Accounting for deaths in neonatal trials: is there a correct approach?

Shalin A Parekh,1 David J Field,1 Samantha Johnson,1 Edmund Juszczak2

INTRODUCTION

The Disability and Perinatal Care report published by the National Perinatal Epidemiology Unit and Oxford Regional Health Authority in 1994 emphasised that data on the neurodevelopmental outcomes of neonates requiring intensive care should be formally collected.1 Over the last 40 years, survival rates of high-risk infants have improved but these have not been matched with parallel improvements in neurodevelopmental outcomes.2–4 Consequently, the focus of neonatal care has shifted increasingly towards reducing long-term morbidity and neurodevelopmental impairment.1 2 Improved long-term neurodevelopment is now considered the ‘Holy Grail’ in neonatology.1 5

These developments have led to a change in focus of perinatal trials, which have moved away from survival as the primary outcome towards using long-term functional outcomes.2 This has raised the question of how to deal with deaths in those trials where neurodevelopmental impairment is of primary interest. In perinatal trials involving the recruitment of high-risk infants, it is inevitable that some will die and quantifying outcome for these infants has led to a range of approaches, none of which are without compromise.6–9

This issue has become even more pertinent since some interventions designed to improve neurodevelopmental outcomes may not necessarily have a biologically plausible effect on mortality. Nevertheless, mortality is significantly higher in neonatology compared with other fields of medicine, particularly among very preterm infants,10 which strongly influences both trial design and analysis.

This review considers approaches that have been taken by trialists regarding the role of death in their outcome measures, the pros and cons of the various approaches, the effect on the outcomes measured and the subsequent interpretation of the findings.

How is neurodevelopmental impairment measured?

There is broad global consensus that neurodevelopmental outcomes should be measured at 18–24 months of age corrected for prematurity. This is a pragmatic compromise between identifying adverse neurodevelopmental outcomes as early as possible, and using tools that are reliable and likely to be predictive of impairments in later life.11 12 This also allows results to become available in a timescale that is not too far removed from the perinatal intervention, while minimising the duration and costs of the trial.

Neurodevelopmental outcomes are usually assessed using validated psychometric instruments which are designed to quantify a child’s developmental progress. These have typically comprised formal standardised tests to assess multiple developmental domains including cognitive, language and motor development, but parent report measures have become increasingly popular as cost-effective alternatives to formal assessments. There are a variety of tools commonly used to assess neurodevelopment at 18 months to 2 years of age in perinatal trials (table 1). One of the most widely used and recently standardised developmental tests is the Bayley Scales of Infant and Toddler Development third edition (Bayley-III),14 which provides separate scores for cognitive, language and motor development. Perinatal trials typically use either the cognitive score or a combination of the domains to assess cognitive development.15 This ‘outcome’ is often combined with other measures of neuromotor and sensory impairment (eg, vision, hearing and cerebral palsy) to establish a ‘broad-spectrum’ assessment of neurodevelopmental outcome (see table 2). Opinions vary as to what combination should be used, but the broad approach to defining neurodevelopmental impairment is clear.11 However, there is no consensus on how death should be incorporated into the analysis of such composite primary outcomes.

DIFFERENT APPROACHES TO ACCOUNTING FOR DEATH IN CLINICAL TRIALS

Use of a composite outcome

One common approach in neonatal trials is to use a composite of death or neurodevelopmental impairment as the primary outcome. Neurodevelopmental impairment for these purposes is usually dichotomised (ie, present or absent). In this approach, a score on a particular psychometric test or combination of measures may be used as a ‘cut-off’ for defining an adverse outcome (see table 1). For example, standardised index scores more than 3 SDs below the normative mean of 100 (ie, scores <53) on the Bayley Scales of Infant Development second edition (BSID-II)13 are generally accepted as defining severe impairment. Therefore, a trial could be based on a primary outcome of infants who either died before 2 years corrected age or had a BSID-II index score <55. Several major national and international studies have used this approach (see table 3).

The main advantage of composite outcomes is that they add statistical efficiency, in terms of an increased number of events and therefore greater statistical power, as demonstrated by the National Institute of Child Health and Human Development trial24 on whole-body hypothermia for hypoxic ischaemic encephalopathy. The authors reported that whole-body hypothermia was associated with a reduction (risk ratio 0.72, 95% CI 0.54 to 0.95; p=0.01) in the primary outcome (death or moderate-to-severe neurodisability) compared with usual care in infants with moderate or severe hypoxic–ischaemic encephalopathy. However, the individual components of the primary outcome were not significant when analysed as secondary outcomes. This shows the benefit of using a composite primary outcome, especially when the components are important outcomes for clinicians and parents alike.

However, there are potential problems in defining the primary outcome in this way. For example, it cannot always be assumed that all components of the composite outcome will be affected by the intervention in the same direction.29 Composites work best when an intervention anticipated to reduce morbidity is also expected to improve survival and this may not always be true; a trial investigating target ranges of oxygen saturation in extremely preterm infants30 illustrates this. The oxygen saturation component of this factorial trial tested the hypothesis that a lower target range of oxygen saturation (85%–89%), as compared with a
higher target range (91%–95%), would reduce the incidence of the composite outcome of severe retinopathy of prematurity or death among infants who were born between 24 20 weeks and 27 6 weeks gestation. The results showed no evidence of a difference in the composite outcome overall. However, the study found that a lower target range of oxygenation (85%–89%), as compared with a higher range (91%–95%), resulted in an increase in mortality and a substantial decrease in severe retinopathy of prematurity among survivors. 10

Treating neurodevelopmental impairment as a dichotomous outcome in analysis

A further layer of complexity of treating neurodevelopmental impairment as a dichotomous outcome is that it effectively becomes ‘all or nothing’. For example, if a study defines a BSID-II index score of <70 as representing moderate-to-severe neurodevelopmental impairment, then a child with a score of 70 would be classified as unimpaired, while a child with a score of 69 would be classified as impaired, even though the difference between these scores is not clinically significant. In addition, in this case, moderate or severe neurodevelopmental impairment is mathematically treated equally as important as death. Clearly, this may be a reasonable compromise, but illustrates the problems that may arise when interpreting study results.

Furthermore, an intervention capable of producing a clinically significant difference in the mean neurodevelopmental outcome of the population may be completely missed. A randomised trial (MOMS) of prenatal versus postnatal repair myelomeningocele to postnatal surgery, 9 used the BSID-II Psychomotor Development Index (PDI) score as a secondary outcome. There was a significant difference in the mean PDI score between the two groups (p=0.03). However, when the proportion of infants who had a PDI score ≥50 was compared, there was no significant difference between the two groups (p=0.15). This was true even when a higher cut-off of 85 was used (p=0.06). Dichotomising a continuous outcome measure using a cut-off may lead to a loss of power. 31 Thus, a significant result on a continuous outcome may no longer be significant when the outcome is dichotomised.

| Assessment | Domains measured | Continuous scores | Classifying neurodevelopmental impairment |
|------------|------------------|-------------------|----------------------------------------|
| Examiner administered tests | i. Cognitive & Language | Standardised (Mean 100; SD 15) Mental Development Index (MDI) & Psychomotor Development Index (PDI) scores | SD-banded cut-offs for moderate (Index score –2 SD to –3 SD) and severe (Index score < –3 SD) impairment |
| Bayley Scales of Infant Development second edition (BSID-II) | ii. Motor | | |
| Bayley Scales of Infant & Toddler Development third edition (Bayley-III) | i. Cognitive | Standardised (Mean 100; SD 15) Cognitive, Language and Motor Composite scores | SD-banded cut-offs for moderate (Composite score –2 SD to –3 SD) and severe (Composite score < –3 SD) impairment |
| Griffiths Mental Development Scales-Revised: Birth to 2 years (GMDS 0–2) | i. Motor | | |
| Parent report measures | i. Non-verbal cognition | Parent Report Composite (PRC) score. | PRC composite score <44 for moderate/severe impairment 19; PRC <31 for severe impairment 19 |
| Parent Report of Children’s Abilities-Revised (PARCA-R) | ii. Language | | |
| Ages and Stages Questionnaires third edition (ASQ-3) | i. Communication | Total scores for Communication, Gross Motor, Fine Motor, Problem Solving and Personal–Social domains | Domain scores are compared with age-appropriate cut-offs for developmental delay to classify moderate/severe impairment (equivalent to standardised scores < –2 SD) |
| Clinical observation | i. Gross Motor | Five level classification system | Moderate impairment Level 2; Severe impairment levels 3, 4 or 5. |
| Gross Motor Function Classification System (GMFCS) | ii. Coordination and Performance | | |

Table 2

| Domain | Moderate disability | Severe disability |
|--------|---------------------|-------------------|
| Motor  | Cerebral palsy with GMFCS Level 2. | Cerebral palsy with GMFCS Level 3, 4 or 5. |
| Hearing| Hearing loss corrected or partially corrected with aids. | No useful hearing even with aids. |
| Vision | Moderately reduced vision but better than severe impairment or blind in one eye with good vision in contralateral eye. | Blind or can only perceive light. |
| Speech & Language | Some words or signs but fewer than 5 or unable to comprehend un-cued command but able to comprehend cued command. | No meaningful words or unable to comprehend cued command. |
| Cognitive function | Score –2 SD to –3 SD below the normative mean. | Score < –3 SD below the normative mean. |
Treating neurodevelopmental impairment as a continuous outcome and imputing a value for death

A number of studies have considered neurodevelopmental impairment as a continuous variable at analysis. Comparative analysis is performed on the scores attained on developmental tests between the intervention and control groups. How should authors account for death in such trials? Some studies have considered imputing a score for those participants who have died. For example, one option is to allocate an arbitrary low value on the Bayley scales for those participants who have died. This may be, for example, equivalent to 3 SDs below the normative mean, the conventional cut-off for severe disability. Despite being consistent with an intention-to-treat analysis using complete data, this approach involves compromises.

Table 3  Recent neonatal and perinatal trials where neurodevelopmental impairment is included in the primary outcome

| Trial                                                                 | Year published | Region                                      | Primary Outcome                                                                                                                                                                                                 | Death as a part of primary outcome |
|-----------------------------------------------------------------------|----------------|---------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| Benefits of Oxygen Saturation Targeting (BOOST-II UK)22               | Ongoing        | UK                                          | Composite of death or serious neurodisability at age 2 years corrected age (serious neurodisability defined as a Bayley-III language or cognitive score <85 or equivalent; or severe visual loss or severe cerebral palsy or deafness). | Yes                               |
| Total Body Hypothermia (TOBY)24                                       | 2009           | Finland, Hungary, Israel, Sweden             | Composite of death or severe neurodisability at 18 months of age (severe neurodisability defined as BSID-II Mental Development Index (MDI) <70, GMFCS Level 3 to 5, or severe visual loss). | Yes                               |
| National Institute of Child Health and Human Development (NICHD) trial on total body hypothermia24 | 2005           | USA                                         | Composite of death or moderate disability at 2 years corrected age (moderate disability defined as per criteria set out in the National Perinatal Epidemiology Unit (NPEU) and Oxford Regional Health Authority report and includes any major disability in neuromotor function, seizures, auditory function, communication, visual function, cognitive function and other physical disability). | Yes                               |
| GRIT (Growth Restriction Intervention Trial)25                        | 2004           | UK, Belgium, Czech Republic, Greece, Hungary, Italy, Netherlands, Poland, Portugal, Saudi Arabia, Slovenia | Composite of death or disability at or after 2 years corrected age (Disability defined as Griffith GQ ≤70 or diagnosis of cerebral palsy or severe visual loss or deafness) | Yes                               |
| INIS (International Neonatal Immunotherapy Study)19                   | 2011           | UK, Argentina, Australia, Belgium, Denmark, Greece, Ireland, New Zealand, Serbia        | Composite of death or major disability at 2 years corrected age. (Major disability defined as per criteria set out in the National Perinatal Epidemiology Unit (NPEU) and Oxford Regional Health Authority report and includes any major disability in neuromotor function, seizures, auditory function, communication, visual function, cognitive function and other physical disability). Cognitive delay defined as PARCA-R <31)19 | Yes                               |
| INNOVO Trial (Neonatal ventilation with INhaled Nitric Oxide versus Ventilatory support withOut inhaled nitric oxide for severe respiratory failure: a multicentre randomised controlled trial, RCT)26 | 2005           | UK, Ireland                                 | Death and disability at 1 year corrected age (Disability defined by set clinical criteria, no psychometric scales used).                                                                               | Yes                               |
| Trial of umbilical and fetal flow in Europe (TRUFFLE): a multicentre randomised study27 | Ongoing        | UK, Austria, Germany, Italy, Netherlands | Survival without neurodevelopmental impairment at 2 years corrected age (neurodevelopmental impairment defined as Bayley-III Cognitive composite score ≤70 or severe visual loss or GMFCS level ≥2 or deafness)                                                                 | Yes                               |
| Neurodevelopmental impairment as a continuous outcome                 |                 |                                             |                                                                                                                                                                                                               |                                    |
| Neonatal ECMO Study of Temperature (NEST): A Randomized Controlled Trial7 | 2013           | UK                                          | Bayley-III Cognitive composite score at 2 years corrected age (24–27 months)                                                                                                                                  | No                                 |
| A RCT of peer-mentoring for first-time mothers in socially disadvantaged areas (the MOMENTS Study)28 | 2011           | UK                                          | BSID-II MDI and Psychomotor Development Index (PDI) at 1 year                                                                                                                                                | No                                 |
| Iodine supplementation study in preterm infants (I2S2)18              | Ongoing        | UK                                          | Neurodevelopmental status at 2 years corrected (for prematurity). Neurodevelopmental status is defined by the three main domains of the Bayley-III scales, that is, cognitive score, language composite score and motor composite score. | Yes                               |
Allocating any single value (ie, using single imputation) to participants who have died is technically problematic since the data may not be missing at random. There is the compromise of assigning a similar, if not identical, score to those participants who have died with those severely disabled and also the scenario where one might impute a score for the deceased, which is higher than the minimum possible neurodevelopmental score for survivors on that scale. Trialists need to guard against this possibility when considering imputation. However, to impute a Bayley score implies that we know what the ‘trade off’ is between level of disability and death and imparts extra-neous value judgements, which could vary from individual to individual.

The use of a single imputation to assign a value to participants who have died also affects the precision and hence the interpretation of results, depending on the value assigned. At a more fundamental level, when planning a study, single imputation for participants who have died at analysis is likely to artificially inflate the overall SD thereby adversely affecting the precision of the results. This will consequently impact on the sample size and the appropriateness of the statistical test used.

### Focusing solely on neurodevelopmental impairment

An alternative to incorporating death into a composite primary outcome would be to consider the developmental test score within survivors only. In this case, the study findings would reflect the impact of the intervention solely on neurodevelopmental outcome, and not on death. However, such an analysis is a non-randomised comparison and therefore subject to an increased risk of bias, the chances of which are affected by the magnitude of the death rate and whether the death rate is differential across the groups being compared. Here the difference in test scores will be easy to interpret; and the child and the distribution across the two groups may not be balanced. In the two-stage statistical model, deaths will be modelled using a binomial test, and survivors modelled using a generalised linear model. The two parts will then be combined to form the appropriate test statistic.

A different approach was used in the MOMs study in which the second primary outcome, at 30 months, was a composite of death or serious neurosensory disability at 2 years corrected age. The trial was stopped early due to significantly increased mortality at 36 weeks post-menstrual age in the group treated with the lower oxygen saturation target. Since deaths in neonatal trials mainly occur in the first few weeks of life, monitoring safety in such trials, typically by an independent Data Monitoring Committee, requires the uncoupling of such composite primary outcomes for this purpose, given that the ‘whole’ primary outcome is not available until much later.

### Other approaches

The on-going OPPTIMUM trial is examining whether prophylactic vaginal progesterone to prevent preterm birth has long-term neonatal or infant benefit. For the analysis of the childhood primary outcome (Bayley-III cognitive composite scale at 2 years of age, a continuous measure), the investigators plan to incorporate deaths in a two-stage statistical model. Their rationale for the inclusion of deaths in the analysis is twofold; the number of deaths may not be negligible and the distribution across the two groups may not be balanced. In the two-stage statistical model, deaths will be modelled using a binomial test, and survivors modelled using a generalised linear model. The two parts will then be combined to form the appropriate test statistic.

### CONCLUSIONS

The recent change of focus within day-to-day neonatal care, with its increasing attention on reducing neurodevelopmental impairment, has been mirrored in the outcomes used in many perinatal trials. Clinical trials have increasingly incorporated neurodevelopment into their primary outcome and this has led to the question of how to deal with death in these studies. A range of possible solutions have emerged, each of which involves pragmatic statistical and clinical compromises, and there does not seem to be a correct approach.

The resources required to run large multicentre trials and the finite population of high-risk neonates limit the potential size and feasibility of neonatal trials. Catastrophic events (ie, the typical negative outcomes of interest) are thankfully uncommon, but this drives up the sample size unless we compromise and create meaningful composites. Further work is needed to clarify how, and to what extent, each of the designs and chosen analysis used to date can affect the findings and the impact of the trial. Where value judgements are needed, views of patient groups should be considered and may enable trialists to make better judgements regarding which approach to choose for their particular trial.

### Contributors

SAP, DJF and EJ were involved in conception of this review article. SAP wrote the initial draft manuscript. SP and SJ designed the tables. SAP, SJ, DJF and EJ were all involved in the design and contributed to subsequent drafts. All authors revised the final manuscript.

### Competing interests

None.

### Provenance and peer review

Commissioned; externally peer reviewed.

---

**The views of families**

The opinion of the ultimate beneficiaries of treatment (patients and families) may well be highly useful in identifying the appropriate ‘trade-off’ between neurodevelopmental impairment and death, necessary when considering all of the above approaches. However, it would not be possible to extrapolate the views of families involved in one trial to those in another, since the risk of death or disability will vary between trials. Hence, parental views of what ‘trade-off’ is acceptable must directly relate to a particular intervention in a specific clinical scenario.
REFERENCES

1. National Perinatal Epidemiology Unit. Disability and perinatal care: measurement of health status at two years: a report of two working groups convened by the National Perinatal Epidemiology Unit and the former Oxford Regional Health Authority; Oxford: National Perinatal Epidemiology Unit, 1994.

2. Stephens BE, Vohr BR. Neurodevelopmental outcome of the premature infant. Pediatr Clin North America 2009;56:631–46.

3. Robertson CMT, Watt M-J, Dinu IA. Outcomes for the extremely premature infant: what is new? and where are we going? Pediatr Neuro 2009;40:189–96.

4. Moore T, Hennessy EM, Myles J, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. BMJ 2012;345:e7976.

5. Strand M, Jobe AH. The multiple negative factors: why? Semin Perinatol 2003;27:343–50.

6. Norman JE, Shennan A, Bennett P, et al. Trial protocol OPPTIMUM—does progesterone prophylaxis for the prevention of preterm labour improve outcome? BMC Pregnancy Childbirth 2012;12:79.

7. Field D. Neonatal ECMO Study of Temperature (NEST): a randomized controlled trial. Pediatrics 2013;132:e1247–56.

8. Williams F, Hume R, Ogston S, et al. A summary of the iodine supplementation study protocol (IS2): a UK multicentre randomised controlled trial in preterm infants. Neonatology 2014;105:282–9.

9. Adzick NS, Thom EA, Spong CY, et al. A randomized trial of prenatal versus postnatal repair of myelomeningocele. N Engl J Med 2011;364:993–1004.

10. Costeloe KL, Hennessy EM, Hadler S, et al. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). BMJ 2012;345:e7976.

11. Marlow N. Measuring neurodevelopmental outcome in neonatal trials: a continuing and increasing challenge. Arch Dis Child Fetal Neonatal Ed 2013;98:F554–F8.

12. British Association of Perinatal Medicine. Report of a BAPM/RCPCCH Working Group: Classification of health status at 2 years as a perinatal outcome Version 1.0; 8 January 2008, London: 2008.

13. Bayley N. Bayley Scales of Infant Development. 2nd edn. San Antonio: Psychological Corporation, 1993.

14. Bayley N. Bayley Scales of Infant and Toddler Development. 3rd edn. San Antonio: Harcourt Assessment; 2006.

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

16. Griffith R. The Griffiths Mental Development Scales from birth to 2 years, manual, the 1996 revision. Henley: Association for Research in Infant and Child Development, Test Agency; 1996.

17. Johnson S, Marlow N, Wilke D, et al. Validation of a parent report measure of cognitive development in very preterm infants. Dev Med Child Neurol 2004;46:389–97.

18. Johnson S, Wilke D, Marlow N. Developmental assessment of preterm infants at 2 years: validity of parent reports. Dev Med Child Neurol 2008;50:58–62.

19. The INN Collaborative Group. Treatment of neonatal sepsis with intravenous immune globulin. N Engl J Med 2011;365:1201–11.

20. Squires J, Twombly E, Bricker D, et al. The Ages and Stages Questionnaires-3 User’s Guide. Baltimore, MD: Brookes Publishing; 2009.

21. Palisano R, Rosenbaum P, Walter S, et al. Development and reliability of a system to classify gross motor function in children with cerebral palsy. Dev Med Child Neurol 1997;39:214–23.

22. Stenson BJ, Tamrow-Mordi WD, Darlow BA, et al. Oxygen saturation and outcomes in preterm infants. N Engl J Med 2013;368:2094–104.

23. Azizpooz D, Strohm B, Edwards AD, et al. Moderate hypothermia to treat perinatal asphyxial encephalopathy. N Engl J Med 2009;361:1349–58.

24. Shankaran S, Lantopk AR, Ehrenkranz RA, et al. Whole-body hypothermia for neonates with hypoxic–ischemic encephalopathy. N Engl J Med 2005;353:1574–84.

25. The GRIT Study Group. Infant wellbeing at 2 years of age in the Growth Restriction Intervention Trial (GRIT): mult centred randomised controlled trial. Lancet 2004;364:513–20.

26. Field D, Elbourne D, Tresdale A, et al. Neonatal ventilation with inhaled nitric oxide versus ventilatory support without inhaled nitric oxide for preterm infants with severe respiratory failure: the INNOV multicentre randomised controlled trial (ISRCTN 17821339). Pediatrics 2005;115:926–36.

27. Lees C. Protocol OPTRT34: trial of umbilical and fetal flow in Europe (TRUFFLE): a multicentre randomised study 2007. http://www.thelancet.com/protocol-reviews/OPTRT-34.

28. Cupples ME, Stewart MC, Percy A, et al. A RCT of peer-mentoring for first-time mothers in socially disadvantaged areas (the MOMENTS Study). Arch Dis Child 2011;96:252–8.

29. Freeman M, Calvert M, Wood J, et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? JAMA 2003;289:2554–9.

30. SUPPORT Study Group of the Eunice Kennedy Shriver NICHD Neonatal Research Network. Target ranges of oxygen saturation in extremely preterm infants. New Engl J Med 2010;362:1959–69.

31. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ 2006;332:1080.

32. Rubin DB. Causal inference through potential outcomes and principal stratification: application to studies with “censoring” due to death. Stat Sci 2006;21:299–309.