BMJ Paediatrics Open

Risk factors for early language delay in children within a minority ethnic, bilingual, deprived environment (Born in Bradford's Better Start): a UK community birth cohort study

Rachael W Cheung [ID] ,[1,2] Kathryn Willan,[2] Josie Dickerson,[2]
Claudine Bowyer-Crane[3,4]

[1]Department of Health Sciences, University of York, York, UK
[2]Better Start Bradford Innovation Hub, Born in Bradford, Bradford Institute for Health Research, Bradford, UK
[3]National Institute of Economic and Social Research, London, UK
[4]Department of Education, University of Sheffield, Sheffield, UK

Correspondence to
Dr Rachael W Cheung; rachael.cheung@york.ac.uk

ABSTRACT

Background Preschool language skills and language delay predict academic and socioemotional outcomes. Children from deprived environments are at a higher risk of language delay, and both minority ethnic and bilingual children can experience a gap in language skills at school entry. However, research that examines late talking (preschool language delay) in an ethnically diverse, bilingual, deprived environment at age 2 is scarce.

Methods Data from Born in Bradford's Better Start birth cohort were used to identify rates of late talking (≤10th percentile on the Oxford-Communicative Development Inventory: Short) in 2-year-old children within an ethnically diverse, predominantly bilingual, deprived UK region (N=712). The relations between known demographic, maternal, distal and proximal child risk factors, and language skills and language delay were tested using hierarchical linear and logistic regression.

Results A total of 24.86% of children were classified as late talkers. Maternal demographic factors (ethnicity, born in UK, education, financial security, employment, household size, age) predicted 3.12% of the variance in children's expressive vocabulary. Adding maternal language factors (maternal native language, home languages) and perinatal factors (birth weight, gestation) to the model predicted 3.76% of the variance. Adding distal child factors (child sex, child age) predicted 11.06%, and adding proximal child factors (receptive vocabulary, hearing concerns) predicted 49.51%. Significant risk factors for late talking were male sex (OR 2.07, 95% CI 1.38 to 3.09), receptive vocabulary delay (OR 8.40, 95% CI 4.99 to 14.11) and parent-reported hearing concerns (OR 7.85, 95% CI 1.90 to 32.47). Protective factors were increased household size (OR 0.85, 95% CI 0.77 to 0.95) and age (OR 0.82, 95% CI 0.70 to 0.96).

Conclusions Almost one in four children living in an ethnically diverse and deprived UK area have early language delay. Demographic factors explained little variance in early vocabulary, whereas proximal child factors held more predictive value. The results indicate further research on early language delay is warranted for vulnerable groups.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Children from minority ethnic, bilingual, deprived populations are at risk of delayed school entry language skills and subsequent lower academic outcomes.
⇒ Existing research on early language delay (late talking) is limited to monolingual, white, mid-socioeconomic status populations.

WHAT THIS STUDY ADDS

⇒ 24.86% of 2 year olds in a minority ethnic, bilingual, deprived population were classified as having late talking status, approximately double that of other cohort studies.
⇒ Demographics explain little variance, whereas proximal child factors such as, for example, receptive vocabulary, may provide more predictive value.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Highlights the potential need for early language intervention in vulnerable groups in the UK.
⇒ Provides estimates to aid public health planning.

INTRODUCTION

Early language delay is associated with poor academic attainment and socioemotional well-being,[1 2] and can be an early manifestation of neurodevelopmental disorders.[3] Vocabulary development is sensitive to negative socioeconomic effects,[4 5] and there is a well-known achievement gap between children from mid-high socioeconomic status (SES) and those from deprived[6] and minority ethnic[7] communities. Bilingual children with language delay are also at risk of under detection.[8] Therefore, minority ethnic children from deprived backgrounds with low English language exposure may be particularly vulnerable to early language delay and low school readiness.[9]

Cohort research on preschool language delay tests for 'late talking' status, defined as children at or below the 10th percentile for expressive vocabulary size expected for 2 years old as compared with population norms.[10–12] Late talking prevalence rates in five English-speaking, predominantly white cohorts (Australia, Canada, UK, USA) were 9.6%–15%,[10 12–15] although another study reported 19.1% (Australia[11]). Two cohorts with more ethnic and linguistic diversity reported similar rates (8.7%–15.4%) but only 20%–30% of these cohorts were from minority ethnic or bilingual communities.[16 17] Research on late talking in deprived populations is also scarce. One UK study reported 23% of 2-year-old skewed towards socioeconomic disadvantage had broader language delay.[18] A US study identified 35.9% of children from low-SES backgrounds were late talking, 28.4% of which were exposed to a second language, as compared with 11.7% of a mid-SES sample, where 14% had exposure to a second language.[19] These limited results indicate that children from minority ethnic, bilingual, deprived homes may be at a higher risk of early language delay.

Of note is that population-level variables are of limited predictive value alone.[20] More proximal child factors may also hold additional predictive value, but are not always measured at population level. Demographic, maternal and distal child factors (such as age and sex) account for ~4%–6.2% of the variance in children's expressive vocabulary, whereas proximal child factors such as receptive vocabulary and concurrent developmental skills account for ~9%–30%.[11 16 21] However, we do not know if these patterns hold also for bilingual, deprived communities in the UK—despite their potential vulnerability to early language delay. The recent Early Language Identification Measure (ELIM) can be administered by health visitors at the 2–2.5 years old universal screening visit in England if concerned about language skills and includes some proximal child factors.[18] However, although the ELIM was tested on a diverse SES sample, ~70% participants were from monolingual English-speaking homes, requiring further research to gauge need in multilingual populations.

This study uses the Born in Bradford's Better Start (BiBBS) data,[22] a community birth cohort that uses routine data collection covering three regions of Bradford, UK. In BiBBS, 85% of participants are from minority ethnic backgrounds and live in deprived communities, with multiple languages spoken at home. Approximately, one-third of mothers report understanding and speaking some, a little or no English during pregnancy.[23] The present study sought to identify predictors of children's vocabulary, rates of late talking as compared with normative data from other cohorts, and potential risk factors (including proximal child language factors) for delay in the 2-year-old children in BiBBS.

METHODS

Population

The BiBBS experimental birth cohort was established to evaluate early-life interventions (birth to 4 years old) in three inner-city areas of Bradford, funded by the National Lottery Community Fund, via the Better Start Bradford programme.[22] BiBBS recruits women during pregnancy and up to 2 weeks post partum, where they complete a baseline questionnaire and consent to linkage and research use of their and their child's routinely collected health, education and Better Start Bradford intervention data. BiBBS includes a representative sample of families living in the Better Start Bradford areas.[23]

From an interim BiBBS pre-COVID-19 data freeze of 2626 pregnancies recruited between January 2016 and November 2019, 2485 children had linkable individual data to early-life interventions (detail on data linkage in online supplemental information).[23] Of these, 1027 attended a language screening and intervention, Talking Together, designed by Bradford-based service BHT Early Education and Training with local speech and language therapists. BHT receive contact information about all 2 years old in the Better Start Bradford areas from the National Health Service (NHS) and invite them to a screening visit (Talking Together Screening). At this visit, practitioners undertake the Oxford-Communicative Development Inventory: Short (CDI-S)[24] and a clinical assessment which includes questions about children's everyday language use. Children who are identified as at risk of language delay on the clinical assessment are referred to a 6-week home-based intervention (Talking Together Intervention; for further information, see Bowyer-Crane et al).[25]

Inclusion and exclusion criteria

Only children who had a Talking Together Screening visit with the outcome measure (CDI-S) in English were included in analyses (n=712). Children were excluded from analyses if they did not have a Talking Together Screening visit, the outcome measure was missing, not in English or the screening language was not recorded (see Figure 1).

Patient and public involvement

Members of the community are involved in the design and conduct of BiBBS cohort research through regular Community Research Advisory Group meetings, composed of Better Start Bradford residents, including recruitment, questionnaires, measures and interpretation and dissemination of findings.
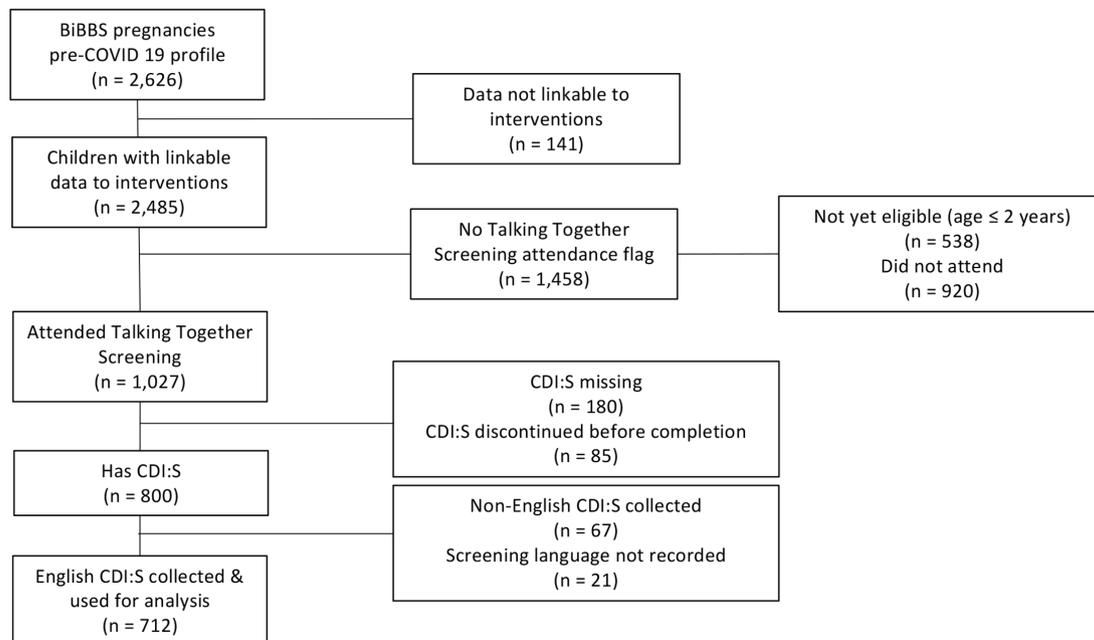
Measures

Potential risk factors

Potential risk factors were identified from prior literature[10 11 13 15 16 21] and are listed in Table 1 with data sources.

Bilingual classification

Children were classified as being from bilingual-English homes based on BiBBS baseline data on

Figure 1 Participant flow diagram. Further details on data linkage and missing data available in online supplemental information.

maternal native language and the languages spoken in the home. Where native language was monolingual-English and languages spoken at home were English only, children were classified as monolingual-English, otherwise children were classified as bilingual-English (see online supplemental information).

Outcome measures

Raw expressive vocabulary scores on the CDI-S[24] collected at Talking Together Screening were used as the outcome measure (please see Supporting Information for receptive vocabulary as outcome measure). The CDI-S is a 100-item parent-reported checklist of words that children understand (receptive) and say (expressive vocabulary). Receptive vocabulary was used as a predictor of expressive vocabulary.[26] Data were collected in the child's home by trained staff from BHT Early Education and Training who supported

parents in completing the measure. Parents were asked which language they were most comfortable using for the screening visit; only data from those who completed the visit and CDI-S in English were included. Maternal self-reported confidence in oral English language ability from the BiBBS baseline data in can be viewed in Table 2.

Late talking was defined as ≤10th percentile for expressive raw scores using the CDI-S[24] and the UK Bilingual Assessment Tool (UKBTAT) normative data,[27] consistent with the literature.[26] The CDI-S published data were used to calculate norms for monolingual English children aged 2 years old. The UKBTAT published data (normed on 2 years old using the CDI-S word list) were used to estimate norms for bilingual-English children with exposure to additional languages.

Table 1	A priori groups of variables used in analyses	
Variable group	**Potential risk factors**	**Data source**
Model 1: Maternal demographic factors	Ethnicity, whether born in the UK, education level,* maternal age,* parental employment,* financial security, size of household	BiBBS baseline questionnaire
Model 2: Maternal language factors	Native language, languages spoken at home	BiBBS baseline questionnaire
Model 3: Perinatal factors	Birth weight, preterm status	Linked health data (maternity records)
Model 4: Distal child factors	Child sex, child age	Linked health data (maternity records), Talking Together service data
Model 5: Proximal child factors	Receptive vocabulary score, parent-reported hearing concerns	Talking Together service data (Oxford CDI-Short; parent-reported questionnaire)

For information on how data was linked, see online supplemental supporting information.
*This variable was derived using the data source listed. See online supplemental supporting information for further details.
BiBBS, Born in Bradford's Better Start; CDI, Communicative Development Inventory.

Table 2 Characteristics of mothers who had Talking Together Screening

Maternal data	Included in analysis (n=712)	Excluded from analysis* (n=315)	Did not have screening (n=1458)
Age at screening (years), mean (SD)	31.52 (5.44)	32.38 (5.42)	NA
Age at BiBBS baseline questionnaire (years), mean (SD)	29.42 (5.43)	30.26 (5.43)	29.10 (5.49)
Index of Multiple Deprivation 2019			
Most deprived, n (%)	593 (83.29)	253 (80.03)	1193 (81.82)
Second most deprived, n (%)	102 (14.33)	53 (16.83)	179 (12.28)
Third most deprived, n (%)	12 (1.69)	5 (1.59)	26 (1.78)
Fourth most deprived, n (%)	0	0	1 (0.07)
Fifth most deprived, n (%)	0	0	2 (0.14)
Missing, n (%)	5 (0.70)	4 (1.27)	57 (3.91)
Self-reported financial security			
Comfortable, n (%)	243 (34.13)	104 (33.02)	482 (33.06)
Doing alright, n (%)	281 (49.47)	114 (36.19)	556 (38.13)
Just getting by, n (%)	108 (15.17)	53 (16.83)	220 (15.09)
Finding it quite difficult, n (%)	31 (4.35)	19 (6.03)	69 (4.73)
Finding it very difficult, n (%)	7 (0.98)	6 (1.90)	27 (1.85)
Missing, n (%)	42 (5.90)	19 (6.03)	104 (7.13)
Ethnicity			
Pakistani heritage, n (%)	477 (66.99)	211 (66.98)	803 (55.08)
White British, n (%)	68 (9.55)	26 (8.25)	190 (13.03)
Central/Eastern European, n (%)	17 (2.39)	7 (2.22)	56 (3.84)
Other,† n (%)	143 (20.08)	66 (20.95)	355 (24.35)
Missing, n (%)	7 (0.98)	5 (1.59)	54 (3.70)
Education			
Degree or equivalent, n (%)	228 (32.02)	78 (24.76)	419 (28.74)
A Levels or equivalent, n (%)	87 (12.22)	25 (7.94)	162 (11.11)
GCSEs or equivalent, n (%)	299 (41.99)	133 (42.22)	596 (40.88)
No qualifications, n (%)	53 (7.44)	43 (13.65)	123 (8.44)
Missing/other/don't know, n (%)	45 (6.32)	36 (11.43)	158 (10.84)
Native language			
English, n (%)	204 (28.65)	64 (20.32)	414 (28.40)
Bilingual including English, n (%)	203 (28.51)	60 (19.05)	373 (25.58)
Bilingual not including English, n (%)	33 (4.63)	22 (6.98)	58 (3.98)
Non-English, n (%)	244 (34.27)	146 (46.35)	507 (34.77)
Missing, n (%)	28 (3.93)	23 (7.30)	106 (7.27)
Confidence in speaking English			
Very well, n (%)	170 (23.88)	67 (21.27)	198 (13.58)
Quite well, n (%)	75 (10.53)	21 (6.67)	85 (5.83)
Some, n (%)	47 (6.60)	22 (6.98)	58 (3.98)
A little bit, n (%)	53 (7.44)	43 (13.65)	80 (5.49)
Not at all, n (%)	5 (0.70)	12 (3.81)	12 (0.82)
Missing, n (%)	362 (14.57)	150 (6.04)	1025 (70.30)
Confidence in understanding English			
Very well, n (%)	210 (29.49)	80 (25.40)	325 (22.29)
Quite well, n (%)	113 (15.8)	43 (13.65)	154 (10.56)

Table 2 Continued

Maternal data	Included in analysis (n=712)	Excluded from analysis* (n=315)	Did not have screening (n=1458)
Some, n (%)	50 (7.02)	41 (13.02)	106 (7.27)
A little bit, n (%)	56 (7.87)	53 (16.83)	173 (11.87)
Not at all, n (%)	2 (0.28)	12 (3.81)	9 (0.62)
Missing, n (%)	281 (39.47)	86 (27.30)	691 (47.39)

*Excluded due to either CDI-S missing, discontinued before completion, non-English CDI-S collected, or screening language not recorded; see Figure 1 for further breakdown.
†Includes African, Bangladeshi, Caribbean, Indian, mixed ethnic groups.
BiBBS, Born in Bradford's Better Start; CDI-S, Communicative Development Inventory: Short; GCSE, General Certificate of Secondary Education; NA, not available.

Statistical analysis

The analyses were preregistered (https://osf.io/e29hs/?view_only=78c6cc6e95a94fc4a1a09fb8c539b94f) and conducted in R. Analysis 1 used hierarchical linear regression analyses containing a priori sets of variables[28] to identify predictors of children's raw expressive vocabulary scores on the CDI-S. Models were built up consecutively with each variable set using forced entry, and the percentage of unique variance was identified at each step. Results are reported as unstandardised estimates and those with $p < 0.05$ reported as significant risk factors. Analysis 2 used a general linear model to predict late talking (delay=1, no delay=0) using all predictors in Analysis 1. Results are reported with ORs and confidence intervals, with p values reported for information.[29]

Missing data

A total of 82.58% of cases were complete for analysis (17.40% missing; Figure 1; Supporting Information, online supplemental table S1). The data were not MCAR (Little's test p<0.001). Although there is no statistical test for MNAR versus MAR data, multiple imputation is considered less biased than complete cases.[30] We therefore used multiple imputation with 18 datasets, similar to the percentage of incomplete cases,[31] using the mice package in R. All results report the pooled estimates of the imputed datasets.

RESULTS

Participants

Characteristics are available in Tables 2 and 3. Families were predominantly in the first (83.29% of sample) or second (14.33%) most deprived decile using the Indices of Multiple Deprivation 2019. A total of 18% of children were classified as monolingual-English and 82% as bilingual-English. The mean age of children at screening

Table 3 Characteristics of children who had Talking Together Screening

Child data	Included in analysis (n=712)	Excluded from analysis* (n=315)	Did not have screening (n=1458)
Age at screening (years), mean (SD)	2.10 (0.11)	2.13 (0.15)	NA
Sex, m:f (% m: % f)	344:368 (48:52)	175:139 (56:44)	718:735 (49:51)
Missing, n (%)	1 (0.14)	1 (0.32)	5 (0.34)
Gestation (weeks), mean (SD)	39.36 (1.84)	39.22 (2.08)	39.31 (1.91)
Prematurity (<37 weeks), n (%)	26 (3.65)	21 (6.67)	76 (5.21)
Low birth weight (<2.5 kg), n (%)	68 (9.55)	35 (11.11)	148 (10.15)
Ethnicity†			
Pakistani heritage, n (%)	435 (61.10)	198 (62.86)	NA
White British, n (%)	47 (6.60)	18 (5.71)	NA
White Other, n (%)	26 (3.65)	10 (3.17)	NA
Other‡, n (%)	135 (18.96)	56 (17.78)	NA
Missing, n (%)	27 (3.79)	12 (3.81)	NA

*Excluded due to either CDI-S missing, discontinued before completion, non-English CDI-S collected, or screening language not recorded; see Figure 1 for further breakdown.
†As child ethnicity was collected by the service, the categories differ slightly to maternal ethnicity. Child ethnicity is not available for those not screened.
‡Includes African, Bangladeshi, Caribbean, Indian, mixed ethnic groups.
CDI-S, Communicative Development Inventory: Short; NA, not available.

was 25.21 months (SD=1.30). Children had a mean raw score of 69.06 (SD=22.01) for receptive vocabulary and a mean raw score of 49.93 (SD=28.82) on the CDI-S.

Language delay at 2 years old

A total of 177 children (24.86%) were of late talking status at the Talking Together screening visit, and 221 (31.04%) of the sample were referred to Talking Together Intervention, based on the clinical assessment.

Analysis 1: predictors of expressive vocabulary using known variables

Full results can be viewed in Table 4. Demographics explained only 3.12% of the variance. Pakistani heritage significantly predicted lower expressive vocabulary across all models; Model 1–Model 4 indicated where children were of Pakistani heritage as compared with white British, they had an average difference of B=−10.56 to −12.33 in expressive vocabulary, although model 5 indicated with the addition of proximal child factors, this effect was less (B=−3.05). This pattern was also present for those of Central/Eastern European and 'Other' relative to white British ethnicities, although these did not reach the significance threshold (model 1–4: Bs=−6.98 to −12.24, model 5: B=−1.86 to −2.84). Factors that predicted a decrease in average expressive vocabulary, but also did not reach the threshold of significance, were having an education level below that of degree, having a financial security level less than 'comfortable' (with the except of 'just getting by', which was associated with a small increase), maternal age at screening and neither parent being employed (see Table 4). Factors that predicted an increase in average expressive vocabulary were not being born in the UK, and an increased size of household—however, all estimates were relatively small, and only size of household was significant (B=0.94–1.36).

Adding maternal factors did not significantly improve the model, contributing 0.32% variance, with model 2 explaining 3.44% overall. Having a mother who was bilingual was associated with a small detriment to expressive vocabulary, as was having bilingual or no English languages at home at baseline, although these estimates were not significant.

Perinatal factors also did not significantly add to the model, contributing an additional 0.32%, with model 3 explaining 3.76% overall. Increased birth weight and gestation estimates were largely associated with a small increase in average vocabulary, although these increases were not significant.

Distal child factors provided a significant increase to the variance of the model, contributing an additional 7.30%, with model 4 explaining 11.06% of the variance overall. Male sex indicated a significant decrease in average expressive vocabulary similar to being of Pakistani heritage (Bs=−7.26 to −12.33), and increased child age indicated a significant increase (Bs=2.96–4.32), although this was less than that of male sex and Pakistani heritage.

Proximal child factors were also a significant addition, contributing an additional 38.45% of the variance. The final model (model 5) explained 49.51% of the overall variance in expressive vocabulary. Receptive vocabulary was associated with a small but significant increase in expressive vocabulary (B=0.82),[1] and parent-reported hearing concerns was associated with a significant decrease in expressive vocabulary exceeding that of all other predictors (B=−19.45). The model was examined for multicollinearity: variance inflation factors were <3[32] and condition indices <30.[33]

Analysis 2: risk factors for late talking

We tested all variables as risk factors for late talking status using a general linear model with odds ratios (Table 5). These were generally consistent with analysis 1. Children from a Pakistani heritage family were over two times more likely to be delayed (OR 2.43, 95% CI 0.99 to 5.97). An increased number of people in the household was a protective factor (OR 0.85, 95% CI 0.77 to 0.95). Of distal child factors, children who were male at birth were two times more likely to be delayed (OR 2.07, 95% CI 1.38 to 3.09), whereas higher age in months at assessment was a protective factor (OR 0.82, 95% CI 0.70 to 0.96). Of proximal child factors, being ≤10th percentile for receptive vocabulary was associated with an eight-fold increase in late talking status (OR 8.40, 95% CI 4.99 to 14.11), and having a parent-reported hearing concern was associated with an eight-fold increase (OR:7.85, 95% CI:1.90 to 32.47).

DISCUSSION

We identified that rates of late talking were higher in a minority ethnic, bilingual, deprived environment and that more proximal child factors might be useful when identifying early language delay in the community. Late talking rates were approximately double that found in a monolingual, mid-high SES UK sample using similar criteria,[12] higher than other monolingual English-speaking populations globally,[10 11] and higher than those that have more ethnic or linguistic diversity[16 17] even when accounting for bilingualism. Our results are similar to the ELIM validation sample, which included some low SES children (using a different language assessment.)[18] Head Start US data suggest it is low SES, rather than bilingualism, that predicts lower language scores when assessing children in both community and heritage languages,[34] consistent with our results. Mechanisms behind low SES and low language skills are not fully understood, but low SES early caregiving environments appear to have less rich language input than in mid-high SES families,[4 5] which correlates with child brain structure and neurophysiological function.[35]

This study provides evidence for potential risk factors and rates of early language delay in ethnically diverse, deprived communities, making use of bilingual normative data where possible. Although

Table 4 Predictors of expressive vocabulary as a continuous variable (N=712)

	Model 1	Model 2	Model 3	Model 4	Model 5
	Est (p value)	Est (p value)	Est (p value)	Est (p value)	Est (p value)
Maternal demographic factors					
Ethnicity (relative to white British)					
Pakistani heritage*	−12.33 (0.002)	−11.35 (0.012)	−10.56 (0.021)	−11.94 (0.007)	−3.05 (0.362)
Central/Eastern European	−12.24 (0.136)	−10.97 (0.193)	−9.71 (0.252)	−9.90 (0.226)	−1.86 (0.763)
Other	−8.78 (0.054)	−7.73 (0.113)	−6.98 (0.154)	−8.01 (0.090)	−2.84 (0.428)
Born in UK					
No	4.11 (0.098)	5.16 (0.116)	4.72 (0.153)	3.82 (0.230)	−0.03 (0.990)
Education (relative to Degree-level)					
A-levels*	−2.08 (0.319)	−2.02 (0.366)	−2.17 (0.298)	−3.40 (0.161)	−2.23 (0.033)
GCSEs	−4.46 (0.402)	−4.67 (0.448)	−4.97 (0.427)	−5.10 (0.301)	−3.56 (0.185)
None	−1.06 (0.607)	−0.76 (0.686)	−0.98 (0.696)	−1.83 (0.492)	−0.16 (0.376)
Financial security (relative to Comfortable)					
Alright	−5.85 (0.431)	−5.62 (0.420)	−6.59 (0.388)	−5.40 (0.160)	3.84 (0.223)
Just getting by	2.27 (0.174)	1.86 (0.169)	2.21 (0.144)	2.24 (0.120)	−0.76 (0.151)
Quite difficult	−3.31 (0.775)	−3.34 (0.893)	−3.87 (0.863)	−5.03 (0.737)	−5.86 (0.970)
Very difficult	−2.05 (0.647)	−2.01 (0.615)	−2.11 (0.557)	−2.64 (0.621)	−2.57 (0.634)
Parental employment					
Neither parent employed	−2.14 (0.618)	−1.85 (0.662)	−1.79 (0.603)	−3.01 (0.580)	−3.05 (0.804)
Size of household (number of individuals)*	1.36 (0.002)	1.35 (0.003)	1.33 (0.003)	1.29 (0.003)	0.94 (0.004)
Maternal age at screening	−0.11 (0.606)	−0.14 (0.500)	−0.12 (0.595)	−0.09 (0.685)	0.02 (0.887)
Maternal language factors					
First language (relative to English)					
Bilingual including English		−1.43 (0.667)	−1.74 (0.602)	−0.13 (0.967)	−0.91 (0.705)
No English		0.62 (0.880)	0.17 (0.966)	1.51 (0.706)	−0.36 (0.905)
Languages at home (relative to English)					
Bilingual including English		−0.28 (0.927)	0.09 (0.975)	−1.71 (0.559)	−1.33 (0.548)
No English		−4.05 (0.269)	−3.57 (0.332)	−4.52 (0.205)	−2.08 (0.436)
Perinatal factors					
Birth weight (kg)			1.13 (0.663)	3.05 (0.226)	2.42 (0.202)
Gestation at birth (weeks)			0.63 (0.409)	0.31 (0.676)	−0.41 (0.461)
Distal child factors					
Child sex					
Male*				−12.33 (<0.001)	−7.26 (<0.001)
Child age (months)*				4.32 (<0.001)	2.96 (<0.001)
Proximal child factors					
Receptive vocabulary (raw score)*					0.82 (<0.001)
Parent-reported hearing concerns					
Yes*					−19.45 (0.001)
R^2	3.12%	3.44%	3.76%	11.06%	49.51%
Δ R^2	NA	+0.32%	+0.32%	+7.30%	+38.45
F for change in R^2 (p value)	1.51 (0.098)	0.52 (0.722)	1.14 (0.321)	28.36 (<0.001)	197.70 (<0.001)

*Indicates p<0.05.
GCSE, General Certificate of Secondary Education.

we did not have the in-depth information that quantifies exposure to each language in bilingual children, and levels of exposure to English affects core language skill,[27] the study provides a practical estimate for minority ethnic, deprived, bilingual community population-level data, as required in front-line health service settings. Additionally, they provide a much-needed baseline for future research, as school-entry language skills have suffered following the pandemic.[36]

Table 5 Risk factors for early expressive vocabulary delay at age 2 (≤10th percentile) (N=712)

	Est	OR	P value	95% CI
Maternal demographic factors				
Ethnicity (relative to white British)				
Pakistani heritage	0.89	2.43	0.054	(0.99 to 5.97)
Central/Eastern European	0.97	2.64	0.196	(0.60 to 11.48)
Other	0.38	1.47	0.437	(0.56 to 3.85)
Born in UK				
No	−0.32	0.73	0.291	(0.40 to 1.32)
Education (relative to degree level)				
A-levels	0.33	1.39	0.318	(0.73 to 2.68)
GCSEs	0.12	1.12	0.625	(0.70 to 1.80)
None	0.04	1.04	0.926	(0.46 to 2.37)
Financial security (relative to comfortable)				
Alright	0.14	1.15	0.538	(0.73 to 1.81)
Just getting by	0.10	1.11	0.748	(0.59 to 2.09)
Quite difficult	0.27	1.31	0.587	(0.49 to 3.47)
Very difficult	−0.63	0.53	0.566	(0.06 to 4.55)
Parental employment				
Neither parent employed	−0.27	0.76	0.522	(0.34 to 1.74)
Size of household (no of individuals)*	−0.16	0.85	0.005	(0.77 to 0.95)
Maternal age at screening	0.02	1.02	0.221	(0.99 to 1.07)
Maternal language factors				
First language (relative to English)				
Bilingual including English	0.10	1.10	0.754	(0.60 to 2.02)
No English	0.29	1.34	0.448	(0.63 to 2.83)
Languages at home (relative to English)				
Bilingual including English	0.19	1.21	0.489	(0.70 to 2.11)
No English	0.47	1.60	0.159	(0.83 to 3.06)
Perinatal factors				
Birth weight (kg)	−0.14	0.87	0.577	(0.54 to 1.41)
Gestation at birth (weeks)	0.02	1.02	0.733	(0.90 to 1.17)
Distal child factors				
Child sex				
Male*	0.73	2.07	0.000	(1.38 to 3.09)
Child age (months)*	−0.20	0.82	0.014	(0.70 to 0.96)
Proximal child factors				
Receptive vocabulary (raw score)*	2.13	8.40	0.000	(4.99 to 14.11)
Parent-reported hearing concerns				
Yes*	2.06	7.85	0.005	(1.90 to 32.47)

*Indicates p<0.05.

Consistent with the existing literature,[11] demographic factors explained little variance in expressive vocabulary. Some risk factors for late talking were consistent with existing literature, such as male sex, hearing concerns and receptive vocabulary, although the CI for hearing concerns was quite wide, suggesting instability. However, unlike prior studies, maternal language and languages spoken at home did not predict higher rates of language delay, potentially because we used bilingual norms. Pakistani heritage as a risk factor for lower expressive vocabulary was consistent with the UK-Millennium Cohort Study.[7] This may in part be explained by larger health inequalities that also disadvantage minority ethnic families,[37] as well as child factors that develop alongside language acquisition, such as motor skills, compounded by socioeconomic factors.[38]

Interestingly, having a larger household was protective, contrary to a Canadian cohort,[10] although the beneficial effect was small. Many Pakistani heritage families have an

intergenerational integrative structure where multiple generations live together as part of a cultural norm;[39] a possible benefit may be a greater variety of speakers, which has previously corresponded with success in word learning.[40] However, these possibilities require further research using in-depth access to the immediate learning environment.[41]

Our results also suggest that assessing individual factors such as receptive vocabulary and hearing concerns may be more useful when detecting early language delay in the community than demographic factors alone. This is consistent with the limited research which suggests that more proximal child factors hold predictive weight[10 21]—further research in this area would likely be fruitful. Screening programmes like ELIM have the advantage of testing some of these and of using existing home-visiting; however, they also place demand on increasingly stretched services—particularly following COVID-19.[42] Any national screening programmes for language delay, therefore, must account for differential local need while planning capacity.

There are several limitations to this study. First, as the sample consisted of only participants who were sufficiently confident in their English ability to complete the Talking Together screening visit in English, families who did not speak sufficient English for the CDI-S are not represented, and may be at even greater risk of under detection of early language delay. Second, the nature of the sample is also a potential threat to validity. The sample only contained those who agreed to the universal screening visit and were from an interventional birth cohort that utilises routinely collected data from the NHS and from commissioned services, and we did not have outcome data for those who did not take part in the screening. This means the prevalence rates of late talking and predictors may not generalise to other populations, and may mean children whose parents are concerned about language skills are over-represented. However, as other community data on preschool language ability for minority ethnic, multilingual, deprived UK populations is scarce, routine data collection can provide some insight into early language delay within a population that is seldom-heard.

In addition, data regarding the home learning environment,[7 21] family history of language delay[10–12 15] or in-depth skills such as phonological development and grammar[43] were not available, all of which are targets for future research. We also do not yet have data on later language skills, although we plan to use national school data to track longer-term outcomes via the BiBBS cohort. Despite some research previously suggesting that late talking children 'catch-up', others find they have persistent difficulties throughout school.[44] Failing to intervene early may thus risk missing a key period for intervention.[45]

CONCLUSION

The achievement gap between low and mid-high SES children starts early in life,[46 47] and the negative effect of SES on child neurocognitive outcomes appears strongest for language when compared with other cognitive domains such as working memory or executive function.[48] Early language intervention may be particularly important in deprived environments;[49] however, to effectively plan services, research in minority ethnic, deprived, bilingual environments is necessary. This study indicates these communities are at risk and further research on early language delay is likely warranted for vulnerable groups in the UK.

ORCID iD
Rachael W Cheung http://orcid.org/0000-0001-5207-240X

REFERENCES

1 Bornstein MH, Hahn CS, Suwalsky JTD. Language and internalizing and externalizing behavioral adjustment: developmental pathways from childhood to adolescence. *Dev Psychopathol* 2013;25:857–78.
2 Conti-Ramsden G, Durkin K, Toseeb U, *et al*. Education and employment outcomes of young adults with a history of developmental language disorder. *Int J Lang Commun Disord* 2018;53:237–55.
3 Sim F, Haig C, O'Dowd J, *et al*. Development of a triage tool for neurodevelopmental risk in children aged 30 months. *Res Dev Disabil* 2015;45–46:69–82.

4   Golinkoff RM, Hoff E, Rowe ML, *et al*. Language matters: denying the existence of the 30-million-word gap has serious consequences. *Child Dev* 2019;90:985–92.

5   Hoff E. The specificity of environmental influence: socioeconomic status affects early vocabulary development via maternal speech. *Child Dev* 2003;74:1368–78.

6   Law J, Charlton J, Dockrell J, *et al*. *Early language development: needs, provision, and intervention for preschool children from socioeconomically disadvantaged backgrounds: a report for the education endowment foundation october 2017*. Public Health Engl, 2017.

7   Zilanawala A, Kelly Y, Sacker A. Ethnic differences in longitudinal latent verbal profiles in the millennium cohort study. *Eur J Public Health* 2016;26:1011–6.

8   Stow C, Dodd B. Providing an equitable service to bilingual children in the UK: a review. *Int J Lang Commun Disord* 2003;38:351–77.

9   Hoff E. Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev Psychol* 2013;49:4–14.

10   Collisson BA, Graham SA, Preston JL, *et al*. Risk and protective factors for late talking: an epidemiologic investigation. *J Pediatr* 2016;172:168–74.

11   Reilly S, Wake M, Bavin EL, *et al*. Predicting language at 2 years of age: a prospective community study. *Pediatrics* 2007;120:e1441–9.

12   Dale PS, Price TS, Bishop DVM, *et al*. Outcomes of early language delay: I. predicting persistent and transient language difficulties at 3 and 4 years. *J Speech Lang Hear Res* 2003;46:544–60.

13   Armstrong R, Scott JG, Whitehouse AJO, *et al*. Late talkers and later language outcomes: predicting the different language trajectories. *Int J Speech Lang Pathol* 2017;19:237–50.

14   Rescorla L, Hadicke-Wiley M, Escarce E. Epidemiological investigation of expressive language delay at age two. *First Language* 1993;13:5–22.

15   Zubrick SR, Taylor CL, Rice ML, *et al*. Late language emergence at 24 months: an epidemiological study of prevalence, predictors, and covariates. *J Speech Lang Hear Res* 2007;50:1562–92.

16   Henrichs J, Rescorla L, Schenk JJ, *et al*. Examining continuity of early expressive vocabulary development: the generation R study. *J Speech Lang Hear Res* 2011;54:854–69.

17   Horwitz SM, Irwin JR, Briggs-Gowan MJ, *et al*. Language delay in a community cohort of young children. *J Am Acad Child Adolesc Psychiatry* 2003;42:932–40.

18   Law J, Charlton J, McKean C, *et al*. Identifying and supporting children's early language needs. Public Health England/Department for Education; 2020.

19   Arriaga RI, Fenson L, Cronan T, *et al*. Scores on the MacArthur communicative development inventory of children from lowand middle-income families. *Applied Psycholinguistics* 1998;19:209–23.

20   Nelson HD, Nygren P, Walker M, *et al*. Screening for speech and language delay in preschool children: systematic evidence review for the US preventive services task force. *Pediatrics* 2006;117:e298–319.

21   Law J, Clegg J, Rush R, *et al*. Association of proximal elements of social disadvantage with children's language development at 2 years: an analysis of data from the children in focus (CIF) sample from the ALSPAC birth cohort. *Int J Lang Commun Disord* 2019;54:362–76.

22   Dickerson J, Bird PK, McEachan RRC, *et al*. Born in Bradford's better start: an experimental birth cohort study to evaluate the impact of early life interventions. *BMC Public Health* 2016;15:711.

23   Dickerson J, Bridges S, Willan K, *et al*. Born in bradford's better start (bibbs) interventional birth cohort study: interim cohort profile. *Wellcome Open Res* 2022;7:244.

24   Hamilton A, Plunkett K, Schafer G. Infant vocabulary development assessed with a British communicative development inventory. *J Child Lang* 2000;27:689–705.

25   Bowyer-Crane C, Nielsen D, Bryant M, *et al*. A randomised controlled feasibility trial and qualitative evaluation of an early years language development intervention: study protocol of the " outcomes of talking together evaluation and results " (otter) project. *Pilot Feasibility Stud* 2019;5:119.

26   Fisher EL. A systematic review and meta-analysis of predictors of expressive-language outcomes among late talkers. *J Speech Lang Hear Res* 2017;60:2935–48.

27   Floccia C, Sambrook T, Delle Luche C, *et al*. Vocabulary of 2-year-olds learning english and an additional language: norms and effects of linguistic distance. In: Floccia C, Sambrook T, Delle Luche C, et al, eds. *Monogr Soc Res Child Dev, 83*. 2018: 1–80.

28   Cohen P, Cohen P, West SG, *et al*. *Applied multiple regression/correlation analysis for the behavioral sciences*. 3rd ed. New York, NY: Routledge, 2015.

29   Wasserstein RL, Lazar NA. The ASA statement on P -values: context, process, and purpose. *Am Stat* 2016;70:129–33.

30   van Ginkel JR, Linting M, Rippe RCA, *et al*. Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *J Pers Assess* 2020;102:297–308.

31   van Buuren S. *Flexible imputation of missing data*. 2nd edn. London: Chapman & Hall/CRC Interdisciplinary Statistics Series, 2018.

32   Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol* 2010;1:3–14.

33   Belsley DA. A guide to using the collinearity diagnostics. In: *Comput Sci Econ Manag, 4*. 1991: 33–50.

34   Hammer CS, Hoff E, Uchikoshi Y, *et al*. The language and literacy development of young dual language learners: a critical review. *Early Child Res Q* 2014;29:715–33.

35   Farah MJ. The neuroscience of socioeconomic status: correlates, causes, and consequences. *Neuron* 2017;96:56–71.

36   Bowyer-Crane C, Bonetti S, Compton S, *et al*. The impact of covid-19 on school starters: interim briefing 1parent and school concerns about children starting school. Education Endowment Foundation: 2021. Available: https://educationendowmentfoundation.org.uk/public/files/Impact_of_Covid19_on_School_Starters_-_Interim_Briefing_1_-_April_2021_-_Final.pdf [Accessed 03 Jan 2021].

37   Evandrou M, Falkingham J, Feng Z, *et al*. Ethnic inequalities in limiting health and self-reported health in later life revisited. *J Epidemiol Community Health* 2016;70:653–62.

38   Kelly Y, Sacker A, Schoon I, *et al*. Ethnic differences in achievement of developmental milestones by 9 months of age: the millennium cohort study. *Dev Med Child Neurol* 2006;48:825–30.

39   Crozier G, Davies J. Family matters: a discussion of the Bangladeshi and Pakistani extended family and community in supporting the children's education. *The Sociological Review* 2006;54:678–95.

40   Rost GC, McMurray B. Speaker variability augments phonological processing in early word learning. *Dev Sci* 2009;12:339–49.

41   Rowe ML. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Dev* 2012;83:1762–74.

42   Morton A, Adams C. Health visiting in England: the impact of the COVID-19 pandemic. *Public Health Nurs* 2022;39:820–30.

43   Bishop DVM, North T, Donlan C. Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *J Child Psychol Psychiatry* 1996;37:391–403.

44   Snowling MJ, Duff FJ, Nash HM, *et al*. Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *J Child Psychol Psychiatry* 2016;57:1360–9.

45   Capone Singleton N. Late talkers: why the wait-and-see approach is outdated. *Pediatr Clin North Am* 2018;65:13–29.

46   Hoff E. How social contexts support and shape language development☆. *Dev Rev* 2006;26:55–88.

47   Huttenlocher J, Waterfall H, Vasilyeva M, *et al*. Sources of variability in children's language growth. *Cogn Psychol* 2010;61:343–65.

48   Farah MJ, Shera DM, Savage JH, *et al*. Childhood poverty: specific associations with neurocognitive development. *Brain Res* 2006;1110:166–74.

49   Foster MA, Lambert R, Abbott-Shim M, *et al*. A model of home learning environment and social risk factors in relation to children's emergent literacy and social outcomes. *Early Child Res Q* 2005;20:13–36.

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Paediatrics Open

1

Supporting Information for Manuscript: 'Risk factors for early language delay in children within a minority ethnic, bilingual, deprived environment (Born in Bradford's Better Start): A UK community birth cohort study'

Bilingual classification

The UKBTAT (Floccia et al., 2018, doi: 10.1111/mono.12349) typically measures the amount of direct exposure to English and another language between child and parent, as well as the amount of overheard speech in English and another language. This is measured using eight questions (includes number of hours in nursery, hours exposed to relatives, hours exposed to each parent, hours of sleep per night, language spoken by each parent to the child, who speaks most to the child, followed by a six-step calculation process outlined in Floccia et al., 2018, Appendix 3). As we did not have the in-depth information required to calculate exposure using the UKBTAT due to the use of a community cohort sample, we used an *a priori* standard of direct exposure and overheard speech for English and another language for all bilingual families, assuming that bilingual children were directly spoken to equally in English and in a second language, and that they overheard their parents speaking English and a second language an equal amount.

Data linkage and sample

The paper uses data from the BiBBS pre-COVID-19 profile (Dickerson et al., 2022, doi: 10.12688/wellcomeopenres.18394.1). Note that because cohort recruitment and data collection is ongoing, the sample within the paper is a data freeze of the data available for linkage in the BiBBS cohort from all pregnancies between 1st Jan 2016 and 30th November 2019, with an estimated due date between 1st April 2016 and 8th March 2020. In order to utilise all data sources, including project data that contained the UK-CDI language measure, project data was linked using the procedures outlined in Dickerson et al., 2022. Full information on this process can be found in the Supplementary Materials for Dickerson et al., 2022 at:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZQIUNC.

Table S1. Missing data by variable used in analyses (total N = 712)

Variable	N missing cases	Percent missing (out of 712)
Ethnicity	7	0.98
Born in UK status	6	0.84
Education level	45	6.32
Employment status	27	3.79
Financial management	42	5.90
Maternal age	3	0.42
Total people in household	7	0.98
Maternal native language	28	3.93
Languages spoken at home	21	2.95
Birthweight status	1	0.14
Preterm status	6	0.84
Child sex	1	0.14
Parent-reported hearing concerns	2	0.28

Derived maternal variables

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Paediatrics Open

2

<u>Maternal age:</u> The original BiBBS cohort dataset lists the age at which mothers took the baseline questionnaire. This information was used and combined with the age of the child at the Talking Together screening to provide maternal age at Talking Together screening.

<u>Maternal education:</u> The original BiBBS cohort dataset has 7 sub-categories for education level as follows:1) No qualifications, 2) 5 or less GCSE (grades A-C) or equivalent, 3) 5 or more GCSE (grades A-C) or equivalent, 4) A levels or equivalent, 5) Degree or equivalent, 6) Don't know, and 7) Other.

      For the purpose of analyses, the above were collapsed into five sub-categories: 1) No qualifications, 2) GCSEs (collapsing category 2 and 3 above), 3) A Levels or equivalent, 4) Degree or equivalent, 5) Missing, don't know, or other (collapsing category 6 and 7 above with missing (NA) entries).

<u>Parental employment:</u> The original BiBBS cohort dataset provides maternal and paternal employment status. For the purpose of analyses, this data was combined to provide a variable that detailed any parental employment. Where at least one parent was employed, then parental employment was coded as 1. Where no parents were employed, then parental employment was coded as 0.

Sensitivity analysis with maternal mental health variables

A sensitivity analysis was carried out, testing the value of maternal mental health variables (Patient Healthcare Questionnaire-8, PHQ-8 [Kronke et al., 2009, doi: 10.1016/j.jad.2008.06.026] and Generalised Anxiety Disorder-7, GAD-7 [Spitze et al., 2006, 10.1001/archinte.166.10.1092]) in the final model when predicting expressive vocabulary at age 2 in the population. These variables did not significantly predict expressive vocabulary. The full results can be viewed in Table S2 below.

Table S2. Sensitivity analysis results: predictors of expressive vocabulary with maternal mental health variables (n = 382)

	Model 1	Model 2	Model 3	Model 4	Model 5
	Estimate (p-value)	Estimate (p-value)	Estimate (p-value)	Estimate (p-value)	Estimate (p-value)
Maternal demographic factors					
Ethnicity (relative to White British)					
Pakistani heritage	-9.90 (.052)	-10.00 (.083)	-8.65 (.138)	-10.93 (.054)	-2.29 (.614)
Central/Eastern European	3.4960 (.792)	4.02 (.767)	5.68 (.675)	6.57 (.618)	14.60 (.164)
Other	-4.78 (.419)	-4.60 (.469)	-3.19 (.618)	-5.31 (.396)	-0.59 (.905)
Born in UK					
No	1.05 (.751)	3.77 (.387)	3.81 (.382)	3.13 (.462)	0.75 (.825)
Education level (relative to Degree)					
A Levels	-5.46 (.270)	-4.93 (.321)	-5.63 (.258)	-7.05 (.146)	-4.03 (.296)
GCSEs	-1.77 (.593)	-0.94 (.785)	-0.86 (.803)	-2.13 (.526)	-1.42 (.594)

Supplemental material
BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)
BMJ Paediatrics Open

3

None	-3.62 (.588)	-1.90 (.782)	-1.68 (.764)	-3.83 (.565)	-3.56 (501)
Financial security (relative to Comfortable)					
Alright	1.08 (.740)	1.37 (.676)	0.99 (.764)	-0.61 (.850)	-0.12 (.964)
Just getting by	-2.91 (.535)	-3.08 (.512)	-3.25 (.490)	-4.18 (.360)	-3.07 (.397)
Quite difficult	2.6924 (.749)	3.24 (.705)	1.67 (.846)	-0.64 (.939)	-2.50 (.707)
Very difficult	-1.24 (.941)	2.11 (.902)	1.73 (.919)	-0.50 (.976)	9.60 (.470)
Parental employment					
Neither parent employed	-1.77 (.804)	-1.68 (.815)	-0.86 (.905)	0.68 (.923)	-1.39 (.878)
Size of household (number of individuals)	1.00 (.065)	0.916 (.095)	0.85 (.123)	0.83 (.122)	0.65 (.127)
Maternal age at screening	-0.17 (.572)	-0.21 (.490)	-0.19 (.517)	-0.08 (.772)	0.04 (.538)
Maternal mental health					
GAD-7 Score	-0.88 (.062)	-0.85 (.073)	-0.75 (.115)	-0.71 (.125)	-0.23 (.526)
PHQ-8 Score	0.42 (.367)	0.35 (.447)	0.26 (.583)	0.11 (.806)	-0.08 (.814)
Maternal language factors					
First language (relative to English)					
Bilingual-English		-2.29 (.583)	-2.67 (.524)	-1.24 (.761)	-2.00 (.538)
Not English		-3.75 (.499)	-4.44 (.424)	-3.09 (.568)	-2.39 (.578)
Languages at home (relative to English)					
Bilingual-English		3.84 (.348)	4.62 (.262)	2.81 (.485)	0.63 (842)
Not English		-1.60 (.756)	-0.89 (864)	-1.42 (.777)	-1.45 (.717)
Perinatal factors					
Birthweight (kg)			3.48 (.356)	5.73 (.121)	4.29 (.144)
Gestation at birth (weeks)			0.64 (.578)	0.18 (.874)	-0.19 (.830)
Distal child factors					
Child Sex					
*Male**				-12.13 (<.001)	-7.89 (.001)
Child age (months)*				3.46 (.002)	2.38 (.008)
Proximal child factors					

4

Receptive vocabulary					0.73 (<.001)
Parent-reported hearing concerns					
Yes					-18.10 (.014)
R²	3.66%	4.38%	5.13%	11.34%	44.37%
Δ R²		+0.72%	+0.75%	+6.21%	+33.03%
F for change in R²		0.68	1.42	12.51	105.36
(*p*-value)		(.605)	(.244)	(<.001)	(<.001)

Receptive vocabulary

An exploratory analysis of receptive vocabulary on the Oxford-CDI: Short as a continuous outcome variable was conducted using all predictors on a multiple imputation dataset as per the main analysis. Results can be viewed in Table S3 below.

Table S3. Predictors of receptive vocabulary as a continuous variable (*N* = 712).

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
	Est (p-value)	*Est (p-value)*	*Est (p-value)*	*Est (p-value)*	*Est (p-value)*
Ethnicity (relative to White British)					
*Pakistani heritage**	-10.72 (.001)	-10.92 (.002)	-10.24 (.003)	-10.88 (.002)	-10.97 (.001)
Central/Eastern European	-9.57 (.125)	-9.53 (.137)	-8.43 (.190)	-8.77 (.168)	-8.38 (.188)
Other	-6.33 (.069)	-6.35 (.089)	-5.66 (.130)	-6.24 (.092)	-6.26 (.090)
Born in UK					
No	4.65 (.014)	4.56 (.065)	4.17 (.092)	3.87 (.114)	3.57 (.146)
Education level (relative to Degree)					
A Levels	1.62 (.561)	1.73 (.536)	1.19 (.671)	0.62 (.824)	0.60 (.828)
GCSEs	0.17 (.932)	0.06 (.977)	0.02 (.993)	-0.17 (.931)	-0.23 (.907)
None	-0.01 (.998)	0.15 (.965)	0.21 (.952)	-0.31 (.926)	-0.33 (.922)
Financial security (relative to Comfortable)					
Alright	-0.85 (.655)	-0.81 (.672)	-0.91 (.633)	-1.52 (.422)	-1.60 (.396)
Just getting by	-0.96 (.710)	-1.11 (.669)	-1.46 (.576)	-1.67 (.517)	-1.73 (.502)
Quite difficult	-1.47 (.729)	-1.39 (.747)	-1.60 (.709)	-1.96 (.645)	-1.87 (.661)
Very difficult	-8.43 (.323)	-8.48 (.323)	-9.21 (.283)	-8.80 (.299)	-7.71 (.365)

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Paediatrics Open

5

Parental employment					
Neither parent employed	3.59 (.249)	3.44 (.271)	3.75 (.230)	4.02 (.193)	4.12 (.181)
Size of household (number of individuals)	0.47 (.164)	0.45 (.192)	0.43 (.210)	0.43 (.206)	0.43 (.203)
Maternal age at screening	-0.17 (.286)	-0.21 (.203)	-0.16 (.332)	-0.14 (.395)	-0.14 (.391)
First language (relative to English)					
Bilingual-English		0.72 (.773)	0.45 (.857)	1.14 (.645)	1.24 (.617)
Not English		2.79 (.371)	2.33 (.454)	2.80 (.364)	3.02 (.327)
Languages at home (relative to English)					
Bilingual-English		-0.12 (.960)	0.20 (.933)	-0.55 (.809)	-0.59 (.796)
Not English		-3.39 (.221)	-2.93 (.293)	-3.21 (.245)	-3.27 (.235)
Birthweight (kg)			-0.43 (.829)	0.54 (.784)	0.45 (.818)
Gestation at birth (weeks)			1.00 (.084)	0.84 (.145)	0.82 (.153)
Child Sex *Male**				-6.21 (<.001)	-6.24 (<.001)
Child age (months)*				1.69 (.009)	1.69 (.008)
Parent-reported hearing concerns Yes					-9.16 (.120)
R^2	3.42%	3.82%	4.43%	7.05%	7.38%
Δ R^2	NA	+0.40%	+0.61%	+2.62%	+0.33%
F for change in R^2 (*p*-value)	1.68 (.052)	0.68 (.606)	2.23 (.107)	9.94 (<.001)	2.48 (.115)

*p <.05

1

Supporting Information for Manuscript: 'Risk factors for early language delay in children within a minority ethnic, bilingual, deprived environment (Born in Bradford's Better Start): A UK community birth cohort study'

Bilingual classification
The UKBTAT (Floccia et al., 2018, doi: 10.1111/mono.12349) typically measures the amount of direct exposure to English and another language between child and parent, as well as the amount of overheard speech in English and another language. This is measured using eight questions (includes number of hours in nursery, hours exposed to relatives, hours exposed to each parent, hours of sleep per night, language spoken by each parent to the child, who speaks most to the child, followed by a six-step calculation process outlined in Floccia et al., 2018, Appendix 3). As we did not have the in-depth information required to calculate exposure using the UKBTAT due to the use of a community cohort sample, we used an *a priori* standard of direct exposure and overheard speech for English and another language for all bilingual families, assuming that bilingual children were directly spoken to equally in English and in a second language, and that they overheard their parents speaking English and a second language an equal amount.

Data linkage and sample
The paper uses data from the BiBBS pre-COVID-19 profile (Dickerson et al., 2022, doi: 10.12688/wellcomeopenres.18394.1). Note that because cohort recruitment and data collection is ongoing, the sample within the paper is a data freeze of the data available for linkage in the BiBBS cohort from all pregnancies between 1st Jan 2016 and 30th November 2019, with an estimated due date between 1st April 2016 and 8th March 2020. In order to utilise all data sources, including project data that contained the UK-CDI language measure, project data was linked using the procedures outlined in Dickerson et al., 2022. Full information on this process can be found in the Supplementary Materials for Dickerson et al., 2022 at:
https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ZQIUNC.

Table S1. Missing data by variable used in analyses (total N = 712)

Variable	N missing cases	Percent missing (out of 712)
Ethnicity	7	0.98
Born in UK status	6	0.84
Education level	45	6.32
Employment status	27	3.79
Financial management	42	5.90
Maternal age	3	0.42
Total people in household	7	0.98
Maternal native language	28	3.93
Languages spoken at home	21	2.95
Birthweight status	1	0.14
Preterm status	6	0.84
Child sex	1	0.14
Parent-reported hearing concerns	2	0.28

Derived maternal variables

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Paediatrics Open

2

Maternal age: The original BiBBS cohort dataset lists the age at which mothers took the baseline questionnaire. This information was used and combined with the age of the child at the Talking Together screening to provide maternal age at Talking Together screening.

Maternal education: The original BiBBS cohort dataset has 7 sub-categories for education level as follows:1) No qualifications, 2) 5 or less GCSE (grades A-C) or equivalent, 3) 5 or more GCSE (grades A-C) or equivalent, 4) A levels or equivalent, 5) Degree or equivalent, 6) Don't know, and 7) Other.

For the purpose of analyses, the above were collapsed into five sub-categories: 1) No qualifications, 2) GCSEs (collapsing category 2 and 3 above), 3) A Levels or equivalent, 4) Degree or equivalent, 5) Missing, don't know, or other (collapsing category 6 and 7 above with missing (NA) entries).

Parental employment: The original BiBBS cohort dataset provides maternal and paternal employment status. For the purpose of analyses, this data was combined to provide a variable that detailed any parental employment. Where at least one parent was employed, then parental employment was coded as 1. Where no parents were employed, then parental employment was coded as 0.

Sensitivity analysis with maternal mental health variables

A sensitivity analysis was carried out, testing the value of maternal mental health variables (Patient Healthcare Questionnaire-8, PHQ-8 [Kronke et al., 2009, doi: 10.1016/j.jad.2008.06.026] and Generalised Anxiety Disorder-7, GAD-7 [Spitze et al., 2006, 10.1001/archinte.166.10.1092]) in the final model when predicting expressive vocabulary at age 2 in the population. These variables did not significantly predict expressive vocabulary. The full results can be viewed in Table S2 below.

Table S2. Sensitivity analysis results: predictors of expressive vocabulary with maternal mental health variables (n = 382)

	Model 1	Model 2	Model 3	Model 4	Model 5
	Estimate (p-value)	Estimate (p-value)	Estimate (p-value)	Estimate (p-value)	Estimate (p-value)
Maternal demographic factors					
Ethnicity (relative to White British)					
Pakistani heritage	-9.90 (.052)	-10.00 (.083)	-8.65 (.138)	-10.93 (.054)	-2.29 (.614)
Central/Eastern European	3.4960 (.792)	4.02 (.767)	5.68 (.675)	6.57 (.618)	14.60 (.164)
Other	-4.78 (.419)	-4.60 (.469)	-3.19 (.618)	-5.31 (.396)	-0.59 (.905)
Born in UK					
No	1.05 (.751)	3.77 (.387)	3.81 (.382)	3.13 (.462)	0.75 (.825)
Education level (relative to Degree)					
A Levels	-5.46 (.270)	-4.93 (.321)	-5.63 (.258)	-7.05 (.146)	-4.03 (.296)
GCSEs	-1.77 (.593)	-0.94 (.785)	-0.86 (.803)	-2.13 (.526)	-1.42 (.594)

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Paediatrics Open

3

None	-3.62 (.588)	-1.90 (.782)	-1.68 (.764)	-3.83 (.565)	-3.56 (501)
Financial security (relative to Comfortable)					
Alright	1.08 (.740)	1.37 (.676)	0.99 (.764)	-0.61 (.850)	-0.12 (.964)
Just getting by	-2.91 (.535)	-3.08 (.512)	-3.25 (.490)	-4.18 (.360)	-3.07 (.397)
Quite difficult	2.6924 (.749)	3.24 (.705)	1.67 (.846)	-0.64 (.939)	-2.50 (.707)
Very difficult	-1.24 (.941)	2.11 (.902)	1.73 (.919)	-0.50 (.976)	9.60 (.470)
Parental employment					
Neither parent employed	-1.77 (.804)	-1.68 (.815)	-0.86 (.905)	0.68 (.923)	-1.39 (.878)
Size of household (number of individuals)	1.00 (.065)	0.916 (.095)	0.85 (.123)	0.83 (.122)	0.65 (.127)
Maternal age at screening	-0.17 (.572)	-0.21 (.490)	-0.19 (.517)	-0.08 (.772)	0.04 (.538)
Maternal mental health					
GAD-7 Score	-0.88 (.062)	-0.85 (.073)	-0.75 (.115)	-0.71 (.125)	-0.23 (.526)
PHQ-8 Score	0.42 (.367)	0.35 (.447)	0.26 (.583)	0.11 (.806)	-0.08 (.814)
Maternal language factors					
First language (relative to English)					
Bilingual-English		-2.29 (.583)	-2.67 (.524)	-1.24 (.761)	-2.00 (.538)
Not English		-3.75 (.499)	-4.44 (.424)	-3.09 (.568)	-2.39 (.578)
Languages at home (relative to English)					
Bilingual-English		3.84 (.348)	4.62 (.262)	2.81 (.485)	0.63 (842)
Not English		-1.60 (.756)	-0.89 (864)	-1.42 (.777)	-1.45 (.717)
Perinatal factors					
Birthweight (kg)			3.48 (.356)	5.73 (.121)	4.29 (.144)
Gestation at birth (weeks)			0.64 (.578)	0.18 (.874)	-0.19 (.830)
Distal child factors					
Child Sex					
*Male**				-12.13 (<.001)	-7.89 (.001)
Child age (months)*				3.46 (.002)	2.38 (.008)
Proximal child factors					

4

Receptive vocabulary					0.73 (<.001)
Parent-reported hearing concerns					
Yes					-18.10 (.014)
R²	3.66%	4.38%	5.13%	11.34%	44.37%
Δ R²		+0.72%	+0.75%	+6.21%	+33.03%
F for change in R² **(*p*-value)**		0.68 (.605)	1.42 (.244)	12.51 (<.001)	105.36 (<.001)

Receptive vocabulary

An exploratory analysis of receptive vocabulary on the Oxford-CDI: Short as a continuous outcome variable was conducted using all predictors on a multiple imputation dataset as per the main analysis. Results can be viewed in Table S3 below.

Table S3. Predictors of receptive vocabulary as a continuous variable (*N* = 712).

	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
	Est (p-value)	*Est (p-value)*	*Est (p-value)*	*Est (p-value)*	*Est (p-value)*
Ethnicity (relative to White British)					
*Pakistani heritage**	-10.72 (.001)	-10.92 (.002)	-10.24 (.003)	-10.88 (.002)	-10.97 (.001)
Central/Eastern European	-9.57 (.125)	-9.53 (.137)	-8.43 (.190)	-8.77 (.168)	-8.38 (.188)
Other	-6.33 (.069)	-6.35 (.089)	-5.66 (.130)	-6.24 (.092)	-6.26 (.090)
Born in UK					
No	4.65 (.014)	4.56 (.065)	4.17 (.092)	3.87 (.114)	3.57 (.146)
Education level (relative to Degree)					
A Levels	1.62 (.561)	1.73 (.536)	1.19 (.671)	0.62 (.824)	0.60 (.828)
GCSEs	0.17 (.932)	0.06 (.977)	0.02 (.993)	-0.17 (.931)	-0.23 (.907)
None	-0.01 (.998)	0.15 (.965)	0.21 (.952)	-0.31 (.926)	-0.33 (.922)
Financial security (relative to Comfortable)					
Alright	-0.85 (.655)	-0.81 (.672)	-0.91 (.633)	-1.52 (.422)	-1.60 (.396)
Just getting by	-0.96 (.710)	-1.11 (.669)	-1.46 (.576)	-1.67 (.517)	-1.73 (.502)
Quite difficult	-1.47 (.729)	-1.39 (.747)	-1.60 (.709)	-1.96 (.645)	-1.87 (.661)
Very difficult	-8.43 (.323)	-8.48 (.323)	-9.21 (.283)	-8.80 (.299)	-7.71 (.365)

Supplemental material

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

BMJ Paediatrics Open

5

Parental employment					
Neither parent employed	3.59 (.249)	3.44 (.271)	3.75 (.230)	4.02 (.193)	4.12 (.181)
Size of household (number of individuals)	0.47 (.164)	0.45 (.192)	0.43 (.210)	0.43 (.206)	0.43 (.203)
Maternal age at screening	-0.17 (.286)	-0.21 (.203)	-0.16 (.332)	-0.14 (.395)	-0.14 (.391)
First language (relative to English)					
Bilingual-English		0.72 (.773)	0.45 (.857)	1.14 (.645)	1.24 (.617)
Not English		2.79 (.371)	2.33 (.454)	2.80 (.364)	3.02 (.327)
Languages at home (relative to English)					
Bilingual-English		-0.12 (.960)	0.20 (.933)	-0.55 (.809)	-0.59 (.796)
Not English		-3.39 (.221)	-2.93 (.293)	-3.21 (.245)	-3.27 (.235)
Birthweight (kg)			-0.43 (.829)	0.54 (.784)	0.45 (.818)
Gestation at birth (weeks)			1.00 (.084)	0.84 (.145)	0.82 (.153)
Child Sex					
Male*				-6.21 (<.001)	-6.24 (<.001)
Child age (months)*				1.69 (.009)	1.69 (.008)
Parent-reported hearing concerns Yes					-9.16 (.120)
R^2	3.42%	3.82%	4.43%	7.05%	7.38%
Δ R^2	NA	+0.40%	+0.61%	+2.62%	+0.33%
F for change in R^2 (*p*-value)	1.68 (.052)	0.68 (.606)	2.23 (.107)	9.94 (<.001)	2.48 (.115)

*p <.05