Do Infants Born Very Premature and Who Have Very Low Birth Weight Catch Up With Their Full Term Peers in Their Language Abilities by Early School Age?

Emily Zimmerman

Purpose: This study examined the extent to which children born preterm (< 37 weeks) and/or who have low birth weight (< 2,500 g) catch up with their full term peers in terms of their language abilities at early school age (≥ 5 to < 9 years).

Method: A systematic literature search identified empirical studies that fit the inclusion criteria. Data from the tests/questionnaires used for meta-analysis spanned the following language categories: total language score, expressive language, receptive language, pragmatics, phonological awareness, and grammar. The means (standard deviations) were extracted from the studies and were converted to mean difference and 95% confidence intervals to test for overall effect.

Results: Sixteen studies met the inclusionary criteria, for a total of 2,739 participants, of which 1,224 were born full term and 1,515 were born preterm. It is important to note that the preterm cohort represented very preterm infants who have a very low birth weight. The meta-analysis found that preterm infants scored significantly worse on total language ($p < .001$), receptive language ($p < .001$), expressive language ($p < .001$), phonological awareness ($p < .001$), and grammar ($p = .03$) than their full term peers. However, preterm infants did not score significantly worse than their peers on their pragmatics ($p = .19$).

Conclusions: Children born VPT and who have VLBW perform worse than their peers on their total language, receptive language, expressive language, phonological awareness, and grammar abilities by early school age. This information is important for speech-language pathologists to consider as children born prematurely reach school age.

It is estimated that 15 million babies are born prematurely each year (Blencowe et al., 2012). Infants born very preterm (VPT; ≤ 33 weeks’ gestation), with a very low birth weight (VLBW; ≤ 1,500 g) or both, comprise 1% to 2% of all live births (Darlow, Cust, & Donoghue, 2003; Draper, Zeitlin, Field, Manktelow, & Truffert, 2007). Although advances in technology have resulted in increased survival rate, infants born preterm remain at an increased risk for poor neurodevelopmental outcomes, such as deficits in cognition, language, and behavior (Aarnoudse-Moens, Weisgles-Kuperus, van Goudoever, & Oosterlaan, 2009; Allen, Cristofalo, & Kim, 2011; Aylward, 2002; Bhutta, Cleves, Casey, Cradock, & Anand, 2002; Goldenberg, Culhane, Iams, & Romero, 2008; Jarjour, 2015; Johnson, 2007; Pritchard et al., 2009).

Studying the development of children born premature is critical but remains very difficult given the complexity of the perinatal and demographic variables that have the potential to influence neurodevelopmental outcomes. The most common of these perinatal variables are gestational age (GA) at birth and birth weight, with lower levels in both posing the greatest risk for future impairments (Anderson & Doyle, 2008; Bhutta et al., 2002; Marlow, Wolke, Bracewell, Samara, & EPICure Study Group, 2005; Taylor, Minich, Bangert, Filipek, & Hack, 2004). Other perinatal variables that can alter neurodevelopmental outcomes include birth location (inborn vs. outborn), bronchopulmonary dysplasia (BPD), postnatal steroid therapy, necrotizing enterocolitis, sepsis, jaundice, neonatal seizures, intraventricular hemorrhage, cerebral gray matter changes, and longer hospitalizations (Anderson, Doyle, & Victorian Infant Collaborative Study Group, 2003; Cusson, 2003; Fily

Disclosure: The author has declared that no competing interests existed at the time of publication.
have highlighted the substantial impairments across multiple developmental domains in children born preterm throughout childhood, including cognitive, language, motor, behavior, and academic achievement (Aarnoudse-Moens, Weisglas-Kuperus, van Goudoever, & Oosterlaan, 2009; Barre, Morgan, Doyle, & Anderson, 2011; Bhutta et al., 2002; de Kievet, Piek, Aarnoudse-Moens, & Oosterlaan, 2009; van Noort-van der Spek, Franken, & Weisglas-Kuperus, 2012). Meta-analyses examining language abilities have mainly focused on the following domains: total language, expressive language, receptive language, semantics (expressive and receptive), and grammar (expressive and receptive) from various standardized tests across all of early childhood (ages 2 to 12 years) and have shown that children born preterm have difficulties with language function compared with their peers (Barre et al., 2011; van Noort-van der Spek et al., 2012). This meta-analysis will expand on the previous studies in the following ways: (a) It will focus only on a critical period, ages 5–9 years, when children are entering school; (b) It will expand on the previously reported language domains to include pragmatics and phonological awareness; (c) It will include standardized and nonstandardized assessments/questionnaires. Therefore, this study aimed to meta-analytically chart the language outcomes of children born preterm (<37 weeks) and/or who have low birth weight (<2,500 g) at early school age (≥5 to <9 years).

Method

Literature Search Procedures

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Liberati et al., 2009) guideline was used for this systematic review. Research studies were found through a comprehensive search using the Northeastern University’s Scholar OneSearch with the following key word combinations: infant-premature/preterm or infant, low birth weight, and cognition, IQ, intelligence, developmental disabilities, child development, language development, phonological skills, grammar, comprehension, receptive language, expressive language, child development disorders, human development, impairments, school age, preschool, school outcomes, outcomes. Narrowing the publication date to 1995–2015 further refined the results. Based on these results and analysis of relevant review articles, 211 studies were found related to preterm infant outcomes that were published after 1995. Each of the 211 studies was screened to determine if it met our inclusion criteria (see Figure 1).

Inclusion Criteria

Inclusion criteria included the following: infants born preterm (<37 weeks) and/or with low birth weight (<2,500 g), after the year 1990, and assessment of language outcomes at early school age (average age at follow-up between ≥5 and <9 years). Exclusion criteria included study interventions to improve language outcomes, articles published prior to 1995, and articles that did not report the mean (standard deviation) of the language outcomes. One hundred seventy-six studies were excluded for one or a combination of the
following reasons: birth year not specified, review, book, infant born before 1990, assessment age, not relevant due to study type or study topic, and intervention applied (see Figure 1). Based on the aforementioned criteria, 35 studies were found to meet these criteria, and the full text was reviewed for eligibility. From here, 19 of the full-text articles were excluded for the following reasons: reported on a cohort older or younger than our inclusion criteria, no control group, contained outcomes that were not comparable to any other study (must have at least two studies to complete meta-analysis), did not report test scores (mean, standard deviation), and outcomes were based on head circumference or SES, not birth group (preterm/full term). In the end, 16 studies were included in the meta-analysis. Both standardized and nonstandardized tests were included for the following reasons: (a) Including both test types allows for a more representative sample of how these children born premature are functioning in their communication abilities at early school age; and (b) Smith et al. (2014) found that standardized tests often do not fully capture language abilities in children born preterm. Last, once the 16 studies were examined based on their population, it was determined that the cohort represented VPT infants (≤ 33 weeks’ gestation) who have VLBW (≤ 1,500 g).

**Data Selection**

The identified studies varied in respect to the following: country, subject number, age, birth weight, birth GA, age at follow-up, and the language assessment used (see Table 1). The 16 included studies (Guarini et al., 2009, 2010, 2016; Hanke et al., 2003; Harvey, O’Callaghan, & Mohay, 1999; Jansson-Verkasalo et al., 2004; Lewis et al., 2002; Lind et al., 2011; Ortiz-Mantilla et al., 2008; Pritchard et al., 2009; Reidy et al., 2013; Smith et al., 2014; Taylor, Klein, Drotar, Schluchter, & Hack, 2006; Wolfe, Vannatta, Nelin, & Yeates, 2015; Wolke et al., 2008; Woodward et al., 2009) used assessments or questionnaires that examined the following language categories: total language score, expressive language, receptive language, pragmatics, phonological awareness, and grammar. In order for a meta-analysis to be completed, the respective language categories had to have at least two studies that examined those outcomes (Valentine, Pigott, & Rothstein, 2010). All extracted data were entered into Review Manager 5.3 (The Cochrane Collaboration, 2014) for statistical analysis. The mean difference (standard deviation) or standard mean differences with 95% confidence interval (CI) for continuous outcomes were selected to estimate the pooled effect size. For this study, the mean difference reflected the absolute difference between the mean values between the preterm and the full term cohorts within a study. A few studies had outcomes that reflected the number of errors indicating that a higher score meant lower functioning. In order to include these data and not have them skew the mean difference toward favoring the wrong group, the number of errors was multiplied by −1 (see superscript “b” in Table 1). Three of the included studies, Smith et al. (2014), Ortiz-Mantilla et al. (2008), and Woodward, Clark, Bora, and Inder (2012), consisted of follow-up years before, during, and/or after our inclusion criteria; therefore, we chose to include these studies but to only focus on the early school age aspect of their studies. For the Ortiz-Mantilla et al. (2008) study, both the 5- and 7-year follow-ups were included, and for the Smith et al. (2014) study, the 7- and 8-year follow-ups were included. Lewis et al. (2002) reported on two different cohorts of premature infants: infants with BPD and infants born with VLBW, and both cohorts were included in the analyses. When a higher order score for participants existed in the study, that score was used rather than the subtests that composed that higher order score. If multiple measures of a single type (e.g., expressive language) were administered to the same participants, then a composite standardized effect size was determined, thereby taking each of these measures into account (see superscript “c” in Table 1). This methodology was also used for studies that included two age groups (Ortiz-Mantilla et al. [2008] and Smith et al. [2014] studies) or two birth cohorts (BPD/ VLBW from the Lewis et al. [2002] study) that met our inclusion criteria. The composite standardized effect size was completed by the Review Manager 5.3 using its calculator function.
<table>
<thead>
<tr>
<th>Author, Year, Country</th>
<th>Preterm, N</th>
<th>Full term, N</th>
<th>Preterm infant birth GA, M (SD)</th>
<th>Preterm infant birth weight (g), M (SD)</th>
<th>Follow-up age</th>
<th>Total language</th>
<th>Expressive language</th>
<th>Receptive language</th>
<th>Grammar</th>
<th>Phonological awareness</th>
<th>Pragmatics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guarini et al., Italy 2009</td>
<td>70</td>
<td>34</td>
<td>29.70 (2.30)</td>
<td>1,137.00 (327.70)</td>
<td>6</td>
<td>³Test di Vocabolario Figurato</td>
<td>³TCGB</td>
<td>&quot;Battery of nine metaphonological tasks&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guarini et al., Italy 2010</td>
<td>68</td>
<td>26</td>
<td>30.44 (2.22)</td>
<td>1,243.00 (264.00)</td>
<td>8</td>
<td>³Test di Vocabolario Figurato</td>
<td>³TCGB</td>
<td>&quot;CFM&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guarini et al., Italy 2016</td>
<td>60</td>
<td>60</td>
<td>28.90 (2.29)</td>
<td>1,177.53 (360.26)</td>
<td>5</td>
<td>⁶BVL 4-12: Naming</td>
<td>⁶BVL 4-12: Lexical Comprehension</td>
<td>&quot;BVL 4-12: Syntactical Comprehension &amp; Sentence Completion&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hanke et al., Germany 2003</td>
<td>60</td>
<td>60</td>
<td>29.00 (2.20)</td>
<td>1,124.00 (259.00)</td>
<td>6</td>
<td>MSVK: Language Comprehension</td>
<td>PPVT-R</td>
<td>&quot;Comprehension of Idiomatic Expressions &amp; Global Coherence in Narrative Production Task&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harvey et al., Australia 1999</td>
<td>30</td>
<td>50</td>
<td>27.00 (2.30)</td>
<td>846.00 (103.00)</td>
<td>5</td>
<td>⁷BNT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jansson-Verkasalo et al., Finland 2004</td>
<td>12</td>
<td>12</td>
<td>29.00 (SD not reported)</td>
<td>1,115.00 (SD not reported)</td>
<td>6</td>
<td>⁷BNT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lewis et al., United States 2002</td>
<td>93</td>
<td>93</td>
<td>BPD = 77.00 (2.10) VLBW = 89</td>
<td>BPD = 970.00 (254.00) VLBW = 30.00 (2.20)</td>
<td>⁸CELF-3: Total Language</td>
<td>⁸CELF-3: Expressive Language</td>
<td>⁸CELF-3: Receptive Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lind et al., Finland 2011</td>
<td>161</td>
<td>161</td>
<td>28.00 (3.00)</td>
<td>1,054.00 (259.00)</td>
<td>5</td>
<td>NEPSY-II: Comprehension of Instructions</td>
<td>NEPSY-II: Phonological Processing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smith et al., United States 2014</td>
<td>57</td>
<td>57</td>
<td>29.80 (1.70)</td>
<td>1,453.00 (331.00)</td>
<td>7, 8</td>
<td>⁸Language Sample: Metalinguistic verb density, morphologically complex word density, low frequency word density, Measure D, NDW-100, NTW-100</td>
<td>⁸Language Sample: Adverb density, conjunction density, complex conjunction density, elaborated noun phrase density, MLU, DSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(table continues)
Table 1. (Continued).

<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Country</th>
<th>Preterm, N</th>
<th>Full term, N</th>
<th>Preterm infant birth GA, M (SD)</th>
<th>Preterm infant birth weight (g), M (SD)</th>
<th>Follow-up age</th>
<th>Total language</th>
<th>Expressive language</th>
<th>Receptive language</th>
<th>Grammar</th>
<th>Phonological awareness</th>
<th>Pragmatics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ortiz-Mantilla et al., 2008</strong></td>
<td>United States</td>
<td>32</td>
<td>32</td>
<td>26.90 (2.17)</td>
<td>976.00 (245.50)</td>
<td>5, 7</td>
<td>TOLD-P:3 Speaking Quotient</td>
<td>TOLD-P:3 Listening Quotient</td>
<td>PAT: Rhyming, Deletion, Blending</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pritchard et al., 2009</td>
<td>New Zealand</td>
<td>102</td>
<td>108</td>
<td>27.90 (2.30)</td>
<td>1,071.00 (315.00)</td>
<td>6</td>
<td>WJ-III Understanding Directions</td>
<td>CELF-4: Language Structure</td>
<td>NEPSY-II: Phonological Processing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reidy et al., 2013</td>
<td>Australia</td>
<td>198</td>
<td>70</td>
<td>27.40 (1.90)</td>
<td>960.00 (222.00)</td>
<td>7</td>
<td>WJ-III Understanding Directions</td>
<td>CELF-4: Language Structure</td>
<td>CELF-4: Pragmatics Profile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taylor et al., 2006</td>
<td>United States</td>
<td>204</td>
<td>176</td>
<td>26.40 (2.00)</td>
<td>810.00 (124.00)</td>
<td>8</td>
<td>WJ-III: Picture Vocabulary</td>
<td>PLS-3: Total Language</td>
<td>PLS-3: Receptive Language</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolfe et al., 2015</td>
<td>United States</td>
<td>20</td>
<td>18</td>
<td>28.05 (2.70)</td>
<td>1,102.20 (256.80)</td>
<td>4, 6</td>
<td>WJ-III Understanding Directions</td>
<td>PKBS</td>
<td>PKBS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wolfe et al., 2008</td>
<td>United Kingdom</td>
<td>241</td>
<td>160</td>
<td>Not reported</td>
<td>Not reported</td>
<td>6</td>
<td>PLS-3: Total Language</td>
<td>PLS-3: Expressive Language</td>
<td>WJ-III Understanding Directions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Woodward et al., 2012</strong></td>
<td>New Zealand</td>
<td>104</td>
<td>107</td>
<td>27.90 (2.30)</td>
<td>1,065.90 (312.60)</td>
<td>6</td>
<td>WJ-III Understanding Directions</td>
<td>PKBS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. GA = gestational age; TCGB = Test di Comprensione Grammaticale per Bambini; CFM = Valutazione delle competenze metafonologiche; BVL 4-12 = Batteria per la Valutazione del Linguaggio in Bambini dai 4 ai 12 anni; MSVK = Marburg Language Comprehension Test for Children; PPVT-R = Peabody Picture Vocabulary Test–Revised; BNT = Boston Naming Test; BPD = bronchopulmonary dysplasia; CELF-3 = Clinical Evaluation of Language Fundamentals–Third Edition; CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; VLBW = very low birth weight; NEPSY-II = A Developmental NEuroPSYchological Assessment–Second Edition; NDW = number of different words; NTW = number of total words; MLU = mean length of utterance; DSS = Developmental Sentence Scoring; TOLD-P:3 = Test of Language Development–Primary: Third Edition; PAT = Phonological Abilities Test; WJ-III = Woodcock Johnson Test of Achievement; PKBS; Preschool and Kindergarten Behavior Scales; SCBE-30 = Social Competence and Behavior Evaluation–Third Edition; PLS = Preschool Language Test.

*Indicates only the early school age aspect of the study that was used. **Indicates numbers of errors that were reported. In order to include these outcomes in the analysis, the errors were multiplied by −1. 

*Indicates that subscores/outcomes were combined for a composite standardized effect size.
Study Heterogeneity
Given the variety of different birth characteristics (preterm, full term), birth weights, birth GAs, follow-up ages, and methods used to assess language development in children born premature and their peers, it was anticipated that there would be considerable heterogeneity. Given this increased heterogeneity, it is unlikely that all studies included in this meta-analysis were functionally equivalent and without a common effect size; therefore, the random effects model was used for all comparisons.

Results
A total of 211 studies were captured in the initial literature search stage. In conclusion, 16 studies, capturing 2,739 participants, remained according to the inclusion and exclusion criteria. A total of 1,224 children born full term and 1,515 children born preterm participated across 16 studies (see Table 1). All of the studies reported the birth GA and birth weight study criteria; however, one study did not report the average birth GA and birth weight of the preterm participants; instead, this study included only preterm infants born at 25 weeks GA or less. When examining the studies that reported these outcomes, the birth GA ranged from 26.40 to 30.44 weeks (average 28.24 ± 1.32 weeks), and the birth weight ranged from 810 g to 1,453 g (average 1,057.68 g ± 143.32 g). On the basis of these averages and ranges, we can label the preterm cohort in this meta-analysis as VPT and VLBW. Participants ranged in follow-up age from 5 to 8 years.

Assessment of Risk Bias
Basic characteristics of all eligible studies are shown in Table 1. Most of the included studies have issues with random sequence generation and allocation concealment. Random sequence generation is selection bias due to inadequate generation of a randomized sequence. There is a high risk of selection bias if the investigators describe a nonrandom component in the sequence generation process, such as sequence generated by date of birth, date (or day) of admission, hospital or clinic record number, or allocation by judgment of the clinician. Because our studies compared preterm infants chosen at birth in a nonrandom sequence and most with a judgment regarding their inclusion by the researchers, they are vulnerable to such bias. Allocation concealment is a selection bias due to inadequate concealment of allocations prior to assignment. Put simply, did the researchers have knowledge of group assignments? The risk of this bias is also high for the included studies, as they often compared a prospective group of preterm infants to retrospective matched controls on the basis of some criteria, and group assignments were known.

Meta-Analysis on Total Language Score
Two studies, which included 611 participants, reported the overall language score. Meta-analysis suggested that preterm infants performed significantly (test for overall effect, \( p < .001 \)) lower than their full term peers at early school age, mean difference (MD) = −13.20; 95% CI [−15.88, −10.51] (see Figure 2).

Meta-Analysis on Receptive Language Score
Nine different studies, which included 1,671 participants, reported the receptive language scores of children born preterm and of those born full term. Meta-analysis suggested that children born preterm performed significantly (test for overall effect, \( p < .001 \)) lower than their full term peers at early school age, MD = −6.10; 95% CI [−8.47, −3.73] (see Figure 3).

Meta-Analysis on Expressive Language Score
Nine different studies, which included 1,522 participants, reported the expressive language scores of children born preterm and of those born full term. Meta-analysis suggested that children born preterm performed significantly (test for overall effect, \( p < .001 \)) lower than their full term peers at early school age, MD = −6.16; 95% CI [−8.49, −3.84] (see Figure 4).

Meta-Analysis on Pragmatics
Three different studies, which included 426 participants, reported the pragmatics scores of children born preterm and of those born full term. Meta-analysis suggested that children born preterm did not perform significantly (test for overall effect, \( p = .19 \)) lower than their full term peers at early school age, MD = −8.30; 95% CI [−20.76, 4.15] (see Figure 5).

Meta-Analysis on Phonological Awareness
Seven different studies, which included 1,281 participants, reported the phonological awareness scores of children born preterm and of those born full term. Meta-analysis suggested that children born preterm performed significantly (test for overall effect, \( p < .001 \)) lower than their full term peers at early school age, MD = −1.46; 95% CI [−1.91, −1.01] (see Figure 6).

Meta-Analysis on Grammar
Five different studies, which included 694 participants, reported the grammar scores of children born preterm and of those born full term. Meta-analysis suggested that children born preterm performed significantly (test for overall effect, \( p = .03 \)) lower than their full term peers at early school age, MD = −4.55; 95% CI [−8.75, −0.34] (see Figure 7).

Discussion
Overall, this meta-analysis revealed that children born VPT and who have VLBW do not catch up with their
full term peers in their language abilities by early school age. These results are consistent with previous meta-analyses that examined language skills in children born premature at follow-up (Barre et al., 2011; Foster-Cohen, Friesen, Champion, & Woodward, 2010). This meta-analysis purposely chose to focus on a brief developmental window, early school age, when it is critical to determine if children are in need of school-based speech-language pathology services. This meta-analysis is unique in its follow-up time frame (school age), the language categories included, and the focus on premature infants born VPT and who have VLBW.

Robust differences were evident between children born VPT/VLBW and full term peers across total language, receptive language, and expressive language outcomes. The VPT/VLBW infants scored 0.78 to 0.99 SD below their peers in the total language analysis, 0.34 to 1.69 SD below their peers in the receptive language analysis, and 0.03 to 1.42 SD below their peers in the expressive language analysis. Considering that for most school districts a child must be 1–2 SDs below the mean to qualify for speech and language services, these delays may not appear substantial. However, Reidy et al. (2013) highlighted in their study that even though these children fall within the “average” range, the effect sizes were large and the proportion of children with language impairment was greater than 20%. It is very important that language outcomes in this population be examined through both standardized and nonstandardized testing, as standardized tests often do not fully capture the language abilities and deficits of the preterm cohort (Smith et al., 2014). For example, Hack et al. (2005) found that when parents completed the Parent Questionnaire for Identifying Children with Chronic Conditions (Stein, Westbrook, & Bauman, 1997), 22% of the parents reported that their preterm-born child had trouble understanding and speaking compared with parents of healthy birth weight children, who only reported 6% and 7% difficulty, respectively. Results from this questionnaire reveal a much larger gap in functional outcomes than are present in the standardized tests often utilized by SLPs.

The mechanisms responsible for the differences in language outcomes are likely substantially mediated by general cognitive deficits that originate in global changes in brain development (Inder, Warfield, Wang, Hüppi, & Volpe, 2005; Kesler et al., 2004; MacKendrick, 2006; Ortiz-Mantilla et al., 2008; Taylor, Burant, Holding, Klein, & Hack, 2002; Wolke & Meyer, 1999). For example, when the study by Wolke et al. (2008) controlled for general cognitive performance, no significant language difficulties were observed in the VPT/VLBW cohort, which is consistent with previous studies (Taylor et al., 2002; Wolke & Meyer, 1999). This finding points to the fact that deficits in language are embedded in cognitive deficits—where it has been shown that children born VPT/VLBW score 11–12 IQ points below their term peers (Bhutta et al., 2002;
Kerr-Wilson, Mackay, Smith, & Pell, 2012). Unsurprisingly, the amount of language/cognitive deficit is highly related to the amount of change evident in brain development for the preemie, specifically related to white matter. For instance, Woodward et al. (2012) found that VPT children without white matter abnormalities showed no neurocognitive deficits relative to their peers; whereas, children born VPT with mild and moderate-to-severe white matter abnormalities had impairments across all neurocognitive measures, with more severe cerebral abnormalities being associated with more cognitive impairments. Thus, their findings suggest that preterm infants who are spared of white matter injury can be expected to have similar levels of preschool and school-age cognitive function as their peers in regard to their intelligence, language, and executive functioning (Woodward et al., 2012).

The meta-analysis did not reveal significant differences in pragmatic scores (0.09 to 0.71 SDs below the M) between preterm and full term infants at early school age. In fact, the study by Wolfe et al. (2015) revealed that their preterm cohort performed better on their pragmatic skills than the full term group. However, the authors noted that this difference could be due to the fact that their term control group all spent some time in the NICU, which may have affected their scores. Interestingly, this study found that executive functioning and social information processing were positively correlated, which is consistent with the literature revealing the interrelatedness of executive functions and social competence across development (Nigg, Quamma, Greenberg, & Kusche, 1999; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006). Children who were born preterm have been reported to perform lower than their peers on global coherence in a narrative production task, arguably due to the higher level executive functioning needed to monitor, inhibit, and control the flow of thoughts (Guarini et al., 2016). When the authors controlled for cognitive functioning, this measure was no longer significant, again suggesting that these measures are closely linked. Although white matter abnormalities significantly affected results of other language measures, it was not found to have a mediating effect on pragmatics, suggesting that other biological factors, cerebral gray matter changes, and/or quality of parental speech may instead have stronger influence (Landry, Miller-Loncar, Smith, & Swank, 2002). Although this meta-analysis did not reveal that preterm infants scored significantly lower than their peers on pragmatics, the parents of children born preterm have reported increased social problems compared with those of the parents of children born full term (Hack et al., 2004; Reijneveld et al., 2006; Wocadlo & Rieger, 2006).

This study was the first to aggregate studies on phonological awareness in the preterm population and found that infants born VPT/with VLBW performed significantly lower on their expressive language scores than their full term peers, random effects model. The BPD/VLBW cohorts in the Lewis et al. (2002) study were combined, as were two age ranges from the Ortiz-Mantilla et al. (2008) study to form composite standardized effect sizes. The expressive language outcomes from the Smith et al. (2014) study were also combined to form one composite standardized effect size. BPD = bronchopulmonary dysplasia; VLBW = very low birth weight.
worse (0.22 to 0.76 SDs below their peers) than their peers in this area. In the Guarini et al. (2009) study, they found that phonological awareness at the syllable level was influenced by preterm birth but that this was not the case at the phoneme level. The authors noted that this inconsistency could be due to the fact that phonemic awareness is still being acquired at the age of 6 years in both preterm and full term infants. The studies by Ortiz-Mantilla et al. (2008) and Wolke et al. (2008) both used the Phonological Abilities Test (Robertson & Salter, 1997) and found that the preterm infants scored lower than full term infants on these measures, yet none of these differences reached significance.

Preterm infants also performed worse than their peers on their grammar scores (0.10 to 1.23 SDs below the M). However, Guarini et al. (2016) found that when they controlled for nonverbal cognition and memory skills, group differences in grammar production skills were no longer evident. The authors suggest a connectivity between cognition and grammar development, which is supported by the literature (de Abreu, Gathercole, & Martin, 2011; Sansavini et al., 2007). Smith et al. (2014) used discourse level language samples to examine grammatical features and found that, in the absence of frank neurological impairment, syntactic skills remain relatively intact. Interestingly, the study by Reidy et al. (2013) found that there was no mediating effect of white matter abnormality on grammar, reiterating that this measure (along with pragmatics described above) may be due to issues related to prematurity beyond white matter abnormalities.

Role of the SLP

SLPs play an essential role with the VPT/VLBW population. During the infant’s NICU stay, the SLPs should inform parents of the risk for future language deficits and educate them on the importance of early language exposure (Leffel & Suskind, 2013; Suskind et al., 2016). This knowledge will empower parents to be better advocates for their child as they reach preschool and school age. NICU SLPs should provide local resources, such as early intervention (EI) upon discharge, and encourage parents to contact these services if they begin to suspect that their child might have delays. Preterm children are more in need of EI services than term children (Hanke et al., 2003), and these programs are effective at improving outcomes with this population. For example, premature children who have been exposed to EI programs have significantly higher cognitive, language, and social scores compared with children who did not receive EI (Als et al., 2004; Gianni et al., 2006). These potent effects of EI are likely due to the fact that this type of support and stimulation is occurring during

Figure 6. Meta-analysis on the phonological awareness scores: Preterm infants scored significantly lower on their phonological skills than their full term peers, random effects model. The subtests from the Guarini et al. (2009, 2010) studies were combined, as were the PAT scores from the Ortiz-Mantilla et al. (2008) and Wolke et al. (2008) studies to form composite standardized effect sizes. PAT = Phonological Abilities Test.

Figure 7. Meta-analysis on the grammar scores: Preterm infants scored significantly lower on their grammar scores than their full term peers, random effects model. The outcomes from the Guarini et al. (2016) and Smith et al. (2014) were combined to form composite standardized effect sizes.
a known critical period of brain development (Volpe, 2009), where the infant’s brain is more plastic and adaptable.

School SLPs and teachers also need to be well informed of the language difficulties VPT/VLBW children may face. Even at this relatively late juncture of development, obtaining information about the child’s early development may help professionals better understand the child’s current level of functioning. This type of information could be critically important in helping the SLP to determine the causality of some of the language deficits evident in the child and to identify children who may not qualify for services through traditional standardized testing. In fact, the study included by Pritchard et al. (2009) found that VPT children were two to three times more likely to be identified by their teachers as performing below average or at delayed levels across seven different curricular areas, including math, reading, spelling, language, and writing skills. Without a doubt, teachers have observed academic/behavioral differences in their classroom in children born preterm—regardless of whether the etiology of these issues is known or if they have been formally identified as needing services. This is an ideal opportunity for SLPs to train teachers in techniques to help with these language delays in order to enhance the child’s classroom learning.

Considerations & Limitations

The results of this meta-analysis and the abundance of research reporting delays in the preterm infant population point to the vital need to reevaluate outcomes of children born VPT/VLBW at school age (Guarini et al., 2009; Salt & Redshaw, 2006; Walther, den Ouden, & Verloove-Vanhörick, 2000). The VPT/VLBW population is extremely complex, and the contributing factors known to be essential for language development must be properly weighed as we follow these infants across development. Roberts, Bellinger, and McCormick (2007) have proposed a model in which biological, demographic, and experiential factors all predict academic difficulties in children born preterm. In addition, reassessing language abilities throughout schooling is key, as early childhood outcomes are not always predictive of long-term consequences (Hintz et al., 2005), and delays may not manifest until later in childhood in those with less severe impairments (Holsti, Grunau, & Whitfield, 2002). It is important to note that children who perform adequately in early childhood may lag behind peers as language demands increase in scope and complexity at school age. Ideally, a set of multifaceted and multidomain evaluations, including tasks beyond standardized tests (Burnett et al., 2015; Watkins & DeThorne, 2000), should be given to VPT/VLBW children entering school and throughout schooling so that SLPs and other developmental specialists may ensure that these children are being supported in the best way to optimize outcomes. Without multifaceted evaluations, including birth history, teacher and parent reports, and informal evaluation by SLPs, VPT/VLBW children may remain in the low average range for language abilities, when they have the potential to catch up with peers given appropriate intervention.

There are several important factors that need to be considered when examining preterm follow-up research. Preterm children have difficulties with working memory (Sansavini et al., 2007; Vicari, Caravale, Carlesimo, Casadei, & Allemd, 2004) and attention issues (Briscoe, Gathercole, & Marlow, 2001; Elgen, Sommerfelt, & Markestad, 2002), both of which could affect their test results and academic abilities. It is likely that these children may need more time or a quieter room to complete the standardized tests given these known concerns. A close examination of between-groups differences is essential. In this meta-analysis, there were several instances where the preterm group performed at or slightly above average, as was evidenced in the receptive language areas with the Ortiz-Martína et al. (2008), Pritchard et al. (2009), and Woodward et al. (2012) studies; in the expressive language areas with the Ortiz-Martína et al. (2008) study; and in the pragmatics area with the Reidy et al. (2013) study. These findings indicate that, while preterm infants perform at an average level, they are lower than their peers who are often performing “above average.” Therefore, the peer group chosen needs to be closely considered and can greatly influence if and to what extent the preterm infant “catches up.” For example, in the included study by Wolfe et al. (2015), they intentionally matched their VLBW cohort with term-born, healthy birth weight infants who spent some time in the NICU (with no major prenatal or perinatal complications), as this is a more appropriate comparison group for isolating the effects of VLBW from other medical confounders. In their analyses, they did not find significant differences between the groups. In addition, individual differences within the preterm cohort need to be considered. In the Ortiz-Martína et al. (2008) study, 10.50% of their preterm cohort scored better than the full term cohort at age 7 years. It appears that a small subset of children born with VLBW develop similarly or better than their full term peers despite their prematurity (Ortiz-Martína et al., 2008; Saigal et al., 2006; Taylor et al., 2002). These “resilient” infants need to be studied in more detail to learn more about the mediating/moderating factors that contributed to them “catching up” and, in some cases, exceeding their full term peers.

This meta-analysis included only 16 studies that followed children born prematurely into early school age. These were the only studies that met our rigid inclusion and exclusion criteria. There were no geographical constraints in this meta-analysis as the included studies spanned many countries (see Table 1). Six of the studies included non-English-speaking children. Therefore, the ability to generalize the findings from this meta-analysis to other non-English-speaking cohorts beyond Italian, German, and Finnish may be limited. In addition, it remains unknown if and to what extent late preterm infants (born between 34 and 36 weeks) catch up with their peers at school age. This is an important question and should be examined in subsequent analyses.

Conclusion

Children born VPT and who have VLBW do not catch up with their full term peers at early school age in
terms of their total language, receptive language, expressive language, phonological awareness, and grammar abilities by early school age. SLPs are in a unique position to educate parents on the long-term outcomes of VPT/VLBW infants and to ensure that they are receiving the proper language resources that are necessary to thrive at school age. Future studies need to be completed to examine these outcomes later in school age and early adulthood and focus on different cohorts of premature infants.

Acknowledgments

No funding was secured for this study. The author would like to thank Kelsey Thompson and Samudragupta Bora for their help with this meta-analysis.

References


