Personal practice

Follow up studies: design, organisation, and analysis

L M M MUTCH,* M A JOHNSON,† AND R MORLEY‡

*Social Paediatric and Obstetric Research Unit, University of Glasgow, †Oxford Region Child Development Project, John Radcliffe Hospital, Oxford, and ‡Medical Research Council Dunn Nutrition Unit, Cambridge

The follow up of infants who were exposed to adverse perinatal events is an important component of perinatal audit. We need to know whether the changes in perinatal care that have undoubtedly resulted in their improved survival have altered the incidence of impairments in their long term health, growth, or neurodevelopmental state. Moreover, recent improvements in diagnostic imaging techniques carried out in the neonatal period may considerably enhance our ability to predict longer term outcome.^{1–3} Their importance lies not only in their use to monitor the eventual outcome for individual children, but because therein could lie some explanation of the mechanisms linking clinical pathological entities to the lesions visualised.⁴

In the absence of reliable validated routine health surveillance of all children, special studies are required to monitor the effects of rapidly changing perinatal practice and to evaluate specific components of care. Studies from any one centre are often too small to answer the questions being posed. The purpose of this paper is to discuss ways in which the design of following studies could be made compatible to enable multicentre collaboration, and thus provide more rapid feedback to neonatal paediatricians on current issues.

Background

In 1981, Kiely and Paneth outlined some of the theoretical problems that beset the design and reporting of follow up studies.⁵ In the same year, Sinclair *et al* expanded on the methodological requirements for a proper evaluation of neonatal intensive care.⁶ This included testing of the efficacy of particular treatments, their effectiveness in the field, the efficiency of these treatments in relation to other calls on resources, and the implications (economic and otherwise) of making them available to all who need them. In this paper we discuss some

of the other issues and practical problems encountered in mounting follow up studies and suggest some solutions, but our main aim is to stimulate further discussion that will eventually encourage more effective use of the vast amount of resource and expertise expended in the name of 'follow up'.

How is the information derived from follow up studies used?

Most reported follow up studies based on hospital populations are descriptive or observational; that is, information is collected on the outcome of cohorts of children with particular characteristics—for example birthweight groups, infants who were ventilated, infants with seizures, and so on. This information tends to be misused in two ways. Firstly, associations between perinatal variables and outcome are observed, and there is then a tendency to ascribe—erroneously—a causal relationship between the two. Descriptive follow up studies can, however, only generate hypotheses that need to be tested in a new cohort. Causal relationships can be determined only by a randomised study design. Secondly, the information collected is compared with other similar observations made in previous years or in other centres. Differences in outcome are ascribed to differences in care, but they may merely reflect differences in the study population (either over time or between centres), or a lack of uniformity in the way outcomes have been defined or ascertained.

How can we improve on study design?

The design of the study must be prospective, with the aims, the population, and the outcomes defined before collection of data starts. The selection of a comparison group must also be considered.

DEFINITION OF THE AIMS

The aim of the study should be defined at the outset and usually falls into one of the following categories: to estimate the incidence of disease, disorder, or impairment in a defined cohort followed up for a specific period; to evaluate the effectiveness or otherwise of a particular therapeutic intervention; or to test a specific aetiological hypothesis.

A descriptive follow up study of the first type can provide important information on the natural history and prognosis of a condition, and can identify the presence or absence of associations between perinatal events and outcome. The evaluation of specific treatments is best achieved by randomised clinical trials with predefined measurements of outcome made by investigators who are preferably 'blind' to the original treatment. The role of each type of study is illustrated in the following example.

A study of the natural history and prognosis of a group of infants with neonatal seizures identified a subgroup of infants born at full term whose seizures occurred in the first 48 hours of life and who had sustained an intrapartum insult.⁷ This gave rise to the aetiological hypothesis that changes in intrapartum management might reduce the number of infants born at full term with early seizures. It was then postulated that the rate of such seizures might provide a useful marker of the quality of intrapartum care.⁸ Two case control studies were set up to investigate the relationship between various adverse outcomes of pregnancy including early neonatal seizures and the quality of care in pregnancy and labour; both showed that there was a strong association between suboptimal care and early seizures.^{9 10} The hypothesis was sustained in a large randomised controlled trial of intrapartum electronic fetal heart rate monitoring.¹¹

DEFINITION OF THE STUDY COHORT

The criteria by which the study cohort is selected should be clearly defined and will be determined by whether the study is investigating prevalence, or the effects of an intervention, or testing a specific aetiological hypothesis.

In a study of the prevalence of morbidity, the population should be geographically defined.⁵ This permits evaluation of care to a total population whose characteristics can be defined, and hence the findings can be compared with other populations for which reference statistics are available, taking account of any differences in these variables. Results from tertiary referral centres are difficult to interpret, because they are confounded by the transfer of mothers and infants between units. An example of the importance of this source of bias was

described by Horwood *et al* who showed that 28% of the reduction in mortality experienced after the opening of a new intensive care centre was accounted for by the transfer in from outside the geographical region of mothers who had been thought, incorrectly, to be at high risk of a poor outcome.¹²

Enrolment in such a study can be on the basis of all births, or all live births. It is, however, important to ensure that a birth is also defined by some other variable such as—for example, all births over 500 g—because there is marked variation in the definition of live birth both internationally and within countries.¹³ For the purposes of international comparison, it is preferable to use standard categories such as those recommended by the World Health Organisation for birthweight groupings.¹⁴

In the United Kingdom, information on all births to women resident in a defined geographical area can be obtained through statutory birth notifications that are held by district medical officers in England and Wales, and chief administrative medical officers in Scotland and Northern Ireland. In most district general hospitals outside the big conurbations, the difference between the hospital based population and the geographically based one is not large. Those infants who are in the geographical cohort, but who were not included in the hospital population, can be easily identified and included. In large cities where infants could be cared for in a wide range of tertiary care centres dictated by the availability of an intensive care cot, it is not possible to evaluate the quality of care from information on the outcome of infants treated in a single unit.

In conducting studies to test aetiological hypotheses, or intervention studies, it is not necessary to use a geographically defined population. Once the study cohort has been defined it is important that all are accounted for in the analysis. In the special case of an antepartum or intrapartum intervention study, the women entered into the study and all the infants born to those women should be included in the evaluation of the intervention. For example, in a trial of a tocolytic agent to prevent preterm labour, the condition of all the infants resulting from all the pregnancies being studied needs to be compared, whether they are born prematurely or not. It could—for example—be argued that postponing delivery might subject the fetus to additional unsuspected hazards.

DETERMINATION OF SAMPLE SIZE

The sample should be large enough to provide an answer, either on its own or in conjunction with other similarly structured studies, to the main question being asked in the study. The involvement

of parents and children in a follow up study that has little chance of producing an answer to the question posed should be regarded as unethical.

In a study of prevalence in a defined population, the study size is obviously constrained by the available eligible entrants. In testing a hypothesis it is important that the sample size is big enough to permit identification of enough subjects with the predefined outcome. For example, there is little point in examining 100 infants for the incidence of an outcome that occurs in only 1/1000. Furthermore, the sample needs to be large enough to detect intergroup differences that are also biologically important.

It is difficult to give general advice because each study has a different perspective; a statistician should be consulted to advise on the necessary calculations based on the reported frequency of a given dichotomous outcome, or some measure of central tendency and dispersion such as the mean and standard deviation of a continuous outcome measure such as growth or a developmental score.¹⁵ It should be borne in mind that the standard deviation for a developmental score is likely to be greater in a 'high risk' population than in the standardisation sample for the test.

In order to study infrequent outcomes (such as retinopathy of prematurity) or a limited population (such as infants weighing less than 1000 g at birth) within a reasonable time span, multicentre collaboration is essential to achieve sufficient numbers. For this to be successful, standard methods for identifying and reporting those outcomes are needed.

DEFINITION OF OUTCOMES

The outcomes to be measured will depend on the aim of the study, but whatever they are they must be specified and clearly defined before the start of the study. The outcome should be relevant to the child or family, or likely to improve our understanding of the aetiology or treatment of the condition being studied. The availability of funds or personnel, or both, may impose constraints on the study, making the choice of the most important and valid outcome measures even more essential.

Commonly reported outcomes are death, neurodevelopmental state, sensory impairment, morbidity, growth, behavioural attributes, and learning ability. It is important that the outcome to be measured can be defined accurately. In the case of death, for example, the definition is simple, whereas neurodevelopmental impairment or emotional problems would require careful definition before a study was undertaken.

In making a full evaluation of the effects of any treatment it is important to consider outcomes that

have economic implications. Measures such as the need for hospital admission, length of stay, the need for increased primary health care contacts, and the use of ancillary services are important factors that could influence whether recommendations based on the findings of a particular study are accepted.

Measuring outcomes

The tests (tools) used in measuring outcome should have both reproducibility and validity. These terms are illustrated in the figure (JC Sinclair, personal communication).

Even with standardised 'tests' inaccuracies, imprecision, and biases can arise. There is a real chance of interobserver variation—and also intra-observer variation—and these need to be tested in a pilot study. It has been shown that further training can help to improve the level of agreement between observers.¹⁶ The problem is accentuated when non-standardised tests, sometimes devised by the examiner, are used before being subjected to rigorous evaluation. There is a large number of tests currently in use for assessing a wide variety of functions at differing ages and we do not intend to discuss specific ones here. There is, however, a need for some consensus on the optimal ages for testing children and the most appropriate tests at each age in order to achieve comparability between studies.

Outcome measures that require intensive assessment by highly trained personnel are not feasible if the study population is large. Conversely, smaller studies on their own, or 'nested' in large trials, may provide the opportunity for in depth examination of unsuspected effects of the treatments being compared.

Other sources of outcome information

In addition to data collected in the study there are several other sources which can be used to find out about the longer term health and development of children. Information on deaths in the study cohort can usually be obtained from routinely collected data held by districts on the child health register, as the death of a child is entered immediately to ensure that parents are not contacted for routine appointments.

Where appropriate, for a small cost, the National Health Service record of individuals held by the Office of Population Censuses and Surveys can be 'flagged' so that the researcher will be informed of the death or emigration of study children, or their movement from one family practitioner committee area to another.

Parents, health visitors, general practitioners, community child health doctors, hospital paediatricians, and teachers as well as research workers

Figure Reproducibility and validity: (a) the shot is valid (it hit the target) and reproducible (it hit the target every time); (b) one shot is valid, but the others went wide (not reproducible); (c) shots all hit the same spot (reproducible), but missed the target (not valid); and (d) shots neither valid nor reproducible.

committed to the study can all provide their own perspectives on the child. The ideal, but expensive, situation would be to draw on all these sources, using them to validate each other. The advantages and disadvantages of the various sources are summarised in table 1. There is a need for validation of inexpensive methods of obtaining information on children's health and development. Such validation studies should be incorporated in more detailed and costly studies.

Reporting outcomes

In order to compare outcomes of studies from different centres the diseases, disorders, or impairments should be defined in a standard way.

It is helpful to consider outcomes in the way described in the *International Classification of Impairments, Disabilities and Handicap*.¹⁷ A disease or disorder causes an impairment, which is defined as '... any loss or abnormality of psychological,

physiological or anatomical structure or function'. A disability resulting from this impairment is '... any restriction or lack of ability to perform an activity in the manner or within the range considered normal for a human being'. Handicap is the disadvantage resulting from an impairment or disability that '... limits or prevents the fulfilment of a role that is normal (depending on age, sex, and social and cultural factors) for that individual'.

Impairments—Of these three terms, impairments are the most amenable to definition. For example, consensus has been reached on the standard description and reporting of the grades of retinopathy of prematurity.¹⁸ A working group has developed a standard format for the recording and reporting of the motor deficit in cerebral palsy.¹⁹ An attempt to provide uniformity in describing and identifying neurological impairments has been proposed by Amiel-Tison and Stewart.²⁰

Table 1 *Summary of methods of follow up*

<i>Sources of information</i>	<i>Advantages</i>	<i>Disadvantages</i>
Routinely collected data (available in UK):	Inexpensive	Lack of uniformity between districts in child health surveillance
Preschool health (school health) system	Accommodate large study populations	Limited detail
Hospital activity analysis		
Parents:		
Parental questionnaires	Inexpensive Accommodate large study populations Unique knowledge of child's abilities and disabilities and social and behavioural attributes	Inaccuracy of recall 'Subjectivity' bias Inadequately validated
Health diaries	Improved accuracy of recall	'Subjectivity' bias Inadequately validated
Health visitors:		
Assessment	No additional funding if part of routine surveillance Can accommodate quite large numbers Specific training feasible	Additional workload Unstandardised 'tools' Interobserver variation
Teachers:		
Assessment	No additional funding if part of routine Can accommodate quite large numbers Use of standardised tests Explores cognitive development	Additional workload
Hospital based personnel:		
Routine outpatient	No additional visits	Lack of uniformity in testing Interobserver variation Limited time for assessment Clinical environment
Special study personnel:		
Standard protocol	Longer time for assessment Can be seen in own surroundings In depth study possible	Additional appointments Cost
Multiple observers	Larger numbers possible	Interobserver variation
Single observer	Relative consistency in observation In depth study possible Finer discrimination possible between groups	Small numbers only Cost Undetected systematic error

Disability—Assigning grades of severity to the disability arising from impairments is fraught with problems. In prevalence studies it is important that the 'cut off' point for ascertaining the condition under review is predefined for comparative purposes, as children with minimal problems may well not be universally recognised, and the reader needs to know the threshold of inclusion.

The description of disability needs to be related to a particular task and the developmental age of the child and must take account of the use and usefulness of aids.²¹ In most currently reported studies, however, there is a lack of uniformity in describing disability in children and this is compounded by the arbitrary categorisation of disabilities into grades of 'mild', 'moderate', and 'severe', which may have different meanings in different studies. Furthermore, a physician's perception of

the severity of a disability will depend on his or her clinical experience.

In a population based study in the Northern region of England a system was developed to categorise the degree of disability in a group of children with established cerebral palsy in which aspects of mobility, physical dependence, social and economic implications and resource use were summed to provide an overall measure of disability.²²

In a recent study of disabilities in children by the Social Survey Division of the Office of Population Censuses and Surveys, considerable attention was paid to all these points; it will be a useful reference source.²³

One possible way forward would be the development of a 'minimum data set' incorporating standard ways of categorising death, accepted definitions of impairments, and guidelines on assigning grades of

severity. This would permit meaningful accurate comparisons between studies.

Handicap—This is a particularly difficult outcome for comparative use, as it depends heavily on subjective factors and is a reflection of the attitudes within the family and society as a whole to people with disabilities. It is best avoided in the types of study described here.

CHOICE OF A COMPARISON GROUP

In a controlled trial the control group should be randomly chosen from those eligible for entry to the study and therefore should be comparable with the index group in every respect other than the intervention being studied.

In an observational study there is no easy answer to the choice of a comparison group.⁵ As previously stated, comparison of the outcome of hospital populations either over time (historical controls) or between centres may result in misleading inferences.

The choice of a comparison group depends on the question being asked. The choice must be between matched and 'unmatched' controls; if the preference is for matched controls, then subjects could be matched on antecedent factors such as maternal age, parity, birth weight, or gestational age, depending on the design of the study. In considering developmental outcomes, however, the over-riding determining factors are sex, social class,²⁴ and preschool experience,²⁵ so there is a rational basis for matching on those variables. Often it is better to match for only exceptional factors and adjust for others in the analysis. Overmatching can introduce difficulties in interpretation by taking out of consideration the very factor being investigated.

In a prevalence study in, for example, a very low birthweight population where the outcome to be measured is relatively uncommon, more than one control for each index case may be necessary to provide more accurate estimates of the increased relative risk experienced by the study subjects. But the advantage to be gained by increasing the number of controls soon falls off¹⁵; statistical advice is needed on this point.

The use of siblings as a comparison may have the advantage that socioeconomic variables and experience of parenting are likely to be the same. In some families, however, there will be no siblings and any finding limited to those cases who had a sibling would not necessarily be relevant for the one child family. Birth order in itself has been shown to have a significant effect on developmental score almost equal to that of the sex of the child.²⁶

At an older age group, classroom controls of the same sex and social class are an attractive alterna-

tive, but this necessarily excludes children not able to attend school who may be represented to a greater extent in the index group.

If a concurrent comparison is carried out between two geographically defined populations with the same ascertainment criteria for measures of outcome, observed differences may indeed reflect differences in care; these may, however, also be the results of known differences in the sociodemographic structure of the populations, or to some as yet unidentified confounding factor. Information should be collected on factors known to be related to the outcome being measured, so that adjustment can be made in the analysis for any marked imbalance.

TRACING CHILDREN

Tracing children can be time consuming, and several sources may need to be tapped.²⁷ High ascertainment rates are extremely important, because biases arise when follow up rates are low and it should be recognised that the lower the follow up rate the more inaccurate the estimate of the true rate of impairment. There is, furthermore, some evidence that children who are difficult to trace may be at a higher risk of an abnormal outcome. The extent to which this is true has recently been quantified in a study of very low birthweight children in the Northern region of England; children not seen on the first attempt had a sevenfold increased risk of impairment compared with those who were easily traced.²⁸

If parents are aware from the time of birth or soon after that their children will form part of a study, contact can be maintained in a number of ways. In the 1946 British cohort used to study child health, birthday cards are used to keep in touch.²⁹ Assess-

Table 2 Information required to tap sources helpful in tracing children

Child:	Full name, date of birth, last address, National Health Service number, telephone number
Mother:	Full name, maiden name, National Health Service number
Father:	First name or initials
General practitioner:	Name, address
Grandparents:	Name, address, telephone number
Useful sources:	Child health registers held by districts Family practitioner committees – registration departments National Health Service Central Registry in Southport (England), in Edinburgh (Scotland), and in Belfast (Northern Ireland) Telephone directories Local education authorities for children of school age Local post offices and electoral rolls

ments in the early months of life can help in establishing rapport with the family. The information that can be helpful in tracing children within the United Kingdom, and agencies and other sources that can be used, are outlined in table 2.

In most follow up studies there are a few children who cannot be traced. Every effort should be made to restrict their number as much as possible. It is important to record the number of untraced children and to compare their known characteristics with those of the children assessed.

Analysis and reporting of results

The selected study groups form the basis for the analysis. Statistical advice should be sought to ensure that the analysis of the data and interpretation of the results are appropriate, and this should be a stated part of the initial design of the study. There are several circumstances that deserve special consideration.

REFUSAL BY PARENTS TO PARTICIPATE

Parents may refuse because they suspect the motives of the assessor, particularly if their child is already the subject of concern on the part of the welfare authorities. A further explanation of the aims of the study will often secure their cooperation. In other cases parents may suspect that their child is not developing normally and do not welcome confirmation of their suspicions. It is also understandable that assessments may be unwelcome when family dynamics are disturbed, or where a child is severely impaired and already having regular assessments. Information from general practitioners and paediatricians may assist in these cases.

REFUSAL OF CHILDREN TO PARTICIPATE

Some children refuse to comply with testing. There is evidence that they do so because they are unable to do a test.³⁰ For those testing small children, it is often difficult to decide whether a child is failing to do a test or refusing to try it. Experience gained in a pilot study will help in defining a 'refusal'. Ounsted *et al* recommend exclusion of such children from the analysis of the variable concerned with a clear statement about their numbers and characteristics,³⁰ and Stewart (AL Stewart, personal communication) recommends assigning them to a grade above failure but below pass (for example, the 'suboptimal range' in the McCarthy test³¹).

EXCLUSIONS AFTER ASSESSMENT

It is difficult to make a general statement about this issue. In a randomised trial, once subjects have been assigned to a treatment they should all be included

in the analysis. In an observational study of the developmental outcome of for example, children whose mothers had had some late onset pregnancy condition, then the inclusion of a child with a genetically determined syndrome would be inappropriate. When language development is a principal outcome of interest, then children whose first language differs from that of the tester are at a disadvantage, and not only in the sphere of language. Decisions about how to deal with these situations should be discussed at the outset of the study.

DISTRIBUTION OF CONFOUNDING VARIABLES

The first priority is to determine the distribution in the study group, and in the population from which the study sample was derived, of characteristics that might affect outcome such as birth weight, gestational age, sex, social class, birth order, and pre-school experience.

DESCRIPTION AND CLASSIFICATION OF OUTCOME VARIABLES

The distribution of the outcome variables in the whole study group and separately in the comparison groups should be examined first. If the results in each group are normally distributed and have the same variance, comparison of means and standard deviations will be appropriate. Information on the number of subjects whose results fall below the 10th centile can be a useful additional way to describe the results especially if they are variables like developmental quotients or anthropometric measures. By itself, comparing groups of children falling below the 10th centile for a test may be a crude measure, and it is analogous to measuring the height of a group of people by placing a bar at the strategic height and seeing how many people could walk under it. Nevertheless, using cut off points of this type can also serve to increase the chance of data dependent choices to highlight or minimise apparent differences.

Where the results are skewed (that is, not normally distributed) statistical advice on the best and most informative approach should be sought. In the absence of a comparison group the findings can be described in terms of both means and standard deviations, and the numbers falling below the 10th centile.

Costing and funding

Follow up studies that do not use routine sources of information are usually costly to mount; they are usually project funded and estimates of their true cost are difficult to come by. It would be valuable for future researchers to have information on how

Table 3 Costing a follow up study

Personnel costs:
Study design, including statistician's time
Tracing children
Training of personnel
Conducting a pilot study
Testing interobserver variation
Organising appointments or questionnaires
Assessing children
Communicating findings to those responsible for clinical care
Data entry and processing
Data analysis (with computing and statistical advice)
Writing up time
Administrative and clerical costs:
Preparation of assessment protocols
Stationery
Photocopying, telephone, and postage
Equipment costs:
Weighing and measuring equipment
Kits for developmental tests
Standard test recording forms
Computer and other office equipment
Travel costs:
Costs for either parents or study personnel

much a study has cost available as part of an appendix to publication. The items to be considered are outlined in table 3.

Estimates of travel costs depend on local circumstances and how far removed in age the subjects are from the age at entry into the study. For example, of a birth cohort contacted first at the age of 4 years, 80% were still living within 25 miles of the district general hospital in which they were born, 12% were between 25 and 100 miles away, and 1% were more than 200 miles away²⁵ (unpublished data).

Ethical perspective on follow up studies

It could be said that a physician has a responsibility to monitor the outcome of any treatment prescribed as part of routine health care: to ensure—for example—that a child's hearing is intact after treatment for an ear infection. The responsibility is perhaps less well defined in the perinatal period where the effectiveness of treatment for the mother is not always interpreted in the light of possible effects on the child.

The follow up systems developed by most special care nurseries provide a continuing clinical service for patients and their parents and have a counselling and supportive role as well as being a means of assessing overall health and development. Carefully designed follow up studies can be incorporated within these systems, but a fine balance needs to be

achieved between fulfilling the needs of individual patients and meeting the rather stringent requirements of a research study.

Ethical approval must be sought for follow up where this is not considered an integral part of continuing clinical care. This is particularly important if a 'normal' control group is enrolled. Parents' consent should be sought after an explanation of the purpose of the study and what is entailed in the follow up assessment. This can be done orally or by letter, and parents can consent either by opting in or opting out.³² General practitioners should be informed that their patient has consented to take part in a study. Results should be communicated promptly so that they are as relevant as possible to current practice; parents and general practitioners are entitled to receive the reports of assessments.

If we are concerned to monitor the effects of changing perinatal practice, then it must also be appreciated that the effect of follow up studies in themselves has never been evaluated fully. On the one hand, benefits may accrue from the early identification of—for example—sensory deficit and the implementation of treatment; on the other hand it may be that the identification of other problems of development, some of which cannot be 'cured', might cause unnecessary anxiety to parents and children. It is therefore important that studies should not be undertaken unless they are properly designed and of sufficient size to answer specific questions.³³ This counsel of perfection could well limit the number of studies but would undoubtedly increase their usefulness.

We acknowledge with thanks the helpful advice of Professor E Alberman and Dr G Escobar who commented on a later draft, and colleagues in the National Perinatal Epidemiology Unit who read earlier drafts; also the help of the department of medical illustration at the University of Edinburgh.

References

¹ Stewart AL, Thorburn RJ, Hope PL, Goldsmith M, Lipscomb AP, Reynolds EOR. Ultrasound appearance of the brain in very preterm infants and neurodevelopmental outcome at 18 months of age. *Arch Dis Child* 1983;**58**:598-604.

² de Vries LS, Dubowitz LMS, Dubowitz V, *et al*. Predictive value of cranial ultrasound in the newborn baby: a reappraisal. *Lancet* 1985;**ii**:137-40.

³ Stewart AL, Reynolds EOR, Hope PL, *et al*. Probability of neurodevelopmental disorders estimated from ultrasound appearance of brains of very preterm infants. *Dev Med Child Neurol* 1987;**29**:3-11.

⁴ Bax MO. Forward from neonatology. *Dev Med Child Neurol* 1987;**29**:1-2.

⁵ Kiely JL, Paneth N. Follow up studies of low birthweight infants: suggestions for design, analysis and reporting. *Dev Med Child Neurol* 1981;**23**:96-9.

⁶ Sinclair JC, Torrance GW, Boyle MH, Horwood SP, Saigal S, Sackett DL. Evaluation of neonatal intensive care programs. *N Engl J Med* 1981;**305**:489-94.

- ⁷ Dennis J. Neonatal convulsions: aetiology, late neonatal status and long term outcome. *Dev Med Child Neurol* 1978;**20**:143-58.
- ⁸ Dennis J, Chalmers I. Very early neonatal seizure rate: a possible epidemiological marker of the quality of perinatal care. *Br J Obstet Gynaecol* 1982;**89**:418-26.
- ⁹ Niswander K, Elbourne D, Redman C, Henson G, Chalmers I, Macfarlane A. Adverse outcome of pregnancy and the quality of obstetric care. *Lancet* 1984;**ii**:827-31.
- ¹⁰ Minchom P, Niswander K, Chalmers I, et al. Antecedents and outcome of very early neonatal seizures in infants born at or after term. *Br J Obstet Gynaecol* 1987;**94**:431-9.
- ¹¹ MacDonald D, Grant A, Sheridan-Pereira M, Boylan P, Chalmers I. The Dublin randomised controlled trial of intrapartum fetal heart rate monitoring. *Am J Obstet Gynecol* 1985;**152**: 524-39.
- ¹² Horwood SP, Boyle MH, Torrance GW, Sinclair JC. Mortality and morbidity of 500-1499 g birthweight infants liveborn to residents of a defined geographic region before and after neonatal intensive care. *Pediatrics* 1982;**69**:613-20.
- ¹³ Mugford M. International comparisons of definitions of vital events. *World Health Stat Q* 1983;**36**:201-12.
- ¹⁴ World Health Organisation. *International classification of diseases*. 9th Revision. Geneva: WHO, 1970.
- ¹⁵ Gore SM, Altman DG. *Statistics in practice*. London: British Medical Association, 1982.
- ¹⁶ Blair E, Stanley FJ. Interobserver agreement in the classification of cerebral palsy. *Dev Med Child Neurol* 1985;**27**:615-22.
- ¹⁷ World Health Organisation. *International classification of impairments, disabilities and handicaps*. Geneva: WHO, 1980.
- ¹⁸ International Committee for the Classification of Retinopathy of Prematurity. An international classification of retinopathy of prematurity. *Pediatrics* 1984;**74**:127-33.
- ¹⁹ Evans P, Johnson A, Mutch L, Alberman E. A standard form for recording clinical findings in children with a motor deficit of central origin. *Dev Med Child Neurol* 1989;**31**:119-27.
- ²⁰ Amiel-Tison C, Stewart A. Follow up studies during the first five years of life: a pervasive assessment of neurological function. *Arch Dis Child* 1989;**64**:496-502.
- ²¹ Saigal S, Rosenbaum P, Stoskopf B, Milner R. Follow-up of infants 501-1500 g birthweight delivered to residents of a geographically defined region with perinatal intensive care facilities. *J Pediatr* 1982;**100**:606-13.
- ²² Jarvis S, Hey E. Measuring disability and handicap due to cerebral palsy. In: Stanley F, Alberman E, eds. *The epidemiology of the cerebral palsies*. Clinics in Developmental Medicine No 87. Spastics International Medical Publications. London: Heinemann Medical Books, 1984.
- ²³ OPCS Surveys of Disability in Great Britain. Report No 3. *The prevalence of disability among children*. London: HMSO, 1989.
- ²⁴ Ounsted MK, Osborn ML, Sleight G, Good FJ. A method of developmental assessment at 4 years and some associated findings. *Early Hum Dev* 1979;**3**:1-19.
- ²⁵ Dennis J, Johnson A, Mutch L, Yudkin P, Johnson P. Acid-base status at birth and neurodevelopmental outcome at four and a half years. *Am J Obstet Gynecol* (in press).
- ²⁶ Morley R, Cole TJ, Powell R, Lucas A. Mother's choice to provide breast milk and developmental outcome. *Arch Dis Child* 1988;**63**:1382-5.
- ²⁷ Hewitt A. In: Illsley R, Mitchell R, eds. *Low birth weight: a medical, psychological and social study*. Chichester: John Wiley, 1984:105-32.
- ²⁸ Wariyar U, Richmond S. Morbidity and preterm delivery: importance of 100% follow up. *Lancet* 1989;**i**:387-8.
- ²⁹ Douglas JWB, Blomfield JM. *Children under five*. London: Allen and Unwin, 1958.
- ³⁰ Ounsted MK, Cockburn J, Moar VA. Developmental assessment at four years: are there differences between children who do or do not cooperate? *Arch Dis Child* 1983;**58**:286-9.
- ³¹ McCarthy scales of children's abilities. New York: Psychological Corporation, 1970.
- ³² Mutch L, King R. Obtaining parental consent - opting in or opting out? *Arch Dis Child* 1985;**60**:979-80.
- ³³ Aylward GP. Developmental assessment: caveats and a cry for quality control. *J Pediatr* 1987;**110**:253-254.

Correspondence to Dr LMM Mutch, Social Paediatric and Obstetric Research Unit, University of Glasgow, 1 Lilybank Gardens, Glasgow G12 8RZ.