Development and validation of a parent report measure for detection of cognitive delay in infancy

Development and Validation of a Parent Report Measure for Detection of Cognitive Delay in Infancy

Graham Schafer[1], DPhil, CPsychol, CSci, AFBPsS; Lucia Genesoni[1], PhD; Greg Boden[2], MD MRCP; Helen Doll[3], MSc DPhil, Rosamond A K Jones[4], MD, FRCPCH; Ron Gray[5], MB ChB FRCPsych FFPH; Eleri Adams[6] MB BS, FRCPCH; Ros Jefferson[7], MB PhD MRCP FRCPCH

Affiliations: [1]School of Psychology & Clinical Language Sciences, University of Reading, UK; [2]Department of Paediatrics, Greenacres Hospital, Port Elizabeth, RSA; [3]Faculty of Medicine and Health Sciences, Norwich Medical School, University of East Anglia, Norwich, UK; [4]Department of Paediatrics, Wexham Park Hospital, Slough, UK; [5]National Perinatal Epidemiology Unit, University of Oxford, Oxford, UK; [6]Newborn Intensive Care Unit, Oxford University Hospitals, Oxford, UK; [7]Dingley Specialist Children's Centre, Royal Berkshire NHS Foundation Trust, Reading, UK

Address correspondence to: Graham Schafer, School of Psychology & Clinical Language Sciences, University of Reading, Harry Pitt Building, Earley Gate, Reading, UK, RG6 7BE; +44 (0)118 378 6221, g.w.schafer@reading.ac.uk.

Word count: 3487

Abstract

Aim

To develop a brief, parent-completed instrument ('ERIC') for detection of cognitive delay in 10-24 month-olds born preterm, or with low birth weight, or with perinatal complications, and to establish its diagnostic properties.

Method

Scores were collected from parents of 317 children meeting ≥1 inclusion criteria (birth weight <1500g; gestational age <34 completed weeks; 5-minute Apgar <7; presence of hypoxic-ischemic encephalopathy) and meeting no exclusion criteria. Children were assessed for cognitive delay using a criterion score on the Bayley Scales of Infant and Toddler Development Cognitive Scale III[1] <80. Items were retained according to their individual associations with delay. Sensitivity, specificity, Positive and Negative Predictive Values were estimated and a truncated ERIC was developed for use <14 months.

Results

ERIC detected 17 out of 18 delayed children in the sample, with 94.4% sensitivity (95% CI [confidence interval] 83.9-100%), 76.9% specificity (72.1-81.7%), 19.8% positive predictive value (11.4-28.2%); 99.6% negative predictive value (98.7-100%); 4.09 likelihood ratio positive; and 0.07 likelihood ratio negative; the associated Area under the Curve was .909 (.829-.960).

Interpretation

ERIC has potential value as a quickly-administered diagnostic instrument for the absence of early cognitive delay in preterm or premature infants of 10-24 months, and as a screen for cognitive delay. Further research may be needed before ERIC can be recommended for wide-scale use.

Short title: Parental report on infant cognitive delay

What this paper adds

- Preliminary report of an instrument designed for premature/low birth weight infants.
- Early Report by Infant Caregivers ('ERIC') - uses parental report to detect cognitive delay.
- ERIC is quick and cheap to administer.
- ERIC has potential to be used to reassure parents (no delay), or as a screen (positive cases).
- ERIC is available for use or for further research, from www.xxx.xxx.ac.uk/xxxxxx [website to be added once the paper is accepted; we do not wish to place ERIC in the public domain in advance].

Development and Validation of a Parent Report Measure for Detection of Cognitive Delay in Infancy

Prematurity and low birth weight are associated with adverse neurodevelopmental outcomes[2]. Extreme prematurity (<28 weeks), and extremely low birth weight (<1000g) are well-documented risk factors for developmental delay[3]. There is also increasing evidence of delay in those born late preterm and early term[4]. Many professional bodies recommend neurodevelopmental follow-up of premature and low birth weight children. Specifically, the British Association of Perinatal Medicine (BAPM) recommends a 2-year neurodevelopmental follow up of children with gestational age (GA) <32 weeks or birth weight (BW) <1500g[5]. Similarly, the American Academy of Pediatrics (AAP) mandates follow-up in cases of extreme prematurity, and recommends surveillance in many other neonatal situations[6]. In practice, however, it is not always possible to follow up all children at risk. Efficient methods of assessment are urgently needed.

Even the widely-used 32 week GA milestone may be insufficiently inclusive as a cut-point for neurodevelopmental follow-up. Higher survival of extremely preterm infants[3] are accompanied by increasing rates of late preterm birth (taken here to mean $34 \leq GA \leq 36$ completed weeks: ~75% of the US preterm population). Two-year-old late preterm infants show increased odds of severe (1.52) and mild (1.43) mental developmental delay[4], whilst adverse neurodevelopmental outcomes in late preterm infant population have also been reported at school age[7, 8]. Because these infants have little or no developmental follow-up, an efficient early cognitive assessment for at-risk infants is desirable and would redress missed opportunities for referral [9, 10] and for early intervention. Indeed, cognitive assessment in infancy can predict later educational performance, particularly in less able groups[11]. Therefore, more attention to also late preterm children is important in the context of clinical care decisions[12]. (NB. By 'at-risk' we refer throughout to prematurity, low birth weight or perinatal complications, rather than to factors relating to genetic, social, or other risks.)

Parental report measures of early development are efficient, and comparable, even preferable, to professionally-administered tests[13]. The only specifically cognitive instrument (the PARCA-r[14]) to be recently validated in at-risk infants, applies only at 2 years. Our group has previously published a parent administered Cognitive Development Questionnaire (CDQ)[15] for use with typically-developing 10-24 month olds. We describe here its further development and validation in revised form as the Early Report by Infant Caregivers ('ERIC'). ERIC is designed as a parental screening instrument for infants at risk for cognitive delay by reason of prematurity, low birth weight, or perinatal complications, and aims to discriminate between individuals with and without cognitive delay, at any time between 10 and 24 months corrected age. It is distinct from other parental report measures in including both a self-report questionnaire and of a series of developmental tasks for the carer to complete with the child. Our reference standard for a diagnosis of delay was the widely-adopted score <80[16] in the Cognitive Scale of the Bayley Scales of Infant and Toddler Development-III[1] ('Bayley-III(C)'), although a range of Bayley-III(C) thresholds was explored.

Method

Development of the Early Report by Infant Caregivers (ERIC)

Candidate items for a test battery were selected from the Cognitive Development Questionnaire (CDQ), based on clinical expertise. CDQ items had been based on existing instruments[17, 18], modified for use by parents, and designed to assess a range of cognitive abilities including learning and memory, problem solving, and conceptual development. The initial item battery consisted of two sections: (1) 18 short games using small household objects or simple toys, yielding 33 scorable items; (2) 16 yes/no questions about the child's everyday competences. Our plan was to use statistical methods to generate a short and therefore user-friendly test by removing items not individually associated with cognitive delay according to Bayley-III (C), and by allowing early stopping of the assessment (by completion of only a subset of the full ERIC) for the youngest children.

Participants and Sampling Plan

Children between the ages of 10 and 24 months (corrected for prematurity) were recruited from three English hospitals between August 2009 and March 2012. Inclusion criteria were gestational age < 34 completed weeks; birth weight <1500g; 5 minute Apgar <7; diagnosis of hypoxic-ischemic encephalopathy; exclusion criteria were life-threatening or life-limiting illness; suspected congenital abnormality or inherited neuromuscular condition; failed auditory screen at birth; severe motoric dysfunction; profound, sustained intellectual impairment; visual impairment requiring appointment of preschool teacher; no English-speaking principal caregiver. Inclusion criteria were based on the BAPM guidance for follow-up, broadened to include <34-weekers[12]. It was anticipated that those excluded would already be receiving more detailed follow-up for clinical reasons.

Based on an assumed prevalence of cognitive delay of 15%, and values of sensitivity and specificity of 80%[14], the target sample size was 300 children, to give ±12% precision around the estimate of sensitivity (i.e., 95% CI of 68%-92%). To have a sample distributed evenly across the age range, we intended to recruit at least 20 children of each monthly age, giving n~300. If the prevalence was lower, precision would be correspondingly reduced (to ±15% at 10% prevalence, and ±20% at 5% prevalence).

Recruitment and study procedure

Ethical approval was granted by the Berkshire Research Ethics Committee, and the Research and Development departments of each hospital. Details were retrieved from neonatal records of all babies meeting any of the inclusion criteria and none of the exclusion criteria. Eligible families were provided by the medical staff with written details of the study and with an opt-out form in case they didn't want to be contacted at all. Unless the families completed the first opt-out form, they were contacted by the researchers as children reached an eligible age. Those interested in participation were sent an Information Sheet, Consent Form, and the initial item battery. An appointment was subsequently made for assessment using the Bayley-III(C)[1]. This assessment took place at the hospital or university, or in the home. Assessments were performed by three trained assessors blind to parental test responses. The test questions were completed by parents during the week prior to the Bayley-III(C) assessment. Parents were asked to administer all items, but advised not to worry if their child did not manage them all. Parents were permitted to spread the test over a week. Travel costs were reimbursed

and a small incentive given (shopping voucher). Families were subsequently told their child's Bayley-III(C) score; if <80 the family doctor was informed and parents contacted for onward referral.

To assess test-retest reliability of the test battery, 5% of parents were given a second complete set of test questions, for completion within two weeks of the first assessment. To establish inter-rater reliability of Bayley-III(C) scoring, a second independent researcher re-scored 15 randomly selected video-recorded Bayley-III(C) assessments; items actually administered (rather than merely credited) were subjected to reliability analysis as described below.

Measures

Socio-demographic characteristics (age, education level, employment status, marital status) for parent(s) or any other principal caregiver were obtained by questionnaire, completed at the time of the Bayley-III(C) assessment. Parity of the birth, number of other children living within the family, and postal code were also recorded. As an indication of parental IQ[19], we asked one parent to complete the National Adult Reading Test (NART[20]). Socio-economic status of parents was calculated from demographic data using UK government guidelines[21]. Finally, we asked whether caregivers had any concern about their child's development (No, Yes, Maybe).

Statistical methods

Data are presented as mean with standard deviation (SD), median (range), or proportion (%) as appropriate for the distribution of data. Statistical significance was taken throughout at 2-sided $p<0.05$, with 95% confidence intervals (CI) used to express the uncertainty in the data. All analyses were conducted within Excel, SPSS version 20, or Stata SE version 12.

Associations between continuous variables were assessed with Pearson's correlation (r). Chi-squared (χ^2) statistics were used to assess the statistical significance of associations between individual test items/questions and delay. Multiple logistic regression and discriminant function analyses were used to explore the associations between multiple items and delay. Items not significantly ($p<0.05$) associated with delay were excluded from ERIC. ERIC standardized scores were obtained by regression of total ERIC score on age (corrected for prematurity).

To define ERIC, we examined the initial test battery item-by-item for statistical association with delay using chi-squared tests (with continuity correction for 2 by 2 tables), multiple logistic regression and discriminant function analysis. In addition, feedback from parents suggested we should allow early stopping for the youngest children. Our item battery had not included stopping points. To explore the possibility of shortening ERIC for the youngest children we ranked the items in ERIC in order of success rate, separately for children in the bands 10-13, 16-19, and 22-24 months corrected age, and inspected the profiles to select those items most applicable to the youngest children. In the youngest group, the mean difference between the standardized scores on these selected games items, plus questions, and their scores on the complete ERIC was calculated, and an estimated total score obtained by adding this difference to their standardized score for the selected items (and 7 questions) only.

Items contributing to ERIC, and Bayley-III(C) administration, were examined for test-retest reliability by calculating (two way mixed) intraclass correlation coefficients (ICC), and Spearman correlation coefficients (ρ), with any systematic difference in score determined by paired t-test. The ICC obtained was interpreted with reference to the following thresholds: ICC <0.40 poor; 0.40-0.75 fair to good; >0.75 excellent agreement. Bland-Altman methodology was also used.

Receiver Operating Characteristics (ROC) curves using non-parametric estimation were used for diagnostic test analyses, to examine the associations between ERIC scores and a diagnosis of delay, with sensitivity (Se., proportion of true positives identified by the test), specificity (Sp., proportion of true negatives identified), positive predictive value (PPV, proportion of those identified being true positives) negative predictive value (NPV, proportion of those identified being true negatives), likelihood ratio positive (LR+, ratio of true positive to false positive proportions) and likelihood ratio negative (LR-, ratio of false negative to true negative proportions) calculated at varying ERIC cut-points. The area under the curve (AUC, a value from 0-1 showing the performance of the test, with higher values indicating better test performance and 0.5 indicating randomness) with 95% CI calculated after 1000 bootstrap replications. In identifying the best cut-point preference was given to increased sensitivity over specificity, maximizing the likelihood of the test identifying truly delayed infants.

Results

Participant recruitment

A total of 1258 children were initially identified by clinicians as potentially eligible. Of these, 62 were immediately excluded (i.e. met ≥1 exclusion criterion), and families of 43 further children declined to take part before being contacted by researchers, leaving 1153 children available for recruitment. Of these, 414 were not approached (229 not contactable with the available details; 185 not at the required age). Of the families of the 739 children approached, 114 (15%) declined to participate, 193 (26%) could not make arrangements to participate, leaving 432 children (58%) recruited (55% of the 782 eligible and contactable).

We compared participant and non-participant data using anonymized lists. There were no large differences between the three groups *n*=414 not approached, *n*=307 not recruited, *n*=432 recruited, in terms of sex, gestational age, birth weight, or Apgar at 5 minutes, although those recruited were more likely to be multiple births (*p*=0.016), to have higher Apgar scores (*p*=0.022), and to live in relatively less deprived areas based on postal code[22] than children not approached/not recruited (*p*<0.001) (Table 1).

Of 432 children recruited, 70 were excluded: 45 failed to attend for the Bayley-III(C) assessment; 25 attended for Bayley-III(C) but failed to complete ERIC fully. Of the 362 completing both assessments, 90 were one of surviving twins/triplets; 45 of these children were randomly removed from the data set leaving *n*=317. Sample characteristics are given in Table 2.

Assessment of cognitive delay using Bayley-III Cognitive Scale

There is debate about the most suitable cutpoint for delay in this population when using Bayley-III(C). We adopted the widely recognized cutpoint of <80 as diagnostic of delay[16],

but also examined a range of cutpoints. Of the 317 children assessed, mean (SD) Bayley-III(C) score was 101.1(14.3), 95% CI 98.5-101.7, with 18 (5.7%) found to be delayed (score <80: Table 3).

Raw battery scores and standardization for age

Raw total scores for the battery were highly correlated with age corrected for prematurity (n=317, r=.73, p<0.001) indicating the need to standardize scores for age. Under the assumption of a simple linear association, age explained 50% of the variance in total raw scores, $F(1, 315) = 338.8$, $p < .001$, with only marginal improvement by addition of higher-order coefficients (≤2% improvement in variance explained). An age-standardized score was therefore defined as:

Age-standardized score = (100x raw score)/(age in days, corrected for prematurity) Eq.1

Item battery reduction: Initial definition of ERIC

The mean (SD) raw score total over all items was 23.4 (9.02). Of the 33 scorable games *items*, 17 had no significant association (i.e., all p>.055 on χ^2 test, two-tailed). These were removed, with the exception of: (1) four items in short series preceding items associated with delay; and (2) a single item 'pointing to body parts' because, unlike the other items, it had both a reasonable association with delay (p=.12), and had been popular with parents; a total of 12 items were thus eventually removed. Similarly, of the 16 *questions*, only seven had significant χ^2 associations with delay (i.e., p<0.05); the remaining nine were removed. Additional multiple logistic regression and discriminant function analysis confirmed that we were not losing useful information by the removal of these items (results not shown). These removals produced an ERIC with 21 scorable items across 11 games and 7 questions and a mean (SD) raw score of 15.1(5.45), 95% CI 14.5-15.7 and mean (SD) standardized score of 2.86(0.79), 95% CI 2.78-2.95. ROC analyses identified a cut-point on the ERIC standardized score (i.e., 100 x total for 21 scorable items divided by age in days, corrected for prematurity) of 2.50 for identifying delay (Bayley-III(C) <80), with sensitivity of 94.4% (17 of 18 delayed infants identified) and specificity of 76.6% (229 of 299 normal infants identified). The AUC was 0.90 (95% CI 0.84 to 0.96).

Test-retest reliability of ERIC and Bayley-III (C)

As might be expected, children scored slightly higher on the second ERIC administration than the first (mean [SD] ERIC standardized score: 2.74 [0.89] vs. 2.57 [0.87]), but with no statistically significant difference (mean difference 0.16, 95% CI -0.11 to 0.44; $t(27) = 1.215$, p=.24): true also for the regression of score difference on score mean (p=.90). Bland-Altman plotting showed no systematic bias. Ranking of the two ERIC scores was similar (ρ=.76, n=28, p<.001, two-tailed), with the ICC between the two administrations being 0.67 (95% CI .41 to .84), indicating at least fair to good agreement. The pairs of Bayley-III(C) scores were very similar (ρ=.995, n=15, p<.001, two-tailed), with the ICC between the two scorers being 0.998 (95% CI .991 to .999), indicating excellent agreement.

Concurrent validity of ERIC

Age-standardized scores were moderately correlated with Bayley-III(C) scores (n=317, both for the full battery r=.408, and for ERIC r=.468, ps<0.001). Bland-Altman plotting showed

no bias in the relation between Z-transformed ERIC standardized scores and Z-transformed Bayley-III(C) scores.

Modification of ERIC for use in younger children

Inspection of the item profiles for the three age-bands of children suggested that for infants <14 months corrected it made sense to stop ERIC after 8 games items, because most items beyond this point scored zero. The mean difference between the *n*=88 10-13 month-olds' standardized scores on the first 8 items, plus 7 questions, and their score on the whole of ERIC was 0.20, this being added to their standardized score to obtain an estimated score. The correlation of this truncated-ERIC score (i.e., for all children, but with the scores for < 14 month-olds estimated following truncation) with the Bayley-III(C) remained high at *r*=0.497, n=317, *p*<0.001. ROC analyses for this version gave diagnostic statistics at least as good (AUC = 0.91 95% CI 0.85 to 0.97) as the 'full' version (described above under Item Battery Reduction). With a cut-point of 2.52 this version (i.e., with truncation for children < 14 months) detected 17 out of 18 delayed children in the sample, with 94.4% sensitivity (95% CI 83.9-100%), 76.9% specificity (72.1-81.7%), 19.8% PPV (11.4-28.2%); 99.6% NPV (98.7-100%); 4.09 LR+; and 0.07 LR-. The associated AUC was .909 (.829-.960). This version was therefore retained as the final ERIC (see also Figure 1 and Table 4).

Use of ERIC in higher risk children as defined by BAPM

Our sample was intentionally inclusive with respect to GA, reflecting our interest in children born before the clinically-relevant cut-point of 34 weeks[12]. To check that ERIC retains its diagnostic properties for at-risk children as defined by BAPM[5], we re-ran the analyses using more stringent inclusion criteria (BW<1500g, or GA<32 completed weeks, or 5-min Apgar<7). In this subsample of *n*=208 children, 13 (6.3%) were delayed; ERIC performed at least as well as for as for those children in the main sample (see Figure 1).

Effects of other variables

We investigated ERIC's performance within subgroups of the sample for moderating effects of variables such as parental SES (two groups divided at the median score) and whether or not parents had any developmental concerns. Within each group the diagnostic performance of ERIC was similar to the performance in the total sample (data not shown).

Discussion

ERIC showed excellent sensitivity (94%) and good specificity (77%) in detection of cognitive delay in a broad sample of 10-24 month olds whose risk of such delay is elevated by reason of prematurity, low birth weight, or perinatal complications. ERIC may thus be useful in detection of delay, or reassurance of absence of delay, in children meeting current minimal criteria for follow-up (e.g., those of <32 weeks GA or < 1500g BW), and indeed preterm children born after 32 weeks but before the clinically-significant cut-point of 34 weeks[12]. ERIC may be administered by post, as in this study, or else left by health visitors with families and collected later. In children of this age, a professionally-administered assessment typically takes 40-90 minutes, whereas scoring ERIC from the completed pamphlet takes 1-2 minutes.

Variance in cognitive performance of preterm infants is high[23] and screening of these children by parents may therefore lead to detection of cases of cognitive delay which would otherwise go undetected. Our data support this view. Of the 18 cases of cognitive delay, four had parents who indicated 'no developmental concerns'; two cases neither fell within the BAPM criteria for follow up, nor our HIE/Apgar inclusion criteria. (And in one of these cases, the parent had reported 'no developmental concerns'.)

There are some limitations to our preliminary findings. First, generalizability of ERIC might be questioned due to the low prevalence of delay found[4]. However, exclusion criteria will have removed the most obvious cases of delay: these children are significant contributors to prevalence rates in most studies[24]. One study reporting prevalence of cognitive delay in children with BW<1000g and who were neurosensorily intact[25] reported 7% prevalence, in line with the rate in the present paper, and suggesting that ERIC has diagnostic value in detection of relatively covert delay. Second, our sample is affected by non-participation of families living in areas of relatively high deprivation. Recruitment in this population is known to be challenging, with participation rates around 50% not uncommon. However, there is no evidence that the performance of ERIC was affected by SES level. Third, ERIC's 77% specificity, coupled with the relatively low prevalence of delay in the sample (and therefore also in the likely population) suggests it will generate around 23% of false positives (95% CI: 18% to 28%). In the present context, the consequence of a 'false positive' would be a referral for a professionally-administered assessment[5, 6]. Given that ERIC has the potential to *replace* such follow-up assessments, a false positive rate of 23% amounts to a reduction in clinical workload of 77%, with consequent benefits. Furthermore, there is a raised risk for cognitive delay in such over-referrals[26], who do not qualify for special education but nonetheless may require other forms of intervention (e.g., parent-training). Specificity of ERIC is therefore acceptable, especially given its high NPV (>99%) and thus low LR- (0.07).

Finally, this preliminary study was necessarily small-scale. Although it is common to undertake test-retest and inter-rater reliability assessment on a sub-sample of the participants, we recognize that the proportions tested in these regards were small. Further work is needed to confirm both the promising psychometric properties of the ERIC instrument and the high ICC values reported here.

Conclusion

Our preliminary study suggests that ERIC may well be suitable as a quickly-administered diagnostic instrument in infants between the ages of 10 and 24 months who are at risk for cognitive delay by reason of prematurity, low birth weight, or perinatal complications. Such an instrument would likely: (1) reassure clinicians and families of the absence of early cognitive delay; and (2) act as a screen for the presence of such delay. However, further research is indicated to evaluate these possibilities, and to evaluate ERIC's potential cost-effectiveness. We invite other researchers to investigate this freely-available instrument.

Acknowledgements

We gratefully acknowledge the support of the parents and children who assisted in our research. We are grateful to Professor Frances Glascoe and Professor Neil Marlow for perceptive, helpful, and encouraging commentaries on previous drafts of this paper. This paper presents independent research funded by the UK National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0807-14202). The bulk of this work was carried out under UK NIHR Research for Patient Benefit Grant PB-PG-0807-14202. Additional support (for researcher salaries) was provided by the Royal Berkshire NHS Foundation Trust and (for participant recruitment activities) the NIHR Thames Valley Comprehensive Local Research Network. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the UK Department of Health.

Reference List

1. Bayley N. Bayley scales of infant and toddler development. San Antonio, TX: Harcourt Assessment, Inc.; 2006.
2. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. JAMA. 2002;288(6):728-37.
3. Nosarti C, Murray RM, Hack M, editors. Neurodevelopmental outcomes of preterm birth: From childhood to adult life. Cambridge: CUP; 2010.
4. Woythaler MA, McCormick MC, Smith VC. Late preterm infants have worse 24-month neurodevelopmental outcomes than term infants. Pediatrics. 2011;127(3):e622-e9.
5. BAPM. Report of a BAPM/RCPCH Working Group: Classification of health status at 2 years as a perinatal outcome. British Association of Perinatal Medicine; 2008 [cited 2012 6/22/12]; 1.0:[Available from: http://www.bapm.org/publications/.
6. AAP. Follow-up care of high-risk infants. Pediatrics. 2004 November 1, 2004;114(Supplement 5):1377-97.
7. Chyi LJ, Lee HC, Hintz SR, Gould JB, Sutcliffe TL. School outcomes of late preterm infants: special needs and challenges for infants born at 32 to 36 weeks gestation. J Pediatr. 2008;153(1):25-31.
8. Morse SB, Zheng H, Tang Y, Roth J. Early school-age outcomes of late preterm infants. Pediatrics. 2009;123(4):E622-E9.
9. Tang BG, Feldman HM, Huffman LC, Kagawa KJ, Gould JB. Missed opportunities in the referral of high-risk infants to early intervention. Pediatrics. 2012;129(6):1027-34.
10. Rosenberg SA, Zhang D, Robinson CC. Prevalence of developmental delays and participation in early intervention services for young children. Pediatrics. 2008;121(6):e1503-e9.
11. Sonnander K. Early identification of children with developmental disabilities. Acta Paediatr. 2000 Sep;89:17-23.
12. Engle WA. A recommendation for the definition of “Late Preterm” (Near-Term) and the Birth Weight–Gestational Age classification system. Semin Perinatol. 2006;30(1):2-7.
13. Saudino KJ, Dale PS, Oliver B, Petrill SA, Richardson V, Rutter M, et al. The validity of a parent-based assessment of the cognitive abilities of 2-year-olds. Br J Dev Psychol. 1998;16:349-63.
14. Johnson S, Marlow N, Wolke D, Davidson L, Marston L, O'Hare A, et al. Validation of a parent report measure of cognitive development in very preterm infants. Dev Med Child Neurol. 2004;46(6):389-97
15. Baker M, Schafer G, Alcock K, Bartlett S. A parentally-administered assessment of cognitive development for children aged between 8 and 24 months. Infant Behav Dev. 2013;36:279-87.
16. Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: Which cut-off should be used? . Pediatr Res. 2014.
17. Bayley N. Bayley scales of infant development. 2nd ed. San Antonio, TX: Psychological Corporation; 1993.
18. Alcock K. Cognitive Ability Questionnaire. 2002.
19. Nelson H, Willison J. Restandardisation of the NART against the WAIS-R. Windsor: NFER-Nelson; 1991.

20. Nelson HE. National Adult Reading Test (NART) Part I. 2nd ed. Windsor, UK: NFER-Nelson; 1982.
21. Office for National Statistics. The National Statistics Socio-economic Classification User Manual. Basingstoke, UK: Palgrave Macmillan; 2005.
22. Noble M, McLennan D, Wilkinson K, Whitworth A, Barnes H. The English Indices of Deprivation 2007. In: Government CaL, editor. London: HMSO; 2008.
23. Aylward GP. Cognitive function of preterm infants: no simple answers. JAMA. 2003;289(6):752-3.
24. Arpino C, Compagnone E, Montanaro ML, Cacciatore D, De Luca A, Cerulli A, et al. Preterm birth and neurodevelopmental outcome: a review. Childs Nerv Syst. [Review]. 2010 Sep;26(9):1139-49.
25. Hack M, Taylor HG, Drotar D, Schluchter M, Cartar L, Andreias L, et al. Chronic conditions, functional limitations, and special health care needs of school-aged children born with extremely low-birth-weight in the 1990s. JAMA: the journal of the American Medical Association. 2005;294(3):318-25.
26. Glascoe FP. Are over-referrals on developmental screening tests really a problem? Arch Pediatr Adolesc Med. 2001;155(4):54-9.

Tables

Table 1. Characteristics of the potentially eligible sample and those recruited (see also Figure 1)

Characteristic	Recruited (n = 432)	Not Recruited (n = 307)	Not Approached (n = 414)	p value
Sex, n (%)				
Male	245 (56.7)	164 (53.6)	230 (55.6)	
Female	187 (43.3)	142 (46.4)	184 (44.4)	.702
N of children at birth, n (%)				
Singleton	314 (72.7)	251 (81.8)	313 (75.6)	
Multiple birth	118 (27.3)	56 (18.2)	101 (24.4)	.016
GA, n (%) weeks				
23-27	53 (12.3)	40 (13.0)	44 (10.6)	
28-31	151 (35.0)	92 (30.0)	131 (31.6)	
32-36	167 (38.7)	126 (41.0)	170 (41.1)	
≥ 37	61 (14.1)	49 (16.0)	69 (16.7)	.709
Mean GA (95% CI) weeks	32.2 (31.8-32.6)	32.5 (32.1-33)	32.7 (32.3-33.1)	.236
BW, n (%) g				
< 1000	53 (12.3)	47 (15.4)	41 (9.9)	
1000-1500	120 (27.8)	61 (20.0)	114 (27.5)	
> 1500	259 (60.0)	197 (64.6)	259 (62.6)	.038
Mean birth weight (95% CI) g	1838 (1754-1922)	1888 (1791-1986)	1884 (1803-1965)	.658
No with missing data	0	2	0	
Apgar 5 min, n (%)				
< 7	54 (13.2)	56 (19.9)	74 (19.7)	
≥ 7	355 (86.8)	225 (80.1)	302 (80.3)	.022
Mean Apgar score (95% CI)	8.6 (8.4-8.8)	8.3 (8.1-8.6)	8.3 (8.1-8.5)	.126
No with missing data	23	26	38	
Mean IMD Score (95% CI)	13.0 (12.0-14.1)	16.1 (14.8-17.4)	16.1 (14.3-15.6)	< .001

Notes: n = number of participants; GA = gestational age; CI = confidence interval; BW = birth weight; IMD = Index of Multiple Deprivation[22]

Table 2. Characteristics of the sample used for validation of ERIC (n=317)

Child Characteristics	N (%)
Sex, n (%)	
Male	173 (54.6)
Female	144 (45.6)
N of children at birth, n (%)	
Singleton	270 (85.2)
Multiple birth	47 (14.8)
Order of birth, n (%)	
Firstborn	178 (56.2)
Second or subsequent child	138 (43.5)
Number with missing data	1 (0.3)
Ethnicity, n (%)	
White	242 (76.3)
Black	10 (3.2)
Asian	35 (11.0)
Mixed	30 (9.5)
Gestational age, n (%), weeks	
23-27	44 (13.9)
28-31	104 (32.8)
32-36	114 (36.0)
≥ 37	55 (17.4)
Preterm Infants (median GA, range)	262 (31, 23-36)
Full term Infants (median GA, range)	55 (40, 37-42)
Birth weight, n (%), g	
< 1000	42 (13.2)
1000-1500	88 (27.8)
> 1500	187 (59.0)
Preterm Infants, *n*=262, Median (range)	1510 (490-3010)
Full term Infants, *n*=55, Median (range)	3600 (2300-4580)
Apgar 5 min., n (%)	
< 7	48 (15.1)
≥ 7	265 (83.6)
No with missing data	4
Preterm Infants, *n*=262, Median (range)	9 (2-10)
Full term Infants, *n*=55, Median (range)	6 (2-10)
HIE, n (%)	
No HIE	281 (88.6)
Mild	20 (6.3)
Moderate	11 (3.5)
Severe	5 (1.6)

Note: HIE = hypoxic-ischemic encephalopathy

Table 3. Numbers of children (%) within the normal range and cognitively delayed by corrected age

Age (corrected)	Normal range Bayley-III(C) ≥80	Delayed Bayley-III(C) <80	Totals
10.0 ≤ Months < 11.0	20	0 (0%)	20
11.0 ≤ Months < 12.0	23	0 (0%)	23
12.0 ≤ Months < 13.0	23	1 (4.2%)	24
13.0 ≤ Months < 14.0	23	0 (0%)	23
14.0 ≤ Months < 15.0	22	0 (0%)	22
15.0 ≤ Months < 16.0	17	4 (19.0%)	21
16.0 ≤ Months < 17.0	20	1 (4.8%)	21
17.0 ≤ Months < 18.0	19	1 (5.0%)	20
18.0 ≤ Months < 19.0	18	4 (18.2%)	22
19.0 ≤ Months < 20.0	23	1 (4.2%)	24
20.0 ≤ Months < 21.0	23	1 (4.2%)	24
21.0 ≤ Months < 22.0	22	1 (4.3%)	23
22.0 ≤ Months < 23.0	24	0 (0%)	24
23.0 ≤ Months < 24.0	22	4 (15.4%)	26
Total	299	18 (5.7%)	317

Note: Bayley-III(C): Bayley Scales of Infant and Toddler Development, 3rd Edition, Cognitive Scale scores

Table 4

Cross-tabulations of final ERIC diagnostic performance against various cut-points on the Bayley Scales of Infant and Toddler Development III Cognitive Scale

ERIC[a]	**Bayley-III(C) Diagnostic Status defined by increasing cut-point**							
Diagnostic Status	**≤70**	**>70**	**<80**[b]	**80+**	**≤80**	**>80**	**<90**	**90+**
Delayed N	11 (91.7%) [Se]	75	17 (94.4%) [Se]	69	26 (86.7%) [Se]	97	43 (75.4%) [Se]	120
Normal range N	1	230 (75.4%) [Sp]	1	230 (76.9%) [Sp]	4	190 (66.2%) [Sp]	14	140 (53.8%) [Sp]
Total	12	305	18	299	30	287	57	260
AUC (95% CI) [c]	0.90 (0.79 to 0.97)		0.91 (0.83 to 0.96)		0.80 (0.69 to 0.88)		0.71 (0.64 to 0.80)	
Cut-point	2.52		2.52		2.83		3.00	

Notes:

[a] Final ERIC: Standardized (for age), abbreviated (by removal of items not associated with delay), and truncated (for children <14months) as described in the text; [b] definition of delay in the present paper; [c] after 1000 bootstrap replications; [Se] Sensitivity; [Sp] Specificity. AUC: Area under the (ROC) curve; Bayley-III(C): Bayley Scales of Infant and Toddler Development 3rd Edition, Cognitive Scale; ROC: Receiver Operating Characteristics.

Figure Legends

Figure 1. ROC curves for ERIC for (a) the total sample (n=317) and (b) the high risk[a] sample (n=208)

[a]High-risk=BW<1500g, or GA<32 completed weeks, or 5-min Apgar<7

Note: These ROC curves plot, for each possible cut-point of the relevant ERIC scale, the true positive proportion (sensitivity) against the false positive proportion (1-specificity). A perfect test would have an area under the curve of 1.0 and the 'curve' would pass through the upper left corner of the plot (100% sensitivity, 100% specificity). The arrows show the threshold for each test which is associated with the best values of sensitivity and specificity. 'Truncated' refers to the instrument which became the final ERIC, and for which administration is terminated early for children aged <14 months. See text for further details.

AUC: .909 for full sample, .927 for high-risk as defined by our more stringent criteria[a]. Sensitivity: 94.4% for full sample, 100% high-risk. Specificity: 76.9% full sample, 74.4% high-risk.

Figure 1a

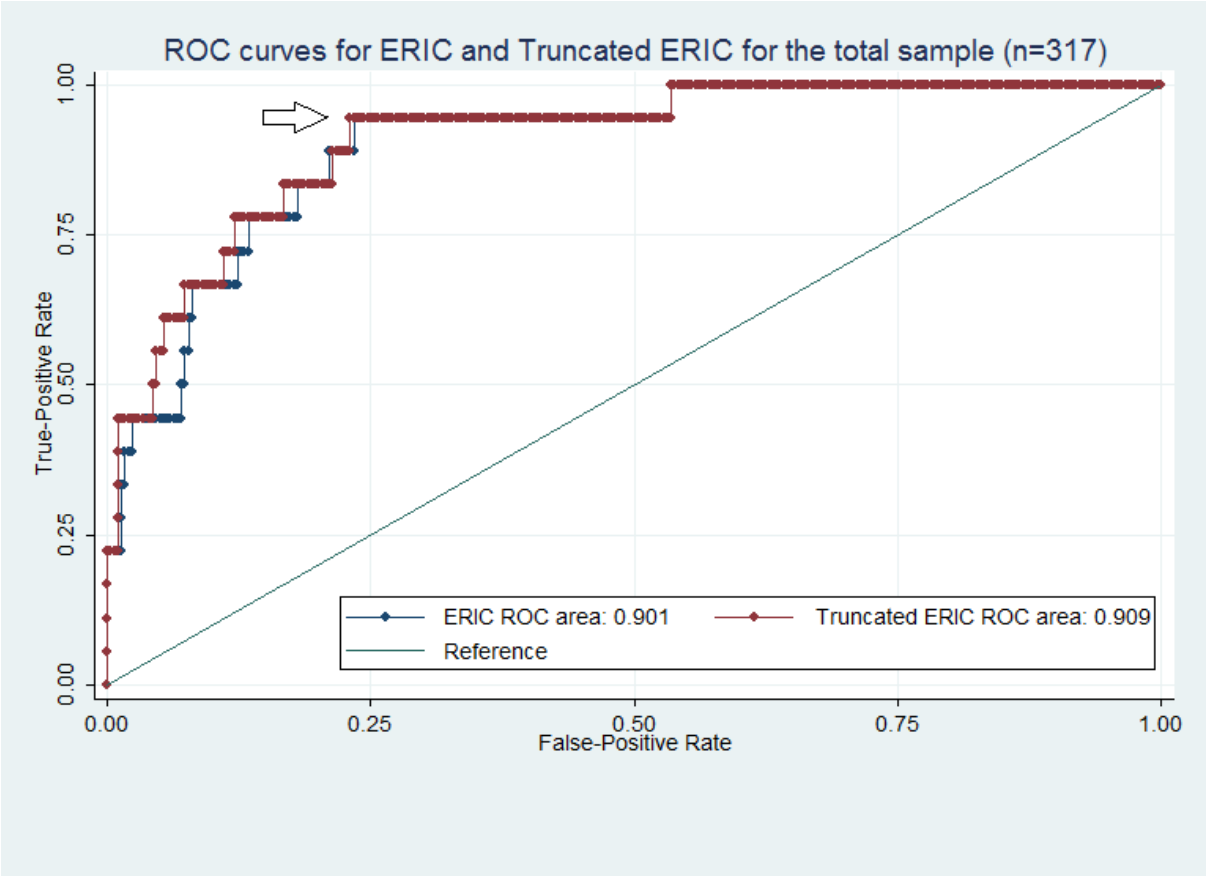
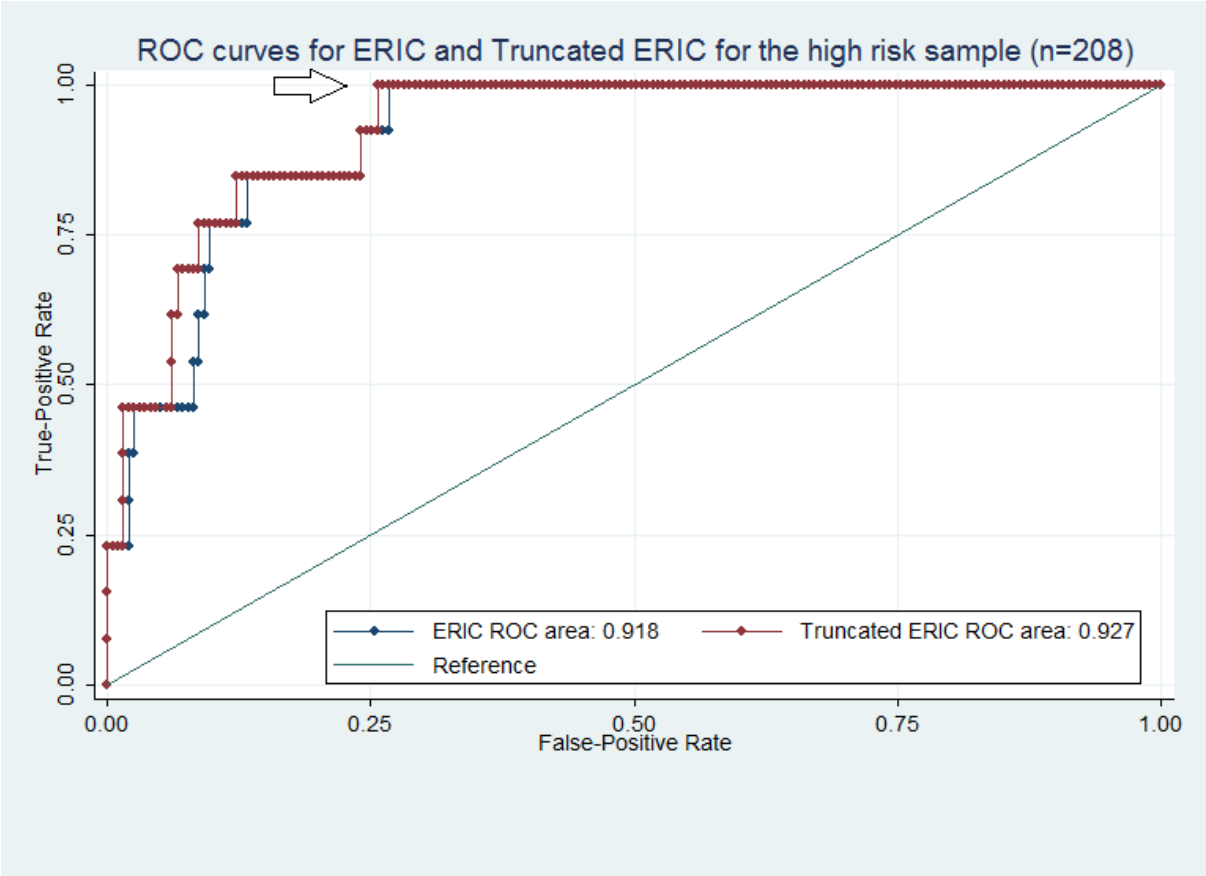

ROC curves for ERIC and Truncated ERIC for the total sample (n=317)
True-Positive Rate
0.00
0.25
0.50
0.75
1.00
False-Positive Rate
0.00
0.25
0.50
0.75
1.00
ERIC ROC area: 0.901
Truncated ERIC ROC area: 0.909
Reference

Figure 1b


ROC curves for ERIC and Truncated ERIC for the high risk sample (n=208)
True-Positive Rate
0.00
0.25
0.50
0.75
1.00
False-Positive Rate
0.00
0.25
0.50
0.75
1.00
ERIC ROC area: 0.918
Truncated ERIC ROC area: 0.927
Reference

Acknowledgements

We gratefully acknowledge the support of the parents and children who assisted in our research. We are grateful to Professor Frances Glascoe and Professor Neil Marlow for perceptive and encouraging commentaries on previous drafts of this paper. This paper presents independent research funded by the UK National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0807-14202), the Royal Berkshire NHS Foundation Trust and the NIHR Thames Valley Comprehensive Local Research Network. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the UK Department of Health.

Funding source: The bulk of this work was carried out under UK NIHR Research for Patient Benefit Grant PB-PG-0807-14202. Additional support (for researcher salaries) was provided by the Royal Berkshire NHS Foundation Trust and (for participant recruitment activities) the NIHR Thames Valley Comprehensive Local Research Network.

Financial disclosure: The authors have no financial relationships relevant to this article to disclose.

Conflict of Interest: The authors have no conflicts of interest relevant to this article to disclose.

Contributors' Statement:

Graham Schafer conceptualized and designed the study; supervised data collection and entry; supervised ethical arrangements; performed initial data consolidation and analysis; and prepared the manuscript.

Lucia Genesoni undertook data collection and entry at all 3 sites; performed initial data consolidation and analysis; and prepared the manuscript.

Greg Boden conceptualized and designed the study; coordinated and supervised data collection at one of the three sites; critically reviewed the manuscript and approved the final manuscript as submitted.

Helen Doll co-designed the study; planned and undertook the analysis; and prepared the manuscript.

Rosamond A K Jones and Eleri Adams coordinated and supervised data collection at one of the three sites; critically reviewed and edited the manuscript; and approved the final manuscript as submitted.

Ron Gray co-designed the study; critically reviewed and edited the manuscript; and approved the final manuscript as submitted.

Ros Jefferson co-designed the study; critically reviewed and edited the manuscript; and approved the final manuscript as submitted.