THE JOURNAL OF PEDIATRICS • www.jpeds.com

Long-Term Stability of Language Performance in Very Preterm, Moderate-Late Preterm, and Term Children

Diane L. Putnick, PhD[1], Marc H. Bornstein, PhD[1], Suna Eryigit-Madzwamuse, PhD[2], and Dieter Wolke, PhD[3]

Objective To investigate whether children born very preterm, moderate-late preterm, and term differ in their average level and individual-difference stability in language performance over time.

Study design Language was assessed at 5 and 20 months and 4, 6, and 8 years of age in 204 very preterm (<32 weeks' gestation), 276 moderate-late preterm (32-36 weeks' gestation), and 268 term (37-41 weeks' gestation) children from the Bavarian Longitudinal Study.

Results Very preterm children consistently performed worse than term-born children, and moderate-late preterm children scored in between. Language performance was stable from 5 months through 8 years in all gestation groups combined, and stability increased between each succeeding wave. Stability was stronger between 5 months and 4 years in very preterm than moderate-late preterm and term groups, but this differential stability attenuated when covariates (child nonverbal intelligence and family socioeconomic status) were controlled.

Conclusions Preterm children, even moderate-late preterm, are at risk for poorer language performance than term-born children. Because individual differences in language performance are increasingly stable from 20 months to 8 years in all gestation groups, pediatricians who attend to preterm children and observe language delays should refer them to language intervention at the earliest age seen. *(J Pediatr 2016;■■:■■-■■).*

See editorial, p •••

Worldwide, more than 15 million infants are born prematurely every year.[1] Very preterm children are at increased risk for delays and deficits in various aspects of language.[2-5] As survival rates for preterm infants have risen as the result of improvements in obstetrics and neonatology,[6] preterm birth has emerged as a risk factor for poor development in an increasing proportion of the population.

According to guidelines from the World Health Organization,[7] preterm birth can be subdivided into very preterm (births before 32 weeks' gestation), moderate preterm (births at 32 and 33 weeks' gestation), and late preterm (births between 34 and 36 weeks' gestation). Language skills are impaired in children born very preterm[8]; however, findings regarding mean differences in language are less consistent for moderate-late preterm compared with term-born children.[4,9-11] In addition to differences between gestational age groups, developmental stability of language (consistency in relative standing over time)[12] is important because it is prognostic of future ability. There is emerging evidence that individual differences in language are stable from toddlerhood in term-born children,[13-16] but it is unclear whether stability in language development during childhood varies across the full gestational age spectrum.[17-19] When pediatricians evaluate language in young children, they need to know the point in early development at which individual differences are predictive of later language performance (or deficits) in children born term and preterm. Therefore, the current study aims to investigate the mean differences and stability of language performance, by using multiple age-appropriate measures, in children from 5 months to 8 years of age following a very preterm, moderate-late preterm, or healthy term birth within a population-based cohort.

Methods

Data were drawn from the prospective Bavarian Longitudinal Study.[20] Participants were children born alive in a geographically defined area of Southern Bavaria (Germany) during a 14-month period who required admission to children's hospitals within the first 10 days after birth ($N = 7505$; 10.6% of all live births). Healthy

AWST	Active Vocabulary Test
CFI	Comparative Fit Index
HSET	Heidelberger Language Development Test
LSVT	Language Comprehension Test
SES	Socioeconomic status

From the [1]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, Bethesda, MD; [2]Centre for Health Research, University of Brighton, Brighton, United Kingdom; and [3]Department of Psychology, and Division of Mental Health and Wellbeing, Warwick Medical School, University of Warwick, Coventry, United Kingdom

D.P. and M.B. were supported by the intramural program of the National Institutes of Health, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. Data acquisition was supported by the German Federal Ministry of Education and Science (PKE24, JUG14, 01EP9504, and 01ER0801 [to D.W. and S.E.-M.]). The authors declare no conflicts of interest.

0022-3476/$ - see front matter. Published by Elsevier Inc.
http://dx.doi.org10.1016/j.jpeds.2016.09.006

infants who were born in the same obstetric hospitals (most born at term), cared for on normal postnatal wards, and discharged with their mothers were recruited as controls (N = 916). Ethical approval was granted by the ethical review board of the University of Munich Children's Hospital and the Bavarian Health Council in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki). Parents provided informed consent within 48 hours of their child's birth. Very preterm children were born between 25 and 31 weeks' gestation, moderate-late preterm children were born between 32 and 36 weeks' gestation, and term children were born between 37 and 41 weeks' gestation.

Figure 1 (available at www.jpeds.com) describes the Bavarian Longitudinal Study participant flow. The full sample was assessed at birth, 5 months, 20 months, and 4 years 8 months of age (hereafter called 4 years); the sample was then reduced (N = 1543) before 6-year and 8-year assessments. All very preterm children were included in the reduced sample. A random sample of children born at >31 weeks' gestation was drawn according to the following stratification variables: sex, family socioeconomic status (SES; low, moderate, high), and degree of neonatal risk (very low, low, moderate, high[3]).

For this study, we removed 17 children who were not German speakers, 142 who were twins or higher-order multiple births (excluded because they have unique reasons for being preterm and have been found to have different language development than singletons[21]), 6 with language data at only one or no assessments, and 78 with physical or developmental disabilities or unknown status (ie, blindness, deafness, or cerebral palsy levels 3-4 [unable to move unaided]).[22,23] Because the focus of this report is to compare children who were born preterm with healthy full-term children, we also removed 556 children who were born at term but were hospitalized at birth with early medical problems and 12 children in the healthy control sample who were born preterm but cared for on normal obstetric wards. The **Table** (available at www.jpeds.com) describes the sample (N = 749) by gestational age group: very preterm (n = 205), moderate-late preterm (n = 276), and full-term (n = 268).[24]

Procedures

Assessments at 5 and 20 months were carried out at term-corrected ages[25] by pediatricians, and at 4, 6, and 8 years at chronological ages by postgraduate clinical psychologists.[26] German versions of assessments were used.

At 5 and 20 months of age, the Griffiths Mental Development Scales[27] hearing and speech subscale was used to evaluate children's age-appropriate receptive and expressive communication. Scores were standardized to M = 100, SD = 15.

At 4 years of age, the Active Vocabulary Test (AWST)[28] and the Language Comprehension Test (LSVT)[29] were used. The AWST is a reliable and valid vocabulary assessment of expressive language ability of children from ages 3 to 6 years.[30] The LSVT was developed for children aged 4-8 years to assess language comprehension. Standardized scores with M = 100 and SD = 15 were used for both the AWST and the LSVT.[31]

At 6 years of age, 4 subscales of the Heidelberger Language Development Test (HSET)[32] were administered to measure: (1) grammatical rules (plural-singular rules); (2) language production (sentence production); (3) grammatical structure (understanding of grammatical structures); and (4) language comprehension (correction of semantically inconsistent sentences). *T* scores were used for each subtest with M = 50 and SD = 10.[26] Next, experimenters observed the quality of children's speech and grammatical correctness during the assessment day and made judgments using consensus ratings based on the Diagnosis of Speech and Language.[26] Finally, prereading skills, including recognition of rhymes, sounds, and knowledge of numbers and letters, were assessed with the use of 4 prereading tasks adapted from the School Maturity Assessment.[26,33]

At 8 years of age, experimenters administered the HSET,[32] observed the Diagnosis of Speech and Language,[26] and administered the Zurich Reading Test[34,35] to assess reading speed and number of reading errors, as well as a Pseudoword Reading Test[36] to measure children's word decoding skills by asking them to read words that have no meaning.

Covariates

Family SES, computed as a weighted composite score of parents' education and occupation and grouped as low, middle, and high,[37] was used as a general covariate. To control for child nonverbal intelligence, we standardized and averaged multiple measures at each age. At 5 and 20 months of age, we used the eye-hand and performance subscales of the Griffiths Mental Development Scales.[27] At 4 years of age, we used the Beery Visual-Motor Integration test[38] and Columbia Mental Maturity Scale.[39,40] At 6 and 8 years of age, we used the Beery Visual-Motor Integration test[38] and the nonverbal index of the Kaufman Assessment Battery for Children.[41,42] These covariates were used as controls for language performance at each age, and residualized language scores were used in a covariate controlled model.

Statistical Analyses

A full analytic plan, details about measurement models, and additional statistical details appear in **Appendix 1** (available at www.jpeds.com).

Results

Full Sample Language Stability Model

We used latent variables to model the shared variance among language measures. This procedure has the advantage of removing measurement error and specific variance for each scale from the latent factor, leaving a more precise and reliable estimate of language ability at each age.[43] Furthermore, the use of latent variables allows for developmentally appropriate changes in the measurement of language as children age. Measurement models supported a single language factor at 4 years of age and second-order factor models with first-order factors for each of the major tests given at 6 and 8 years of age (**Appendix 2**; available at www.jpeds.com). By using these factors, stability of individual differences was modeled from 5- and 20-month language scales to 4-, 6-, and 8-year

FLA 5.4.0 DTD ■ YMPD8657_proof ■ October 10, 2016