Correcting for prematurity affects developmental test scores in infants born late and moderately preterm

Authors: Shalin A Parekh[a], BSc, MBBch, MRCPch; Elaine M Boyle[a], MD, PhD; Alexa Guy[a], BSc, MRes; Samarita Blaggan[a], MSc; Bradley N Manktelow[a], PhD; Dieter Wolke[b], PhD, CPsychol, AFBPsS; Samantha Johnson[a] PhD, CPsychol, AFBPsS.

Affiliations: [a]Department of Health Sciences, University of Leicester, Leicester, UK; [b]Department of Psychology and Health Sciences Research Institute and Division of Mental Health and Wellbeing, Warwick Medical School, University of Warwick, Coventry, UK.

Corresponding author (address for reprints): Dr Samantha Johnson, Senior Research Fellow, Department of Health Sciences, 22-28 Princess Road West, Leicester LE1 6TP United Kingdom. Tel: 0116 252 5798; Fax: 0116 252 3272; Email: sjj19@le.ac.uk

Running head: Correcting for late prematurity in developmental tests**.**

Conflicts of interest: None.

Manuscript word count: 3103

Number of figures and tables: 3

Abbreviations: BSID-II Bayley Scales of Infant Development 2nd Edition; Bayley-III Bayley Scales of Infant and Toddler Development 3rd Edition ; LMPT Late and Moderately Preterm; LAMBS Late and Moderately preterm Birth Study ; MDI Mental Development Index.

Abstract

Background: Corrected age is typically applied when assessing the development of children born <32 weeks gestation. There is no consensus as to whether corrected age should be applied when assessing children born late/moderately preterm (LMPT; 32-36 weeks gestation).

Aims: This study explored the impact of corrected age on developmental test scores in infants born LMPT.

Study design: 221 LMPT infants were assessed at two years corrected age using the Bayley-III Cognitive and Language scales, from which Cognitive and Language composite scores were derived (Normative Mean 100; SD 15). Assessments were then re-scored using chronological age. Bayley-III composite scores <80 were used to define developmental delay. Paired samples t-tests were used to assess the difference in mean test scores derived using corrected versus chronological age, and McNemar's tests to assess the difference in the proportion of infants with developmental delay using corrected versus chronological age.

Results: Mean corrected age scores were significantly higher than chronological age scores (Cognitive: 2.1 points; 95% CI 1.6, 2.5; Language 2.5; 95% CI 2.1, 2.8). Overall, significantly more LMPT infants were classified with developmental delay when chronological (18.3%) versus corrected (15.0%) age (p=0.016).

Conclusions: Correcting for prematurity results in significantly higher developmental test scores and a significantly lower prevalence of developmental delay in LMPT infants and may affect eligibility for intervention services. Researchers and clinicians should be aware that the use of corrected age may impact on developmental test scores at both an individual and population level among infants born LMPT.

Keywords: developmental assessment; preterm; developmental delay; corrected age; neurodevelopmental outcomes.

Late and moderately preterm births (LMPT; 32-36 weeks gestation) constitute up to 84% of all preterm births.[1] Children born at these gestations are at higher risk for developmental delay, cognitive deficits, attention problems and special educational needs than their term-born peers.[2-5] These problems are evident in infancy and recent population-based studies have shown that children born LMPT are at twice the risk for neurodevelopmental disability compared with term-born controls at two years of age.[6]

Developmental assessments in the early years are regarded as important for identifying children at risk and for targeting early intervention.[7, 8] As such, developmental tests at two years are widely used to assess neurodevelopmental outcomes and to ascertain eligibility for intervention services.[9] [10] For assessing the development of infants born very (< 32 weeks) and extremely (< 28 weeks) preterm, it is common practice to correct for gestation at birth up to two years of age. However, there is uncertainty as to whether to use corrected age when assessing children born LMPT. Mounting evidence regarding the increased risk for impairments across multiple developmental domains leads one to question whether corrected age should be applied when assessing these infants. This is an important practical consideration as the application of corrected age may affect group mean scores on standardized developmental tests, the identification of children with developmental delay and subsequent eligibility for early intervention services. Moreover, it has been shown to impact on the statistical significance of study results in which developmental outcomes are compared between LMPT infants and term-born controls.[11]

There is a lack of studies that have assessed how corrected age affects test scores and the identification of developmental delay in children born LMPT. The Bayley Scales of Infant and Toddler Development 3rd Edition (Bayley-III)[12] is one of the most recently standardized developmental tests and is widely used in clinical practice and in perinatal and pediatric research. There is broad global consensus that developmental outcomes should be measured at 18-24 months of age, at which point standardized tests have greater reliability than earlier measures for identifying adverse developmental outcomes.[13, 14] The aim of this study was to determine the effect of correcting for prematurity on Bayley-III developmental test scores and the identification of developmental delay in infants born LMPT at two years of age.

METHOD

Participants

Participants for this study were drawn from the Late and Moderately Preterm Birth Study (LAMBS), a prospective geographical population-based study of outcomes following LMPT birth. The study was conducted in the East Midlands region of the United Kingdom and recruited 1113 infants born at 32^{+0} to 36^{+6} weeks of gestation from 1st September 2009 to 31st December 2010. This cohort was followed up at two years of age using parent questionnaires. At this time a sub-sample of 253 infants was recruited to LAMBS-II, a sub-study in which infants' cognitive and language development was formally assessed by a study psychologist with the aim of validating a parent questionnaire assessing the same developmental domains. Parents of children in the full LAMBS cohort were contacted via email or telephone and invited to participate in LAMBS-II. A home visit was arranged for those who expressed interest and written parental consent was obtained at the start of the visit, prior to data collection. A feedback letter summarizing their child's test results was sent to parents after the assessment. LAMBS and LAMBS-II were approved by the Derbyshire National Health Service Research Ethics Committee (Ref: 09/H0401/25). This report comprises secondary analysis of LAMBS-II data.

Measures

Infant development was assessed using the Bayley Scales of Infant and Toddler Development, 3rd Edition (Bayley-III).[12] This is a standardized, norm referenced test of development consisting of separate scales to assess cognitive, language and motor development. Each scale comprises a series of developmental play activities and the child's performance is scored based on the total number of items completed appropriately. This raw score is then compared with age standardized normative reference data derived from the standardization sample in order to obtain a scaled score (Mean 10; SD 3) and a composite score (Mean 100; SD 15) for each scale.

For this study, the Bayley-III cognitive and language scales were administered in a single session by one of two study psychologists (AG; SB) who were formally trained in test administration and scoring prior to commencing the study. Throughout the study, 10% of Bayley-III assessments were scored independently by both examiners to assess inter-rater reliability; this was shown to be excellent with 97% agreement across test items on all assessments. All Bayley-III assessments were administered and scored using the child's corrected age (to 40 weeks gestation). Subsequently, the assessments were re-scored using the start point and norm reference table appropriate for the child's chronological age.

A Bayley-III composite cognitive or language score more than 2 SD below the standardized mean of 100 (i.e., score < 70) has conventionally been used to define moderate to severe developmental delay. However, a number of studies have shown that the Bayley-III produces higher scores than corresponding Mental Development Index (MDI) scores obtained using the 2nd Edition of the Bayley Scales of Infant Development (BSID-II)[15] and that the conventional cut-off of scores <70 underestimates developmental delay relative to the BSID-II.[16-21] Given growing concern regarding constitution of the Bayley-III standardization sample and the underestimation of developmental delay, a cut-off score of 80 is recommended for defining moderate to severe developmental delay when using Bayley-III composite scores.[21] We therefore defined developmental delay for this study as follows: Bayley-III cognitive composite score <80 as cognitive delay; Bayley-III language composite score <80 as language delay; and a cognitive *or* language composite score <80 as any developmental delay.

Statistical analyses

Data were analysed using SPSS software (Version 22 IBM Corporation). Paired sample t-tests were used to assess the difference in mean Bayley-III cognitive and language composite scores derived using corrected versus chronological age. A change score was computed by subtracting the child's chronological age composite score from that derived using corrected age for each scale. The number of infants with developmental delay was cross tabulated and McNemar's tests for paired data were used to assess the difference in the proportion of children with developmental delay classified using corrected versus chronological age. Subgroup analyses using paired sample t-tests were also conducted to explore the difference in mean cognitive and language composite scores by corrected versus chronological age where re-scoring of the test resulted in the application of a different start point or norms table.

RESULTS

Study sample

Recruitment and sampling are described in Figure 1. Of the 1113 LMPT children in the LAMBS cohort, the parents of 394 were contacted in succession and invited to participate in LAMBS-II, of which 253 were recruited. Of the remaining 141 children not recruited, the parents of 49 (35%) declined to participate and for 92 (65%) children assessments could not be scheduled within the study period. An analysis of non-responders previously reported[22] showed that there were no significant differences between the sample recruited to LAMBS-II (n=253) and the rest of the LMPT cohort (n=860) in terms of sex (p=0.63), gestational age (p=0.81) and birth weight (p=0.67); however, the 253 infants recruited to LAMBS-II were less likely to be multiple births (p <0.001) and had a lower level of socio-economic deprivation (p <0.001) compared with the rest of the LMPT birth cohort.

Of the 253 children recruited to LAMBS-II, 12 (5%) were not tested as only 1 child from each multiple birth was assessed for this study. Completed Bayley-III cognitive or language assessments were not available for an additional 20 (7.9%) children; of these 3 (1.2%) had severe developmental delay and 17 (6.7%) did not complete the assessment due to performance issues (i.e., lack of attention, tiredness or unwillingness to cooperate with the examiner). The final sample for the present report thus comprised 221 LMPT children for whom full Bayley-III cognitive or language assessments were available and could be re-scored using chronological age (Figure 1). Of these 221 children, 213 completed both the Bayley-III cognitive and language scales, 7 completed only the cognitive scale and 1 only the language scale. Thus 220 (213+7) Bayley-III cognitive assessments and 214 (213 +1) Bayley-III language assessments were analysed.

<<Figure 1>>

Characteristics of the final sample of 221 LMPT children are as follows. The mean gestational age was 35 weeks 3 days (SD 8 days; range 32 weeks 1 day to 36 weeks 6 days), mean birthweight 2465g (SD 530g; range 1150g to 4960g) and 113 (51%) children were male. Their mean corrected age at the time of the Bayley-III assessment was 25 months 12 days (range 24 months 0 days to 27 months 24 days) and mean chronological age 26 months 13 days (range 24 months 28 days to 29 months 15 days).

Effect of corrected versus chronological age on Bayley-III scores

Descriptive statistics for Bayley-III scores are shown in Table 1. Mean cognitive and language composite scores were within the average range for both corrected and chronological age for these LMPT children (cognitive 96.4 and 94.3; language 99.8 and 97.3, respectively). Overall, mean composite scores for corrected age were significantly higher than scores for chronological age, with a mean difference of 2.1 (95% CI 1.6 to 2.5) points on the cognitive scale and 2.5 points (95% CI 2.1 to 2.8) on the language scale (Table 1). Contrary to expectations, some children had lower standardized test scores for corrected age than chronological age, with the difference in test scores derived using corrected and chronological age ranging from -10 to +10 on the cognitive scale and -3 to +9 on language scale.

<TABLE 1>

Effect of corrected versus chronological age on the prevalence of developmental delay

Overall, the use of corrected age resulted in a significantly lower prevalence of children being classified with cognitive delay as well as any developmental delay (Table 1). Thirty two (15.0%) children were classified with any developmental delay when scored using corrected age, compared with 39 (18.3%)

when scored using chronological age (Table 1). In the cognitive domain, 13 (5.9%) had delay when assessed using corrected age compared with 20 (9.1%) using chronological age. Although the prevalence of language delay was lower using corrected versus chronological age (13.1% vs. 15.4%), the difference was not statistically significant (Table 1).

Effect of different norms tables and start points

Age standardized scores are derived from raw scores using age-specific norms tables. For 15 children, re-scoring the test using chronological age required the application of the same norms table as used for the corrected age and there was no change in test scores on both the cognitive and language scales for these children. However, for children where a different norms table was required to derive chronological age to that required for corrected age, scores ranged from 10 points lower to 10 points higher for the cognitive scale and from 3 points lower to 9 points higher for the language scale (Table 2).

We also explored the effect on standardized test scores when chronological versus corrected age resulted in the application of a different start point for administering and scoring the test. Overall, corrected age scores were significantly higher than chronological age scores for both the cognitive and language scales, regardless of whether the same or different start point was used (Table 2). However, in some cases, and contrary to expectations, applying corrected age resulted in lower test scores (i.e., a negative change score) than using chronological age. It was observed that this occurred in cases in which a different (lower) start point was required for corrected age assessments, which resulted in scores up to 10 points and 3 points lower on the cognitive and language scales respectively (Table 2). See Appendix A for a description of how the application of corrected age can result in lower standardized scores.

<TABLE 2>

DISCUSSION

There is growing concern regarding an excess of neurodevelopmental sequelae in infants born at 32-36 weeks of gestation compared with their term-born peers. Long-term follow up is increasingly recommended for this large group of infants[23], but there is no consensus over whether corrected age should be applied when assessing children born LMPT. We have shown that the use of corrected age when administering and scoring Bayley-III assessments at two years of age results in significantly higher test scores compared to using chronological age in this population. The application of corrected age also results in a significantly smaller proportion of LMPT born children being identified as developmentally delayed at two years of age.

Our results are consistent with those of previous studies conducted by Romeo and colleagues, in which they assessed 62 infants born late preterm (33-36 weeks gestation) using the BSID-II MDI. The authors reported that the mean MDI score at 18 months corrected age was 96.2 (SD 9.9) whilst the mean MDI calculated using chronological age was significantly lower at 87.2 (SD 9.8).[24] Among this sample there was only one child (1.6%) with an MDI score <85 using corrected age, whilst 31 (50%) children had MDI scores <85 using chronological age.[11, 24] In our larger study using the newer Bayley-III, the difference in test scores derived using corrected versus chronological age was smaller in magnitude (2.1 to 2.5 points). We also identified a higher proportion of children with developmental delay from scores calculated using corrected age (15%) and a smaller proportion of children (18.3%) from scores calculated using chronological age compared with that of Romeo and colleagues. The differences between the results of the two studies may be explained by assessment at different ages (18 months vs. 24 months) and by differences in the structure of the two tests, particularly as the revised Bayley-III comprises reversal and discontinue rules to determine which test items are administered to individual children, in contrast to the more rigid ‘item sets’ used in the BSID-II. In addition, the Bayley-III comprises separate scales for cognitive and language development whilst the BSID-II comprised a single MDI score for assessing development across these domains. Furthermore, there are mounting reports that the Bayley-III underestimates developmental delay relative to the BSID-II[25-27] which may explain the larger proportion of children identified with developmental delay by Romeo et al. Mean Bayley-III scores for our sample were close to the normative test mean (100) which may be surprising given that there is growing evidence of an increased risk for developmental delay, cognitive impairments and language difficulties among infants born LMPT.[6] This again may be accounted for by the well-documented inflation in test scores using the Bayley-III relative to the BSID-II.[18, 26, 27] Using a theoretical model to explore the effects of corrected age on standardized test scores, Wilson-Ching and colleagues recently demonstrated that Bayley-III test scores calculated using corrected and chronological age can differ where children are born one to two months preterm, thus supporting our findings relating to the Bayley-III.[28]

The application of corrected age leads to a developmental advantage in test scores, effectively making the test easier than when chronological age is used. This was borne out in the present study in which corrected age scores were significantly higher than those derived using chronological age. This was particularly the case for children who were close to the age boundaries for deriving normative scores and where application of corrected age resulted in the use of a different norms table. Indeed, the use of age-banded normative scores on Bayley-III test results has recently been shown to affect rates of delay and may result in misclassifications of developmental abilities depending on the child's age relative to age-banded cut-offs.[29] Contrary to expectations, we also found that the application of corrected age did not always result in a higher cognitive or language score, and in many cases lower scores were obtained (cognitive: n=16 (7.2%); language: n=1 (0.5%)). This is a result of the structure of the test in which a drop in scores using corrected age may arise when a different (lower) start point is used. Although the proportion of children with developmental delay generally decreased when using corrected age, for some children the application of corrected age resulted in a disadvantage in test scores. Thus, for some, the application of corrected age results in a greater likelihood of being classified with developmental delay.

These results highlight important considerations for examiners using the Bayley-III. Decisions about whether to apply corrected age can affect an individuals' test score and thus their eligibility for intervention services. Although the mean difference in test scores is small, on a population level this has a significant effect as the application of corrected age increases group mean scores by two to three points. This can affect the statistical significance of between group differences in cohort studies, particularly in the LMPT population in which differences from term-born controls are around this magnitude. Indeed, in some studies in which corrected age has been applied to assess long term outcomes for late preterm infants the authors have reported no significant differences from term-born controls.[11, 30] It is therefore important to consider the use of corrected age when planning studies and in interpreting test results.

The strengths of this study include the large sample of 221 children in comparison to previous studies in this area, and the representativeness of the sample compared with the whole LMPT population in terms of sex, gestational age and birthweight. However, three children who could not complete the test due to severe developmental delays were excluded and children recruited to the study were from less socio-economically deprived families, both of which are factors which may affect sample representation in terms of underestimation of the prevalence of developmental delays. We used the Bayley-III which is one of the most recently standardized and widely used developmental tests, particularly in preterm and clinical populations, and thus these results will have ecological validity for examiners using this measure in clinical practice. Furthermore, there was excellent inter-rater reliability of Bayley-III assessments performed in this study. A potential limitation of the study is that assessments were re-scored using chronological age rather than re-administered. However, this is appropriate as fewer items need to be administered for chronological age assessments and re-scoring is possible without further administration, allowing direct comparison. Moreover, re-administering the test would not provide a direct comparison as different test conditions and practice effects may impact on children's performance. At present there are no long term follow-up data available for this cohort and thus we are unable to assess the relative predictive accuracy of corrected versus chronological age Bayley-III scores for later childhood outcomes. This should be the focus of future studies in this area. Future research should also focus on the difference in test scores obtained using corrected vs chronological age in other clinical populations at risk for poor neurodevelopmental outcomes.

In conclusion, correcting for prematurity when assessing infants born at 32-36 weeks of gestation leads to small but significantly higher mean developmental test scores and a lower prevalence of developmental delay compared to using chronological age. This may affect eligibility for intervention services, even among infants born LMPT. Examiners should be aware of the impact of using corrected versus chronological age on developmental test scores at both an individual and population level when planning clinical evaluations and research studies.

Acknowledgements

The authors thank all the children and families that took part in the LAMBS-II Study. This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under the Research for Patient Benefit Scheme (PB-PG-1010-23049). The views expressed in this paper are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The funder was not involved in study design, data collection, analysis and interpretation of the data, preparation of the manuscript or decision to submit for publication.

Contributors' statements

Shalin Parekh rescored the Bayley-III assessments using chronological age, analyzed the data, drafted the initial manuscript, reviewed and revised the manuscript and approved it for submission. Elaine Boyle was the principal investigator, conceptualized and designed the study, critically reviewed and revised the manuscript and approved it for submission. Samarita Blaggan and Alexa Guy were responsible for data collection and management, reviewed and revised the manuscript and approved it for submission. Bradley Manktelow conceptualized and designed the study, supervised data analyses, critically reviewed and revised the manuscript and approved it for submission. Dieter Wolke and Samantha Johnson conceptualized and designed the study, and critically reviewed and revised the manuscript and approved it for submission.

Conflict of interest statement: the authors have no conflicts of interest to disclose

References

[1] Shapiro-Mendoza CK, Lackritz EM. Epidemiology of late and moderate preterm birth. Seminars in fetal & neonatal medicine. 2012;17:120-5.

[2] McGowan JE, Alderdice FA, Holmes VA, Johnston L. Early childhood development of late-preterm infants: a systematic review. Pediatrics. 2011;127:1111-24.

[3] Guy A, Seaton SE, Boyle EM, Draper ES, Field DJ, Manktelow BN, et al. Infants born late/moderately preterm are at increased risk for a positive autism screen at 2 years of age. The Journal of pediatrics. 2015;166:269-75 e3.

[4] Morse SB, Zheng H, Tang Y, Roth J. Early school-age outcomes of late preterm infants. Pediatrics. 2009;123:e622-e9.

[5] Quigley MA, Poulsen G, Boyle E, Wolke D, Field D, Alfirevic Z, et al. Early term and late preterm birth are associated with poorer school performance at age 5 years: a cohort study. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2012;97:F167-F73.

[6] Johnson S, Evans TA, Draper ES, Field DJ, Manktelow BN, Marlow N, et al. Neurodevelopmental outcomes following late and moderate prematurity: a population-based cohort study. Archives of disease in childhood Fetal and neonatal edition. 2015;100:F301-F8.

[7] Guevara JP, Gerdes M, Localio R, Huang YV, Pinto-Martin J, Minkovitz CS, et al. Effectiveness of developmental screening in an urban setting. Pediatrics. 2013;131:30-7.

[8] Glascoe FP. Screening for developmental and behavioral problems. Ment Retard Dev Disabil Res Rev. 2005;11:173-9.

[9] World Health Organization (WHO) and The The United Nations Children's Fund (UNICEF). Early Childhood Development and Disability : A discussion paper. WHO: World Health Organization and The United Nations Children's Fund (UNICEF); 2013.

[10] United States Department of Education. Early Intervention Program for Infants and Toddlers with Disabilities. In: Education USDo, editor. **76**: Federal Register; 2011. p. 60140–309.

[11] Romeo DM, Di Stefano A, Conversano M, Ricci D, Mazzone D, Romeo MG, et al. Neurodevelopmental outcome at 12 and 18 months in late preterm infants. European journal of paediatric neurology : EJPN : official journal of the European Paediatric Neurology Society. 2010;14:503-7.

[12] Bayley N. Bayley Scales of Infant and Toddler Development (Third Edition). San Antonio: Harcourt Assessment; 2006.

[13] Marlow N. Measuring neurodevelopmental outcome in neonatal trials: a continuing and increasing challenge. Archives of Disease in Childhood - Fetal and Neonatal Edition. 2013;98:F554-F8.

[14] British Association of Perinatal Medicine. Report of a BAPM/RCPCH Working Group Classification of health status at 2 years as a perinatal outcome Version 1.0; 8 January 2008. London2008.

[15] Bayley N. Bayley Scales of Infant Development (Second Edition). San Antonio: Psychological Corporation; 1993.

[16] Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW, Victorian Infant Collaborative G. Underestimation of developmental delay by the new Bayley-III Scale. Archives of pediatrics & adolescent medicine. 2010;164:352-6.

[17] Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. The Journal of pediatrics. 2012;160:553-8.

[18] Vohr BR, Stephens BE, Higgins RD, Bann CM, Hintz SR, Das A, et al. Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. The Journal of pediatrics. 2012;161:222-8 e3.

[19] Acton BV, Biggs WS, Creighton DE, Penner KA, Switzer HN, Thomas JH, et al. Overestimating neurodevelopment using the Bayley-III after early complex cardiac surgery. Pediatrics. 2011;128:e794-800.

[20] Msall ME. Overestimating neuroprotection in congenital heart disease: problems with Bayley III outcomes. Pediatrics. 2011;128:e993-4.

[21] Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? Pediatr Res. 2014;75:670-4.

[22] Blaggan S, Guy A, Boyle EM, Spata E, Manktelow BN, Wolke D, et al. A parent questionnaire for developmental screening in infants born late and moderately preterm. Pediatrics. 2014;134:e55-62.

[23] Boyle JD, Boyle EM. Born just a few weeks early: does it matter? Archives of Disease in Childhood - Fetal and Neonatal Edition. 2013;98:F85-F8.

[24] Romeo DM, Guzzardi S, Ricci D, Cilauro S, Brogna C, Cowan F, et al. Longitudinal cognitive assessment in healthy late preterm infants. Eur J Paediatr Neurol. 2012;16:243-7.

[25] Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? Pediatric Research. 2014;75:670-4.

[26] Moore T, Johnson S, Haider S, Hennessy E, Marlow N. The bayley-III cognitive and language scales: how do scores relate to the bayley ii? Archives of Disease in Childhood. 2011;96:A39-A40.

[27] Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW. Underestimation of developmental delay by the new Bayley-III Scale. Arch Pediatr Adolesc Med. United States2010. p. 352-6.

[28] Wilson-Ching M, Pascoe L, Doyle LW, Anderson PJ. Effects of correcting for prematurity on cognitive test scores in childhood. J Paediatr Child Health. 2014;50:182-8.

[29] Veldhuizen S, Rodriguez C, Wade TJ, Cairney J. Misclassification due to age grouping in measures of child development. Archives of Disease in Childhood. 2014.

[30] Gurka MJ, LoCasale-Crouch J, Blackman JA. Long-term cognition, achievement, socioemotional, and behavioral development of healthy late-preterm infants. Arch Pediatr Adolesc Med. 2010;164:525-32.

Figure legend

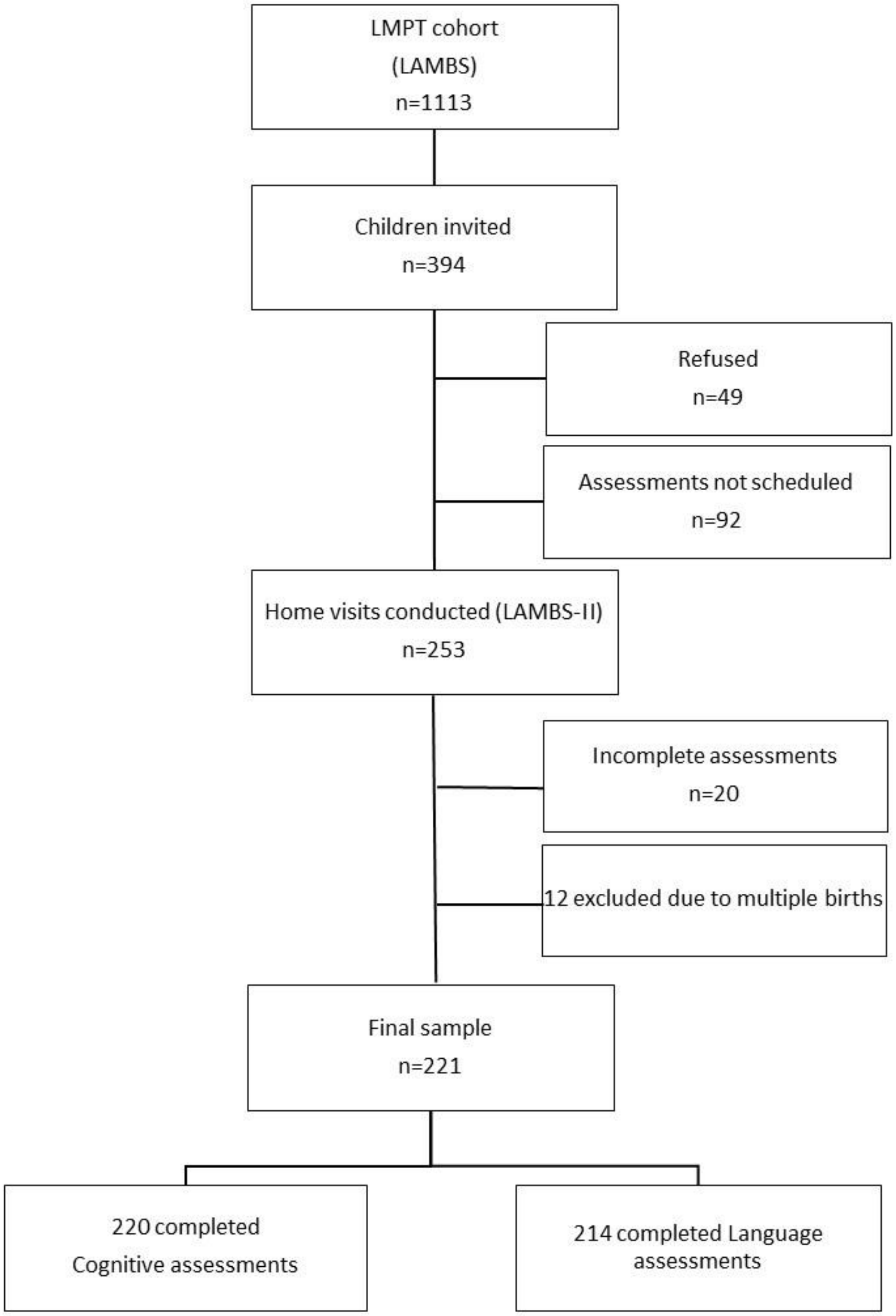

Figure 1. Participant recruitment and follow-up for the LAMBS-II Study.

Table 1: Bayley-III test scores and prevalence of developmental delay derived using corrected and chronological age.

	Corrected age Mean (SD)	**Chronological age Mean (SD)**	**Change in score Mean (95% CI)**	**P**
Mean Cognitive composite score*	96.4 (11.3)	94.3 (11.2)	2.1 (1.6 to 2.5)	<0.001
Mean Language composite scores**	99.8 (17.0)	97.3 (16.1)	2.5 (2.1 to 2.8)	<0.001
	N (%)	**N (%)**		
Cognitive delay (cognitive score <80)*	13 (5.9%)	20 (9.1%)	-	0.016
Language delay (language score <80)**	28 (13.1%)	33 (15.4%)	-	0.063
Any delay (cognitive or language <80)◊	32 (15.0%)	39 (18.3%)	-	0.016

*N=220 for this analysis; ** N=214 for this analysis; ◊ N= 213 for this analysis.

Table 2: The effect of different norms tables and start points on Bayley-III test scores in children born LMPT.

Domain		N	Corrected age Mean (SD)	Chronological age Mean (SD)	Change Score Mean (95% CI)	Change Score Range	P
Effect of change in norms table[1]							
Cognitive	Same norms table	15	98.7 (12.6)	98.7 (12.6)	0	0	-
	Different norms table	205	96.2 (11.2)	94.0 (11.0)	2.2 (1.8 to 2.7)	-10 to +10	<0.001
Language	Same norms table	15	106.9 (17.2)	106.9 (17.2)	0	0	-
	Different norms table	199	99.2 (16.9)	96.6 (15.8)	2.6 (2.3 to 3)	-3 to +9	<0.001
Effect of change in start point[2]							
Cognitive	Same start point	101	98.3 (11.4)	95.3 (10.9)	3.0 (2.5 to 3.5)	0 to +5	<0.001
	Different start point	119	94.8 (11.1)	93.5 (11.4)	1.3 (0.6 to 2)	-10 to +10	<0.001
Language	Same start point	96	101.3 (15.8)	99.2 (15.5)	2.0 (1.6 to 2.5)	0 to +8	<0.001
	Different start point	118	98.6 (17.8)	95.8 (16.5)	2.8 (2.3 to 3.3)	-3 to 9	<0.001

[1]Re-scoring of the test using chronological age may result in the use of a different norms table to that used for corrected age. [2]Applying corrected age may result in the use of a different start point for administering and scoring the test to that used for corrected age.

Appendix 1: The effect of start point on Bayley-III scores

The start point for a Bayley III-assessment is determined by the age of the child at the time of testing. Start points correspond with letters of the alphabet from A-Q. If the child gets the first three items in the age-appropriate start point correct, then the examiner proceeds administering the items in a forward sequence and the child is automatically credited (i.e., scored 1) for all the items prior to that start point, even though the preceding items have not been administered. However, if the child fails any of the first three items, the examiner applies the reversal rule and goes back to the preceding start point to administer the items from there.

We noted that assessments scored using chronological age resulted in higher test scores for some children. For example, assessment at around two years corrected age may be administered using start point M, and using start point N for their corresponding chronological age. If the child starting at N (chronological age) gets the first three items correct, s/he is automatically credited for all the items preceding N and these would be scored 1 when calculating the raw score. However, if the child failed any items prior to start point N, these would be counted as zeros when calculating the raw score from start point M (corrected age), resulting in a lower score for corrected than chronological age. Using the procedure to derive composite score equivalents, a 5-point difference in raw scores can result in a 10-point difference in composite scores. This is shown in Table A1 in which the following scores are derived from the same assessment:

A) Assessed at corrected age = 25 months 6 days; start point = M; raw score = 55; scaled score = 6; composite score = 80.

B) Assessed at chronological age = 26 months 0 days; start point =N; raw score = 60; scaled score = 8; composite score = 90.

.

Table A1: Example of derivation of Bayley-III Cognitive Scale scores for corrected and chronological age

Start point	**Cognitive Scale Item Number**	**Cognitive Scale Raw Score**
Start Point L/M	45	1
Corrected age:	46	1
25 months 6 days	47	1
	48	0
	49	1
	50	0
	51	1
	52	0
	53	0
	54	0
	55	1
Start Point N	56	1
Chronological age:	57	1
26 months 0 days	58	1
	59	0
	60	1
	61	0
	62	1
	63	0
	64	0
	65	0
	66	0
	67	0