

Routine educational outcome measures in health studies: Key Stage 1 in the ORACLE Children Study follow-up of randomised trial cohorts

David R Jones,¹ Katie Pike,² Sara Kenyon,² Laura Pike,¹ Brian Henderson,³ Peter Brocklehurst,⁴ Neil Marlow,⁵ Alison Salt,⁶ David J Taylor²

¹Health Sciences Department, University of Leicester, Leicester, UK

²Cancer and Molecular Medicine Department, University of Leicester, Leicester, UK

³Centre for Evaluation and Monitoring, University of Durham, UK

⁴National Perinatal Epidemiology Unit, University of Oxford, Oxford, UK

⁵Academic Division of Neonatology, Institute for Women's Health, University College London, London, UK

⁶Great Ormond Street Hospital for Children and Institute of Child Health, University College London, London, UK

Correspondence to

Professor David R Jones, Health Sciences Department, Adrian Building (Room 214d), University of Leicester, University Road, Leicester LE1 7RH, UK; drj@le.ac.uk

Accepted 2 July 2010
Published Online First
7 October 2010

ABSTRACT

Objectives Statutory educational attainment measures are rarely used as health study outcomes, but Key Stage 1 (KS1) data formed secondary outcomes in the long-term follow-up to age 7 years of the ORACLE II trial of antibiotic use in preterm babies. This paper describes the approach, compares different approaches to analysis of the KS1 data and compares use of summary KS1 (level) data with use of individual question scores.

Participants 3394 children born to women in the ORACLE Children Study and resident in England at age 7.

Methods Analysis of educational achievement measured by national end of KS1 data (KS1) using Poisson regression modelling and anchoring of the KS1 data using external standards.

Results KS1 summary level data were obtained for 3239 (95%) eligible children; raw individual question scores were obtained for 1899 (54%). Use of individual question scores where available did not change the conclusion of no evidence of treatment effects based on summary KS1 outcome data.

Conclusions When accessible for medical research purposes, routinely collected educational outcome data may have advantages of low cost and standardised definition. Here, summary scores lead to similar conclusions to raw (individual question) scores and so are attractive and cost-effective alternatives.

INTRODUCTION

Educational attainment and performance measures are widely used in paediatric studies, as proxy measures of development,¹ usually collected specifically for individual studies. Routinely collected, statutory attainment measures are relatively rarely used as outcomes in health services research. In England an extensive array of educational data is collected routinely. All state schools must be involved in the end of Key Stage assessments² at ages 7, 11 and 14, currently a mixture of teacher assessments and statutory national tests, and there are nationally standardised examinations at older ages.

The ORACLE trials^{3 4} evaluated the effects of prescription of erythromycin or co-amoxiclav for women with either preterm rupture of the membranes or spontaneous preterm labour (SPL) with intact membranes and no overt infection. The ORACLE Children Study (OCS)^{5 6} sought follow-up information for surviving ORACLE children at 7 years of age in the UK using a parent-report postal questionnaire. The

What is already known on this topic

- ▶ Measures of educational attainment are widely used in paediatric studies, as proxy outcome measures of development or as covariates and prognostic factors.
- ▶ Examples of the use of routinely collected, statutory educational measures as outcomes in health studies are rare.

What this study adds

- ▶ Use of routinely collected educational measures as outcome data in medical studies is feasible, and may be cost-effective compared with special collection of outcome data within a study.
- ▶ Biases in their assessment are unlikely to be associated with treatment group in randomised, blind (masked) trials and thus between-treatment comparisons should be valid.
- ▶ In this study, conclusions were not changed by alternative approaches to analysis of the data or by use of scores for individual questions instead of summary data.

primary outcome was defined as the presence of any level of functional impairment⁷: results for the primary outcome have been reported elsewhere.^{5 6}

Secondary outcomes included a range of medical, behavioural and educational outcomes. This paper investigates the use of statutory National Curriculum tests at Key Stage 1 (KS1) data as measures of educational attainment at age 7 in the long-term follow-up of the ORACLE II trial. Approaches to obtaining KS1 data for use in the study, and the analysis and interpretation of these data are described. In particular, different approaches to analysis of the KS1 data are explored, the use of raw score (on individual questions) and aggregated data compared, and the impact of missing data on validity of results obtained is assessed.

POPULATION AND METHODS

Study population

Here we consider 3394 children born to women in the OCS by 2007 and resident in England since KSI is not routine elsewhere (see figure 1).

Data collection

The intention to carry out a subsequent follow-up evaluation was indicated when written informed consent was sought in the original trial. The study questionnaire was sent to parents/carers shortly before the child was 7 years old, along with a form seeking parental consent for us to request KS1 test returns from the children's schools. After consent, raw score data were sought by direct contact with the school at the time KS1 assessments were being undertaken, teachers being asked to complete a questionnaire with the child's KS1 marks for each question. If this was not returned, KS1 level data were obtained from the local education authority.

At the end of KS1 all children are awarded a level for each of reading, writing and mathematics. Levels 4 and 3 are above average, and level 2 is the modal level; children may also attain level 1, be 'working towards level 1' or not be entered ('disapplied') by the teacher. Since the OCS began, the content and administration of the tests has varied between years, with tests now seen as supplementary to teachers' assessments rather than the sole basis of the KS1 assessment.² Raw score data from tests are much more detailed and extensive than the

summary level scores, comprising separate marks for about 110 questions for each child. All outcome data were double entered and validated.

In addition to the individual KS1 test returns provided by the OCS children's schools, for all eligible OCS children in England KS1 data on level attained (but not raw scores) were provided as an anonymised data file by the Department for Education and Science (now superseded by the Department for Education (DfE)), categorised by treatment group.

Statistical analysis

KS1 data were dichotomised into level 2 or above/below level 2, and Mantel-Haenszel ORs and 95% CIs obtained, stratifying by test year, and investigating inclusion of covariates to adjust for birth weight, gestational age and other factors. Retention of most KS1 categories, and conducting log linear and ordinal regression analyses are compared with this basic analysis. Alternative approaches combined KS1 data from several years by standardisation with external weights. Consistency of raw score and level data was also assessed. Stata⁸ commands for models and detailed results are provided in the supplementary materials (<http://www2.le.ac.uk/departments/health-sciences/research/ships/jones-sup-mat/arch-dis-child-2010>).

RESULTS

Full KS1 level data provided by DfE were available for 3239 (95%) of the 3394 children in the OCS SPL cohort who had

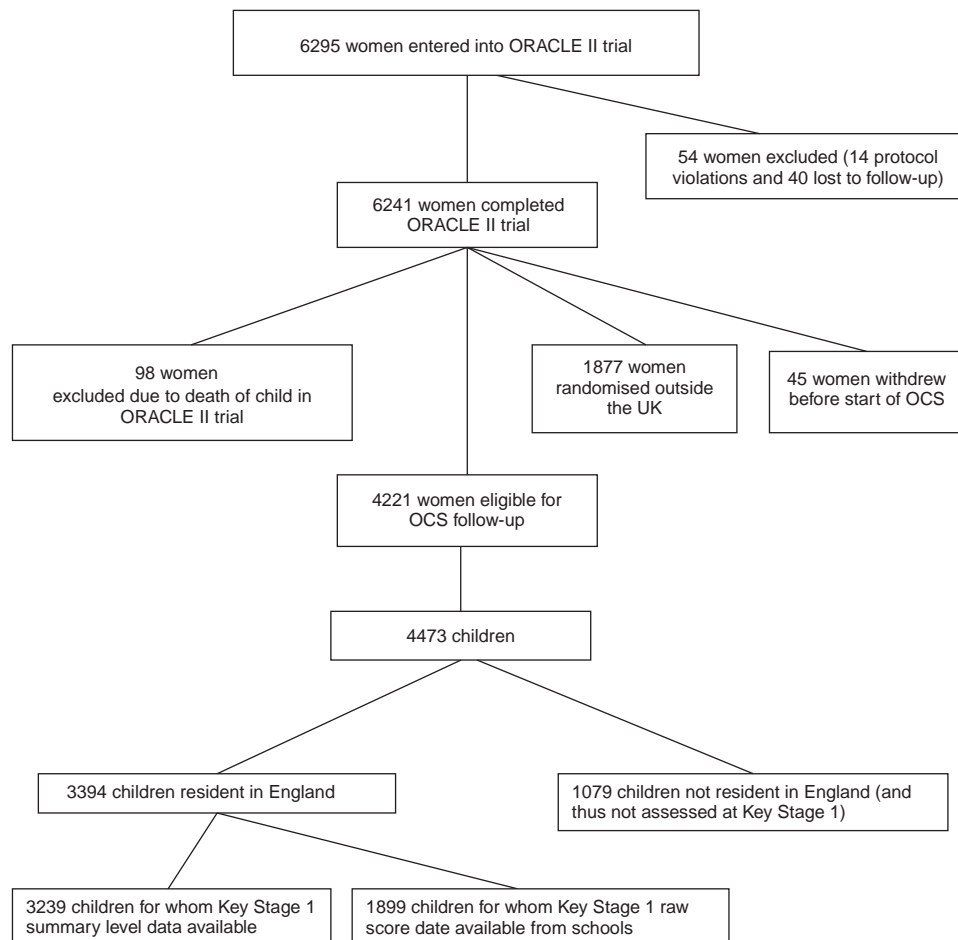


Figure 1 Flowchart for progression through ORACLE II trial and extended follow-up on educational outcomes (Key Stage 1) in ORACLE Children Study (OCS).

been entered into the tests by 2007, but KS1 data were obtained from schools after parental consent for only 1899 (54%) children, predominantly because of lack of parental consent rather than school non-response. Table 1 shows that women consenting to collection of KS1 tend to be older, more likely to be white and less likely to live in the most socially deprived⁹ areas. However, their children have worse neonatal morbidity. The supplementary materials provide more details, and show that the treatment groups were reasonably balanced in respect of response.

Table 2 shows the proportions of children not achieving level 2 or above in the dataset obtained from schools and parents, alongside results in the DfE-supplied data from the principal report of the study.⁶ Although the proportions reaching level 2 or above are higher in the parental/school results, as expected given the over-representation of disadvantaged groups in the parental non-responders, there are no clear differences between

treatment groups in either set of results. In national normative data for 2002–2007, the proportions failing to achieve level 2 or above were 15% for reading, 18% for writing and 11% for mathematics. In contrast, corresponding proportions for the OCS children were 23%, 25% and 15%, respectively.

Table 3 shows RRs for treatment effects from Poisson regression models¹⁰ fitted to DfE and parental/school datasets without dichotomisation at level 2 but adjusting for test year. The results for the DfE dataset are again those in table 6 of the main study report. Neither analysis provides evidence of effect on the proportions of children achieving each level in reading, writing or mathematics at KS1 for either antibiotic. There are no material differences between the DfE and the parental/school results except for slightly wider CIs in the latter smaller dataset. OR estimates in table 2 and RRs in table 3 are generally similar, but CIs from the Poisson models, which retain more of the data, are narrower.

Table 1 Characteristics of groups consenting or not consenting to collection of Key Stage 1 (KS1) data from child's school

	Consent to KS1 data being collected	Contact made but no consent
Number of women	1776 (100%)	1418 (100%)
Maternal age at trial entry (years, median (IQR))	27.3 (23.3–31.5)	25.0 (21.4–29.0)
Gestation age at trial entry (days, median (IQR))	219 (201–232)	219 (201–232)
Multiple births	131 (7.4%)	79 (5.6%)
Number of children	1899 (100%)	1493 (100%)
Delivery within 48 h of trial entry	241 (12.7%)	125 (8.4%)
Delivery within 7 days of trial entry	339 (17.9%)	189 (12.7%)
Gestational age at delivery (days, median (IQR))	265 (242–277)	268 (249–279)
Birth weight (g, median (IQR))	2920 (2210–3400)	2940 (2350–3390)
Males	1018 (53.6%)	786 (52.6%)
Admission to neonatal unit	588 (31.0%)	390 (26.1%)
Ventilated	185 (9.7%)	105 (7.0%)
Respiratory distress syndrome	202 (10.6%)	121 (8.1%)
Oxygen at 28 days	91 (4.8%)	57 (3.8%)
Positive blood culture	40 (2.1%)	31 (2.1%)
Necrotising enterocolitis (suspected or proven)	19 (1.0%)	15 (1.0%)
Abnormal cerebral ultrasonography	28 (1.5%)	20 (1.3%)
Social deprivation: number (%) in lowest quartile		
Income	910 (47.9%)	898 (60.1%)
Education	854 (45.0%)	846 (56.7%)
Child poverty	884 (46.6%)	877 (58.7%)
Ethnicity	1546 (100%)	1693 (100%)
White	1449 (93.7%)	1311 (77.4%)

Numbers (%) except where otherwise indicated (In two cases not tabulated, no contact was made with parents/carers to seek consent).

Table 2 Educational attainment in reading, writing and mathematics at Key Stage 1: data from (1) DfE and (2) schools, with parental consent and direct from parents

	Anonymous data from DfE			Data from schools/parents				
	Erythromycin	No erythromycin	Co-amoxiclav	No co-amoxiclav	Erythromycin	No erythromycin	Co-amoxiclav	No co-amoxiclav
	1641	1598	1608	1631	963	936	918	981
Numbers (%) failing to achieve level 2 or higher								
Reading	377 (23.0%) 1.00 (0.85 to 1.18)	367 (23.0%)	366 (22.8%) 0.98 (0.83 to 1.15)	378 (23.2%)	165 (17.1%) 1.06 (0.83 to 1.35)	153 (16.3%) 0.84 (0.66 to 1.06)	142 (15.5%)	176 (17.9%)
Writing	413 (25.2%) 0.97 (0.83 to 1.13)	413 (25.8%)	395 (24.6%) 0.91 (0.77 to 1.06)	431 (26.4%)	188 (19.5%) 1.04 (0.82 to 1.30)	178 (19.0%) 0.88 (0.70 to 1.11)	168 (18.3%)	198 (20.2%)
Maths	239 (14.6%) 1.04 (0.85 to 1.26)	225 (14.1%)	230 (14.3%) 1.00 (0.82 to 1.21)	224 (13.7%)	95 (9.9%) 1.14 (0.83 to 1.55)	82 (8.8%) 0.84 (0.62 to 1.15)	79 (8.6%)	98 (10.0%)

Mantel-Haenszel ORs (95% CIs)

Table 3 Educational attainment in reading, writing and mathematics at Key Stage 1: data from (1) DfE and (2) schools, with parental consent and direct from parents

	Anonymous data from DfE		School/parental data	
	Erythromycin	Co-amoxiclav	Erythromycin	Co-amoxiclav
Reading	1.00 (0.96 to 1.04)	0.99 (0.95 to 1.03)	1.03 (0.97 to 1.08)	0.96 (0.91 to 1.01)
Writing	1.00 (0.97 to 1.04)	0.99 (0.95 to 1.02)	1.03 (0.98 to 1.08)	0.96 (0.91 to 1.00)
Maths	0.99 (0.96 to 1.03)	0.99 (0.95 to 1.03)	1.00 (0.94 to 1.05)	0.97 (0.92 to 1.03)

Overall RR and 95% CIs are from Poisson models for level achieved (scaled 1–6) adjusting for test year, 2002–2007.

There is no evidence of overdispersion. If interaction effects for reading in the DCFS data are included, main effect estimates are not importantly changed from those shown. More details are available in the supplementary materials.

Several alternative models for RRs, and the results of exploring the impact of inclusion of covariates (as well as treatment effect terms) in the models, are explored in the supplementary material. There is again no clear evidence of treatment effects in the alternative models. Smoking in the family, being male, being born at low gestational age and receiving oxygen at 28 days are related to poorer KS1 grades, but conclusions concerning treatment effects are almost unchanged after adjustment.

A subset of schools provided both raw score and KS1-level data. Although this subset cannot be regarded as representative of the whole study data, it is valuable to compare analyses of raw and level data within it to see whether and how conclusions are modified if raw score data are available. The level recorded from the teacher's assessment of each child (based on an amalgam of test results and ongoing classroom performance) is compared with 'highest equivalent level (HEL)' derived from all raw test scores, being the highest level achieved in any test taken by the child (it being possible for a child to sit—eg, both a level 2 and a level 3 test) in table S1 in the supplementary material. Teacher-assessed levels and HELs agree for 1191/1452 (82%) children. HELs are higher than teacher-assessed levels for 159 (11%) children and lower for 91 (6%) children. Results of Poisson models for treatment effects, variously adjusting for calendar or test year and combinations of test taken, are shown in the supplementary material. Treatment effect estimates are broadly similar to those in table 3, and no differences between treatment groups are apparent.

The distribution of test scores may vary from year to year; this may reflect true variations in performance between cohorts or variation of test characteristics or administration over time. Several analyses above have included adjustment for calendar or test year, since a single summary measure over time is required. Another approach is to anchor and scale year-by-year KS1 data by reference to an external standard believed not to vary over time. In table S2 the results of using Performance Indicators in Primary Schools¹¹ scores as an external standard are shown. ORs for treatment effects obtained are very similar to those obtained from the unanchored data in table 2.

DISCUSSION

Although many clinical trials and studies make use of educational outcome measures, the current OCS project is unusual in seeking to make use of statutory, routinely collected educational outcome data rather than specially collected data. When accessible for medical research purposes, such routine data may have advantages of low cost and standardised definition across all centres in a country, but use of routine outcome data also raises general issues of validity, quality and relevance.

Disadvantaged groups were over-represented among parents not consenting to our requesting KS1 data from their child's school, and the data were not forthcoming from all of

the schools contacted. Nonetheless, there was no evidence of substantially imbalanced response between the randomised treatment groups; although there was an over-representation of older, white mothers from higher status areas, and sicker babies in the responders it is thus unlikely that this would have biased our findings. Indeed balance could be expected since very few of the participants were unblinded to their treatment group. However, in this study an almost complete set of the KS1 level data was also available at relatively little cost through provision of anonymised data files by DfE.

KS1 data do not provide a direct measure of cognitive development, but from this study it is apparent that the treatments in the trial do not have differential longer-term effects on educational outcomes. Another plausible explanation for the absence of substantial observed between-group differences is that KS1 level data are insufficiently sensitive to detect them. Raw score KS1 data are a much richer source of information than KS1 level data but there is no direct indication here that they allow a more sensitive comparison of treatment effects that results in different conclusions; this would need to be checked in future studies. Furthermore, non-ignorable missing data will generally, as here, be an important issue if raw score data are to be used; avoiding the problem by using relatively simple but complete summary data may be preferable to seeking imputation method solutions.¹²

The ways in which the KS1 results are derived have been modified year-to-year since 2005, emphasising the importance of allowing for year in analyses. Stratification by year, external standardisation across years and multilevel modelling have all been applied. However, the move from test-based to principally teacher-based assessment of KS1 introduces the possibility of biases not just by gender and social class, but also between teachers, school and local authorities. Extensive educational attainment tests results are also routinely collected in many other (developed) countries,¹³ including the USA, and could in principle be employed as outcome measures in medical and healthcare trials in the same way.

This project has demonstrated the feasibility of using routinely collected educational measures as outcome data in medical studies, an attractive possibility on cost grounds since such data are collected anyway for educational review purposes. As any systematic biases are unlikely to be associated with treatment group in randomised, blind (masked) trials it is in such studies where the potential value of routine educational outcomes will be greatest, provided the available educational measure is demonstrably valid for evaluation of the trial intervention. Finally, it should be noted that wider use of educational outcomes in medical trials may be expected to promote cross-disciplinary learning, with educational research gaining from perspectives and techniques common in medical research but less so in educational research (at least in the UK), and vice versa.

Acknowledgements The authors thank Peter Tymms (Durham) and Tony Cline (UCL) for helpful comments about educational aspects of the OCS.

Funding Funded by UK Medical Research Council; sponsored by University Hospitals of Leicester. The sponsors had no role in the preparation of the manuscript or the decision to submit it.

Competing interests None.

Contributors DRJ, KP, SK, PB, NM, AS and DJT contributed to the OCS study design, developed the protocol and contributed to conduct of the study, interpretation of results, and drafting the paper. BH provided PIPS data and expertise. KP, DRJ and LP performed the statistical analyses and drafted the paper. SK led the OCS study and DRJ the statistical aspects; they are joint guarantors. NM receives funding from the Department of Health's NIHR Biomedical Research Centres funding scheme at UCLH/UCL.

Ethical approval OCS was approved by the West Midlands ethics committee in 2002.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Richards M**, Hardy R, Kuh D, et al. Birth weight and cognitive function in the British 1946 birth cohort: longitudinal population based study. *BMJ* 2001;**322**:199–203.
2. **National Curriculum Teacher Assessments and Key Stage Tests**. http://www.direct.gov.uk/en/Parents/Schoolslearninganddevelopment/ExamsTestsAndTheCurriculum/DG_4016665 (accessed 28 April 2010).
3. **Kenyon SL**, Taylor DJ, Tarnow-Mordi W. ORACLE Collaborative Group. Broad-spectrum antibiotics for preterm, prelabour rupture of fetal membranes: the ORACLE I randomised trial. *ORACLE Collaborative Group. Lancet* 2001;**357**:979–88.
4. **Kenyon SL**, Taylor DJ, Tarnow-Mordi W. ORACLE Collaborative Group. Broad-spectrum antibiotics for spontaneous preterm labour: the ORACLE II randomised trial. *ORACLE Collaborative Group. Lancet* 2001;**357**:989–94.
5. **Kenyon S**, Pike K, Jones DR, et al. Childhood outcomes after prescription of antibiotics to pregnant women with preterm rupture of the membranes: 7-year follow-up of the ORACLE I trial. *Lancet* 2008;**372**:1310–18.
6. **Kenyon S**, Pike K, Jones DR, et al. Childhood outcomes after prescription of antibiotics to pregnant women with spontaneous preterm labour: 7-year follow-up of the ORACLE II trial. *Lancet* 2008;**372**:1319–27.
7. **Saigal S**, Rosenbaum P, Stoskopf B, et al. Comprehensive assessment of the health status of extremely low birth weight children at eight years of age: comparison with a reference group. *J Pediatr* 1994;**125**:411–17.
8. **StataCorp**. Stata Statistical Software: Release 10. College Station, Texas, USA: StataCorp LP, 2007.
9. **Office for National Statistics**. Neighbourhood Statistics. <http://www.neighbourhood.statistics.gov.uk/dissemination> (Accessed 28 April 2010).
10. **Clayton DG**, Hills M. Statistical models in epidemiology. Oxford: Oxford University Press, 1993.
11. **Tymms P**. Are standards rising in english primary schools? *Br Educ Res J* 2004;**30**:477–94.
12. **Rubin DB**. Multiple imputation for non-response in surveys. Chichester: Wiley, 2004.
13. **UNESCO Institute for Statistics Montreal**. Education Counts: World Education Indicators 2006. <http://www.uis.unesco.org/template/publications/wei2006/WEI2006-FINALwc.pdf> (Accessed 28 April 2010).