Predictors of early precocious talking: A prospective population study

JEMMA SKEAT, MELISSA WAKE, SHEENA REILLY, PATRICIA EADIE, LESLEY BRETHERTON, EDITH L. BAVIN and OBIOHA C. UKOUMUNNE

Journal of Child Language / Volume 37 / Issue 05 / November 2010, pp 1109 - 1121
DOI: 10.1017/S030500090999016X, Published online: 30 October 2009

J. Child Lang. **37** (2010), 1109–1121. © Cambridge University Press 2009
doi:10.1017/S030500090999016X

BRIEF RESEARCH REPORT

Predictors of early precocious talking: A prospective population study*

JEMMA SKEAT

Murdoch Childrens Research Institute, Department of Paediatrics, University of Melbourne, Speech Pathology Department, Royal Children's Hospital

MELISSA WAKE

Murdoch Childrens Research Institute, Centre for Community Child Health, Royal Children's Hospital, Department of Paediatrics, University of Melbourne

SHEENA REILLY

Murdoch Childrens Research Institute, Department of Paediatrics, University of Melbourne, Speech Pathology Department, Royal Children's Hospital

PATRICIA EADIE

Murdoch Childrens Research Institute

LESLEY BRETHERTON

Psychology Department, Royal Children's Hospital, Department of Psychology, University of Melbourne

EDITH L. BAVIN

School of Psychological Science, La Trobe University

AND

OBIOHA C. UKOUMUNNE

Department of Paediatrics, University of Melbourne, Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute

(*Received* 27 *February* 2008 – *Revised* 8 *December* 2008 – *First published online* 30 *October* 2009)

[*] The ELVS study is supported by Project Grant 237106 from the Australian National Health and Medical Research Council (NHMRC) and small grants from the Murdoch Childrens Research Institute and the Faculty of Health Sciences, La Trobe University. Authors JS and OCU each hold a NHMRC training fellowship (Capacity Building Grant in Population Health # 436914). Ethical approval for the study was obtained from Human Research Ethics Committees at the Royal Children's Hospital (#23081) and La Trobe University (#03-32). Address for correspondence: Jemma Skeat, Murdoch Childrens Research Institute, 2nd Floor, Gantry Building, RCH, Flemington Road, Parkville, Australia 3141.

ABSTRACT

This study examines potential predictors of 'precocious talking' (expressive language ⩾90th percentile) at one and two years of age, and of 'stability' in precocious talking across both time periods, drawing on data from a prospective community cohort comprising over 1,800 children. Logistic regression was used to examine the relationship between precocious talking and the following potential predictors: gender, birth order, birth weight, non-English speaking background, socioeconomic status, maternal age, maternal mental health scores, and vocabulary and educational attainment of parents. The strongest predictors of precocity (being female and having a younger mother) warrant further exploration. Overall, however, it appears that precocity in early vocabulary development is not strongly influenced by the variables examined, which together explained just 2·6% and 1% of the variation at 1;0 and 2;0 respectively.

INTRODUCTION

It has been argued that 'typical' language development or the 'average' child is a myth (Bates, Dale & Thal, 1995). Yet, even accepting this variability, some children stand out as substantially stronger or weaker in their early language skills than their peers. Studying the pathways of these children may provide a better understanding of the factors that support or impede language development. Because of the obvious clinical and policy implications, considerable effort has been directed towards potential determinants of 'late' or 'slow' language development, such as family history of speech and language delay, gender, and perinatal factors (Nelson, Nygren, Walker & Panoscha, 2006). Much less has been published about 'precocious' language abilities (i.e. language that is well above the norm for age). Thal, Bates, Goodman & Jahn-Samilo (1997) argued that examining this subgroup of children is 'relevant to theoretical questions such as whether late and early status are equally stable, and whether the same factors that predict late status are the ones that predict continued precocity' (p. 8).

Definitions of precocious language ability vary in the few studies which have examined it. Robinson, Dale & Landesman (1990) and Crain-Thorensen & Dale (1992) classified children as 'linguistically precocious' if they performed at or greater than two standard deviations above the normed mean on at least one of three language assessments at one year, eight months (1;8). Fenson, Dale, Resnick, Bates, Thal & Prthick (1994) suggested that, in terms of vocabulary development, children scoring at or above the 90th percentile (an arbitrarily chosen cut-point) on the MacArthur-Bates Communicative Development Inventories (CDI; Fenson *et al.*, 1993) could

be considered to be 'precocious'. Thal *et al.* (1997) used this definition for children in their study of 'early talkers', comparing these children to those scoring at or below the 10th percentile ('late talkers').

The extent to which early linguistic precocity is an indicator of an advanced trajectory of language development from an early age, or simply a demonstration of uneven rates of language acquisition, is unclear. Thal *et al.* (1997) found that about one-third of early talkers at 1;1 retained that status seven months later, while nearly two-thirds of precocious talkers at 1;8 remained precocious at around 2;2. There is also evidence that some infants with exceptional language abilities at 1;8 retain robust language skills up to at least 4;6 (Crain-Thorenson & Dale, 1992; Robinson, Dale & Landesman, 1990).

Very few published studies have examined factors that might predict a precocious language outcome in young children. Two studies were published in Thal *et al.* (1997); both examined the relationship between 'subject and family variables', including age of child (measured in days), birth order, gender, ethnicity, exposure to a second language, mild medical complications, socioeconomic status (SES), maternal/ paternal education, maternal/paternal vocation, and the outcome talking status (categorised as 'early talkers', 'late talkers' and 'middle-range talkers'). They additionally examined factors that predicted stability in precocious status. Their findings suggest that when vocabulary was measured in groups of infants who were around 1;8 years and again at 2;2 years of age, early talkers tended to be older (measured in days). Having a higher SES was associated with early talking at 1;8, although this relationship disappeared after post hoc Tukey tests were applied and was not maintained to age 2;2. In the second study, when infants were studied at an earlier age (1;1; range 10–16 months), Thal *et al.* (1997) found a significant association between 'early talking' and being first born, maternal vocation (whether higher or lower is unclear) and lower SES. The latter two associations, however, disappeared when Bonferroni corrections were applied. Children of mothers whose vocation was ranked as lower on an occupational scale were more likely to be stable early talkers between 1;8 and 2;2, but no factors were significantly associated with stability between 1;1 and 1;8 (Thal *et al.*, 1997).

Thal *et al.* (1997) concluded that when predicting precocious status with respect to demographic factors 'the effects are relatively small – at least within the [middle class] socioeconomic range that we have sampled here' (p. 35). They additionally hypothesised that factors that predict the stability of precocity may become more evident at an older age. It should be noted that the sample size used in both studies was relatively small ($N = 185$ for the first study and $N = 217$ for the second study). It may be that the variability observed in children's early language abilities means that searching for single factors which influence language development is less relevant than

exploring how complex combinations of factors work together to support or impede language development. Bates *et al.* (1995) argued that it is 'likely that a combination of factors will be required to explain individual differences' to the extent observed in early language development (p. 25). To achieve this, larger studies are required within an epidemiological framework.

AIMS

The aim of this study was to investigate the predictive strength of child, parent and family factors that we hypothesised might be associated with precocious talking outcomes for one- and two-year-old children. Specifically, drawing on a large, established prospective community cohort, we explored whether putative predictive factors were associated with precocious talking status at one year of age, two years of age and at both ages (i.e. stable precocity).

METHOD

Participants

The participants were children in the Early Language in Victoria Study (ELVS) cohort, a prospective, longitudinal study that was established in 2002 to study language development from infancy. Sampling methods have been previously described (Reilly *et al.*, 2006; Reilly *et al.*, 2007). Briefly, a community sample of 1,911 children was recruited from six local government areas (LGAs) in metropolitan Melbourne in the state of Victoria, Australia. LGAs were selected to represent differing levels of socioeconomic status, from advantage to disadvantage, based on Australian Bureau of Statistics census data. Families were sequentially recruited mainly through routine maternal and child health visits, which are attended by more than 80% of families with eight-month-old infants. The primary caregiver completed questionnaires at baseline (eight months) and then each birthday; data reported here are drawn from the eight-month, one-year and two-year questionnaires and measures included in this study are detailed in Table 1.

Data were included in these analyses if the child's age at the time of completing the one-year questionnaire was within the range 11.5 to 13.5 months and at the time of the two-year questionnaire was within the range 23.5 to 25.5 months. Children were excluded from the study if they did not provide data for at least one time-point on the main outcome measure used to classify children on precocious talking status (see below). This resulted in a sample of 1,813 children used in this study (95% of the full cohort). These participants did not differ markedly from those in the full cohort with regard to the predictor variables used in this study. Fifty-three (2·9%) children were premature and 45 (2·5%) individual children were twin births. Mean maternal age when children were 8 months of age was

TABLE 1. *Potential predictor variables for precocious talking as collected in the Early Language in Victoria Study*

Variables	Explanation
Collected at eight months	
Gender	
Birth weight	In kilograms.
Birth order	Grouped into 'first born' and 'later born' based on having a sibling or not at eight months of age.
Non-English speaking background	Whether or not the parent reported English as the main language spoken at home to the child.
Maternal and paternal education levels	Parent report of highest level of educational attainment.
Socioeconomic status	Quantified using the SEIFA Index of Relative Advantage codes provided by the Australian Bureau of Statistics, assigned at the level of census collection district for the population of Victoria in 1996. Scores are standardized so that the national Australian mean is 1,000 (*SD* 100). Higher scores on the SEIFA correspond to 'less disadvantaged' status.
Maternal age	In years, as reported when the child was eight months of age.
Collected at one year	
Maternal mental health score	Non-specific psychological distress scale – Kessler-6 (Kessler & Mroczek, 1994). Self-report of mother's feelings during the 'past two weeks', using Likert-type scales for six items. Higher summed scores (out of a possible total of 24) represent poorer mental health, with a score greater than four considered clinically relevant.
Maternal and paternal vocabulary scores	Modified version of the Mill Hill Vocabulary Scale (Raven, 1997), a 44-item multiple-choice test of word knowledge completed by each parent individually.
Collected at two years	
Maternal mental health score	Non-specific psychological distress scale – Kessler-6 (Kessler & Mroczek, 1994), as at 1;0. Scored as above.

31.9 years (*SD* 4·42, range 16–45 years). Children providing CDI data at one year of age ($N=1{,}734$) had a mean age of 12.1 months and those providing CDI data at two years of age ($N=1{,}691$) had a mean age of 24.1 months.

Outcome measure

Expressive vocabulary production score on the CDI, reported by ELVS families when infants were one year and two years old, was used to define the outcome of precocious talking status. Specifically, parents completed the CDI 'Words and Gestures' form at 12 months, and the CDI 'Words and Sentences' form at 24 months; at both ages the vocabulary production score was used for this analysis. Vocabulary is a reasonable proxy measure for

precocious language development overall because it is known that children precocious in expressive vocabulary are likely to be precocious in other language areas (McGregor, Sheng & Smith, 2005). Precocity was defined at one and two years as those children with vocabulary scores in the top 10% of the ELVS sample; the top 10% was chosen as this was the cut-point suggested by the authors of the CDI (Fenson *et al.*, 1994). Initial analyses suggested that the published CDI norms (Fenson *et al.*, 1993) under-rated the relative ability of children in the ELVS sample. When the 90th percentile of the norms was used, only 4·5% and 6·8% of subjects at one and two years, respectively, were classified as precocious. As the community sample used in this study included over 1,800 children, we used our own internal norms based on the raw scores achieved by our sample to determine precocity. For the main analyses, raw scores on the CDI were ranked with boys and girls pooled together before dichotomising at the 90th percentile. In a subsidiary analysis an alternative definition of precocity separately ranked the raw vocabulary scores of boys and girls and classified the top 10% of each group as precocious.

For the analysis of 'stable' precocious talkers, children were classified as stable if they were precocious at both one and two years. Children who were classified as precocious at one time-point, but did not provide data at the other, were classified as missing on 'stable' precocious talking status.

Potential predictive factors

The following factors were selected a priori from the broader ELVS dataset as being potentially predictive of a precocious language outcome: gender, birth weight, birth order, non-English speaking background, maternal education level, paternal education level, neighbourhood socioeconomic status, maternal age, maternal mental health, maternal vocabulary and paternal vocabulary. Details of these variables are shown in Table 1.

ANALYSIS

Logistic regression was used to describe the nature and strength of the relationships between the potential predictive factors and precocious talking status. Unadjusted analyses were undertaken, in which each of the variables (covariates) shown in Table 1 was used on its own to predict precocious talking status in separate models. In the subsequent adjusted multivariable analyses, variables with p values that were less than 0·2 in the unadjusted analyses were simultaneously included as covariates in the model. The relationship between continuous covariates and the log odds of precocious talking status was examined for non-linearity by drawing locally weighted scatterplots (Cleveland, 1979) and fitting logistic regression models with

fractional polynomials (Royston, Ambler & Sauerbrei, 1999). There was no evidence of marked non-linearity.

Logistic regression analyses were undertaken for each of the outcomes of precocious talking status at one year and two years, and stable precocious talking status at both time-points. In order to incorporate the measures of maternal mental health taken at one and two years, the one-year mental health score and the *change* in scores between one and two years were used as covariates (in the analysis of two-year-old precocious talking status only). This circumvents the problem of the two-year mental health score being a mediator of the potential effect of the one-year score and aids interpretation of the odds ratios for these variables (De Stavola *et al.*, 2006). The squared Pearson correlation measure of explained variation (R^2) was calculated for the logistic regression analyses (Hosmer & Lemeshow, 2000). All analyses were implemented using Stata 9.2 (StataCorp, 2005).

Prior to analysis, scores were age-adjusted for 53 children who were born prematurely. We conducted sensitivity analyses in which correlation between outcomes of children from twin sets was allowed for by fitting marginal logistic regression models using Generalised Estimating Equations with information sandwich standard errors (Hanley, Negassa, deB Edwardes & Forrester, 2003). These provided almost identical results to the ordinary logistic regression, so we report the latter here.

RESULTS

Gender differences within the ELVS sample

Both the 12- and 24-month scores had a positive skew distribution (especially the 12-month score which was strongly skewed). At 1;0, boys in the ELVS sample had a median CDI vocabulary raw score of 3 (Interquartile Range [IQR]: 0 to 6) and girls had a median raw score of 4 (IQR 1 to 9). The mean (standard deviation) words on the CDI measure for boys was 4·7 (6·2) versus girls 6·8 (11·0) (mean diff = −2·2, 95% CI: −3·0 to −1·3). At 2;0, the median raw vocabulary score for boys was 215 (IQR: 99 to 344) and for girls was 282 (IQR: 162 to 399). The mean (*SD*) number of words for boys versus girls was 234·7 (160·3) versus 286·8 (159·6) (mean diff = −52·1, 95% CI: −67·4 to −36·9). Thus, girls were producing more words on average at both ages. The Spearman's rank correlation between one- and two-year CDI percentile rank scores was 0·37 for boys and 0·40 for girls.

Classification of precocity

At 1;0, the top 10% of the ELVS sample (based on CDI raw scores) comprised 173 children who scored 15 words or more. These children were

TABLE 2. *Multivariable logistic regression of precocious talking status at 1;0 and 2;0, and stable precocious talking status*

Outcome Predictor	Odds ratio	95% CI	*p* value
Precocious talking status at 1;0		*N*=1704	
Female sex	1·85	1·33 to 2·56	<0·001
Birth order (sibling vs. no sibling)	0·97	0·69 to 1·36	0·88
Non-English speaking background	1·88	1·01 to 3·51	0·05
SEIFA disadvantage total score (per 100 unit increase)	0·91	0·70 to 1·18	0·47
Maternal vocabulary score (per 5 unit increase)	0·96	0·81 to 1·15	0·67
Maternal age (years)	0·93	0·90 to 0·97	<0·001
Precocious talking status at 2;0		*N*=1504	
Female sex	1·41	1·00 to 2·00	0·05
Birth weight in kilograms	0·84	0·61 to 1·17	0·31
Birth order (sibling vs. no sibling)	0·82	0·57 to 1·18	0·28
Non-English speaking background	0·34	0·08 to 1·42	0·14
Maternal vocabulary score (per 5 unit increase)	1·00	0·81 to 1·24	0·99
Paternal vocabulary score (per 5 unit increase)	1·15	0·93 to 1·43	0·19
Maternal age (years)	0·99	0·95 to 1·03	0·60
Stable precocious talking status (1;0–2;0)		*N*=1611	
Female sex	1·58	0·90 to 2·78	0·11
Paternal vocabulary score (per 5 unit increase)	1·39	1·01 to 1·91	0·04
Maternal age (years)	0·91	0·85 to 0·97	0·007

classified as 'precocious talkers', and had mean raw scores of 25·4 words (*SD* 16·2; range 15 to 123); 65 (38%) were boys. At 2;0, 173 children were classified as precocious talkers based on a score within the top 10% of the ELVS sample (a score of 495 words or more). These children had mean raw scores of 556·3 words (*SD* 46·99, range 495 to 679); 73 (42%) were boys. Fifty-eight infants were stable precocious talkers, representing 37% (58/156) of the infants who were precocious at 1;0 and who also provided data at 2;0, and 3·2% (58/1785) of the total sample (95% CI 2·4% to 4·1%). This is greater than the proportion that would be expected by chance alone (1%).

Predictors of precocious talking

Table 2 shows the results of the multivariable models fitted to precocious talking status at 1;0 and 2;0, and stable precocious talking status between 1;0 and 2;0 in the main analysis. Odds ratios and 95% confidence intervals for SEIFA disadvantage scores are reported per 100 unit increase, which is the standard deviation across postcodes of the SEIFA scores in the Australian population. The estimates for maternal and paternal vocabulary scores are reported per 5 unit increase, which is the approximate standard

deviation of the vocabulary measures in the ELVS cohort. These are reported to aid interpretation of the odds ratios, as they are a more meaningful comparison than one unit increases for these measures.

Prediction of precocity at 1;0. As shown in Table 2, the factors associated with one-year-old precocious talking status in the multivariable model were gender, maternal age and, with weaker evidence, non-English speaking background. Girls had 85% higher odds of being precocious (i.e. OR of 1·85). Maternal age was strongly associated with precocious talking status at this age, with younger mothers more likely to have precocious talking children. A mother who was one year younger than another had a 7% increase in the odds of having a precocious talking child; a mother five years younger than another had a 43% increase in the odds of having a precocious talking child. The multivariable model accounted for only 2·6% of the total variation (R^2).

Prediction of precocity at 2;0. At two years, gender was the only variable for which there was (albeit moderate) evidence of an association in the multivariable analysis. Table 2 shows that the odds of a girl being a precocious talker were 41% higher than those of a boy. The 95% confidence interval, however, shows that in the population there may be as little as no increase in the odds or as much as a doubling of the odds. The total variation explained by this model was low ($R^2 = 1{\cdot}0\%$).

Prediction of stable precocity. In the multivariable model shown in Table 2, maternal age was associated with stable precocious talking. The results suggest that younger mothers had greater odds of having children who were stable precocious talkers compared to older mothers. There was some evidence that a higher level of paternal vocabulary was associated with higher odds of having a stable precocious talker. The total variance explained by the models (R^2) was only 0·5%.

Using the alternative definition of precocity status, where raw vocabulary scores of boys and girls were ranked separately, provided virtually identical results with respect to the variables that emerged as the strongest predictors. The only difference was that gender was no longer a strong predictor as, by definition, the cut-points for precocity were gender-specific.

DISCUSSION

This study suggests that the psychosocial and environmental variables examined play little role in relation to children's early precocious vocabulary development, explaining at most 2·6% of the variance in precocity. We have confidence in these findings, given the size and epidemiological quality of the available data. Our findings are consistent with those of Thal *et al.* (1997), who reported only weak relationships between demographic factors and precocity. They are also consistent with a previous analysis of 'late

talking' outcomes (CDI $\leqslant$10th percentile) in this cohort at two years of age (Reilly *et al.*, 2007), in which a wide range of putative predictors accounted for only 4% of the variance. Reilly *et al.* (2007) concluded that their findings provide evidence for a 'strong biological trajectory' (p. 1447) in language development in the first two years of life. This conclusion has support from other research (e.g. Dale *et al.*, 1998), suggesting that genetic factors play a role in late talking at age two.

There were some differences between those factors that did contribute to late talking as opposed to precocity. Neither of the factors that were associated with late talking at 2;0 and that were replicated in this study (non-English speaking background [$p=0{\cdot}004$] and maternal education $\leqslant$12 years [$p=0{\cdot}04$]) were significantly associated at the 5% level with precocious talking at 2;0. Additionally, maternal age, significantly associated with precocity here, was not significantly associated with late-talking status at the 5% level (Reilly *et al.*, 2007).

One factor that does appear to contribute to both late talking and precocity in the early years is gender. Previous research has identified that late talking and clinical populations tend to contain a greater proportion of boys. Our study found that girls had greater odds of precocious talking. Between one and two years, however, the relationship between precocious talking and gender weakened. If the association continues to decrease over time it might support a conclusion that, while male gender is a risk factor for being in the bottom tail of language development, female gender confers only a short term 'advantage' in language development, an advantage that is not maintained.

Both lower and higher maternal age have been postulated as a risk factor for poor language development (Nelson *et al.*, 2006), and the results are not consistently in favour of these associations (Reilly *et al.*, 2007; Tomblin, Smith & Zhang, 1997; McCue Horwitz *et al.*, 2003). Our findings suggest that the younger the mother (in the ELVS sample), the more likely her child is to be a precocious talker at one year and a stable precocious talker. Results from locally weighted scatterplots and the fractional polynomials regression method did not show a non-linear effect on the log odds scale, suggesting that the effect was consistent across all ages. The means by which there may be a positive effect of younger maternal age in terms of vocabulary development is yet to be determined. Our findings were not explained by maternal education or SES, which might be hypothesised as confounding this finding, but there may be other unmeasured confounders that partially explain the effect. The use of parental (mostly maternal) report to gather data on expressive vocabulary development may have had an impact; it is possible that mothers' interpretation of their child's vocabulary may vary according to their own age. If this was the case, however, one might expect that the effect of maternal age would persist at

two years, which it does not for these data. It is possible that younger maternal age is associated with benefits in communication development in the first year of life, but that this benefit is not maintained beyond this age. Further research with this sample will be needed to determine whether (either of younger or older) maternal age predicts precocious outcomes when children get older and, arguably, their development is more stable.

Our results also indicated that only one-third of infants who were precocious at 1;0 remained precocious at 2;0, despite moderate correlations between raw and percentile ranked CDI scores at 1;0 and 2;0. Other research has suggested that vocabulary production as measured by the CDI between one and two years has a correlation of around 0·4 (Feldman, Dollaghan, Campbell, Kurs-Lasky, Janosky & Paradise, 2000). Although vocabulary performance is moderately correlated within the group, it is not sufficiently so that there is a distinct group of children that remain in the top 10% at both time-points. Children appear to easily fall out of artificially created cut-points (e.g. 'precocious' vs. 'not precocious') at this age, at least when using expressive vocabulary from parent report as the indicator. Thal *et al.* (1997) reported that 36% of their sample of 44 children remained 'precocious' between an average of 1;1 and 1;8, while 60% of their older sample of 25 precocious-talking children were stable between 1;8 and 2;2. They concluded that stability of precocity increases over time (i.e. children become more stable as they get older). Our data, however, suggest that the period over which stability is measured (a twelve-month, rather than four- and seven-month time periods) may also be a factor.

A strength of this study is the size of the cohort of children from which a relatively large number of precocious infants were identified, compared with previous studies which have been limited by small sample sizes. Additionally, infants in this study were identified within a prospectively sampled community cohort, and we were able to compare them to large numbers of age peers with 'non-precocious' language development. A potential weakness is the use of parent reported vocabulary as our only indicator of precocious language development. Nevertheless, this measure is often used and vocabulary is considered a good predictor of overall language competence.

CONCLUSIONS

Precocity in expressive language in the first two years of life is unstable for the majority of children classified in this group, and does not appear to be strongly predicted by psychosocial factors such as SES, birth order, or parental education or vocabulary. The multivariable models fitted in this study explained very little variation in children's vocabulary development.

REFERENCES

Bates, E., Dale, P. S. & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (eds), *Handbook of child language*, 96–151. Oxford: Basil Blackwell.

Bavin, E. L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., Barrett, Y. & Ukoumunne, O. C. (2008). The Early Language in Victoria Study: predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language* **35**, 687–701.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–36.

Crain-Thorenson, C. & Dale, P. S. (1992). Do early talkers become early readers? Linguistic precocity, preschool language, and emergent literacy. *Developmental Psychology* **28**(3), 421–29.

Dale, P., Simonoff, E., Bishop, D. V. M., Eley, T., Oliver, B., Price, T., *et al*. (1998). Genetic influence on language delay in two-year-old children. *Nature Neuroscience* **1**, 324–28.

De Stavola, B. L., Nitsch, D., dos Santos Silva, I., McCormack, V., Hardy, R., Mann, V., Cole, T. J., Morton, S. & Leon, D. A. (2006). Statistical issues in life course epidemiology. *American Journal of Epidemiology* **163**, 84–96.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E. & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development* **71**(2), 310–22.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J. & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development* **59**(5).

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. & Reilly, J. S. (1993). *MacArthur Communicative Development Inventories user's guide and technical manual*. San Diego, CA: Singular.

Hanley, J. A., Negassa, A., deB Edwardes, M. D. & Forrester, J. E. (2003). Statistical analysis of correlated data using Generalized Estimating Equations: an orientation. *American Journal of Epidemiology* **157**, 364–75.

Hosmer, D. W. & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley.

Kessler, R. & Mroczek, D. (1994). *Final version of our non-specific psychological distress scale* (*Memo*). Ann Arbour, MI: Institute for Social Research.

McCue Horwitz, S., Irwin, J. R., Briggs-Gowan, M. J., Bosson Heenan, J. M., Mendoza, J. & Carter, A. S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child and Adolescent Psychiatry* **42**(8), 932–40.

McGregor, K. K., Sheng, L. & Smith, B. (2005). The precocious two-year-old: status of the lexicon and links to the grammar. *Journal of Child Language* **32**, 563–85.

Nelson, H. D., Nygren, P., Walker, M. & Panoscha, R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the US Preventive Services Task Force. *Pediatrics* **117**, 298–319.

Raven, J. C. (1997). *Mill Hill vocabulary scale*. Oxford: JC Raven Ltd.

Reilly, S., Eadie, P., Bavin, E. L., Wake, M., Prior, M., Williams, J., Bretherton, L., Barrett, Y. & Ukoumunne, O. C. (2006). Growth of infant communication between 8 and 12 months: A population study. *Journal of Paediatrics and Child Health* **42**, 764–70.

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., Eadie, P., Barrett, Y. & Ukoumunne, O. C. (2007). Predicting language at age 2 years: A prospective community study. *Pediatrics* **120**, e1441–e1449.

Robinson, N. M., Dale, P. S. & Landesman, S. (1990). Validity of Stanford-Binet IV with linguistically precocious toddlers. *Intelligence* **14**, 173–86.

Royston, P., Ambler, G. & Sauerbrei, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International Journal of Epidemiology* **28**, 964–74.

StataCorp (2005). Stata Statistical Software (Version 9.2). College Station, TX: Stata Corporation.

Thal, D. J., Bates, E., Goodman, J. & Jahn-Samilo, J. (1997). Continuity of language abilities: An exploratory study of late- and early- talking toddlers. *Developmental Neuropsychology* **13**(3), 239–73.

Tomblin, J., Smith, E. & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders* **30**(4), 325–44.