Evidence of selection bias in preterm survival studies: a systematic review

D J Evans, M I Levene

Abstract

Objective—To determine by how much selection bias in preterm infant cohort studies results in an overestimate of survival.

Design—Systematic review of studies reporting survival in infants less than 28 weeks of gestation published 1978–1998. Studies were graded according to cohort definition: A, stillbirths and live births; B, live births; C, neonatal unit admissions. Proportions of infants surviving to discharge were calculated for each week of gestation.

Results—Sixty seven studies report data on 55 cohorts (16 grade A, 23 grade B, 16 grade C). Studies that are more selective report significantly higher survival between 23 and 26 weeks of gestation (grade C > grade B > grade A, p < 0.01), exaggerating survival by 100% and 56% at 23 and 24 weeks respectively.

Conclusion—To minimise the potential for overestimating survival around the limits of viability, future studies should endeavour to report the outcome of all pregnancies for each week of gestation (terminations, miscarriages, stillbirths, and all live births).

Keywords: preterm; survival; selection bias

The collection of valid data relating to the survival of preterm infants is important for a number of reasons. It is necessary for clinicians to counsel pregnant women accurately with regard to potential survival. Neonatal and perinatal mortality is often used as an outcome measure in studies evaluating effectiveness of interventions. There is an increasing tendency to compare survival data between institutions as an audit of performance of obstetric and neonatal services, despite problems identified with this approach.1

A large number of publications present survival data for preterm infants related to each week of gestation. Inspection of the survival figures shows great heterogeneity in the estimates of survival, particularly in the 23–27 week gestational age groups.

There are a number of potential explanations for such wide variation in reported survival. The differences between studies may be simply a reflection of differences in the population under study (clinical heterogeneity). Factors that may vary across studies are the sociodemographic characteristics of the cohort, the time period of the study, or uptake of interventions known to improve survival, such as antenatal steroids and postnatal surfactant treatment.

It is also possible that systematic error within studies will vary and cause variation in reported survival. Specifically, there is potential for selection bias in preterm cohort studies that only report survival in infants admitted to neonatal units and, to a lesser extent, in studies that include live births but exclude stillbirth data. Cohort studies comprising exclusively neonatal unit admissions may ignore live births that could not be (or were not) resuscitated and will therefore overestimate survival. Studies that report all live births but do not report stillbirth data may overlook pregnancies that were not monitored during labour in the expectation that the pregnancy would result in stillbirth because of poor prognostic factors. Such selection bias results in the reporting of data relating to infants and fetuses with a better prognosis.
and is likely to lead to an overestimate of survival.

Selection bias may also be present in cohorts that are defined by hospital institution, rather than by geographical location. It is difficult to predict its effect on survival. Tertiary centres, which are often the origins of preterm cohorts in the literature, may claim a higher quality of care but also may have a concentration of higher risk pregnancies, depending on referral patterns.

The objective of this study was to determine whether the selected factors influence the reported estimate of survival and, if it does, to gain some idea of the magnitude of the effect. The specific hypothesis under test was that reported survival would be higher in studies with increased potential for selection bias. The secondary objective was to determine how comprehensively clinical factors influencing survival, such as antenatal steroid and surfactant use, were reported in the literature.

**Methods**

**CRITERIA FOR CONSIDERING STUDIES**

The objective of the literature search was to identify all cohort studies reporting survival in preterm infants and published within the period 1978–1998. The search was restricted to the developed countries within the following geographical boundaries: North America, Western Europe, Japan, Australia, and New Zealand. This was to maintain a degree of conformity in standards of perinatal care; these countries reported infant mortality of 4–7 per 1000 births in 1998. Only data relating to infants of less than 28 completed weeks of gestation were considered and the studies had to report survival in one week gestational age bands. Studies exclusively reporting the outcome of multiple pregnancies were disregarded.

**LITERATURE SEARCH**

Electronic databases were searched using a combination of free text (textword) and subject headings (MeSH), if supported by the database. Table 1 gives the Medline search strategy.

### Table 1  Search strategy for Medline (Ovid)

| 001 | exp infant, low birth weight/ |
| 002 | exp infant, premature/ |
| 003 | (low adj1 birth/weight).ti,ab.sh. |
| 004 | (preterm adj3 infant$).ti,ab.sh. |
| 005 | (preterm adj3 bab$).ti,ab.sh. |
| 006 | (premature adj5 infant$).ti,ab.sh. |
| 007 | (premature adj5 bab$).ti,ab.sh. |
| 008 | exp infant mortality/ |
| 009 | death.ti,ab,sh. |
| 010 | surviv$.ti,ab,sh. |
| 011 | exp cohort studies/ |
| 012 | exp gestational age/ |
| 013 | orb1–7 |
| 014 | orb8–10 |
| 015 | 11 or 12 |
| 016 | 13 and 14 and 15 |

Other methodological factors that may influence the estimate of survival were recorded. These included any exclusion criteria (such as congenital malformations), whether the cohort was geographically or hospital defined, the method by which gestational age was assessed, and the postnatal age at which mortality was defined.

Clinical factors, which might influence survival of preterm infants, were also recorded from the manuscripts of the selected studies. These included the period during which the cohort was born, any sociodemographic and ethnic details of the population reported, and the use of antenatal steroids and postnatal surfactant treatment.

**ANALYSIS**

For the purposes of this review, survival was defined as the proportion of live births surviving until discharge from hospital. For each study and for each week of gestation, the proportion of liveborn preterm infants surviving until discharge was calculated. Combined estimates of survival for each grade of cohort study were then calculated by combining proportions from studies in a weighted average.

The hypothesis that there would be a gradient in the estimates of survival, increasing from grade A through to grade C studies, was tested by \( \chi^2 \) analysis to detect a gradient in qualitatively ordered proportions.

Studies that defined survival at other times—for example, expected date of delivery, 28 days, 6 months postnatal age—were included because it was assumed that death rates would be greatest in the first week of life. Thus survival to any time outside the early neonatal period and within the first year of life was assumed to approximate. The effect of this assumption was tested by reanalysing the data from only those studies that reported survival to discharge.

**Results**

**DESCRIPTION OF SELECTED COHORT STUDIES**

From the literature search, 191 articles published in 1978–1998 reported survival of preterm infants. Thirty one (16%) were not appraised (non-English language), 56 (29%) reported data relating to birth weight and not gestational age, and 37 (19%) reported survival in broad categories, rather than one week gestational age bands. The remaining 67 published articles reported data on 55 preterm cohorts.
some text
Table 4 Characteristics of preterm cohorts from grade C studies reporting outcome of neonatal unit admissions

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cohort birth period</th>
<th>Exclusions (details in text)</th>
<th>Geographical cohort</th>
<th>Antenatal steroids</th>
<th>Postnatal surfactant</th>
<th>Age at survival</th>
<th>Proportion of live births surviving for each week of gestation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martens(^{37})</td>
<td>1991–1996</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>neo</td>
<td>57</td>
</tr>
<tr>
<td>Emley(^{57})</td>
<td>1990–1994</td>
<td>No</td>
<td>No</td>
<td>Yes (23%)</td>
<td>No</td>
<td>dis</td>
<td>0</td>
</tr>
<tr>
<td>Bohn(^{51})</td>
<td>1991–1993</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>dis</td>
<td>3</td>
</tr>
<tr>
<td>Hussain(^{51})</td>
<td>Aug 80–Jul 93</td>
<td>No</td>
<td>No</td>
<td>Yes*</td>
<td>No</td>
<td>dis</td>
<td>27</td>
</tr>
<tr>
<td>Wilson(^{41})</td>
<td>Sep 90–Oct 92</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>dis</td>
<td>0</td>
</tr>
<tr>
<td>Marlow(^{50})</td>
<td>Jul 86–Dec 93</td>
<td>No</td>
<td>No</td>
<td>Yes (20%)</td>
<td>Yes*</td>
<td>dis</td>
<td>27</td>
</tr>
<tr>
<td>Lamont(^{49})</td>
<td>1979–1980</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>dis</td>
<td>43</td>
</tr>
<tr>
<td>Cooper(^{47})</td>
<td>1984–1985</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>dis</td>
<td>1</td>
</tr>
<tr>
<td>Philip(^{51})</td>
<td>1976–1979</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>dis</td>
<td>3</td>
</tr>
</tbody>
</table>

*Percentage not specified.

dis, discharge from hospital; neo, neonatal period (28 days); 2 y, 2 years of age.

 Most cohorts defined survival at discharge from hospital (75%). Four cohorts reported survival during the neonatal period (28 days), and 10 cohorts reported survival rates ranging from 6 months to 8 years (tables 2–4).

ANALYSIS OF SURVIVAL DATA

Figure 1 depicts the combined estimates of survival (proportion of liveborn infants surviving to discharge from the neonatal unit) for each week of gestation and for each grade of cohort study. For infants of 23–26 weeks of gestation, there was a significant trend towards increased survival in cohorts with a higher risk of selection bias (grade C > grade B > grade A; p < 0.01). The value of $\chi^2$ was greatest for lower gestations, particularly at 23 and 24 weeks. Grade C studies exaggerated survival compared with grade A studies, by 100% at 23 weeks (20% × 10%), 56% at 24 weeks (42% × 27%), 18% at 25 weeks (52% × 44%), and 13% at 26 weeks (62% × 55%).

Excluding from the analysis cohorts where survival was defined at times other than hospital discharge did not affect the observed trend in survival. An identical trend in $\chi^2$ values was observed (highest value $\chi^2 = 49.2$ at 24 weeks, ranging to lowest value $\chi^2 = 9.0$ at 27 weeks).

Owing to the paucity of reported data, it was not possible to examine any further associations between trends in survival and other factors such as antenatal steroid usage, surfactant treatment, and the geographical nature of the cohort.

Discussion

The results of this systematic review indicate that the potential for selection bias in preterm infant cohort studies is associated with higher estimates of survival, most pronounced at the lower gestations.

A POSSIBLE MECHANISM FOR SELECTION BIAS

Studies that report the outcome of admissions to the neonatal unit are ignoring the outcome of neonates who could not be resuscitated. In this case, the outcome of the neonates admitted would be better than the total population of live births as they represent infants who survived until neonatal unit admission. Studies reporting the outcome of all liveborn infants also give a higher estimate of survival than those reporting outcomes of all pregnancies, despite survival being expressed as a proportion of live births—that is, using the same denominator. Selective obstetric monitoring and intervention during favourable pregnancies may account for the apparent improvement in survival. This effect appears to be more pronounced at lower gestational ages, which would concur with the notion that obstetricians are more likely to be selective about the care of pregnancies around the limit of viability.

There is potential for ambiguous classification of a delivery around the limit of viability. A delivery before 24 weeks of gestation may be treated as a miscarriage on a gynaecology ward. Conversely, it may be classed as a neonatal death on a labour ward if the labour was aggressively managed and signs of life were sought and found present after birth. It is also possible for late terminations of pregnancy to be missed if they occur outside of the maternity unit. Only two studies stated that the investigators actively sought to check admissions to gynaecological wards.

![Figure 1: Survival of preterm infants associated with gestational age and methodological grade of cohort study. Bars indicate 95% confidence intervals. NICU, Neonatal intensive care unit.](http://fn.bmj.com/ArchDisChildFetalNeonatalEd.1stpublishedas10.1136/fn.84.2.F79)
Table 5  Recommended data to be included when reporting preterm cohort studies

<table>
<thead>
<tr>
<th>Cohort characteristics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution or geographical area</td>
<td></td>
</tr>
<tr>
<td>Time period of birth cohort</td>
<td></td>
</tr>
<tr>
<td>Maternal sociodemographic details</td>
<td></td>
</tr>
<tr>
<td>Ethnic characteristics</td>
<td></td>
</tr>
<tr>
<td>Clear recording of all previous publications (including abstracts) relating to cohort</td>
<td></td>
</tr>
</tbody>
</table>

Coexistent perinatal therapies
- Antenatal steroid uptake
- Postnatal surfactant usage

Reporting of outcomes
- Report outcome of all pregnancies (terminations, miscarriages, stillbirths, live births)
- Outcome reported in one week gestational age bands
- Reporting of outcome
- Postnatal surfactant usage
- Antenatal steroid uptake

ARE THERE OTHER PLAUSIBLE EXPLANATIONS?
There is still a large degree of variation in survival within each grade of study and therefore other factors may interact with, or exist instead of, selection bias. There is an observed increase in survival over the period of this systematic review, presumably reflecting improvements in perinatal care. This trend has also been seen within individual studies that compared cohorts from different periods using the same methods. Could the improved survival over time be confounding the association between selection bias and survival? There is no evidence that more grade A studies (live births and stillbirths) were conducted in the first half of the period than the other grades ($\chi^2 = 1.84$ (df2) $p > 0.2$). In an attempt to examine whether this observed trend over time confounds the trend associated with the grade of the study cohort, the cohorts were stratified into two equal groups (1977–1986 and 1987–1996). The stratification was performed by taking each mid-point of the period during which a cohort was born. The gradient associated with grade of study cohort was still observed to be present for the 1987–1996 period but not for the earlier period. It was not possible to test more precisely whether selection bias and year of birth act independently because studies did not report survival for individual years, and the periods during which cohorts were born overlapped considerably.

The effect of clinical heterogeneity, arising from different rates of antenatal steroid administration, surfactant treatment, and the different socioeconomic and ethnic status of the populations under study, on survival was not examined. In general, reporting of these factors was poor, and therefore they were not analysed.

WHAT ARE THE WEAKNESSES OF THIS SYSTEMATIC REVIEW?
Although the intention was to be systematic in the identification of studies and extraction of data, several problems were encountered. Studies in languages other than English were not appraised because of a lack of resources. Data relating survival to birth weight, rather than gestational age, were ignored and could not be combined with the gestational age specific data. This was because the intention was to provide information of use to obstetricians and prospective parents before delivery, when the birth weight would not be accurately known. It was noted that the studies excluded for this reason were predominantly North American in origin, representing a non-random sample of published studies.

The literature search was confined to published studies. Publication bias may be more pronounced for recent cohorts. Many neonatal units now routinely record information on neonatal outcomes, but this information may be viewed by editors and reviewers as uninteresting and unworthy of publication, unless the outcomes are unusually good.

Many cohorts are the subjects of more than one publication, which can lead to duplication bias. Every attempt was made to ensure that this did not occur by considering the cohorts as the unit of analysis rather than the study. Partial duplication, where cohorts overlap but no study completely encompasses all patients, was a problem in three studies.

It was not possible to test whether other clinical factors, such as surfactant treatment, have an effect on survival independent of methodological factors. This would require a systematic review using individual patient data.

IMPLICATIONS FOR FUTURE RESEARCH
Those concerned with the provision and evaluation of perinatal care are required to continuously collect data on survival and outcome of preterm infants. The outcomes of all pregnancies should be reported for each week of gestation. There should be sufficient information to avoid duplication bias—that is, sufficient descriptive data to identify a cohort uniquely. If the study reports longer term outcomes from a cohort where the short term outcomes had previously been described, this fact should be made explicit. It is important that data are comprehensive and include factors such as use of antenatal steroids and postnatal surfactant. It would be desirable to develop conformity of definitions between studies—for example, exclusion of congenital malformations and age at which survival is assessed. Such considerations could form the basis of a minimum dataset for future publications (table 5).

IMPLICATIONS FOR CLINICIANS
In the modern era of evidence based medicine, it is logical and appropriate that the decisions surrounding the provision of expensive intensive care to extremely preterm neonates are based on the best available external evidence. This process may involve parents, obstetricians, neonatologists, developmental pediatricians, and health economists. All these parties need to be aware of the potential bias and the effect that this will have on reported survival figures. It is well known that selection bias is a source of non-random error in cohort studies, but this is the first systematic review to attempt to quantify its effect on studies of preterm survival. The effect appears to be substantial around the limit of viability, with the potential to overestimate survival by 100% and 56% at 23 and 24 weeks respectively. Although not tested in this review, there is every reason to suspect that this effect also exists when the subsequent neurodevelopmental outcome of these infants is considered.
10 Costeloe K, EPICure Study Steering Group. EPICure: sur-
15 Nicholl MC, Gilews WB. The short-term outcome of single-