Standardisation of the Parent Report of Children's Abilities-Revised (PARCA-R): a norm-referenced assessment of cognitive and language development at 2 years of age.

Samantha Johnson[*,1], Vasiliki Bountziouka[1], Peter Brocklehurst[2], Louise Linsell[3], Neil Marlow[4], Dieter Wolke[5], Bradley N Manktelow[1]

[1]Department of Health Sciences, University of Leicester, George Davies Centre, University Road, Leicester, LE1 7RH, UK; [2]Institute of Applied Health Research, Public Health Building, University of Birmingham, Edgbaston, Birmingham, B15 2TT; [3]National Perinatal Epidemiology Unit, University of Oxford, Richard Doll Building, Roosevelt Drive, Headington, Oxford, OX3 7LF; [4]Institute for Women's Health, University College London, Huntley Street, London, WC1E 6AU; [5]Department of Psychology, University of Warwick, Coventry, CV4 7AL

***Corresponding Author:** Samantha Johnson, Professor of Child Development, Department of Health Sciences, University of Leicester, George Davis Centre, University Road Leicester, LE1 7RH; Tel: +44 (0)116 252 5798; Email: sjj19@leicester.ac.uk

Disclosure statement: The authors have no financial relationships or conflicts of interest to disclose.

Word count: 3500

Number of tables: 5

Keywords: cognitive development; language development; parent report; developmental assessment; screening

Background: To develop age- and sex-standardised scores for the Parent Report of Children's Abilities-Revised (PARCA-R) in order to assess children's cognitive and language development at 24 to 27 months of age.

Methods: Anonymised data from PARCA-R questionnaires completed by parents of 2 year-old children in three previous studies were obtained to form a standardisation sample (n=6402) representative of the UK general population. Anonymised data were obtained from three further studies to assess the external validity (n=709) and clinical validity (n=1456) of the standardised scores. The L(lamda)M(mu)S(sigma) method was used to develop age- and sex-specific standardised scores for three scales (non-verbal cognitive development; language development; total Parent Report Composite (PRC)) for children in four 1-month age bands spanning 23·5 to 27·5 months of age.

Findings: For all PARCA-R scales, mean (SD) standardised scores approximated 100 (15) in both sexes and all age groups. These were independent of socio-economic status. Standardised scores were close to 100 (15) in the external validation sample, demonstrating the validity of the scores. Children born very preterm or with neonatal sepsis had, respectively, standardised scores for the total PRC scale 0.47 SD and 0.73 SD lower on average than the normative mean. These were equivalent to a standardised score of 93 (95% Confidence interval (CI): 91 to 94) and 89 (95% CI: 88 to 91) respectively, thus demonstrating clinical validity.

Interpretation: The PARCA-R provides a norm-referenced, standardised assessment of cognitive and language development at 24-27 months of age. The questionnaire is available non-commercially with translations currently available in 14 languages, thus providing clinicians and researchers with a cost-effective tool for assessing development and identifying children with delay.

Funding: This study was funded by Action Medical Research (Ref: GN2580).

Increased awareness of the long term effects of early life events has highlighted the importance of early childhood development for future health and wellbeing. Intervention offered in the first three years of life, during a period of rapid brain maturation, may have lifelong effects on development.[1] As such, identifying children at risk of developmental delay is crucial to ensure they receive timely intervention to promote their development and reduce the risk of long term disability.[2] Developmental screening in early childhood is therefore a central tenet of healthcare systems worldwide.[2-4]

Valid and reliable norm-referenced measures are needed to assess development and screen for delay in early childhood. This is particularly the case for identifying children with delayed cognitive and language development, for which routine clinical assessments lack sensitivity.[5] The Parent Report of Children's Abilities-Revised (PARCA-R)[6] is a brief questionnaire that takes 10-15 minutes for parents to complete to assess cognitive and language development at 24 months of age. The PARCA-R has concurrent validity with examiner administered developmental tests[6-9] and excellent test-retest reliability.[6] Cut-off scores with diagnostic utility for identifying children with moderate to severe developmental delay have also been derived from studies of clinical populations at high risk for developmental disorders.[6-8 10-12] The PARCA-R is widely used as an outcome measure in observational studies and clinical trials[13-18] and is recommended for routine use in the UK to screen for developmental delay in children born preterm.[19] To date, it has been translated into 14 languages and has been validated for use in Italian[11] and Dutch[9] and in samples in Australia and New Zealand.[8 10]

Although cut-off scores exist for identifying children at risk of delay, these were derived from small studies resulting in wide confidence intervals around cut-points and with cut-off scores that vary widely between different populations.[6 7 11 12] Moreover, these are only

available for identifying children with moderate to severe delay in clinical populations. The utility of the PARCA-R for assessing cognitive and language development of children across the whole developmental spectrum in the general population has not yet been explored. At present, there are no standardised scores for comparing a child's developmental level with that of the norm. This limits the ability of the PARCA-R to identify children with subtle delays and to quantify progress across the full spectrum of development, thus limiting its use as a continuous outcome measure and as a universal screening tool.

The aim of this study was to standardise the PARCA-R to enable professionals to precisely quantify a child's developmental level and identify advanced development or delays of any severity among children in the general population. The objectives of the study were to: (1) develop age- and sex-standardised PARCA-R scores; (2) assess the external validity of the standardised scores; (3) assess the clinical validity of the standardised scores.

Methods

Standardisation sample

Anonymised data from multiple studies were obtained in order to produce a standardisation sample representative of the UK general population in sex, gestational age, multiple birth, ethnicity and socio-economic status. The standardisation sample comprised completed PARCA-R questionnaires for 6196 children assessed at 2 years of age and born between 2010 and 2013 to mothers participating in the INFANT randomised controlled trial of computerised interpretation of fetal heart rate during labour.[13] As this sample was not representative of the gestational age range of births in the general population due to the small proportion (2·5%) of children born preterm ($<37^{+0}$ weeks' gestation), the standardisation sample was supplemented using anonymised data from two other studies. These additional

data included 186 randomly selected 2-year-old children born at 32^{+0} to 36^{+6} weeks' gestation between 2009-2010 whose parents completed the PARCA-R for the Late and Moderately Preterm Birth Study (LAMBS)[14], and 20 children born at 27^{+0} to 31^{+6} weeks of gestation between 2009-2011 whose parents completed the PARCA-R in the Preterm and After (PANDA) Study.[20] Thus the total standardisation sample comprised anonymised data for 6402 children. The distribution of sex, ethnicity (white vs. other) and quintiles of Index of Multiple Deprivation (IMD)[21] scores for the standardisation sample were compared with data from the Office for National Statistics 2011 Census[22], whilst the distribution of preterm and multiple births for the standardisation sample were compared with gestation specific birth data from the Office for National Statistics for England and Wales[23] and from the Information Services Division for Scotland[24] for children born in 2011.

External validation sample

To examine the external validity of the standardised scores, anonymised data from a sample of 709 children born at term ($\geq 37^{+0}$ weeks' gestation) whose parents completed the PARCA-R for the LAMBS Study was used.[14]

Clinical validation samples

Given the well documented association of very preterm birth and neonatal sepsis with adverse neurodevelopmental outcomes, anonymised PARCA-R data from follow-up studies in these populations were used to examine the clinical validity of the standardised scores. First, to examine performance in a very preterm population, PARCA-R data from the remaining 692 children born $<32^{+0}$ weeks of gestation in 2009-2011 in the PANDA study[20] were used. Second, PARCA-R data from 764 children with suspected or proven neonatal sepsis born in 2001-2007 in the UK arm of the International Neonatal Immunotherapy Study

(INIS)[15] were used. Detailed information about these samples has been published previously.[15,20]

Measures

Cognitive and language development were assessed using the PARCA-R questionnaire completed by parents when children were as close as possible to 24 months chronological age (24 months corrected age for children born very preterm in the clinical validation samples). The PARCA-R comprises 34 forced choice items to assess non-verbal cognition, and a 100-word vocabulary checklist and 18 forced-choice items to assess sentence complexity, the latter two of which comprised the UK short form adaptation of the MacArthur Communicative Development Inventories.[25] Development of the PARCA-R has been detailed previously.[6 26] The questionnaire and scoring instructions can be obtained from www.parca-r.info. Summed scores for the two sub-scales of non-verbal cognition (NVC; range 0–34) and language development (range 0–124) were computed, and a total Parent Report Composite (PRC; range 0–158) score was computed by summing these sub-scale scores. Scores for missing questions in the NVC sub-scale were substituted with the average of the score for completed questions if ≤4 questions were missing; if >4 questions were missing the questionnaire was excluded from the analysis. Unchecked or unanswered items for the language sub-scale were set to zero.

Statistical methods

An extension of the LMS (Lambda for skewness, Mu for median, Sigma for the coefficient of variation (CV)) method[27], with the implementation of the Beta-inflated distribution at 0 and 1 that allows kurtosis to be modelled, was used to develop the standardised scores. Raw scores for each of the PARCA-R sub-scales and the total PRC score in the standardisation sample

were regressed against chronological age, separately for males and females. The predicted values of the median along with the CV, the skewness and the kurtosis, were used to calculate the cumulative distribution function and convert the raw PARCA-R scores to z-scores, which are normally distributed with a mean (SD) of 0 (1). The z-scores were then standardised to a mean (SD) of 100 (15) in keeping with standard psychometric tests using the formulae: 100+z-score*15. Norms tables were then produced for obtaining standardised scores from PARCA-R raw scores separately for males and females for 4 age bands (23 months (mo) 16 days (d) to 24mo 15d; 24mo 16d to 25mo 15d; 25mo 16d to 26mo 15d; 26mo 16d to 27mo 15d) using the z-score corresponding to the mid-point of each age band (i.e. 24 months; 25 months; 26 months; 27 months).

To evaluate the external validity of the standardised scores when applied to a different population, the equations derived from the LMS models in the standardisation sample were applied to the PARCA-R data in the external validation sample. Using these equations, z-scores were calculated and rescaled to a mean (SD) of 100 (15). External validity of the standardised scores would be demonstrated if the observed mean (SD) in the external validation sample is close to the expected mean (SD) of 100 (15) (or mean 0 (SD 1) in z-scores). In addition, standardised scores for the PARCA-R data in the external validation sample were assigned using the norms tables (see Tables S8 to S15, supplementary appendix). If the standardised scores from the two methods are similar this will confirm the appropriateness of the norms tables for deriving the standardised scores in clinical practice.

To evaluate the clinical validity of the standardised scores, the equations derived from the LMS models in the standardisation sample were applied to the PARCA-R data in the two clinical validation samples. Given the high risk for neurodevelopmental disabilities among

these populations, the mean standardised scores would be expected to be less than 100. For these samples, standardised scores were derived using corrected rather than chronological age as is conventional when assessing very preterm-born children.

All analyses were performed using R version 3.4.3[28] incorporating the package "gamlss"[29] to perform the LMS modelling and calculate centiles. Detailed information on the methodology for the development of the standardised scores can be found elsewhere (*manuscript currently under review*). The study was approved by the University of Leicester Research Ethics Sub-Committee for Medicine and Biological Sciences (Ref: 13832).

Role of funding source

This study was funded by Action Medical Research (Ref: GN2580). The study sponsor had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Representativeness of the standardisation sample

Characteristics of the standardisation sample and that of the general population are shown in Table 1. Children were assessed at a mean age of 25 months 1 day (range: 23mo 16d to 27mo 15d). The standardisation sample was representative of the general population in terms of sex, IMD, and multiple births, however it was not possible to determine representativeness of ethnicity as this was not recorded for 14% of children. Although the proportion of very preterm born children in the standardisation sample was less than in the general population

overall, this was only amongst the youngest age groups; the proportion in the oldest age groups matched that of the general population (Table S1, supplementary appendix).

INSERT TABLE 1

Development of the standardised scores

For the total standardisation sample, sex-specific z-scores calculated from the LMS models approximated 0 (1) with 95% of data points within the [-2, +2] z-score range for each scale, indicating a good model fit. In addition, the fit was good in each age group as the mean z-scores were very close to the expected mean (SD) of 0 (1), although the spread of the data was slightly wider than expected in the older age groups (Table 2).

INSERT TABLE 2

For all scales, in both sexes and all age groups, mean (SD) predicted raw scores were very close to the mean (SD) raw scores observed in the data, and the mean (SD) standardised scores were all close to 100 (15) (Table 3). Although mean standardised scores tended to increase with lower deprivation in both sexes, differences between IMD quintiles were not statistically significant (Table S2, supplementary appendix). The norms tables for deriving standardised scores from raw scores for use in clinical practice are provided in Tables S8 to S15 (supplementary appendix).

Acknowledging that the proportion of preterm born children in the standardisation sample was approximately 1-3% less than in the UK population, depending on sex and age group, we also calculated the fitted centiles using corrected age instead of chronological age for these children. This resulted in a shift in the distribution of very preterm children which then matched the distribution in the UK population, however this did not affect the fitted centiles

which remained similar to those calculated using chronological age (data not shown), thus confirming validity of the standardised scores derived using chronological age.

INSERT TABLE 3

External validity of the standardised scores

The external validity of the standardised scores was assessed in 709 term-born children at a mean age of 24 months 19 days (range: 23m 16d to 27mo 15d). This sample predominantly comprised children of white ethnic background (83%) who lived in the least deprived areas in the UK (44% in the 4th and 5th IMD quintile) (Table S3, supplementary appendix).

The predicted raw scores using the equations derived from the standardisation sample were very similar to those observed in the external validation sample, demonstrating the external validity of the scores (Table 4). Moreover, the mean (SD) standardised scores approximated 100 (15) for all scales and for the youngest age groups, whereas they show greater variation in the oldest age groups. The small differences in scores in the validation sample were expected and can be attributed to the differences in characteristics between the standardisation and validation samples and to differences in the distribution of the raw scores between the two samples.

Moreover, the standardised scores generated from applying the equations derived from the standardisation process to the validation sample (as described above) matched very closely the standardised scores that were derived from the more simple process of using the norms tables (Tables S8 to S15, supplementary appendix). This demonstrates both the external validity of the scores and the accuracy of the norms tables for deriving the standardised scores in clinical practice. Any observed differences between the scores generated using the

two methods were of minor importance as they did not exceed 1 point and may be attributed to the small differences in the raw scores between the two samples (Tables S4 and S5, supplementary appendix).

INSERT TABLE 4

For 11% of males and 12% of females in the external validation sample, the main language spoken at home was not English. These children had similar standardised language and PRC scores to English speaking children, and slightly higher non-verbal cognitive scores, however the differences were not statistically significant (Table S6, supplementary appendix).

Clinical validity of the standardised scores

Clinical validity of the standardised scores was assessed in 692 children born very preterm ($<32^{+0}$ weeks' gestation) assessed at a mean corrected age of 24 months 9 days (range: 23mo 16d to 27mo 14d) in the PANDA Study[20] and 764 children with neonatal sepsis assessed at a mean corrected age of 24 months 8 days (range: 23mo 16d to 27mo 15d) in the INIS Study.[15] Characteristics of these participants are shown in Table S7 (supplementary appendix).

Applying the equations derived from the LMS models in the standardisation sample to these clinical populations resulted in mean standardised scores that were significantly lower than the normative mean of 100 on all PARCA-R scales (Table 5). The mean PRC standardised score in the total PANDA sample was 7 (95% confidence interval (CI): -9 to -5) points lower than the normative mean, and in the total INIS sample was 10 (95% CI: -12 to -8) points lower. As shown in Table 5, the magnitude of the deficit differed by sex and age and between the two clinical populations.

INSERT TABLE 5

Discussion

The newly developed standardised scores presented here make the PARCA-R a norm-referenced measure that can be used to assess children's cognitive and language development, separately and combined, at 24 to 27 months of age. As the validity and reliability of the PARCA-R have previously been demonstrated[6 8 9 11], this study significantly advances the utility of the questionnaire as both a clinical assessment and as an outcome measure for research.

Previous validation studies have derived cut-off scores for identifying children at risk for developmental delay and have reported high levels of sensitivity (> 80%).[6-8 10 12] However, these studies have only developed cut-off scores for use in preterm populations and for identifying children at risk of moderate to severe developmental delay (scores > 2 SD below the mean on examiner administered tests), and in cognitive and language development combined. In addition, the relatively small samples sizes and the frequently reported low positive predictive values, which result in high rates of false positive screens or over-referrals, have garnered concerns regarding the accuracy of these cut-off scores for use in clinical practice.

Our development of the standardised scores addresses these issues allowing clinicians and researchers to precisely measure a child's developmental level and classify delay of any severity relative to the norm. Commensurate with other psychometric tests, standardised scores were developed with a normative mean of 100 and SD of 15 and allow an assessment of development ranging from < -3SD to > +3SD. Professionals can therefore use the PARCA-R to aid in identifying children with either advanced or delayed cognitive and/or language development. This study also significantly improves the utility of the PARCA-R for

research by providing a continuous outcome measure that can be used to quantify development and detect small differences in development between individuals or groups of children.

A developmental assessment carried out by a trained professional is typically considered the gold standard. However, the extensive resources required to administer such tests frequently prohibit their use on a large scale, particularly in low and middle income countries. As the PARCA-R is freely available for use, this provides a cost-effective standardised measure of children's development and is thus a valid and reliable alternative to examiner administered tests. Indeed, in a recent systematic review the PARCA-R was identified as the best developmental screening tool for use with preterm children and was recommended for use to screen for developmental delay at 2 years corrected age in children born preterm.[19] As development of the standardised scores extends the use of the PARCA-R as an assessment for all children in the general population, it could therefore be used for developmental surveillance in other clinical populations and in universal screening programmes. We are not aware of other parental assessments of cognitive and language development that have been age-standardised, not least in such a large sample. Moreover, the PARCA-R standardisation sample of 6402 children spanning 4 months of age far exceeds that of the current (3rd) edition of the Bayley Scales of Infant and Toddler Development which included up to 200 children across the same age range in the standardisation sample.[30] We have also developed an online version of the questionnaire and a pre-programmed calculator for deriving the standardised scores (see www.parca-r.info).

The strengths of this study lie in the development of an age- and sex- standardised parent report measure of cognitive and language development for use at two years of age, and the

large standardisation sample on which the normative data are based. Although data from three different studies had to be collated to form the standardisation sample, ultimately this was representative of the UK population in terms of sex, socio-economic status (IMD) and multiple births. We acknowledge that the ratio of children with white/other ethnic background was different compared with the UK population. However, there was no difference in mean standardised scores between children who did and did not come from homes in which English was the first language within the external validation sample, suggesting that the under-representation of children from non-white backgrounds may not have affected development of the standardised scores. In addition, despite collating data from multiple studies, the proportion of children born very preterm (<32 weeks' gestation) was still under-represented in the two youngest age groups in the standardisation sample. This arose as a result of the PARCA-R being completed at 24 months corrected age for very preterm born children in the original studies, resulting in these children having generally higher chronological age at the time of their assessment. Nonetheless, whilst a sensitivity analysis conducted using corrected rather than chronological improved the distribution of preterm born children in the standardisation sample, it did not alter the fitted centiles and thus the standardised scores, therefore confirming the robustness of our results. As the PARCA-R was designed to assess development *at* 24 months of age, the age range within which it can be used is limited compared with other measures as standardised scores can only be obtained for children aged 24 to 27 months at the time of assessment. It is also important to note that standardised scores should be interpreted with caution for children aged 26 to 27 months given the smaller number in these groups in the standardisation sample. Finally, although the PARCA-R has been translated into 14 other languages to date (see www.parca-r.info), standardised scores have only been developed for the original English version in the UK population. Application of the standardised scores to other languages and populations could

be explored as the subject of future research. Moreover, whilst the concurrent validity of the PARCA-R has been established, a recommendation for research into the predictive validity of the PARCA-R for identifying children at risk for later learning difficulties and special educational has been made.[19]

In summary, the PARCA-R provides an age- and sex-specific standardised assessment of children's cognitive and language development at two years of age and can be used as a cost-efficient developmental assessment for clinical and research purposes.

Data sharing

Individual participant data are not available for sharing. This study comprises secondary analysis of data from multiple sources, therefore the University of Leicester are not the custodians of the data.

Conflict of Interest

We declare that we have no conflicts of interest.

Contributors

Samantha Johnson contributed to study design and data interpretation, drafted the first version of the manuscript, critically revised the manuscript for important intellectual content, and approved the final version for submission.

Vasiliki Bountziouka analysed and interpreted the data, drafted the first version of the manuscript, critically revised the manuscript for important intellectual content, and approved the final version for submission.

Peter Brocklehurst contributed to study design and data interpretation, critically revised the manuscript for important intellectual content, and approved the final version for submission.

Louise Linsell contributed to study design and data interpretation, critically revised the manuscript for important intellectual content, and approved the final version for submission.

Neil Marlow contributed to study design and data interpretation, critically revised the manuscript for important intellectual content, and approved the final version for submission.

Dieter Wolke contributed to study design and data interpretation, critically revised the manuscript for important intellectual content, and approved the final version for submission.

Bradley Manktelow contributed to study design and data interpretation, critically revised the manuscript for important intellectual content, and approved the final version for submission.

References

1. Black MM, Walker SP, Fernald LC, et al. Early childhood development coming of age: science through the life course. Lancet 2017;**389**(10064):77-90.
2. American Academy of Pediatrics Committee on Children With Disabilities. Developmental surveillance and screening of infants and young children. Pediatrics 2001;**108**:192-95.
3. Richter LM, Daelmans B, Lombardi J, et al. Investing in the foundation of sustainable development: pathways to scale up for early childhood development. Lancet 2017;**389**(10064):103-18.
4. Department of Health. Healthy Child Programme: Pregnancy and the first five years of life. London, 2009.
5. Wong HS, Cowan FM, Modi N, et al. Validity of neurodevelopmental outcomes of children born very preterm assessed during routine clinical follow-up in England. Archives of disease in childhood Fetal and neonatal edition 2017.
6. Johnson S, Marlow N, Wolke D, et al. Validation of a parent report measure of cognitive development in very preterm infants. Developmental Medicine and Child Neurology 2004;**46**:389-97.
7. Johnson S, Wolke D, Marlow N, et al. Developmental assessment of preterm infants at 2 years: validity of parent reports. Dev Med Child Neurol 2008;**50**(1):58-62.
8. Martin AJ, Darlow BA, Salt A, et al. Performance of the Parent Report of Children's Abilities-Revised (PARCA-R) versus the Bayley Scales of Infant Development III. Arch Dis Child 2013;**98**(12):955-8.
9. Vanhaesebrouck S, Theyskens C, Vanhole C, et al. Cognitive assessment of very low birth weight infants using the Dutch version of the PARCA-R parent questionnaire. Early Hum Dev 2014;**90**(12):897-900.

10. Martin AJ, Darlow BA, Salt A, et al. Identification of infants with major cognitive delay using parental report. Dev Med Child Neurol 2012;**54**(3):254-9.

11. Cuttini M, Ferrante P, Mirante N, et al. Cognitive assessment of very preterm infants at 2-year corrected age: performance of the Italian version of the PARCA-R parent questionnaire. Early Hum Dev 2012;**88**(3):159-63.

12. Blaggan S, Guy A, Boyle EM, et al. A parent questionnaire for developmental screening in infants born late and moderately preterm. Pediatrics 2014;**134**(1):e55-62.

13. Group IC. Computerised interpretation of fetal heart rate during labour (INFANT): a randomised controlled trial. Lancet 2017;**389**(10080):1719-29.

14. Johnson S, Evans TA, Draper ES, 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**(4):F301-F08.

15. Group IC, Brocklehurst P, Farrell B, et al. Treatment of neonatal sepsis with intravenous immune globulin. N Engl J Med 2011;**365**(13):1201-11.

16. Australia B-I, United Kingdom Collaborative Groups, Tarnow-Mordi W, et al. Outcomes of Two Trials of Oxygen-Saturation Targets in Preterm Infants. N Engl J Med 2016;**374**(8):749-60.

17. Marlow N, Greenough A, Peacock JL, et al. Randomised trial of high frequency oscillatory ventilation or conventional ventilation in babies of gestational age 28 weeks or less: respiratory and neurological outcomes at 2 years. Archives of disease in childhood Fetal and neonatal edition 2006;**91**(5):F320-6.

18. Abbott J, Berrington J, Bowler U, et al. The Speed of Increasing milk Feeds: a randomised controlled trial. BMC pediatrics 2017;**17**(1):39.

19. National Institue for Health and Care Excellence. Developmental follow-up of children and young people born preterm. Full Guideline. NG72, 2017.

20. Field D, Spata E, Davies T, et al. Evaluation of the use of a parent questionnaire to provide later health status data: the PANDA study. Archives of disease in childhood Fetal and neonatal edition 2016;**101**(4):F304-8.

21. Government DoCaL. The English Indices of Deprivation 2010, 2011.

22. Statistics OfN. Office for National Statistics Census 2011 Table Links. Secondary Office for National Statistics Census 2011 Table Links 2011. https://www.nomisweb.co.uk/census/2011/detailed_characteristics.

23. Statistics OfN. Gestation-specific Infant Mortality in England and Wales: 2011. Secondary Gestation-specific Infant Mortality in England and Wales: 2011 2011. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/bulletins/pregnancyandethnicfactorsinfluencingbirthsandinfantmortality/2013-10-10#background-notes.

24. Scotland ISDNNS. Data Tables; Maternity and Births. Secondary Data Tables; Maternity and Births 2016. https://www.isdscotland.org/Health-Topics/Maternity-and-Births/Publications/data-tables.asp.

25. Fenson L, Pethick S, Renda C, et al. Short-form versions of the MacArthur Communicative Development Inventories. Applied Psycholinguistics 2000;**21**:85-115.

26. Saudino K, Dale PS, Oliver B, et al. The validity of parentbased assessment of the cognitive abilities of 2-year-olds. British Journal of Developmental Psychology 1998;**16**:349-63.

27. Cole TJ, Green P, J., . Smoothing reference centile curves: The LMS method and penalized likelihood. . Statistics in Medicine 1992;**11**(10):1305-19.

28. Team RC. *R: A language and environment for statistical computing [program]*. Vienna, Austria: R Foundation for Statistical Computing, 2017.

29. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc C-Appl 2005;**54**:507-44.

30. Bayley N. *Bayley Scales of Infant and Toddler Development, 3rd Edition.*: Harcourt Assessment, Inc, 2006.

Table 1: Distribution of socio-demographic and birth characteristics in the standardisation sample and the general population[a].

	Standardisation sample	General population	Difference % (95% CI) Standardisation-UK
Child's sex, n (%)	6402	757686	
Male	3321 (51·9)	386833 (51·1)	0·8 (-0·4; 2·0)
Female	3081 (48·1)	370853 (48·9)	-0·8 (-2·0; 0·4)
Ethnic background, n (%)	6402	3789571[b]	
White	5009 (78·2)	2956304 (78)	0·2 (-0·8; 1·2)
Other	508 (7·9)	833267 (22)	-14·1 (-14·7; -13·4)
Missing	885 (13·8)	-	13·8 (13·0; 14·7)
IMD Quintiles, n (%)	6402	8012452	
1st Q (most deprived)	1651 (25·8)	2070160 (25·8)	0·0 (-1·1; 1·0)
2nd Q	1284 (20·1)	1701987 (21·2)	-1·2 (-2·2; -0·2)
3rd Q	1081 (16·9)	1478481 (18·5)	-1·6 (-2·5; -0·6)
4th Q	1217 (19·0)	1370451 (17·1)	1·9 (0·9; 2·9)
5th Q (least deprived)	1078 (16·8)	1391373 (17·4)	-0·5 (-1·4; 0·4)
Missing	91 (1·4)	-	1·4 (1·1; 1·7)
Preterm birth, n (%)	6402	772542	
Full term (37-42 weeks)	6039 (94·3)	717040 (92·8)	1·5 (0·9; 2·1)
Late & moderately preterm (32-36 weeks)	343 (5·4)	45889 (5·9)	-0·6 (-1·1; 0·0)
Very preterm (<32 weeks)	20 (0·3)	9613 (1·2)	-0·9 (-1·1; -0·8)
Multiple birth, n (%)	6402	772750	
Singleton birth	6234 (97·4)	748281 (96·8)	0·5 (0·1; 0·9)
Multiple birth	168 (2·6)	24469 (3·2)	-0·5 (-0·9; -0·1)

IMD: Index of Multiple Deprivation. [a] Distribution of sex, ethnicity and IMD quintiles obtained from the Office for National Statistics 2011 Census[22]; Distribution of preterm and multiple births obtained from the Office for National Statistics for England and Wales[23] and from the Information Services Division for Scotland[24] for children born in 2011. [b]The 2011 Census refers to children aged 0-4 years.

Table 2: Mean (SD) of z-scores derived from modelling the PARCA-R non-verbal cognitive and language scales and the total PRC scale in the standardisation sample (n=6402) for males and females overall and by age group.

	n	Non-verbal cognition	Language development	Parent Report Composite
Males				
Total	3321	0·003 (0·99)	-0·006 (1·00)	-0·008 (1·00)
Age group, months				
23mo 16d to 24mo 15d	1136	-0·03 (0·94)	-0·03 (0·98)	-0·03 (0·99)
24mo 16d to 25mo 15d	1395	0·02 (1·01)	0·001 (0·99)	0·002 (0·99)
25mo 16d to 26mo 15d	570	0·04 (1·01)	0·04 (1·04)	0·02 (1·03)
26mo 16d to 27mo 15d	220	-0·05 (0·98)	-0·02 (1·08)	-0·05 (1·05)
Females				
Total	3081	-0·005 (0·99)	0·0003 (1·00)	-0·003 (1·00)
Age group, months				
23mo 16d to 24mo 15d	1049	-0·02 (0·98)	0·02 (1·00)	0·001 (1·00)
24mo 16d to 25mo 15d	1329	-0·009 (0·99)	0·0005 (0·98)	0·002 (0·98)
25mo 16d to 26mo 15d	511	0·03 (1·02)	-0·04 (1·03)	-0·02 (1·05)
26mo 16d to 27mo 15d	192	0·02 (1·01)	0·002 (1·04)	-0·03 (1·05)

Table 3: Mean (SD) PARCA-R raw and standardised scores for males and females by age group in the standardisation sample (n=6402).

		Non-verbal cognitive			Language development			Parent Report Composite		
		Raw score		Standardised	Raw score		Standardised	Raw score		Standardised
	n	Observed	Predicted	score	Observed	Predicted	score	Observed	Predicted	score
Males *(n=3321)*										
Age group, months										
23mo 16d to 24mo 15d	1136	26·9 (3·5)	26·6 (6·9)	100 (14)	57 (31)	58 (29)	100 (15)	84 (33)	85 (36)	100 (15)
24mo 16d to 25mo 15d	1395	27·4 (3·7)	27·3 (7·1)	100 (15)	62 (31)	62 (31)	100 (15)	89 (33)	90 (38)	100 (15)
25mo 16d to 26mo 15d	570	27·8 (3·6)	27·7 (7·2)	101 (15)	66 (32)	65 (33)	100 (15)	93 (33)	94 (40)	100 (15)
26mo 16d to 27mo 15d	220	27·8 (3·5)	27·8 (7·2)	99 (15)	65 (33)	67 (34)	100 (16)	93 (34)	95 (41)	99 (16)
Females *(n=3081)*										
Age group, months										
23mo 16d to 24mo 15d	1049	28·1 (3·3)	27·8 (6·9)	100 (15)	73 (30)	71 (35)	100 (15)	102 (32)	100 (43)	100 (15)
24mo 16d to 25mo 15d	1329	28·4 (3·3)	28·2 (7·0)	100 (15)	77 (29)	77 (38)	100 (15)	105 (31)	105 (45)	100 (15)
25mo 16d to 26mo 15d	511	28·8 (3·3)	28·6 (7·1)	100 (15)	80 (31)	80 (39)	100 (16)	109 (32)	110 (47)	100 (16)
26mo 16d to 27mo 15d	192	29·1 (3·0)	28·9 (7·2)	100 (15)	84 (31)	83 (41)	100 (16)	113 (32)	115 (49)	100 (16)

Table 4: Mean (SD) PARCA-R raw and standardised scores for males and females by age group, in the external validation sample (n=709).

		Non-verbal cognitive			Language development			Parent Report Composite		
		Raw score		Standardised	Raw score		Standardised	Raw score		Standardised
	n	Observed	Predicted	score	Observed	Predicted	score	Observed	Predicted	score
Males										
Total	350	27·6 (3·6)	27·0 (7·0)	102 (15)	60 (32)	60 (30)	100 (16)	88 (33)	88 (38)	100 (15)
Age group, months										
23mo 16d to 24mo 15d	184	27·4 (3·3)	26·7 (6·9)	102 (14)	58 (32)	58 (29)	100 (16)	86 (33)	85 (37)	101 (15)
24mo 16d to 25mo 15d	128	27·8 (3·7)	27·2 (7·0)	102 (16)	61 (31)	62 (31)	100 (15)	89 (33)	89 (38)	100 (15)
25mo 16d to 26mo 15d	25	27·4 (4·3)	27·6 (7·1)	99 (18)	58 (31)	65 (33)	98 (17)	86 (33)	93 (40)	97 (16)
26mo 16d to 27mo 15d	13	28·2 (4·3)	27·8 (7·2)	102 (19)	79 (27)	67 (34)	106 (13)	107 (29)	95 (41)	105 (14)
Females										
Total	359	28·4 (3·7)	28·6 (7·1)	101 (16)	73 (32)	74 (36)	99 (16)	101 (33)	103 (44)	99 (16)
Age group, months										
23mo 16d to 24mo 15d	198	28·2 (4·0)	27·9 (6·9)	101 (16)	70 (31)	72 (35)	99 (16)	98 (33)	101 (43)	99 (16)
24mo 16d to 25mo 15d	125	28·5 (3·3)	28·2 (7·0)	100 (15)	77 (32)	76 (37)	101 (16)	105 (33)	105 (45)	101 (16)
25mo 16d to 26mo 15d	27	29·3 (3·5)	28·6 (7·1)	103 (18)	72 (31)	80 (40)	95 (17)	101 (33)	110 (47)	96 (16)
26mo 16d to 27mo 15d	9	28·0 (2·9)	28·9 (7·1)	94 (14)	76 (36)	83 (41)	97 (17)	104 (37)	115 (49)	95 (17)

Table 5: PARCA-R standardised scores for males and females in the clinical validation samples.

		Standardised score; mean (95% CI)		
		Non-verbal cognition	Language development	Parent Report Composite
Males				
PANDA				
Total	342	91 (89; 93)	94 (93; 96)	94 (92; 96)
Age group, months				
23mo 16d to 24mo 15d	234	92 (90; 94)	96 (94; 98)	96 (93; 98)
24mo 16d to 25mo 15d	91	90 (86; 94)	91 (88; 95)	90 (86; 94)
25mo 16d to 26mo 15d	14	81 (72; 91)	88 (78; 98)	88 (78; 98)
26mo 16d to 27mo 15d	3	89 (69; 109)	89 (68; 110)	89 (67; 110)
INIS				
Total	437	87 (85; 88)	89 (87; 91)	89 (87; 90)
Age group, months				
23mo 16d to 24mo 15d	309	87 (85; 89)	89 (86; 91)	89 (86; 91)
24mo 16d to 25mo 15d	88	87 (83; 91)	90 (86; 94)	90 (86; 94)
25mo 16d to 26mo 15d	23	81 (73; 89)	88 (80; 96)	87 (79; 94)
26mo 16d to 27mo 15d	17	89 (79; 98)	88 (79; 98)	89 (80; 97)
Females				
PANDA				
Total	350	91 (89; 93)	92 (90; 94)	92 (90; 94)
Age group, months				
23mo 16d to 24mo 15d	232	92 (89; 94)	93 (91; 96)	93 (91; 95)
24mo 16d to 25mo 15d	96	89 (85; 93)	89 (86; 92)	89 (85; 92)
25mo 16d to 26mo 15d	18	89 (80; 97)	88 (81; 96)	88 (80; 95)
26mo 16d to 27mo 15d	4	88 (70; 106)	104 (88; 120)	101 (86; 117)
INIS				
Total	327	87 (85; 89)	91 (89; 93)	90 (89; 92)
Age group, months				
23mo 16d to 24mo 15d	237	90 (87; 92)	93 (91; 96)	92 (90; 95)
24mo 16d to 25mo 15d	62	82 (76; 88)	86 (81; 91)	86 (81; 91)
25mo 16d to 26mo 15d	17	79 (68; 90)	82 (72; 91)	81 (72; 91)
26mo 16d to 27mo 15d	11	78 (64; 92)	89 (77; 101)	85 (73; 97)