liaw & Wiener, 2002). We repeated the imputation process 300 times to produce 300 imputed datasets. For each imputed dataset, we computed a random forest with 500 unbiased conditional inference trees using the package *party* (Hothorn, 2005; Strobl et al., 2007, 2009). Variable importance was estimated using conditional permutation importance without replacement (Strobl et al., 2009), where larger variable importance values indicate a closer relationship between the predictor and outcome. We averaged variable importance values across the 300 imputed datasets. We also completed a complete case sensitivity analysis (see *Sensitivity analyses*).

Aim 2: Predictor set selection

We used the estimated variable importance values from Aim 1 to help select a constrained set of predictors. We evaluated variables with the highest variable importance values for the extent to which they accord with previous research on predictors of language outcome (e.g., Sansavini et al., 2021; Zambrana et al., 2014) and would likely be easy to understand and answer by any caregiver. For each data collection wave, we selected 8 questions to include in the models, which approximates 1 minute administration time.

We used SuperLearner with 10-fold cross-validation to estimate the predictive accuracy of our sets of predictors in predicting our language outcome. SuperLearner utilizes various prediction algorithms and weights them according to their accuracy in a new single algorithm (van der Laan et al., 2007). We used a complete case analysis to estimate predictive accuracy. In our SuperLearner algorithm we included the following base learners: elastic net regression, random forests, extreme gradient boosting, Bayesian additive regression trees, multivariate adaptive regression splines (see Appendix S3 for details).

For each set of predictors (8, 12, 24 and 36 months), we started with 8 measures and iteratively excluded 1 variable, by lowest variable importance, to determine the optimal number of variables to include in our final set. For each final set of predictors, we calculated AUC, sensitivity and specificity with 95% confidence intervals (CIs). For the purposes of interpreting results we consider sensitivity and specificity >70% as “fair” and >80% “good” classification accuracy (we decided these thresholds in our preregistration prior to starting the analysis, based on previous studies using ELVS data, Eadie et al., 2021; McKean et al., 2016). We also ran this analysis cumulatively by data collection wave.

Sensitivity analyses

As a sensitivity analysis, we repeated the Aim 2 analysis (predictor set selection) with the cut-off of low and typical language being 1.25 and 2 SDs below the mean CELF-4 score. We also repeated the Aim 2 analysis with 7-year language (wave 8) as the outcome (1.25, 1.5, and 2 SDs below the mean cut-offs), to see if our results were robust to including a larger sample as fewer participants had been lost to attrition by age 7. This included all participants for whom a CELF-4 core language score was available at 7 years (n=1208, 63.2% of baseline participants).

We ran a complete case analysis of the Aim 1 analysis (random forests) to see whether our results can be attributed to patterns in the real data and not imputed data.

Additional analyses

We ran univariate logistic regressions on all variables included in the final predictor sets, as well as a constrained number of predictors which we expected from previous literature might have a relationship with the outcome, to check that the effects were in the directions we expect and allow comparison with other literature.

We assessed the classification accuracy of the preferred predictor sets in the subgroup of participants where an additional language to English was spoken at home or to the child at any ages between 8-36 months.

In a separate study we will use data from Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC, Sanson & Johnstone, 2004) to replicate the current study (Gasparini et al., 2022). Therefore, we ran SuperLearner models using only those variables ranked highly in the Aim 1 analysis that LSAC also collected.

Results

Participant characteristics

At 11 years, 1071 individuals were lost to follow-up and 839 were retained. Comparing the two groups shows that the retained participants were overrepresented with girls, families where English was the only main language spoken to the child, mothers with higher education levels, and higher advantage based on postcode (Appendix S4).

The included sample had a mean 11-year CELF-4 score of 100.6 (SD=12.89, range: 40.0-129.0). The 1.5SD below the mean cut-off was at a raw score of 81 and resulted in 791 (94.3%) participants in the typical language group and 48 (5.7%) in the low language group. A raw score of 81 corresponds to 1.25SD below the mean using the population-based standardized scores, showing that our sample has on average better language skills than the wider population.

Aim 1: Individual variable importance

All variables from each data collection wave were ranked by their estimated variable importance and the top 10 variables from each wave are reported in Table 1.

Table 1. Top 10 predictors from each wave and their estimated variable importance

Question	**Levels or scale**	**Source**	**Variable importance**
Wave 1 (8 months)			
What was the last year of school your partner completed?**	Year 10 or less Year 11	ELVS	.0000636

Question	**Levels or scale**	**Source**	**Variable importance**
	Year 12		
How difficult do you think your life is at present?**	No problems or stresses Few problems or stresses Some problems or stresses Many problems or stresses Too many problems or stresses	ATP	.0000505
Child's age	Numeric (months)	ELVS	.0000173
I help my child learn by talking and showing him/her new things**	Not very often Sometimes Often	BITS	.0000140
Does your child throw a ball?**	Yes No	CDI-WG	.0000067
Does your child blow raspberries?**	Yes No	ELVS	.0000055

Question	**Levels or scale**	**Source**	**Variable importance**
When you are not paying attention to your child, does he/she try to get your attention?**	Not Yet Sometimes Often	CSBS	.0000043
Smacks lips in a "yum yum" gesture to indicate when something tastes good	Not Yet Sometimes Often	CDI-WG	.0000042
Caregiver age**	Numeric (years)	ELVS	.0000039
About how long were you pregnant with this baby?**	Numeric (weeks)	ELVS	.0000019
Wave 2 (12 months)			
I am this child's:*	Natural/biological father Stepfather Other	ELVS	.0007905
Does your child string sounds together?**	Not Yet Sometimes Often	CSBS	.0001533
Does your child use sounds or words to get attention or help?**	Not Yet Sometimes Often	CSBS	.0000678

Question	**Levels or scale**	**Source**	**Variable importance**
In the last two weeks, about how often did you feel so sad nothing would cheer you up?**	None of the time A little of the time Some of the time Most of the time All of the time	Kessler-K6	.0000673
About how many of the following objects does your child use appropriately: cup, bottle, bowl, spoon, comb or brush, toothbrush, washcloth, ball, toy vehicle, toy telephone?**	None 1-2 3-4 5-8 over 8	CSBS	.0000210
Child understands (and says) “give”**	Understands Understands and says Neither	CDI-WG	.0000149
Child understands (and says) “don’t”**	Understands Understands and says Neither	CDI-WG	.0000104
Child understands “give me a kiss”	Yes No	CDI-WG	.0000055
Child understands (and says) “drink”	Understands Understands and says Neither	CDI-WG	.0000046
Does your child sing?**	Yes	CDI-WG	.0000029

Question	Levels or scale	Source	Variable importance
	No		
Wave 3 (24 months)			
About how many of the following objects does your child use appropriately: cup, bottle, bowl, spoon, comb or brush, toothbrush, washcloth, ball, toy vehicle, toy telephone?**	3-4 5-8 over 8	CSBS	.0005823
Does your child use sounds or words to get attention or help?**	Not yet Sometimes Often	CSBS	.0001384
Do you have any concerns about how your child behaves?**	No A little Yes	PEDS	.0001242
Child says “spoon”**	Says Does not say	CDI-WS	.0000440
About how many different words or phrases does your child understand without gestures?**	1-3 4-10 11-30 over 30	CSBS	.0000421
Child says “face”**	Says Does not say	CDI-WS	.0000282

Question	**Levels or scale**	**Source**	**Variable importance**
Child says “quack quack”**	Says Does not say	CDI-WS	.0000265
Does your child talk about objects that are not present?*	Not yet Sometimes Often	CDI-WS	.0000254
Child says “teddy bear”	Says Does not say	CDI-WS	.0000199
Child says “bubbles”	Says Does not say	CDI-WS	.0000196
Wave 4 (36 months)			
Mark the sentence that sounds most like the way your child talks:**	This dolly big This dolly big and this dolly little	CDI-III	.0005135
Has your child begun to combine words yet?**	Not yet Sometimes Often	CDI-III	.0004320
Child says “circle”**	Says Not yet	CDI-III	.0002802
Child says “accident”**	Says Not yet	CDI-III	.0001717
Child says “kangaroo”	Says	CDI-III	.0000811

Question	**Levels or scale**	**Source**	**Variable importance**
	Not yet		
Child says “forget/forgot”**	Says Not yet	CDI-III	.0000633
I play with my child and show him/her things**	Not very often Sometimes Often	BITS	.0000524
How do you think you are coping?**	Not at all A little Fairly well Very well Extremely well	ATP	.0000422
Can your child answer questions?*	Yes No	CDI-III	.0000345
Child says “hurry”	Says Not yet	CDI-III	.0000317

*Variable was included in SuperLearner (Aim 2) analysis

**Variable was included in the final/preferred model

Corrections made on 5 May 2023, after first online publication: For Wave 3 (24 months) in Table 1, ‘Do you have any concerns about how your child uses his/her arms and legs?’ has been corrected to ‘Do you have any concerns about how your child behaves?’ in this version.

Aim 2: Predictor set selection

Here we justify why we deviated from the order of variable importance values when choosing which predictors to include in the SuperLearner models. We explain which predictors we included in the final sets.

In wave 1 (8 months) we excluded "Smacks lips in a "yum yum" gesture to indicate when something tastes good". This was because we already included "Does your child blow raspberries?" and we considered both to target oromotor skills. This meant we could include gestational age at birth, which is known to impact language development (Rudolph, 2017). We also excluded child's age, which is not useful as a predictor. The most accurate predictor set included all 8 of the variables included in the SuperLearner models.

In wave 2 (12 months) we deprioritized questions about understanding the phrases/words "give me a kiss" and "drink", as the vocabulary items "give" and "don't" (also verbs and often used as an imperative) were already included. Instead, we included "Does your child sing?", capturing speech production, pragmatic and more general cognitive skills. Although the question "I am this child's natural/biological father, stepfather or other" had the highest variable importance score, it had more missing data than the other variables and we considered that clinicians might not feel comfortable asking this in all contexts. Therefore, we dropped this variable first (instead of in order of variable importance) and found that the most accurate predictor set excluded this variable and included the remaining 7.

For wave 3 (24 months) we were satisfied with the top 8 in order of variable importance. The most accurate predictor set included the top 7 variables.

In wave 4 (36 months) we excluded the vocabulary item "kangaroo" as it might disadvantage recent migrant families to Australia, and we already included the multisyllabic nouns "circle" and "accident". This meant we could include "Can your child answer

questions?" which captures more general language use. However, the most accurate predictor set excluded this final question and included just the top 7 variables.

Table 1 indicates the 8 variables we included in the SuperLearner models and the variables we selected for the final predictor set. Table 2 shows the AUC, sensitivity and specificity of each of these final predictor sets, by data collection wave. The SuperLearner model is the weighted average of all prediction algorithms, while Discrete is the single algorithm with the best predictive ability. The SuperLearner model was the most accurate for the 8-, 12- and 24-month sets and the discrete model (extreme gradient boosting, see Appendix S3 for more details) was more accurate for the 36-month set. The 24- and 36-month sets reached "fair" accuracy levels, while accuracy levels of the 8- and 12-month sets were unsatisfactory.

Table 2. Estimated AUC, sensitivity and specificity (95% CIs) by wave obtained using SuperLearner. "N" indicates number of participants included after missing data were removed. "Model" indicates if SuperLearner or discrete was the most accurate model

	n	**Model**	**AUC**	**Sensitivity**	**Specificity**
Wave 1 (8 months)	740	SuperLearner	.64 (.54, .74)	.65 (.47, .80)	.62 (.59, .66)
Wave 2 (12 months)	822	SuperLearner	.63 (.54, .72)	.54 (.39, .69)	.64 (.60, .67)
Wave 3 (24 months)	804	SuperLearner	.77 (.68, .86)	.73 (.57, .85)	.77 (.74, .80)
Wave 4 (36 months)	739	Discrete	.79 (.70, .89)	.75 (.58, .88)	.84 (.81, .87)

Combining the variables between data collection waves did not substantially improve accuracy (see results in Appendix S5). Table 3 shows the final predictor sets that yielded fair accuracy (24 and 36 months).

Table 3. Final predictor sets that yielded fair accuracy (at ages 24 and 36 months)

Question	Source
Wave 3 (24 months)	
About how many of the following objects does your child use appropriately: cup, bottle, bowl, spoon, comb or brush, toothbrush, washcloth, ball, toy vehicle, toy telephone?	CSBS
Does your child use sounds or words to get attention or help?	CSBS
Do you have any concerns about how your child behaves?	PEDS
Child says “spoon”	CDI-WS
About how many different words or phrases does your child understand without gestures?	CSBS
Child says “face”	CDI-WS
Child says “quack quack”	CDI-WS
Wave 4 (36 months)	
Mark the sentence that sounds most like the way your child talks: “This dolly big” or “This dolly big and this dolly little”	CDI-III
Has your child begun to combine words yet?	CDI-III
Child says “circle”	CDI-III
Child says “accident”	CDI-III
Child says “forget/forgot”	CDI-III

Question	Source
I play with my child and show him/her things	BITS
How do you think you are coping?	ATP

Corrections made on 5 May 2023, after first online publication: For Wave 3 (24 months) in Table 3, 'Do you have any concerns about how your child uses his/her arms and legs?' has been corrected to 'Do you have any concerns about how your child behaves?' in this version.

Sensitivity analyses

Sensitivity analyses of the SuperLearner models using alternative outcome measures yielded the same interpretation of results (see Appendix S6).

The random forests complete case analysis resulted in substantial data loss (205 participants with missing data at 8 months, 512 at 12 months, 512 at 24 months and 583 at 36 months) and in mostly different variables being ranked highly by variable importance (Appendix S7).

Additional analyses

Univariate logistic regression results showed that effects were all in the expected directions according to previous literature: in general, a skill or supportive environment is associated with being in the typical language group at 11 years (Appendix S8).

The multilingual subgroup analysis had very small numbers of participants (n=39) and low precision so results cannot be considered meaningful (Appendix S9).

The analyses using variables included in LSAC yielded similar results to our primary analysis: unsatisfactory (AUC<70%) accuracy for variables collected at 8 and 12 months and fair (AUC>70%) for those collected at 24 and 36 months (Appendix S10).

Discussion

We have identified two short (<1 minute) sets of questions that can be asked at 24 or 36 months with "fair" accuracy (≥73% sensitivity and specificity) for predicting which children will have low language skills at 11 years (CELF-4 core language score ≤81). These measures relate to children's vocabulary, morphosyntax, symbolic play, pragmatic skills, behavior, parent-child interactions and parental stress. [Corrections made on 5 May 2023, after first online publication: In the preceding sentence, 'pragmatic- and motor-skills' has been corrected to 'pragmatic skills, behavior' in this version]. Measures at 8 and 12 months of age yielded unsatisfactory accuracy in predicting the dichotomous 11-year language outcome.

We expected the 8- and 12-month measures to yield higher accuracy, based on a previous ELVS study that identified 7 12-month predictors yielding >70% accuracy for 5-year language (McKean et al., 2016). Considering that both used ELVS data, the differing results might be attributed to us selecting a later language outcome (11 years). Our results suggest that parent-reported measures collected at or before 12 months are inadequate for predicting persisting language difficulties. This may be because children's communication skills are less meaningful at this stage for predicting later outcomes. Alternatively, it may be more difficult for parents to assess their child's preverbal comprehension abilities than their later productive skills, making parent-reported measures less suitable at such early ages.

As expected, the final predictor sets at 24 and 36 months related to children's sentence complexity, vocabulary, language use, parent-child interactions and parental stress

(related to SES), according with previous research (Sansavini et al., 2021). Children's use of objects and general behavior were also predictive at 24 months, warranting further consideration of symbolic play and behavior for predicting language outcome. [Corrections made on 5 May 2023, after first online publication: In the preceding sentence, 'arms and legs' has been corrected to 'general behavior', 'motor skills' has been corrected to 'behavior' and the reference to Gonzalez et al. (2019) regarding motor skills has been removed here and in the References section]. The 24- and 36-month predictor sets yielded lower accuracy than Borovsky and colleagues (2021), who showed that vocabulary size, semantic network structure and morphosyntax measures yielded >90% accuracy. Some methodological differences may explain the difference in results, including that our study had a later outcome age, larger sample size and all participants were assessed using the same standardized language test (n=839 at 11 years, and n=1208 at 7 years tested using the CELF-4, while Borovsky et al., 2021, had n=476 at 4-7 years assessed using one of parental, school record or clinician report). While our predictor sets have lower predictive accuracy, they are also shorter and so would be much quicker to complete and to score, making it more feasible to implement widespread (if results replicate and generalize in our next study, Gasparini et al., 2022).

Strengths and limitations

The Early Language in Victoria Study recruited participants from across whole communities, spanning a range of SES strata. This sampling approach is a strength to our aim of identifying language disorders across the whole population. In contrast, many language studies recruit through speech/language services, which can bias samples toward participants with more severe impairments and better-serviced communities.

This study's only exclusion criterion was of children with developmental disability or morbidity so significant that it was already evident at age 8 months (the age of recruitment to ELVS). While it could be of interest to determine how predictive values might change if children with conditions (e.g., Autism, ADHD) were excluded, this does not reflect real-world clinical services because diagnosis of such conditions in Australia still usually occurs not only after the age of our baseline, but often after the ages at which all our predictor sets were collected for this study.

Another strength is how we operationalized language outcome. Our primary language outcome was at 11 years of age, striving to capture children with language difficulties persisting throughout childhood. The language outcome was measured using the CELF-4, a standardized tool that can be used for diagnostic purposes. Our results remained robust when we operationalized the language outcome at a different age (7 years) when the sample size was larger, and when we set alternative cut-off points for dichotomizing the language outcome. This strengthens the evidence that our predictors can classify children who will have persisting language difficulties.

Our use of machine-learning methods offers many strengths. Random forests is non-parametric, robust to correlated predictors and can manage many variables with relatively few observations (Tagliamonte & Baayen, 2012). Hence, we could reduce a range of predictors to a manageable set. SuperLearner minimizes the effects of arbitrary methodological choices when selecting a prediction algorithm and setting parameters, by running and evaluating multiple options (McNamara et al., 2022). Both random forests and SuperLearner implement cross-validation, which avoids overfitting (Lever et al., 2016).

A limitation is the wide 95% CIs of sensitivity estimates, due to the low population prevalence of Language Disorder. A larger sample size would yield more precise sensitivity.

Moreover, attrition over the 10 years from baseline to 11-year outcome resulted in a more socio-economically privileged sample compared to the group lost to follow-up. Due to the sample size and not being representative of the wider population, our multilingual subgroup analysis failed to yield meaningful results.

The complete case sensitivity analysis using random forests identified a largely different set of predictors than the primary analysis. This could suggest the primary analysis contains artefacts due to imputed values or that the complete case analysis is biased and less precise due to high levels of missing data. We are comfortable interpreting results based on the analysis with imputed values because we averaged results across 300 different imputations, minimizing the effects of individual imputed values, and the SuperLearner analysis used complete case analysis and yielded satisfactory results for the 24- and 36-month predictor sets. Multiple imputation using data adaptive approaches like SuperLearner remains an open statistical challenge, as it is difficult to define an imputation model that is compatible with the prediction model (Dashti et al., 2021). Hence, we used a complete case analysis to estimate the accuracy of predictor sets, but this might have introduced bias or reduced precision by excluding participants with missing data.

Future directions

Next, we will test the accuracy of the predictor sets we identified in a separate population-based cohort, to assess how our results generalize to the wider population (Gasparini et al., 2022). The sets of predictors should also be assessed for their efficacy and reliability when collected in real-world conditions.

We opted for a population-wide approach in identifying predictors, but future work could focus on identifying precise predictors in subgroups. Namely, a future study should evaluate the accuracy of the measures we identified on a larger group of multilingual

children. Multilingual children are at risk both of over- and under-diagnosis of Language Disorder (Grimm & Schulz, 2014) and so would especially benefit from accurate early identification and support. Future work could also focus on predicting which children with a given condition (e.g., Autism) are at greater risk of poor language outcomes.

Parent-reported measures are a useful resource because they currently yield comparable accuracy to direct assessments in the early years (Wallace et al., 2015), are inexpensive and low burden. However, technological advances may improve precision by enabling accurate, direct measures of predictors that can be collected on a home device and automatically coded (e.g., Chang et al., 2021; identifying Autism). Polygenic risk scores may also allow us to identify genetic variance largely unaccounted for by phenotypic measures and improve accuracy (Mountford et al., 2022).

Conclusion

We have identified two short sets of questions that can be asked at 24 or 36 months in under 1 minute with "fair" accuracy (≥73% sensitivity and specificity) for predicting which children will have low language skills at 11 years. The next step is to replicate our findings in a separate population-based cohort to assess whether these sets of predictors have adequate accuracy for predicting language outcome in the wider population.

Key points:

- Identifying young children who will have persisting Language Disorder throughout childhood can allow provision of targeted support while avoiding over-servicing.
- We used machine-learning methods to identify sets of 7 questions asked at ages 2 or 3 years with ≥73% sensitivity and specificity for predicting low language outcome at 11 years.
- These results should be replicated in a separate cohort to evaluate their generalizability.

References

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & and the CATALISE-2 consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLOS ONE*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Borovsky, A., Thal, D., & Leonard, L. B. (2021). Moving towards accurate and early prediction of language delay with network science and machine learning approaches. *Scientific Reports*, *11*(1), 8136. https://doi.org/10.1038/s41598-021-85982-0

Chang, Z., Martino, J. M. D., Aiello, R., Baker, J., Carpenter, K., Compton, S., Davis, N., Eichner, B., Espinosa, S., Flowers, J., Franz, L., Harris, A., Howard, J., Perochon, S., Perrin, E. M., Babu, P. R. K., Spanos, M., Sullivan, C., Walter, B. K., … Sapiro, G. (2021). Computational Methods to Measure Patterns of Gaze in Toddlers With Autism Spectrum Disorder. *JAMA Pediatrics*, E1–E10. https://doi.org/10.1001/jamapediatrics.2021.0530

Conti-Ramsden, G., & Durkin, K. (2012). Postschool Educational and Employment Experiences of Young People With Specific Language Impairment. *Language, Speech, and Hearing Services in Schools*, *43*(4), 507–520. https://doi.org/10.1044/0161-1461(2012/11-0067)

Dale, P. S. (2007). *MacArthur-Bates Communicative Development Inventory—III* (3rd edition). Paul H. Brookes Publishing Company.

Dashti, S. G., Lee, K. J., Simpson, J. A., White, I. R., Carlin, J. B., & Moreno-Betancur, M. (2021). *Handling missing data when estimating causal effects with Targeted Maximum Likelihood Estimation*. arXiv. https://doi.org/10.48550/arXiv.2112.05274

Eadie, P., Bavin, E. L., Bretherton, L., Cook, F., Gold, L., Mensah, F., Wake, M., & Reilly, S. (2021). Predictors in infancy for language and academic outcomes at 11 years. *Pediatrics*, *147*(2), e20201712. https://doi.org/10.1542/peds.2020-1712

Eadie, P., Conway, L., Hallenstein, B., Mensah, F., McKean, C., & Reilly, S. (2018). Quality of life in children with developmental language disorder: Quality of life in children with DLD. *International Journal of Language & Communication Disorders*, *53*(4), 799–810. https://doi.org/10.1111/1460-6984.12385

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, S., & Bates, E. (2006). *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual* (Second). Singular Publishing Group.

Ganna, A., & Ingelsson, E. (2015). 5 year mortality predictors in 498 103 UK Biobank participants: A prospective population-based study. *The Lancet*, *386*(9993), 533–540. https://doi.org/10.1016/S0140-6736(15)60175-1

Gasparini, L., Shepherd, D. A., Wang, J., Wake, M., & Morgan, A. T. (2022). *Estimating the accuracy of early life predictors of 11-year language outcome: A replication study* [Preregistration]. https://osf.io/jk32c/

Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research*, *7*(4), 821–841. https://doi.org/10.1007/s12187-013-9230-6

Hothorn, T. (2005). Survival ensembles. *Biostatistics*, *7*(3), 355–373. https://doi.org/10.1093/biostatistics/kxj011

Lever, J., Krzywinski, M., & Altman, N. (2016). Model selection and overfitting. *Nature Methods*, *13*(9), 703–704. https://doi.org/10.1038/nmeth.3968

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

McKean, C., Law, J., Mensah, F., Cini, E., Eadie, P., Frazer, K., & Reilly, S. (2016). Predicting meaningful differences in school-entry language skills from child and family factors measured at 12 months of age. *International Journal of Early Childhood*, *48*(3), 329–351. https://doi.org/10.1007/s13158-016-0174-0

McKean, C., Reilly, S., Bavin, E. L., Bretherton, L., Cini, E., Conway, L., Cook, F., Eadie, P., Prior, M., Wake, M., & Mensah, F. (2017). Language outcomes at 7 years: Early predictors and co-occurring difficulties. *Pediatrics*, *139*(3), e20161684. https://doi.org/10.1542/peds.2016-1684

McKean, C., Wraith, D., Eadie, P., Cook, F., Mensah, F., & Reilly, S. (2017). Subgroups in language trajectories from 4 to 11 years: The nature and predictors of stable, improving and decreasing language trajectory groups. *Journal of Child Psychology and Psychiatry*, *58*(10), 1081–1091. https://doi.org/10.1111/jcpp.12790

McNamara, M. E., Zisser, M., Beevers, C. G., & Shumake, J. (2022). Not just “big” data: Importance of sample size, measurement error, and uninformative predictors for developing prognostic models for digital interventions. *Behaviour Research and Therapy*, *153*, 104086. https://doi.org/10.1016/j.brat.2022.104086

Mountford, H. S., Braden, R., Newbury, D. F., & Morgan, A. T. (2022). The Genetic and Molecular Basis of Developmental Language Disorder: A Review. *Children*, *9*(5), 586. https://doi.org/10.3390/children9050586

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical

presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Pearson Education. (2008). *Clinical Evaluation of Language Fundamentals–Fourth Edition Technical Report*. https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/celf-4/celf-4-technical-report.pdf

R Core Team. (2020). *R: A language and environment for statistical computing* (3.6.3). R Foundation for Statistical Computing. https://www.Rproject.org/

Reilly, S., Cook, F., Bavin, E. L., Bretherton, L., Cahir, P., Eadie, P., Gold, L., Mensah, F., Papadopoullos, S., & Wake, M. (2018). Cohort Profile: The Early Language in Victoria Study (ELVS). *International Journal of Epidemiology*, *47*(1), 11–20. https://doi.org/10.1093/ije/dyx079

RStudio Team. (2020). *RStudio: Integrated development environment for R* (1.3.959). RStudio, PBC. http://www.rstudio.com/

Rudolph, J. M. (2017). Case History Risk Factors for Specific Language Impairment: A Systematic Review and Meta-Analysis. *American Journal of Speech-Language Pathology*, *26*(3), 991–1010. https://doi.org/10.1044/2016_AJSLP-15-0181

Rudolph, J. M., & Leonard, L. B. (2016). Early Language Milestones and Specific Language Impairment. *Journal of Early Intervention*, *38*(1), 41–58. https://doi.org/10.1177/1053815116633861

Sansavini, A., Favilla, M. E., Guasti, M. T., Marini, A., Millepiedi, S., Di Martino, M. V., Vecchi, S., Battajon, N., Bertolo, L., Capirci, O., Carretti, B., Colatei, M. P., Frioni, C., Marotta, L., Massa, S., Michelazzo, L., Pecini, C., Piazzalunga, S., Pieretti, M., … Lorusso, M. L. (2021). Developmental Language Disorder: Early Predictors, Age for

the Diagnosis, and Diagnostic Tools. A Scoping Review. *Brain Sciences*, *11*(5), 654. https://doi.org/10.3390/brainsci11050654

Sanson, A., & Johnstone, R. (2004). "Growing Up in Australia" takes its first steps. *Family Matters*, *67*, 46–53.

Semel, E., Wiig, E. H., & Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals- Fourth Edition, Australian Standardised Edition* (4th edition). PsychCorp.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, *8*(1), 25. https://doi.org/10.1186/1471-2105-8-25

Strobl, C., Hothorn, T., & Zeileis, A. (2009). Party on! A new, conditional variable-importance measure for random forests available in the party package. *The R Journal*, *1*(2), 14–17.

Tagliamonte, S. A., & Baayen, R. H. (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178. https://doi.org/10.1017/S0954394512000129

van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1), Article 25. https://doi.org/10.2202/1544-6115.1309

Wake, M., Tobin, S., Girolametto, L., Ukoumunne, O. C., Gold, L., Levickis, P., Sheehan, J., Goldfeld, S., & Reilly, S. (2011). Outcomes of population based language promotion for slow to talk toddlers at ages 2 and 3 years: Let's Learn Language cluster randomised controlled trial. *BMJ*, *343*(aug18 2), d4741–d4741. https://doi.org/10.1136/bmj.d4741

Wallace, I. F., Berkman, N. D., Watson, L. R., Coyne-Beasley, T., Wood, C. T., Cullen, K., & Lohr, K. N. (2015). Screening for speech and language delay in children 5 years old and younger: A systematic review. *Pediatrics*, *136*(2), e448–e462. https://doi.org/10.1542/peds.2014-3889

Yew, S. G. K., & O'Kearney, R. (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies: SLI and emotional and behavioural disorders. *Journal of Child Psychology and Psychiatry*, *54*(5), 516–524. https://doi.org/10.1111/jcpp.12009

Zambrana, I. M., Pons, F., Eadie, P., & Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset. *International Journal of Language & Communication Disorders*, *49*(3), 304–316. https://doi.org/10.1111/1460-6984.12073