Objective This review summarizes the literature to date concerning age adjustment in developmental assessment and illustrates relevant issues for clinicians and researchers in this area. Methods Pubmed, Medline, Premedline, and PsycInfo databases were used to search the following terms: assessment, prematurity, age adjustment, and adjusted age. Additional sources were obtained through the references listed in the primary articles. Results Differences in clinical opinions persist with regard to whether, what type, and how long age adjustments are to be made for premature infants. Research is inconclusive regarding the most appropriate method for accounting for prematurity. Conclusions Previous data may no longer describe the current population of premature infants. Current research is needed to inform developmental assessment practices for this population. Meanwhile, clinicians should remain aware of the issues regarding whether and how to adjust for prematurity. The authors recommend a multifaceted approach to clinical decision making and provide suggestions for future research.

Key words prematurity; age adjustment; developmental assessment.
in the most accurate prediction" of later functioning (p. 1186). Age adjustment may seem simple in theory, and many clinicians routinely use some calculation to account for early birth; however, the issues of accounting for premature birth are plentiful. Research has yet to advise clinicians on the most appropriate course of action.

**Age Adjustment Practices in Developmental Assessment**

There is much debate on the use of adjusted versus unadjusted scores; to determine developmental expectations, some authors prefer full adjustment (subtracting from chronological age the number of days or weeks premature), no adjustment (use of chronological age), or some midpoint (Ouden, Rijken, Brand, Verloove-Vanhorick, & Ruys, 1991). Most measures of development generally encourage adjustment of a child’s “test age” if he or she had a gestational age of fewer than 38 weeks at birth (e.g., the second edition of the Denver Developmental Screener Test by Frankenburg et al., 1990; the Mullen Scales of Early Learning by Mullen, 1993). Because the Bayley Scales of Infant Development—second edition (BSID-II; Bayley, 1993) is often the primary diagnostic tool used by psychologists to determine initial eligibility for early intervention (Gauthier et al., 1999), a particular focus on issues concerning age adjustment for this measure is presented.

The BSID-II has two domains from which standard scores may be derived: the Mental scale and the Motor scale. The standard scores for both scales, called the Mental Development index (MDI) and the Psychomotor Development index (PDI), have means of 100 and standard deviations of 15. Scores below 85 are considered delayed and often prompt further assessment or intervention. For each scale, predetermined items sets are based on the child’s age at testing. With regard to prematurity, the BSID-II instructs clinicians to subtract from the chronological age, the amount of time that the child was premature to arrive at the “adjusted age.” An additional footnote indicates that clinicians may want to use the item set that corresponds to the child’s “adjusted age” if the child is under 2 years old (Bayley, 1993, p. 41). No further direction is provided in the manual concerning this practice; clinical recommendation is to administer the item set appropriate to the child’s level of functioning (Matula, Gyrurke, & Aylward, 1997).

Thus, implementation is left to the clinician to make on a case-by-case basis and include decisions regarding the item set and normative group selected. For example, a 12-month-old child born 3 months premature could be given the item set delineated for a 9-month-old child and compared to norms based on 9-month-old children in the normative sample. A second alternative would be for the clinician to administer the item set based on the chronological age (12 months) and use the normative sample that corresponded to the adjusted age (9 months) to derive the standard score. Finally, the chronological age could be used to determine both the item set and the norm sample, with clinical consideration given to the degree of prematurity (e.g., tempering the feedback and recommendations based on knowledge of premature development).

Table 1 displays the results obtained when the three aforementioned approaches are applied to a 12-month-old child with an adjusted age of 9 months. Resulting scores range from significant delay to within-normal limits and span 31 points, well beyond the average standard error of 5 points. Thus, the same 12-month-old child, receiving credit for the same number of items, looks vastly different depending on the clinician’s approach. With one approach the child may be eligible for early intervention services via Public Law 99–457; with other approaches, he or she may not. The clinician invariably derives different scores, warranting different feedback to parents and different recommendations for services depending on the decision to use item sets and norms based on adjusted age or chronological age.

Inconsistent application of adjusted test scores across various agencies leads to scores that are not truly comparable (Ross & Lawson, 1997). Furthermore, as illustrated with the previous example, each approach has implications for intervention eligibility and feedback to parents. Thus, consideration of this issue is important in the follow-up of premature infants. Numerous questions remain about the clinical use of age-adjusted scores: Is age adjustment appropriate? If so, is it appropriate for all assessment domains or for all premature infants? How much should scores be adjusted to compensate for

<table>
<thead>
<tr>
<th>Item set</th>
<th>Norms</th>
<th>Raw score</th>
<th>Standard score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 month</td>
<td>9 month</td>
<td>66</td>
<td>80</td>
<td>Mild delay</td>
</tr>
<tr>
<td>12 month</td>
<td>9 month</td>
<td>75</td>
<td>101</td>
<td>Within-normal limits</td>
</tr>
<tr>
<td>12 month</td>
<td>12 month</td>
<td>75</td>
<td>70</td>
<td>Significant delay</td>
</tr>
</tbody>
</table>

Raw score is based on 5 credited items, the minimum score required to obtain a basal. Once a basal is established, credit is automatically given for the items that preceded the start of the item set administered. Thus, given 5 credited items, the child would receive a raw score of 66 using the 9-month item set and a raw score of 75 using the 12-month item set.
premature status and for how long? Which score is more predictive of later functioning? These questions have not been adequately resolved. There is limited research specifically addressing the question of the predictive validity of age-adjusted scores. Much of the research dates from over 10 years ago and may not represent the current population of premature infants. Though it is limited in scope, the research presented herein represents the information available to inform clinical decisions.

Theoretical Basis for Age Adjustment

The use of full adjustment versus no adjustment is based on two theoretical viewpoints on infant maturation. Adjusting for the degree of prematurity is based on a biological perspective, assuming that early development proceeds as a function of time since conception. The implicit assumption in this viewpoint is that the developmental status of a preterm child will lag behind that of a full-term child due to maturational differences, at least temporarily. This rationale was first advocated by Gesell and Amatruda (1947), who stated that development was determined by maturation of the central nervous system, independent of environmental influence. This theory led to the concept of “catch-up” growth, the idea that premature infants would meet the developmental level of their full-term counterparts within the first years of life, after complete maturation of their central nervous system. Adjusting for the degree of prematurity was purported to remove the transient delay seen in children born prematurely until they catch up with their full-term peers.

The preference for chronological age reflects an environmental perspective, which places higher importance on the role of external factors (e.g., parental stimulation, quality of nourishment, medical care) in the development of premature infants. Some have purported that social variables (e.g., home environment, parent-child interaction) are more predictive of development than medical factors (e.g., low birth weight, gestational age) at least for different ethnic groups (Zahr, 1999). Reliance on the chronological age, or time since birth, follows this environmental perspective, assuming that external influences are more important than biological maturation in subsequent development. Regardless of a clinician’s theoretical framework, it is unclear to what extent medical factors have negative effects beyond gestational age or to what extent environmental factors accelerate or inhibit developmental functioning of premature babies.

Is Age Adjustment Appropriate?

To what degree does maturation or environment influence and subsequently predict future developmental outcomes for premature babies? And consequently, is it more appropriate to correct for prematurity or to use chronological age when assessing for developmental delay? Parmelee and Schulte (1970) were among the first to investigate this question, in their study of 25 full-term infants, 20 small-for-date infants (full-term but less than 2,500 g), and 26 premature infants (mean gestational age = 35 weeks). Using nerve conduction velocities as a measure of neural development, they reported that premature infants were neurologically less mature at birth than the other two groups. Subsequent developmental assessment, 40 weeks after birth, found that the premature infants had an average developmental quotient (DQ) of 88 when not adjusted for amount of prematurity. Although this score is within the average range of development, it is significantly lower than the average DQ of 99 and 96 for the full-term and small-for-date infants, respectively.

When the scores for premature infants were adjusted to reflect their degree of prematurity, their average DQ rose to 99. Thus, the researchers concluded that development was dependent on time since conception and that neurological maturation was not augmented by extrauterine life but rather based on biological processing (Parmelee & Schulte, 1970). They advocated adjusting age based on prematurity, especially during the first years of life. Although they stated that “as the child becomes older, this correction factor becomes proportionately smaller compared with total age” (p. 25), they offered no data to inform at what age the adjustment was no longer necessary. Furthermore, they demonstrated insufficient follow-up to ensure that these infants really were developmentally equivalent to those in the comparison groups (Blasco, 1989). It may be that the appropriateness of age adjustment depends on the particular skills assessed, the degree of prematurity, and the current age of the child. These issues we turn to next.

Is Age Adjustment Appropriate for All Domains Assessed?

Some have stated that adjusted scores are more appropriate for certain types of skills. For example, using the Dutch version of the BSID (Van der Meulen & Smrkovsky, 1983), Lems and colleagues (1993) compared 36 healthy premature infants (mean gestation age = 32 weeks) and 21 full-term infants, all from the Netherlands, throughout
the first year of life. The authors generated “deviation scores” to account for problems in interpreting the performance of a child whose developmental index fell outside the test’s limits (MDI < 50 or MDI > 150). One deviation score was calculated for all infants (i.e., full-term and premature infants) by subtracting the child’s chronological age from the BSID developmental age equivalent. For premature infants, a second deviation score was calculated by subtracting the child’s adjusted age from the BSID developmental age equivalent. Positive deviation values reflected advanced development, whereas negative values reflected delayed development relative to the normative group. The authors compared the deviation scores for the full-term infants to both deviation scores generated for the premature infants at five points during the first year (12, 18, 24, 39, and 52 weeks). For the mental scale, the authors found that both deviation scores calculated for premature babies were significantly below those of the full-term group. Thus, they recommended a full adjustment during the first year when assessing mental skills. For the motor scale, a significant group effect was found when comparing the mean deviation score of full-term infants to the mean deviation score based on chronological age for premature infants; adjusted age deviation scores failed to reach significance. The authors commented on the increasing variability in deviation scores for the premature infants, particularly that between 39 and 52 weeks, which may have obscured a group effect at these ages. They concluded that full adjustment might not be needed during the second half of the first year when assessing motor skills.

In contrast, other studies have advocated that age adjustment is more important for motor, rather than mental, skills. Ross (1985) compared 1-year-old full-term infants (n = 46) with premature infants (n = 46). Mean scores across both groups were within normal limits; however, premature infants scored significantly lower on motor skills than they did on mental skills. Because of this discrepancy, Ross concluded that prematurity had a greater impact on gross and fine-motor development than it did on mental development, at least in the first year of life.

With a sample of 46 healthy infants (half of which were born between 29 and 32 weeks gestation), Palisano (1986) used motor-specific assessment measures (i.e., Peabody Developmental Motor Scale, Gross Motor Scale) to assess infants’ skills at 12, 15, and 18 months. His results showed that when being assessed by chronological age, the premature infants had significantly lower gross and fine-motor scores than their full-term counterparts. Adjusted-age scores did not differ significantly from those of the full-term group. Thus, he supported the use of age adjustment for motor skills until 18 months (Palisano, 1986). The full adjustment employed in this and other studies eliminated differences between the groups, but generally the data do not attest to the predictive appropriateness of adjusted scores. In other words, were the differences noted true differences or merely transient development lag?

Based on their study of 100 high-risk preterm infants (< 32 weeks gestation), Allen and Alexander (1990) concluded that the motor delay noted in preterm infants was better attributed to transient developmental lag versus true deficit. They established the infants’ ages of acquisition for 12 motor milestones—for example, rolling over, sitting with/without support, crawling, pulling to stand—during regular follow-up visits, recording the chronological age and the age adjusted for prematurity. Based on chronological ages, their sample generally achieved motor milestones an average of 2 to 3 months later than full-term infants. When adjusted ages were used, the developmental trajectory approximated normative development. They noted that all infants had typical motor skill attainment between 12 and 24 months, suggesting that the use of chronological age before this time would inappropriately diagnose persistent motor dysfunction.

Ouden and colleagues (1991) used an approach similar to Allen and Alexander (1990). They compared the age of developmental milestone attainment for 555 premature infants (< 32 weeks gestation) to that of 550 full-term comparison infants, assessing all infants at four points within their first 2 years of life. They compared the percentage of infants in both groups who achieved a positive score on age-specific developmental items—for example, looks at own hands, holds head in sitting position, jabbers, crawls, drinks from cup—taken from a Dutch neurodevelopmental screener used widely in the Netherlands. Across developmental items administered in the first year, the percentage of premature infants that attained positive scores ranged from 80% to 99% when adjusted ages were used. Only 30% to 80% had positive scores when chronological ages were used. At 2 years of age, more than 90% had positive scores across the age-specific items regardless of whether adjusted ages were used. Because the majority of children had no handicap at 2 years of age, the researchers concluded that adjustment in the first year eliminated undue referrals for intervention but that such adjustment was no longer necessary in the second year (Ouden et al., 1991).
**Do Premature Infants “Catch Up” With Regard to Cognitive Skills?**

Taken together, the data suggest that motor delays noted in premature development might be resolved by 2 years. Thus, clinicians could expect most premature infants to “catch up” to their full-term peers in terms of motor development by their first and second birthdays. However, the data are more limited regarding whether and when catch-up cognitive development occurs.

The results of long-term follow-up with regards to cognitive catch-up by preemies are mixed. Ment et al. (2003) recently described an 8-year follow-up for a sample of 296 very-low-birth-weight (VLBW) preemies (< 1,250 g) assessed at 3, 4.5, 6, and 8 years of age (corrected at each assessment point for the amount of prematurity). The authors reported improved cognitive performance as assessed by the Peabody Picture Vocabulary Test–Revised (PPVT-R), from 3 to 8 years of age, with nearly half of their sample having gains of 10 points or more. Of those VLBW preemies with scores in the borderline range of functioning (70–80) at 3 years, 71% had scores within normal limits (> 80) at 8 years. Children who received special services and had mothers with fewer than 12 years of education showed greater PPVT-R increases, as did children in two-parent households (Ment et al., 2003).

Though this study was not designed to assess the predictive utility of age-corrected scores, some have taken these results to suggest continued correction of expectations through school age. However, prudent application of these results is advised. The authors of the PPVT-R caution that stability of scores decreases as the length of time between testing increases (Dunn & Dunn, 1981). Without a contemporary comparison group, it is unclear how the reported gains would compare with those of full-term, normal-birth-weight individuals. Furthermore, the children’s test age was continually adjusted for degree of prematurity; thus, the scores were not compared to age-matched normative tables. Additionally, the PPVT-R was chosen because it could be used across the age range and by individuals with motor involvement. However, other authors have noted that receptive vocabulary may not be as adversely affected by the biologic risk associated with prematurity as would other nonverbal or analytic skills (Aylward, 2003). The fact that the children’s test age was corrected makes it difficult to determine if their scores were truly “within average limits” or whether this was a function of the test age adjustment. No information was provided regarding developmental services received. Therefore, it is difficult to determine how this affected the noted increases.

Hack and colleagues (Hack et al., 2002) reported a 20-year follow-up for 242 VLBW infants (mean birth weight = 1,179 g; mean gestation = 29 weeks). These VLBW individuals had significantly lower cognitive scores, lower achievement scores, and higher incidence of grade repetition than did a cohort of normal-birth-weight adults with similar demographic characteristics. Furthermore, a significantly lower percentage of VLBW infants had graduated from high school as adolescents. VLBW individuals were significantly less likely to be enrolled in 2- or 4-year college programs than were the comparison individuals. The authors concluded that VLBW individuals continued to evidence difficulties that persisted into adulthood, affecting educational and occupational achievement (Hack et al., 2002). At this time, it is still unclear if and to what degree preemies (especially VLBW preemies) catch up, especially in regards to cognitive development. Even if it were assumed that continual catch-up of cognitive skills occurred throughout school age, it would not invariably support continued adjustment throughout this time.

**Is Age Adjustment Appropriate for All Premature Infants?**

Researchers have attempted to better understand the implications of early birth and lower birth weight. Barrera, Rosenbaum, and Cunningham (1987) compared the performance of a group of full-term infants (n = 24) to that of two groups of premature infants: VLBW (less than 1,500 g, n = 20) and high birthweight (HBW; 1,500–2,000 g, n = 39). The VLBW infants had a mean gestational age of 29 weeks, whereas the mean of the HBW infants was 34 weeks. Using the BSID (Bayley, 1969), all infants were assessed at 4, 8, 12, and 16 months of age. The VLBW infants had the lowest scores on the mental and motor domains, for all assessment points. The difference between the HBW and the full-term group, using chronological age, disappeared at 12 months. In contrast, the unadjusted scores for the VLBW infants continued to be significantly lower than those of the full-term group through 16 months of age. Thus, the HBW group seemed to catch up to developmental expectations, but the VLBW group did not. When adjusted ages were used, there were no significant differences between the three groups.

The VLBW infants had significantly earlier gestational ages; they were born on average 5 weeks earlier than the HBW preemies. Thus, it is unclear whether they simply would require more time to catch up and longer use of adjusted age or if the developmental lag...
represents a more persistent deficit that would warrant intervention. For HBW preemies, age adjustment seemed appropriate within the first year but was unnecessary afterward (Barrera et al., 1987). VLBW preemies, however, failed to show catch-up within this period. Information is not available to suggest whether and when this resolution occurs. Continued age adjustment may actually mask persistent developmental concerns leading to overconfidence in the child’s abilities; it may delay access to necessary early intervention; or it may falsely inform or reassure caregivers and professionals working with the child.

How Much Should Scores Be Adjusted?

Aylward (1986) stated that the question is not whether to adjust but how to best separate maturational delay from “true” developmental dysfunction. The Barrera et al. study (1987) demonstrated that full adjustment may not be the most appropriate allowance for premature birth, nor may it serve the interests of all premature infants. Miller, Dubowitz, and Palmer (1984) investigated the usefulness of adjusted and unadjusted DQs (via the Griffith scales) for infants of fewer than 34 weeks gestation. Assessing 114 infants at 6 and 9 months, they compared the adjusted and unadjusted DQs in discriminating those premature infants who were neurologically impaired at 12 months (n = 36) from those who were not (n = 78). They concluded that the unadjusted DQ more readily distinguished the impaired infant than the adjusted DQ. While the unadjusted DQ was more sensitive to impairment (Miller et al., 1984), the specificity was unclear. This has led some to advocate a partial adjustment for at least some duration during the child’s early infancy (Blasco, 1989; Lems et al., 1993; Mauk & Ting, 1987). For this approach, a 12-month-old born 2 months premature would conceptually be considered an 11-month-old for assessment purposes (versus a 10-month-old used in full adjustment). At this time, it is unclear whether this approach (i.e., partial adjustment) is helpful in balancing the issues evident in accounting for premature birth.

Do Age-Adjusted Scores Predict Later Functioning?

Siegel (1983) attempted to determine when preterm development began to approximate that of full-term and which scores (adjusted or chronological) best predicted subsequent development. She followed two cohorts of infants—full term (n = 92) and premature (n = 100, mean gestational age of 30 weeks)—for 3 to 5 years after birth and assessed them at eight points using various developmental and cognitive measures. Comparing the first edition of the BSID (Bayley, 1969) with the McCarthy Scales of Children’s Abilities, she noted that at 8 months of age, the children’s unadjusted scores were often better predictors of the McCarthy scores at 5 years. However, the differences between the correlations for the adjusted and unadjusted scores were nonsignificant at later ages, leading her to conclude, “It does not matter very much which scores are used” (p. 1186) because perhaps neither score was appropriate for predictive purposes. It may be that adjusted scores overestimate a child’s ability, whereas unadjusted scores underestimate the child’s level (Siegel, 1983), leading to overreferral to intervention services. It may also be that neither score captured the variability in this group of infants and the variability in their rates of development.

In a follow-up study of extremely low-birth-weight Japanese children (ELBW, less than 1,000 g), Surgita, Iai, Inoue, and Ohta (1990) described the correlation between toddler developmental scores and the full scale IQ score from the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R). They used the Revised K-Form Developmental Test (RF-DT) at 12, 18, 24, 30, and 36 months and the WPPSI-R full-scale IQ score at 5.5 years, comparing the correlations of scores based on full adjustment and no adjustment. The RF-DT was created in Japan and generates motor, cognitive, and language scores as well as a total IQ. The children had an average gestational age of 27 weeks and an average birth weight of 823 g. Both adjusted and unadjusted IQ scores were significantly correlated with WPPSI-R full-scale IQ at 5.5 years. Though the sample size was small, the researchers concluded that after 1 year of age, “there were no significant differences between the predictive value of the adjusted and unadjusted IQ scores as to later IQ scores” (p. 337). Thus, they concurred with Siegel (1983) that it did not matter which scores were used.

Discussion

Though instrument manuals and conventional practice may lead a clinician to assume that adjustment until 2 years is standard, the research is not sufficient to fully address the practice of age adjustment, specifically regarding whether and how to apply corrections. There is support for age adjustment to reduce undue anxiety and overreferral for at least some premature infants for a period of early development. However, accounting for early birth may depend on factors other than prematurity
Age Adjustment for Premature Infants

Porting age-adjustment practices.

Clinicians’ awareness of the state of the research is meant to. Instead, the intent of the authors is to increase the reduction of overreferral. The aforementioned review suggests how to balance the issues of early identification versus practice are left to their own conclusions to determine whether infants received early intervention (and how much), or they fail to statistically control for this variable. The goal of early intervention is to augment infant development; therefore, if participants did receive such services, their developmental trajectories were likely adjusted. Thus, attention to the amount and type of intervention services received is recommended.

Suggestions for Future Research

Certain limitations in the reviewed literature should be addressed in future research. First and foremost, although theoretical debate has been published in recent years (i.e., Matula et al., 1997; Ross & Lawson, 1997), conclusions about the appropriateness of age adjustment has been based on data collected during the 1970s and 1980s. Little research has looked at the effect of adjusted scores on more current samples of premature infants. Because medical interventions with this population have advanced in recent decades, premature infants may be qualitatively different now from those of earlier decades. Furthermore, the majority of the research from the 1990s was conducted with international samples (e.g., Netherlands, Japan); the appropriateness of generalizing these results is not known due to differences in measurement or populations characteristics.

Suggestions for Clinicians

The implications of developmental assessment are salient to the lives of children whose eligibility for early-intervention services may be affected by the particular assessment paradigm utilized when adjusting for prematurity. Too much adjustment, or for too long, may remove or preclude children from services that would have later been deemed beneficial. However, too little or no adjustment may lead to overreferral to early-intervention services and undue parental anxiety for a child who may catch up. Balancing these issues can be daunting. For those infants born later than 28 weeks without significant complications, adjustment for motor skills until the end of the first year is advised. Conclusions are elusive regarding correction for mental skills as well as for all corrections for infants with more significant histories. Until more research can inform how much clinicians should adjust for prematurity, for which groups of premature infants, and for how long, we concur with others who have advocated reporting both scores (e.g., Barrera et al., 1987; Lems et al., 1993) to monitor the child’s progression and his or her approximations to typical development. It is prudent for clinicians to carefully consider the impact that the adjusted test scores may have on the child’s scores, as illustrated here, and to communicate the amount and type of adjustment that was made during development. Similarly, research findings should be presented with clear identification of whether and how adjustment for prematurity was applied to the sample.
Although the studies of Ment et al. (2003) and Hack et al. (2002) add to the understanding of the complex outcomes for premature birth, they do not address the questions raised in the process of age adjustment in developmental testing. Ment and colleagues’ findings give hope that continued cognitive gains are apparent in VLBW preemies, but it is unclear whether the researchers’ sample fully caught up to their full-term, normal-birth-weight peers. Hack and colleagues’ findings suggest that some impairment persists into adulthood, but it is unclear whether difficulties persist in absence of appropriate identification and intervention. Future outcome studies may include an investigation of which scores—those based on corrected or uncorrected ages—more accurately predicted long-term functioning.

Because the BSID-II is used primarily with children who are suspected of developmental delay in clinical and research settings, its use in research with this sample is particularly in need of validation. Cross-site consensus on adjusting for prematurity is needed to appropriately attend to the needs of premature infants and their families. This consensus should be predicated on research findings based on current population of premature infants. Though it is beyond the scope of this review to point out particular designs appropriate to address these concerns, Aylward (2002) provides helpful guidelines to those interested in advancing this area of premature infant follow-up.

Acknowledgments

The authors wish to thank Russell Hoffmann and Kristen Kullgren for their supportive feedback and comments on previous drafts of this manuscript. Their thoughtful and professional critiques improved the quality of the manuscript and generated needed discourse on this topic.

Received September 9, 2003; revisions received December 30, 2003; accepted March 3, 2004

References


Age Adjustment for Premature Infants


