www.jpeds.com • THE JOURNAL OF PEDIATRICS

Language Abilities in Children Who Were Very Preterm and/or Very Low Birth Weight: A Meta-Analysis

Natalie Barre, BSc(Hons), Angela Morgan, PhD, Lex W. Doyle, MD, FRACP, and Peter J. Anderson, PhD

Objective To conduct a meta-analysis to characterize differences in language ability between children born very preterm (VPT, <32 weeks' gestational age), with a very low birth weight (VLBW, <1500 g), or both and in term-born control children.

Study design Electronic databases were systematically searched, and 12 studies met the inclusion criteria. Effect sizes were calculated to compare VPT/VLBW children and control children.

Results VPT/VLBW children performed between 0.38 and 0.77 SD below control subjects in the areas of expressive and receptive language overall and expressive and receptive semantics. Results for expressive and receptive grammar were equivocal. Subgroup analysis of school-aged children revealed similar results. No studies assessing phonological awareness, discourse, or pragmatics were identified.

Conclusions Language ability is reduced in VPT/VLBW children. When considering only school-aged children, this reduction is still present, suggesting that their difficulty appears to be ongoing. Rigorous studies examining a range of language subdomains are needed to fully understand the specific nature of language difficulties in this population. *(J Pediatr 2011;158:766-74)*.

Infants born very preterm (VPT, <32 weeks' gestational age [GA]), with a very low birth weight (VLBW, <1500 g), or both comprise 1% to 2% of all live births.[1,2] With increased survival rates, particularly of infants born extremely preterm (EPT, <28 weeks' GA) or of extremely low birth weight (ELBW, <1000 g),[3,4] the numbers of high risk VPT/VLBW survivors have increased in the past few decades. VPT/VLBW children are at particular risk for a range of impairments, including language dysfunction; however, findings have been inconsistent. Some studies report no differences between VPT/VLBW groups and control children,[5-7] but other studies report large differences.[8-10] Methodological inconsistencies are a complicating factor, making this literature difficult to interpret.

Adequate language skills are fundamental to daily interpersonal communication and social functioning, with poor language skills significantly influencing friendship quality.[11] Further, language-impaired children perform more poorly in reading, with language scores at early school age predicting later reading accuracy and comprehension,[12] which influences academic achievement. Half as many language-impaired adolescents achieve a high level of qualifications when finishing school compared with peers.[13] Thus, it is vital to identify and characterize language impairment in VPT/VLBW school-aged children, because early detection and intervention of language impairment can lead to earlier treatment and improve long-term outcomes.[14,15]

Language can be categorized in different modalities and in different subdomains. In early language development, expressive language (production) and receptive language (comprehension) are commonly considered separately, with receptive language ability preceding language expression.[16] Language may also be divided in the categories of semantics, grammar, phonological awareness, discourse, and pragmatics. Semantics is the meaning of words and sentences, and grammar refers to language structure, such as word order and use of tense. Semantics is usually operationalized as vocabulary, which predicts later intelligence in children.[17] Semantic and grammatical proficiency are thought to develop concomitantly.[18] Phonological awareness, or the understanding of speech sounds,[19] is particularly important in the initial stages of language development[20,21] and for reading.[22,23] Discourse refers to passages of text or conversation, where adequate expression or comprehension involves integrating information across sentences for coherence.[24] Discourse skills are important in understanding the overall message from larger amounts of information.[25] Finally, pragmatics refers to the use of language that is appropriate to the conversational or social context and is the basis of choosing polite language or slang and understanding non-literal language such as humor and

CELF	Clinical Evaluation of Language Fundamentals
ELBW	Extremely low birth weight
EPT	Extremely preterm
GA	Gestational age
PLS	Preschool Language Scale
PPVT-R	Peabody Picture Vocabulary Test-Revised
VLBW	Very low birth weight
VPT	Very preterm
WJ-III	Woodcock-Johnson III Tests of Achievement

From the Murdoch Children Research Institute, Victoria, Australia (N.B., A.M., L.D., P.A.); University of Melbourne, Victoria, Australia (N.B., A.M., L.D., P.A.); and Royal Women's Hospital, Victoria, Australia (L.D.)

Supported by Australia's National Health & Medical Research Council (Career Development Award for A.M. [ID–607315] and Senior Research Fellowship for P.A. [ID – 628371]). The authors declare no conflicts of interest.

 10.1016/j.jpeds.2010.10.032

metaphor.[26,27] The aim of this study was to systematically review reported studies of the language abilities of VPT/VLBW children, aged >24 months, compared with term-born peers.

Methods

The literature was searched systematically with the electronic databases MEDLINE, PsycINFO, CINAHL, and ERIC, for articles published between 1990 and November 2009 inclusive. Search terms are listed in **Table I** (available at www.jpeds.com). The searches were restricted to English-language publications in peer-reviewed journals. The broad keyword search had high sensitivity and yielded a total of 3386 items.

Studies were examined to determine whether they met these inclusion criteria: (1) examined a VPT (≤32 weeks' GA), VLBW (≤1500 g), or both sample that was representative of a cohort, in which at least part of the sample was born in 1990 or later; (2) had a term-born (>36 weeks' GA) control group; (3) assessed the children at 24 months of age or older; (4) reported language outcomes; and (5) did not report on a subset of children from another publication (to avoid duplicate counting). Studies were excluded when the VPT/VLBW group was selected on the basis of medical complications (such as intraventricular hemorrhage) or without mention of the sample's representation. In two studies in which the method of sample selection was unclear, the authors were contacted, and both were subsequently excluded.[28,29] There was no explicit upper age limit.

Language outcomes were examined, and studies were included in the meta-analysis when means and SDs were reported. Language outcomes were divided in separate subdomains of language (outlined further below), and a meta-analysis was conducted in each of these domains separately. For two studies that fulfilled the inclusion criteria, further examination of the items that made up the language variables (through author contact) revealed that the studies were inappropriate for inclusion, with variables either containing items that were inappropriate for that language variable[30] or non-language items (auditory attention).[31] When a study reported on different measures from the same subdomain of language,[32] only one measure was included. Measures that could not be represented by means and SDs were excluded.[10] The final number of studies included in this review was 12, the details of which are summarized in **Table II**.

The outcomes reported in each study were divided initially in these subdomains: expressive language, receptive language, phonological awareness, discourse, and pragmatics. Expressive language and receptive language were further divided in semantics and grammar. When this was not possible within an individual study, the variables were classified separately as expressive or receptive language only. Verbal IQ and scores from rapid naming tasks were excluded because they primarily involve other cognitive functions and may confound the results. Orthographic tasks were also excluded because the focus of this review was not on literacy skills. These definitions were used to classify the variables: (1) expressive language (n = 3[8,10,33]): Tasks that measure the participant's overall verbal expression (unable to be further classified in semantics or grammar); such measures include expressive language indices from the Preschool Language Scale (PLS-3)[34] or the Clinical Evaluation of Language Fundamentals (CELF-III);[35] (2) expressive semantics (n = 7[8,32,36-40]): Tasks that quantify the meaning of a participant's verbal expression; such tasks include category fluency (in which the participant is asked to name as many items in a semantic category as possible), or expressive vocabulary tests (in which the participant is asked to define words, or name pictures); (3) expressive grammar (n = 1[32]): Tasks that measure the structure of the expressed sentence or word form; such variables include sentence length; (4) receptive language (n = 4[8-10,33]): Tasks that measure the participant's overall understanding of verbal expression, at the level of the word or sentence (unable to be further classified in semantics or grammar); such measures include receptive language indices from the PLS or the CELF; (5) receptive semantics (n = 2[8,41]): Tasks that measure the level of understanding of the meaning of verbal information; such tests include the Peabody Picture Vocabulary Test-Revised (PPVT-R,[42] in which the participant is asked to point to the picture of the named object); (6) receptive grammar (n = 1[43]): Tasks that measure the level of understanding of the structure of sentences or word form; such variables include measures of understanding and following directions; (7) phonological awareness (n = 0): Tasks that measure the perception and manipulation of speech sounds; (8) discourse (n = 0): Tasks that measure the comprehension or production of language in the context of passages of information; and (9) pragmatics (n = 0): Tasks that measure the appropriate use or comprehension of language in social contexts.

Within each subdomain, all studies were initially examined, then a subgroup analysis was conducted only on studies of school-aged children. We focused on this age group because of the large variability in the development of these skills in younger children. In contrast, stability in language scores for normally developing and language-impaired school-aged children has been shown from the age of 4 to 8 years.[44] This analysis comprised only those studies that assessed children at 5 years and older. One study assessed children ranging from age 4 years 6 months to 5 years 6 months,[41] and this study was included in the subgroup analysis. Therefore, the school-aged group includes children who are at varying levels of schooling, but are still likely to demonstrate stable language ability.

Statistical Analyses

With Review Manager software (RevMan, The Nordic Cochrane Center, The Cochrane Collaboration, Copenhagen, Denmark) version 5.0,[45] means, SDs, and sample sizes for each group were used to calculate effect sizes, or standardized mean differences for each subdomain, measured in units of SD. Effect sizes were calculated by using the formula for Hedge's g, weighted by the inverse variance. Effect sizes of 0.2 SD were considered small, 0.5 SD moderate, and 0.8 SD large.[46] When separate results were given for subgroups within the

Table II. Summary of articles included in the meta-analysis, ordered by birth year

Author	VPT/VLBW years of age*	Birth year(s)	VPT/VLBW	VPT/VLBW, n	Full-term control, n	Language test (language subdomain)	Recruitment source of VPT/VLBW group	Reported exclusion criteria	Attrition rate of VPT/VLBW group	Selection of control subjects
Bohm et al[37]	5.5 (0.04) corrected age	1988-1993	≤1500 g	172	125	Category fluency (Expressive–Semantics)	Multicenter/ Geographical	Severe learning disability, cerebral palsy, or visual impairment that hindered testing	27%	Randomly sampled from eligible interested families, individually matched on birth date and hospital
Luu et al[8]	12.2 (0.4) chronological age	1989-1992	600-1250 g	330 to 365†	110	WISC-III Vocabulary (Expressive–Semantics) PPVT-R (Receptive–Semantics) CELF-III Expressive Language (Expressive) CELF-III Receptive Language (Receptive)	Multicenter	None reported	16%-24%†	Community or telemarketing list, frequency matched on age, race, sex, maternal education, zip code
Harvey et al[41]	5.17 (4.5-5.5) chronological age	1990-1991	<1000 g	30	50	PPVT-R (Receptive–Semantics)	Single center	Major neurological disability, lived outside the region	37%	VLBW child's preschool/daycare class, two individual matches on sex
Foulder-Hughes & Cooke[39]	7.48 (variability not reported) unknown whether corrected	1991-1992	<32 weeks	280	210	WISC-III Vocabulary (Expressive–Semantics)	Multicenter	Mothers not residing in area, children attending special schools	23%	Classmates, individually matched on birth date and sex
Taylor et al[40]	8.7 (0.6) chronological age	1992-1995	<1000 g	204	176	WJ-III Picture Vocabulary (Expressive–Semantics)	Single center	Congenital abnormalities, infections	14%	Classmates, individually matched on age, sex, race
Esbjørn et al[38]	5.06 (0.15) corrected age	1994-1995	<28 weeks or <1000g	193	75	WPPSI-R Vocabulary (Expressive–Semantics)	Geographical	Non-native Danish, no full-scale IQ (mainly because of disability)	28%	Central Office of Civil Registration, individually matched on age, sex, parental education, residence
Hanke et al[9]	6.2 (0.67) chronological age	1994-1995	≤1500 g	60	60	Marburg Language Comprehension Test for Children (Receptive)	Single center	Children not admitted to neonatal intensive care unit	26%	Kindergartens, individually matched on age, sex, parental education
Wolke et al[10]	6.3 (5.16-7.25) unknown whether corrected	1995	<26 weeks	199- 204†	159	PLS-3 Expressive Communication Scale (Expressive), Auditory Comprehension Scale (Receptive)	Geographical	Children not in mainstream school	35%	Classmates, individually matched on age, sex, birth date
Aarnoudse-Moens et al[36]	5.9 (0.4) Unknown whether corrected	1998-1999	≤30 weeks	50	50	Category fluency (Expressive–Semantics)	Single center	Children not admitted to neonatal intensive care unit, mental or motor handicaps that hindered testing	82%‡	Local elementary schools, not matched
Foster-Cohen et al[32]	2 (0.04) corrected age	1998-2000	<33 weeks and/or <1500g	90	102	MacArthur-Bates Communicative Development Inventory:[64] Vocabulary Production total score (Expressive–Semantics), sentence length in morphemes (Expressive–Grammar)	Single center	Children not admitted to neonatal intensive care unit, congenital abnormalities, non-English–speaking parents	10%	Hospital records, individually matched on sex and birth date

(continued)

Table II. Continued

Author	VPT/VLBW years of age*	Birth year(s)	VPT/VLBW	VPT/VLBW, n	Full-term control, n	Language test (language subdomain)	Recruitment source of VPT/VLBW group	Reported exclusion criteria	Attrition rate of VPT/VLBW group	Selection of control subjects
Pritchard et al[43]	6 (0.04) corrected age	1998-2000	<33 weeks and/or <1500g	102	108	WJ-III Understanding Directions (Receptive–Grammar)	Single center	Children not admitted to neonatal intensive care unit, congenital abnormalities, non-English–speaking parents, blind	5%	Hospital records, individually matched on sex and birth date
Woodward et al[33]	4 (0.04) corrected age	1998-2000	<33 weeks and/or <1500g	100	105	CELF-P Expressive Language (Expressive) CELF-P Receptive Language (Receptive)	Single center	Children not admitted to neonatal intensive care unit, congenital abnormalities, non-English–speaking parents, blind	7%	Hospital records, individually matched on sex and birth date

CELF-III, Clinical Evaluation of Language Fundamentals-Third Edition[35]; *CELF-P*, Clinical Evaluation of Language Fundamentals-Preschool Edition[65]; *PPVT-R*, Peabody Picture Vocabulary Test-Revised[42]; *PLS-3*, Preschool Language Scale-Third Edition (UK)[34]; *WISC-III*, Wechsler Intelligence Scale for Children-Third Edition[66]; *WJ-III*, Woodcock Johnson III Tests of Achievement[55]; *WPPSI-R*, Wechsler Preschool and Primary Scale of Intelligence-Revised.[67]

*Years are mean (SD) or mean (range).

†Numbers differ for each test.

‡Participants were selected from the larger cohort and is thus not attrition.

VPT/VLBW group,[32] the means and SDs were pooled, weighted by sample size. Random effects meta-analysis was performed, with results presented as effect sizes and 95% CIs. Heterogeneity between studies was measured by using I^2, which represents the percentage of variation in the effect size explained by heterogeneity between studies,[47] in which 25% is thought to be mild, 50% moderate, and 75% high.[48] This statistic has been suggested for use by the Cochrane Collaboration[49] and is preferred over the χ^2 or R statistics.[47]

Means and SDs used in the meta-analyses were all unadjusted for co-variates. One study also reported means and SEs adjusted for child age and parental education[38]; however, the unadjusted means were preferred because once adjusted, they are no longer descriptive statistics. The other disadvantage of using adjusted results is that different studies use different adjustments, making it difficult to combine results across studies.

To account for publication bias, in which unpublished studies are more likely to find insignificant differences in groups, fail-safe N statistics were computed.[50] The fail-safe N statistic is a measure to detect how many published studies with non-significant results are needed to nullify a significant result[51] and therefore was not calculated for non-significant results. It is a value based on the effect sizes, the sample sizes, and the number of studies. It is thought that the fail-safe N must exceed 5k + 10, in which k is the number of studies that are combined to form the overall effect size, to indicate a lack of publication bias.[50] A spreadsheet template was downloaded[52] for use in Microsoft Excel (Microsoft Inc, Redmond, Washington), to calculate fail-safe Ns with Cohen's d values, (which either did not differ from the Hedge's g values, or differed by 0.01 because of rounding).

Methodological Quality

The methods of the included studies varied. Because of the small number of studies in each language subdomain, and the high quality of studies included on the basis of the inclusion criteria, sensitivity analyses based on a formal rating system were not undertaken. However, methodological quality pertaining to these 4 areas was examined and displayed in Table II: (1) Sampling: the ideal cohort includes children born within a discrete geographic region; single center studies introduce bias; (2) Excluded children; most studies excluded children with severe cerebral palsy, congenital malformations, neurosensory impairment, or non-native speaking families; such exclusion is generally accepted when assessing language or other cognitive outcomes; excluding further children by selecting a subset of children from this group will reduce the representativeness of the sample; (3) Attrition rate; those studies with higher attrition rates are less representative of the original population; (4) Term-born control group; an adequate control group is one that has been recruited perinatally, or consists of siblings of the cohort, without differing significantly on a range of demographic variables (often achieved by matching); convenience control groups are less representative.

Results

Table III depicts the number of studies, total sample sizes, effect sizes, CIs, I^2 statistics, and fail-safe N statistics for each language subdomain, categorized by age at assessment. When the category contained only 1 study, a meta-analysis could not be conducted, and so the statistics pertain only to the individual study identified with the systematic review.

Expressive Language

The results of the 3 studies[8,10,33] that reported overall expressive language scores indicated that the VPT/VLBW group scored lower than control children (Table III, Figure 1), with mild to moderate heterogeneity across the studies. The fail-safe N statistic of 88 is robust, higher than 5k + 10, or 25. Thus, it is unlikely that the significant effect size is a result of publication bias.

The somewhat heterogeneous effect sizes may be explained by differences in birth weight and gestational age, with the studies of more immaturely born children[8,10] having greater effect sizes. This is consistent with findings in the literature that children born VPT/VLBW perform better than children born EPT/ELBW.[53,54] Age at testing may also have been a factor, with Woodward et al[33] assessing children at a younger age than the other two studies. Another consideration is that different measures were used, for example, the Expressive Communication measure from the PLS-3 (UK)[34] in Wolke et al[10] and the CELF-P and CELF-III used by Woodward et al[33] and Luu et al,[8] respectively. Attrition rates were lower in Woodward et al[33] (Luu et al[8] was 24% for tests in this language subdomain), and the recruitment source differed (Table II). Finally, exclusion criteria differ. Wolke et al[10] excluded children not attending mainstream schools, and Woodward et al[33] excluded children with non-English speaking parents.

For the two school-age groups,[8,10] a meta-analysis again revealed significantly lower scores from the VLBW/VPT group compared with the control group (Table III). The effect sizes from the two studies were not heterogeneous, and a robust fail-safe N statistic was found.

Expressive–Semantics

The 7 studies[8,32,36-40] that reported on tasks that measure expressive semantics indicated that the VPT/VLBW group scored significantly lower than control children (Table III, Figure 1). The effect sizes across the studies were homogenous, and a robust fail-safe N statistic was found.

Combining the results of the 6 school-age studies, the effect size remained very similar to the overall meta-analysis results (Table III), with little heterogeneity and a robust fail-safe N statistic.

Expressive–Grammar

Only one study was identified as reporting on scores that reflected grammatical expression. Foster-Cohen et al[32] tested

Table III. Meta-analysis results categorized by language subdomain and age at assessment

Language sub-domain	n studies	Sample sizes VPT/VLBW	Sample sizes Term	Effect size (*g*)	95% CI	I^2	Fail safe N
Expressive							
School-age	2	530	269	−0.71†	−0.86-−0.55	0%	58.34
Young children	1	100	105	−0.41*	−0.69-−0.14	NA	2.18
Total	3	630	374	−0.63†	−0.80-−0.45	42%	88.23
Expressive–Semantics							
School-age	6	1264	746	−0.40†	−0.50-−0.31	5%	152.59
Young children	1	90	102	−0.23	−0.51-0.06	NA	NA
Total	7	1354	848	−0.38†	−0.48-−0.29	9%	176.66
Expressive–Grammar							
School-age	0	–	–	–	–	–	–
Young children	1	90	102	−0.23	−0.52-0.05	NA	NA
Total	1	90	102	−0.23	−0.52-0.05	NA	NA
Receptive							
School-age	3	591	329	−0.83†	−0.97-−0.69	0%	141.10
Young children	1	100	105	−0.54†	−0.82-−0.26	NA	4.35
Total	4	691	434	−0.77†	−0.94-−0.60	41%	200.99
Receptive–Semantics							
School-age	2	391	160	−0.59†	−0.79-−0.40	0%	24.21
Young children	0	–	–	–	–	–	–
Total	2	391	160	−0.59†	−0.79-−0.40	0%	24.21
Receptive–Grammar							
School-age	1	102	108	−0.44*	−0.72-−0.17	NA	2.74
Young children	0	–	–	–	–	–	–
Total	1	102	108	−0.44*	−0.72-−0.17	NA	2.74
Phonological awareness	0	–	–	–	–	–	–
Discourse	0	–	–	–	–	–	–
Pragmatics	0	–	–	–	–	–	–

Negative effect sizes indicate lower VPT/VLBW scores.
NA, Not applicable.
**P* < .01.
†*P* < .001.

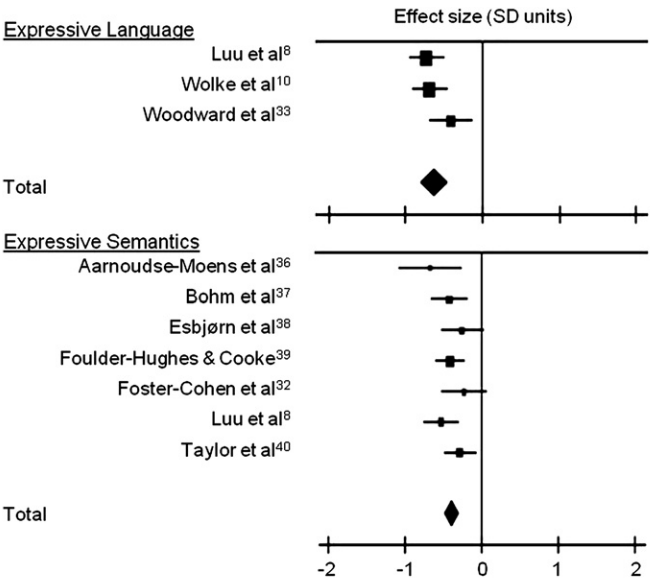

Figure 1. Forest plots of the effect sizes for the subdomains of expressive language and expressive semantics. The effect size for each study is represented by a square with CI bars. The size of the box indicates the relative weight of the study. The total meta-analysis result for each subdomain is represented by the diamond. Negative values indicate lower scores for the VPT/VLBW group.

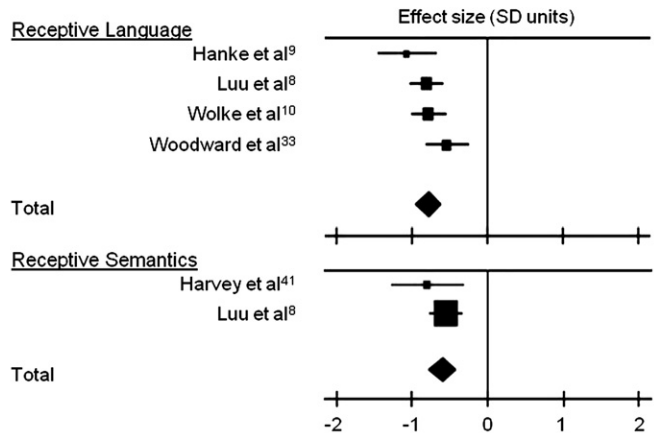

Figure 2. Forest plots of the effect sizes for the subdomains of receptive language and receptive semantics. See **Figure 1** label for explanation.

EPT and VPT groups and the statistics for each were combined; a non-significant difference (**Table III**) was found in the length of their longest sentences compared with those of control children. However, this score was only based on those children who were combining words, which was 58% of the children born EPT, 87% of the children born VPT, and 82% of the children born full term.

Receptive Language

Four studies[8-10,33] were identified as reporting overall receptive language indices. The meta-analysis revealed a significant difference between VPT/VLBW children and control children (**Table III**, **Figure 2**). The individual effect sizes were mildly to moderately heterogeneous, and a robust fail safe N statistic was found.

The studies with the highest[9] and the lowest[33] effect sizes, contributing to heterogeneity, did not differ substantially in their cohort's GA, recruitment source, or year of birth (**Table II**). Woodward et al[33] had a lower attrition rate and excluded children with non-English speaking parents, but the heterogeneity may also be explained by the tests used. The test used by Hanke et al[9] included pragmatics in the total score, unlike the other tests in this subdomain, which primarily included semantic and grammar measures.

The large effect size remained when the school-age studies[8-10] were combined (**Table III**), with no heterogeneity and a robust fail safe N statistic.

Receptive–Semantics

There were only two studies[8,41] reporting results for VPT/VLBW children and control children in the subdomain of receptive semantics; both were school-age studies. There was a significant reduction in scores in VPT/VLBW participants compared with those in control children (**Table III**, **Figure 2**). The reported effect sizes were homogenous, and a robust fail safe N statistic was found.

Receptive–Grammar

Only one study[43] reported on grammar comprehension. This study followed up on the children reported in Foster-Cohen et al[32] and Woodward et al[33] and were thus ≤33 weeks' GA, <1500 g birth weight, or both. Assessed at age 6 years, they were tested by using the Understanding Directions subtest of the Woodcock-Johnson III Tests of Achievement (WJ-III),[55] which required them to listen to long sentences and point to answers. The preterm group scored lower than control children (**Table III**). The fail safe N statistic introduces doubt about the absence of publication bias.

Phonological Awareness, Discourse, and Pragmatics

No studies in the area of phonological awareness, discourse, or pragmatics were identified in this meta-analysis.

Discussion

We demonstrate that VPT/VLBW children perform less well than control children on overall expressive and receptive language measures and in the more specific subdomains of expressive and receptive semantics. The results for expressive and receptive grammar were equivocal, with only one study in each subdomain and a low fail safe N statistic for the single significant effect size. No studies in the subdomains of phonological awareness, discourse, or pragmatics were identified.

The effect sizes identified when examining overall expressive and receptive language showed some heterogeneity across studies. A number of plausible factors that may have explained heterogeneous effect sizes within these two analyses were identified, but the small number of studies prevents any definitive conclusion. A large number of other factors can also contribute to heterogeneity across studies, including demographic differences, child differences (such as rates of developmental disability), and participation in early intervention.

Finally, the only two subdomains to produce heterogeneous findings were overall expressive and receptive language, which are composite indices. Thus, breaking down overall expressive and receptive test measures into further subdomains may result in more homogenous results and thus a clearer picture of functioning in each subdomain.

Subgroup analyses conducted on children at school age suggest that poorer language function is present during later stages of language development. Although longitudinal studies are required to examine the trajectories of VPT/VLBW children's language development, this finding indicates that these children may have ongoing language difficulties.

These results mirror those from meta-analyses conducted in other cognitive domains in VPT/VLBW children compared with term-born children. Medium to large effect sizes were found in the areas of academic achievement[56] and overall cognitive ability,[57] and small to medium effect sizes were found in executive functioning.[56] The large effect sizes in this meta-analysis, particularly in school-age children, suggest that VPT/VLBW children show a greater deficit relative to their peers in language than in executive function, an area that has generated considerable attention.

Many studies have considered the affect of sociodemographic factors on outcomes. This meta-analysis used unadjusted means and SDs to calculate overall effect sizes and therefore could be influenced by these factors. However, this may not be the most likely explanation for the results, because some studies examined differences with and without adjustment, and found no change in most or all of their measures.[32,33,36,43] Additional studies matched their groups on demographic factors such as parental education (Table II). The findings were not considered in the context of IQ scores, because controlling for IQ takes out part of the variability in language scores because IQ tests assess language.[58,59] For correcting age for prematurity (Table II), examination of the individual studies' effect sizes reveals that this difference could not have explained the heterogeneity in the overall expressive and receptive subdomains. Additionally, an age difference of 1 to 3 months would be unlikely to alter scores when standardized because age groupings of 6 to 12 months are used on many tests. Although there is currently no consensus on this issue,[60] it has been recommended on the basis of statistical modeling that corrected age be used until age 8.5 years.[61]

Statistically significant differences of 0.38 to 0.77 SD place VPT/VLBW children 5.7 to 11.6 points lower on a test with a mean of 100 and SD of 15, which is in the normal range but clearly at the lower end. However, there is large variability in this population. We did not examine rates of language impairment in this meta-analysis, and many studies did not report it. However, on the basis of the results, we would expect higher rates of impairment in VPT/VLBW children than in the term-born population. As such, there is a need for closer surveillance of language ability in these children.

Language in this meta-analysis was subdivided to more accurately identify specific areas of weakness for VPT/VLBW children. Language interventions are structured to target different language subdomains and modalities.[14,62,63] Rigorous studies are needed in the subdomains of grammar, phonological awareness, discourse, and pragmatics to determine whether VPT/VLBW children also perform less well in these areas.

The limitations of this meta-analysis should be considered. Although the strict inclusion criteria identified high-quality studies, it limited the number of identified studies, restricting the use and interpretability of meta-analytic techniques. The results are only based on measures that could be represented with means and SDs, excluding measures reported as odds ratios. Also, we did not assess the orthographic domain, thus excluding an important adjunct to language development. At the outcome level, the absence of acceptable studies precluded an analysis of the effect of birth weight/GA or rate of impairments. Finally, no studies were identified for a number of subdomains. We recommend that future studies include a control group and test a representative cohort. We suggest that subdomain language scores be reported in addition to overall "language" scores. Summary scores tend to mask differential effects that may be present because of the heterogeneous nature of language.

To conclude, VPT/VLBW children have significantly poorer language function compared with control children. These language difficulties are still present throughout primary school, a time when language development becomes more stable and adult-like. We combined data across language tests, but within theoretically and clinically relevant language subdomains. However, 3 areas were not represented by any data. Longitudinal follow-up studies with robust methodology are required to fully assess language as a broad cognitive ability in VPT/VLBW children. ■

We thank Katherine Lee of the Murdoch Childrens Research Institute for her advice on statistical analyses.

Submitted for publication Jun 30, 2010; last revision received Sep 28, 2010; accepted Oct 20, 2010.

Reprint requests: Natalie Barre, BSc(Hons), Victorian Infant Brain Studies, Murdoch Children's Research Institute, Level 2 Main Building, Royal Children's Hospital, Flemington Road, Parkville, Victoria 3052, Australia. E-mail: n.barre@alumni.unimelb.edu.au

References

1. Darlow BA, Cust AE, Donoghue DA. Improved outcomes for very low birthweight infants: evidence from New Zealand national population based data. Arch Dis Child Fetal Neonatal Ed 2003;88:F23-8.
2. Draper ES, Zeitlin J, Field DJ, Manktelow BN, Truffert P. Mortality patterns among very preterm babies: a comparative analysis of two

European regions in France and England. Arch Dis Child Fetal Neonatal Ed 2007;92:F356-60.
3. Bode MM, D'Eugenio DB, Forsyth N, Coleman J, Gross CR, Gross SJ. Outcome of extreme prematurity: a prospective comparison of 2 regional cohorts born 20 years apart. Pediatrics 2009;124:866-74.
4. Lemons JA, Bauer CR, Oh W, Korones SB, Papile LA, Stoll BJ, et al. Very low birth weight outcomes of the National Institute of Child Health and Human Development Neonatal Research Network, January 1995 through December 1996. NICHD Neonatal Research Network. Pediatrics 2001;107:e1.
5. Luoma L, Herrgard E, Martikainen A, Ahonen T. Speech and language development of children born at ≤32 weeks' gestation: a 5-year prospective follow-up study. Dev Med Child Neurol 1998;40:380-7.
6. Atkinson J, Braddick O. Visual and visuocognitive development in children born very prematurely. Prog Brain Res 2007;164:123-49.
7. Yliherva A, Olsen P, Suvanto A, Jarvelin MR. Language abilities of 8-year-old preterm children among the northern Finland 1-year birth cohort for 1985-1986. Logoped Phoniatr Vocol 2000;25:98-104.
8. Luu TM, Ment LR, Schneider KC, Katz KH, Allan WC, Vohr BR. Lasting effects of preterm birth and neonatal brain hemorrhage at 12 years of age. Pediatrics 2009;123:1037-44.
9. Hanke C, Lohaus A, Gawrilow C, Hartke I, Kohler B, Leonhardt A. Preschool development of very low birth weight children born 1994-1995. Eur J Pediatr 2003;162:159-64.
10. Wolke D, Samara M, Bracewell M, Marlow N. Specific language difficulties and school achievement in children born at 25 weeks of gestation or less. J Pediatr 2008;152:256-62.
11. Durkin K, Conti-Ramsden G. Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. Child Dev 2007;78:1441-57.
12. Botting N, Simkin Z, Conti-Ramsden G. Associated reading skills in children with a history of Specific Language Impairment (SLI). Read Writ 2006;19:77-98.
13. Conti-Ramsden G, Durkin K, Simkin Z, Knox E. Specific language impairment and school outcomes. I: identifying and explaining variability at the end of compulsory education. Int J Lang Commun Disord 2009; 44:15-35.
14. Law J, Garrett Z, Nye C. Speech and language therapy interventions for children with primary speech and language delay or disorder. Cochrane Database Syst Rev 2009:CD004110.
15. Girolametto L, Wiigs M, Smyth R, Weitzman E, Pearce PE. Children with a history of expressive vocabulary delay: Outcomes at 5 years of age. Am J Speech Lang Pathol 2001;10:358-69.
16. Bates E, Dale PS, Thal D. Individual differences and their implications for theories of language development. In: Fletcher P, MacWhinney B, eds. The handbook of child language. Oxford: Blackwell Publishers; 1995. p. 96-151.
17. Bornstein MH, Haynes OM. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. Child Dev 1998; 69:654-71.
18. Bates E, Goodman JC. On the inseparability of grammar and the lexicon: evidence from acquisition. In: Tomasello M, Bates E, eds. Language development: the essential readings. Oxford: Blackwell Publishers; 2001. p. 134-62.
19. Schatschneider C, Francis DJ, Foorman BR, Fletcher JM, Mehta P. The dimensionality of phonological awareness: an application of item response theory. J Educ Psychol 1999;91:439-49.
20. Baddeley A, Gathercole S, Papagno C. The phonological loop as a language learning device. Psychol Rev 1998;105:158-73.
21. Christiansen MH, Onnis L, Hockema SA. The secret is in the sound: from unsegmented speech to lexical categories. Dev Sci 2009;12:388-95.
22. Wocadlo C, Rieger I. Phonology, rapid naming and academic achievement in very preterm children at eight years of age. Early Hum Dev 2007;83:367-77.
23. Elbro C. Early linguistic abilities and reading development: a review and a hypothesis. Read Writ 1996;8:453-85.
24. Mar RA. The neuropsychology of narrative: story comprehension, story production and their interrelation. Neuropsychologia 2004;42:1414-34.
25. Zwaan RA, Radvansky GA. Situation models in language comprehension and memory. Psychol Bull 1998;123:162-85.
26. Adams C. Practitioner review: the assessment of language pragmatics. J Child Psychol Psychiatry 2002;43:973-87.
27. Lindell AK. In your right mind: right hemisphere contributions to language processing and production. Neuropsychol Rev 2006;16:131-48.
28. Duncan N, Schneider P, Robertson CMT. Language abilities in five-through seven-year-old children born at or under 28 weeks gestational age. J Med Speech Lang Pathol 1996;4:71-9.
29. Kern S, Gayraud F. Influence of preterm birth on early lexical and grammatical acquisition. First Lang 2007;27:159-73.
30. Chen P-S, Jeng S-F, Tsou K-I. Developmental function of very-low-birth-weight infants and full-term infants in early childhood. J Formos Med Assoc 2004;103:23-31.
31. Wang L-W, Wang S-T, Huang C-C. Validity of the Clinical Adaptive Test (CAT)/Clinical Linguistic and Auditory Milestone Scale (CLAMS) as a screening instrument for very low birth weight infants in Taiwan. J Dev Behav Pediatr 2005;26:412-8.
32. Foster-Cohen S, Edgin JO, Champion PR, Woodward LJ. Early delayed language development in very preterm infants: evidence from the MacArthur-Bates CDI. J Child Lang 2007;34:655-75.
33. Woodward LJ, Moor S, Hood KM, Champion PR, Foster-Cohen S, Inder TE, et al. Very preterm children show impairments across multiple neurodevelopmental domains by age 4 years. Arch Dis Child Fetal Neonatal Ed 2009;94:F339-44.
34. Zimmerman IL, Steiner VG, Pond RE. Preschool Language Scale-3. San Antonio, TX: Harcourt Brace Jovanovich; 1992.
35. Semel E, Wiig EH, Secord W. Clinical evaluation of language fundamentals—third edition. San Antonio, TX: Psychological Corporation; 1995.
36. Aarnoudse-Moens CSH, Smidts DP, Oosterlaan J, Duivenvoorden HJ, Weisglas-Kuperus N. Executive function in very preterm children at early school age. J Abnorm Child Psychol 2009;37:981-93.
37. Bohm B, Smedler AC, Forssberg H. Impulse control, working memory and other executive functions in preterm children when starting school. Acta Paediatr 2004;93:1363-71.
38. Esbjørn BH, Hansen BM, Greisen G, Mortensen EL. Intellectual development in a Danish cohort of prematurely born preschool children: specific or general difficulties? J Dev Behav Pediatr 2006;27:477-84.
39. Foulder-Hughes LA, Cooke RWI. Motor, cognitive, and behavioural disorders in children born very preterm. Dev Med Child Neurol 2003;45: 97-103.
40. Taylor HG, Klein N, Drotar D, Schluchter M, Hack M. Consequences and risks of <1000-g birth weight for neuropsychological skills, achievement, and adaptive functioning. J Dev Behav Pediatr 2006;27:459-69.
41. Harvey JM, O'Callaghan MJ, Mohay H. Executive function of children with extremely low birthweight: a case control study. Dev Med Child Neurol 1999;41:292-7.
42. Dunn L, Dunn LPPVT. Peabody Picture Vocabulary Test–Revised Form. Circle Pines, MN: American Guidance Service; 1981.
43. Pritchard VE, Clark CAC, Liberty K, Champion PR, Wilson K, Woodward LJ. Early school-based learning difficulties in children born very preterm. Early Hum Dev 2009;85:215-24.
44. Tomblin JB, Zhang X, Buckwalter P, O'Brien M. The stability of primary language disorder: four years after kindergarten diagnosis. J Speech Lang Hear Res 2003;46:1283-96.
45. Review Manager (RevMan). 5.0 ed. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration; 2008.
46. Cohen J. A power primer. Psychol Bull 1992;112:155-9.
47. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539-58.
48. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
49. Higgins JPT, Green S. Cochrane Handbook for systematic reviews of interventions. Version 5.0.2 (updated September 2009). The Cochrane Collaboration; 2009. Available at www.cochrane-handbook.org; 2009. Accessed December 8 2009.
50. Rosenthal R. Meta-analytic procedures for social research. Newbury Park, CA: Sage Publications; 1991.

51. Hartung J, Knapp G, Sinha BK. Statistical meta-analysis with applications. New Jersey: John Wiley & Sons; 2008.
52. Fail safe N. Available at http://www.stat-help.com/spreadsheets.html. Accessed Dec 9, 2009
53. McGrath M, Sullivan M. Birth weight, neonatal morbidities, and school age outcomes in full-term and preterm infants. Issues Compr Pediatr Nurs 2002;25:231-54.
54. Taylor HG, Klein N, Minich N, Hack M. Middle-school-age outcomes in children with very low birthweight. Child Dev 2000;71:1495-511.
55. Woodcock R, McGrew K, Mather N. Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside; 2001.
56. Aarnoudse-Moens CSH, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. Pediatrics 2009;124:717-28.
57. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. JAMA 2002;288:728-37.
58. Taylor HG. Children born preterm or with very low birth weight can have both global and selective cognitive deficits. J Dev Behav Pediatr 2006;27:485-6.
59. Dennis M, Francis DJ, Cirino PT, Schachar R, Barnes MA, Fletcher JM. Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. J Int Neuropsychol Soc 2009;15:331-43.
60. Wilson SL, Cradock MM. Review: accounting for prematurity in developmental assessment and the use of age-adjusted scores. J Pediatr Psychol 2004;29:641-9.
61. Rickards AL, Kitchen WH, Doyle LW, Kelly EA. Correction of developmental and intelligence test scores for premature birth. Aust Paediatr J 1989;25:127-9.
62. Cirrin FM, Gillam RB. Language intervention practices for school-age children with spoken language disorders: a systematic review. Lang Speech Hear Serv Sch 2008;39:S110-37.
63. Adams C. Intervention for children with pragmatic language impairments. In: Norbury CF, Tomblin JB, Bishop DV, eds. Understanding developmental language disorders: from theory to practice. East Sussex: Psychology Press; 2008.
64. Fenson L, Dale P, Reznick J, Thal D, Bates E, Hartung J, et al. The MacArthur Communicative Development Inventories. San Diego, CA: Singular Publishing Group; 1993.
65. Wiig EH, Secord W, Semel E. Clinical Evaluation of Language Fundamentals - Preschool Edition. San Antonio: Psychological Corporation; 1992.
66. Wechsler D. Weschler Intelligence Scale for Children. 3rd ed. London: Psychological Corporation, Harcourt Brace and Co; 1992.
67. Wechsler D. Wechsler Preschool and Primary Scale of Intelligence-Revised. New York: The Psychological Corporation; 1989.

Table I. Search terms by database

Database
MEDLINE (ISI): Major Topic terms
Infant, low birth weight OR infant, premature OR infant, premature, diseases OR premature birth
AND
Achievement OR cognition OR cognition disorders OR communication OR communication disorders OR developmental disabilities OR educational status OR intelligence OR language development OR language disorders OR learning OR psycholinguistics OR psychological tests
PsycINFO (CSA): Descriptors
Birth weight OR premature birth
AND
Achievement OR childhood development OR cognition OR cognitive development OR communication disorders OR developmental disabilities OR education OR linguistics OR language OR linguistics OR psychological assessment OR verbal communication
CINAHL (EBSCO): Subject Heading terms
Infant, low birth weight OR infant, premature OR infant, premature, diseases OR outcomes of prematurity
AND
Achievement OR cognition disorders OR communication OR developmental disabilities OR educational status OR intelligence OR language development OR language disorders OR language tests OR mental processes OR speech and language assessment OR psycholinguistics OR psychological tests
ERIC (CSA): Descriptors
Body weight OR premature infants
AND
Academic ability OR academic aptitude OR academic education OR cognitive ability OR cognitive development OR cognitive processes OR cognitive science OR cognitive tests OR communication disorders OR developmental disabilities OR early childhood education OR intelligence OR language OR language impairments OR language research OR linguistics OR special education OR speech language pathology OR verbal ability OR verbal tests

The "explode" function was used where applicable. References were managed using Endnote software.