Population-Based Estimates of In-Unit Survival for Very Preterm Infants

WHAT’S KNOWN ON THIS SUBJECT: Survival estimates for preterm infants are vital for counseling parents, informing care, and planning services. Widely use estimates of in-unit survival derived from a large UK population for infants born at <33 weeks’ gestational age have been available since 1999.

WHAT THIS STUDY ADDS: These survival charts have been updated and will be of use to clinicians, parents, and managers. An alternative method for graphical representation of survival probabilities is offered: contour survival plots.

abstract

BACKGROUND: Estimates of the probability of survival of very preterm infants admitted to NICU care are vital for counseling parents, informing care, and planning services. In 1999, easy-to-use charts of survival according to gestation, birth weight, and gender were published in the United Kingdom. These charts are widely used in clinical care and for benchmarking survival, and they form the core of the Clinical Risk Index for Babies II score. Since their publication, the survival of preterm infants has improved, and the charts therefore need updating.

METHODS: A logistic model was fitted with gestational age, birth weight, and gender. Nonlinear functions were estimated by using fractional polynomials. Bootstrap methods were used to assess the internal validity of the final model. The final model was assessed both overall and for subgroups of infants by using Farrington’s statistic, the c-statistic, Cox regression coefficients, and the Brier score.

RESULTS: A total of 2995 white singleton infants born at 23+0 to 32+6 weeks’ gestation in 2008 through 2010 were identified; 2751 (91.9%) infants survived to discharge. A prediction model was estimated and good model fit confirmed (area under receiver-operating characteristics curve = 0.86). Survival ranged from 27.7% (23 weeks) to 99.1% (32 weeks) for boys and from 34.5% (23 weeks) to 99.3% (32 weeks) for girls. Updated charts were produced showing estimated survival according to gestation, birth weight and gender, together with contour plots displaying points of equal survival.

CONCLUSIONS: These survival charts have been updated and will be of use to clinicians, parents, and managers. Pediatrics 2013;131:e425–e432
Accurate estimates of the survival for very preterm infants are vital for counseling parents, informing care, and planning services.\textsuperscript{1,2} There are a large number of prediction models that have been proposed to estimate survival for very preterm or low birth weight deliveries.\textsuperscript{3,4} Such models are useful because they can potentially provide more accurate estimates of the probability of survival than clinical assessment alone.\textsuperscript{5–7} However, these models are often derived by using data from individual hospitals rather than from a population based on residence. The organization of neonatal intensive care units (NICUs) into providers offering different levels of care means that predictions derived from a hospital-based cohort may not accurately reflect survival in the whole population. To obtain a complete picture of in-unit mortality survival, estimates derived from a cohort defined according to residence, rather than place of treatment, are required.

In 1999, residence-based estimates of the survival to discharge of very preterm infants were published in the United Kingdom in the form of easy-to-use charts.\textsuperscript{8} The charts, sometimes referred to as the “Draper Grid,” show survival according to gestational age, birth weight, and gender. They were derived by using data from The Neonatal Survey (TNS), an ongoing population-based survey of NICUs in the East Midlands and Yorkshire regions of England.\textsuperscript{9} All neonatal services in the regions contribute to TNS, and NICUs in adjacent regions also permit data collection on eligible infants. The current survey was established in 1990 and now covers an area that has \~120 000 births each year, with 29 consultant-led and 14 midwife-led delivery units (6 collocated with consultant units). Information is collected on all infants admitted to an NICU who are born at \textless33 weeks’ gestational age to mothers resident in the study area. Seven part-time neonatal nurses prospectively collect the data during regular visits to the 29 NICUs, with audits and validation checks undertaken to ensure data collection is complete.

White singleton infants born at 23\textsuperscript{+0} to 32\textsuperscript{+6} weeks’ gestational age between January 1, 2008, and December 31, 2010, were selected for this analysis. Infants were excluded if they had lethal congenital anomalies, missing data on gender or indeterminate gender, or implausible birth weight for gestational age (defined as being \textgreater3 SDs from the median for their gestational age and gender). Gestational age was defined according to the following hierarchy: mother certain of dates (most reliable); early dating scan (<20 weeks’ gestational age); late dating scan (>20 weeks’ gestational age); and postnatal examination (least reliable). If the difference between maternal dates and early scan was \textgreater7 days, the early scan was used to determine the period of gestation.

**Statistical Methods**

A logistic model was developed to predict the probability of survival by using gestational age in days, birth weight, and gender as predictors. Two-way interactions were investigated and were included if significant at the 10% significance level. An indicator variable for infants born at 23\textsuperscript{+0} to 23\textsuperscript{+6} weeks’ gestational age was included to allow for the differences in this group compared with other preterm births. Nonlinear functions were modeled by using fractional polynomials\textsuperscript{21} and the final form selected using the change in deviance. The internal validity of the chosen model was investigated through the use of bootstrapping techniques. Five hundred bootstrap samples were selected with replacement and the model selection process repeated with the c-statistic (area under the receiver-operating characteristic curve)\textsuperscript{22} and Cox’s regression coefficients\textsuperscript{23} monitored for each sample. Similarly, bootstrap methods were used to determine whether interactions were required in the model.

The overall fit of the final model was assessed by using Farrington’s extension to Pearson’s \(\chi^2\) test\textsuperscript{24,25} and the Brier score\textsuperscript{26} as well as inspection of deviance residuals and the DFBETA statistics.\textsuperscript{27} The Hosmer and Lemeshow goodness-of-fit test was not used because it is known to be unreliable.\textsuperscript{28} In addition to assessing overall model fit, it is important to ensure that the discriminatory performance and calibration of the final model are also adequate. The
discriminatory performance of the final model was quantified by using the c-statistic, and the calibration was assessed according to Cox’s calibration regression. Table 1 provides details regarding the methods used to describe the goodness-of-fit and predictive power of the model. The predictive performance of the model was assessed both overall and according to subgroups based on gestational age and whether birth weight was above or below the 50th percentile.

Predicted survival percentages were calculated separately for male and female infants according to week of gestational age at birth and birth weight in increments of 250 g and then reproduced on updated Draper Grids.

The predicted survival at the midpoint of each cell (week of gestational age plus 4 days and birth weight midpoint) was displayed together with the lowest and highest predicted survival within that cell. To aid interpretation, the second, 10th, 50th, 90th, and 98th UK birth weight centiles were added to the plot.

In addition to the familiar grids, survival contour plots were also created; these are sometimes referred to as “isosurv” plots and have been suggested as a method for displaying survival probabilities. Contours were drawn on these plots to denote the points of equal survival (25%, 50%, 75%, 90%, 95%, and 98% survival) according to birth weight and gestational age. The

<table>
<thead>
<tr>
<th>TABLE 1 Methods Used to Describe the Goodness-of-Fit and Predictive Power of the Model</th>
</tr>
</thead>
</table>
| The Brier score (B) is defined as:
| B = 1/n \sum (x_i - p_i)^2 |
| A low value for the Brier score indicates that there is close agreement between observed and predicted outcomes.
| The Farrington statistic is an extension of the Pearson \( \chi^2 \) statistic and is given by:
| \( \chi^2_F = \sum_{i=1}^{N} \frac{(y_i - m_i p_i)^2}{m_i p_i (1 - p_i)} + \sum_{i=1}^{N} \frac{-(1 - 2p_i)^2}{m_i (1 - p_i)} \) |
| The variance can be calculated and the standardized statistic compared with the standard normal distribution.
| For Cox’s calibration regression, the following model is used:
| \( x_i = \left(1 + \exp \left(-\left\{\alpha + \beta \log \left(\frac{m_j p_j}{1 - m_j p_j}\right)\right\}\right)^{-1} \) |
| If the model is perfectly calibrated then \( \alpha = 0 \) and \( \beta = 1 \). To examine the significance of these values, 3 tests are performed:
| 1. \( \alpha = 0 \) and \( \beta = 1 \): overall reliability – perfect prediction
| 2. \( \alpha = 0 \) \( \beta = 1 \): perfect calibration
| 3. \( \beta = 1 \) \( \alpha \): overall refinement – the correct amount of variation in the model
| Note: the calibration of the overall model by definition is perfect. The calibration of subgroups, however, can be assessed to ensure adequate calibration within groups.
| The c-statistic (area under the ROC curve) is:
| \( c = \frac{0.5(\sum x_i^2 + \sum r_i) - \sum (r_i x_i = 1)}{\sum x_i (n - \sum x_i) + 1} \) |
| The c-statistic, the value of which ranges from 0 to 1, quantifies how well the model discriminates between those who survived and those who did not. A value of 0.5 indicates the model is no better than chance alone, whereas 1 demonstrates there is perfect discrimination.

Where:
- \( x_i \) is the observed outcome for a given infant \( i \) (\( x_i = 1 \) if infant \( i \) survived to discharge and \( x_i = 0 \) otherwise);
- \( p_i \) is the predicted probability of survival for infant \( i \);
- \( n \) is the total number of infants;
- \( N \) is the number of unique combinations of values of the covariates;
- \( m_j \) is the number of infants in covariate combination \( j \);
- \( p_j \) is the predicted probability of survival for infants in covariate combination \( j \);
- \( r_i \) is the value of rank of \( p_i \) when all values ordered from lowest to highest.

ROC, receiver-operating characteristic.

...
the 10% significance level in 2.2% of the bootstrap samples (mean P value = .79). Therefore, no interactions were included in the final model.

The average c-statistic was 0.86 (range: 0.85 to 0.86) in the bootstrap samples, indicating good overall discrimination. The Cox calibration coefficients had a mean intercept of 0.014 (range: 0.28 to 0.09) and a mean slope of 0.99 (range: 0.98 to 1.00), indicating good calibration. Details of the final model are given in the Appendix.

Model Validation

The model fit statistics for the whole data set and for subgroups are shown in Table 3. Overall, the final selected model was well fitted to the data (c-statistic: 0.86; Farrington’s statistic: P = .44; Brier score = 0.058). The model fit was generally good for all of the subgroups except that the discriminatory power of the model was poor for the subgroups based on gestational age (c-statistic: 0.62 to 0.70).

The predictive performance of the final model was investigated for singleton infants of non-white ethnic origin, even though they were not included in the estimation of survival rates: n = 812 (South Asian, n = 438; black, n = 165; other, n = 209); number surviving = 742; predicted number surviving = 731.4; c-statistic = 0.875. The predictive performance was also investigated for multiple births of all ethnic groups: n = 1284; number surviving = 1188; predicted number surviving = 1178.8; c-statistic = 0.896.

Graphical Representation

The estimated survival percentages are given separately for male and female infants in updated versions of the Draper Grid (Fig 1). The contour plots in Fig 2 provide an alternative method for presenting the probability of survival. Due to a categorical variable being used to identify those born at 23 weeks’ gestational age, the contours have been smoothed by using splines at the boundary between 23 and 24 weeks.

DISCUSSION

The probabilities of survival estimated in this article were derived from a large UK population with >120 000 births per year (~17% of all births in England and Wales). The 2 regions (East Midlands and Yorkshire) are considered representative of England and Wales, comprising a mixture of urban and rural populations. In 2010, the neonatal mortality rate in the study area was 3.2 per 1000 live births compared with 3.0 per 1000 live births in the whole of England and Wales.

The use of a cohort based on the mothers’ residence avoids the biases that can occur when estimates are derived from hospital-specific data due to UK neonatal care providers being organized into networks of NICUs providing different levels of care, similar to the pattern of neonatal care seen in most other countries. Survival rates for individual units, therefore, reflect local policies on care, admission, and transfer; the observed survival for these data for TNS units ranged from 85% to 100%. The use of a cohort based on residence, such as TNS, is required to obtain a complete understanding of survival. However, although it is likely the case that the survival probabilities reported here

### Table 2: Observed Survival of Singleton Infants of White Ethnic Origin According to Gender and Completed Week of Gestational Age for the Current Study Cohort (2008–2010) and for the Cohort Analyzed for the Mortality Charts published in 1999 (1994–1997)

<table>
<thead>
<tr>
<th>Gestational Age (completed weeks)</th>
<th>2008–2010</th>
<th>1994–1997</th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>4 (28.6)</td>
<td>3 (13.0)</td>
</tr>
<tr>
<td>24</td>
<td>25 (48.1)</td>
<td>7 (20.0)</td>
</tr>
<tr>
<td>25</td>
<td>47 (73.4)</td>
<td>23 (45.1)</td>
</tr>
<tr>
<td>26</td>
<td>72 (77.4)</td>
<td>39 (55.7)</td>
</tr>
<tr>
<td>27</td>
<td>115 (83.3)</td>
<td>69 (75.8)</td>
</tr>
<tr>
<td>28</td>
<td>125 (88.7)</td>
<td>107 (80.4)</td>
</tr>
<tr>
<td>29</td>
<td>150 (95.0)</td>
<td>137 (89.5)</td>
</tr>
<tr>
<td>30</td>
<td>221 (95.3)</td>
<td>151 (95.0)</td>
</tr>
<tr>
<td>31</td>
<td>305 (99.7)</td>
<td>243 (97.6)</td>
</tr>
<tr>
<td>32</td>
<td>413 (98.8)</td>
<td>316 (97.8)</td>
</tr>
<tr>
<td>Total</td>
<td>1484 (91.2)</td>
<td>1095 (85.5)</td>
</tr>
</tbody>
</table>

### Table 3: Goodness-of-Fit, Calibration, and Discriminatory Statistics for the Final Model: Overall and by Subgroups Based on Gestational Age at Birth and Birth Weight

<table>
<thead>
<tr>
<th>Variable</th>
<th>Overall</th>
<th>Birth Weight</th>
<th>Gestational Age (Weeks&lt;sup&gt;≤&lt;/sup&gt;)&lt;sup&gt;≥&lt;/sup&gt;</th>
<th>23&lt;sup&gt;≤&lt;/sup&gt; to 26&lt;sup&gt;≤&lt;/sup&gt;</th>
<th>27&lt;sup&gt;≤&lt;/sup&gt; to 29&lt;sup&gt;≤&lt;/sup&gt;</th>
<th>30&lt;sup&gt;≤&lt;/sup&gt; to 32&lt;sup&gt;≤&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of infants</td>
<td>2995</td>
<td>1753</td>
<td>1242</td>
<td>422</td>
<td>830</td>
<td>1743</td>
</tr>
<tr>
<td>No. (%) surviving</td>
<td>2751 (91.9)</td>
<td>1599 (91.2)</td>
<td>1152 (92.8)</td>
<td>281 (88.6)</td>
<td>754 (90.8)</td>
<td>1718 (98.5)</td>
</tr>
<tr>
<td>Predicted no. surviving</td>
<td>2751</td>
<td>1599</td>
<td>1152</td>
<td>281</td>
<td>754</td>
<td>1718</td>
</tr>
<tr>
<td>Farrington’s statistic</td>
<td>P = .44</td>
<td>P = .14</td>
<td>P = .74</td>
<td>P = .78</td>
<td>P = .22</td>
<td>P = .63</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.0583</td>
<td>0.0629</td>
<td>0.0517</td>
<td>0.1962</td>
<td>0.0786</td>
<td>0.0152</td>
</tr>
<tr>
<td>Cox regression Intercept</td>
<td>NP*</td>
<td>−0.075</td>
<td>0.149</td>
<td>−0.049</td>
<td>0.398</td>
<td>−0.150</td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td>0.371</td>
<td>1.054</td>
<td>1.005</td>
<td>1.153</td>
<td>0.981</td>
</tr>
<tr>
<td>Overall reliability Calibration</td>
<td>P = .88</td>
<td>P = .77</td>
<td>P = .88</td>
<td>P = .57</td>
<td>P = .92</td>
<td></td>
</tr>
<tr>
<td>Refinement</td>
<td>P = .74</td>
<td>P = .71</td>
<td>P = .64</td>
<td>P = .42</td>
<td>P = .69</td>
<td></td>
</tr>
</tbody>
</table>

* These would be 0 and 1 by definition and are not presented (NP).
will not hold for individual units, even in our study area, the probabilities reported are useful as a benchmark against which variations may indicate important differences in performance or indeed practice.38 Published survival rates require regular updating as survival improves for preterm infants. The survival probabilities derived from TNS data were originally published in 19998 and subsequently updated in 2003.12 The 2003 survival estimates from TNS were validated by using national data from the Netherlands from 2000 to 2007.10 However, improvements in neonatal survival have been reported in this population over the past 10 years,14 although some studies have suggested that this improvement may have now slowed or reached a plateau.31,39 These improvements mean that the estimates required further updating.

Direct comparison of survival in this article with that reported recently from other countries is difficult because rates are usually given for all births rather than just admissions, for networks of NICUs as opposed to geographically defined populations, or are not given according to gender of the infant. However, a crude comparison suggests that the rates reported here are similar to those from the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network: in 2007, survival to discharge for live births was 92% at 28 weeks; 91% at 27 weeks; 84% at 26 weeks; 71% at 25 weeks; 54% at 24 weeks; and 21% at 23 weeks.40 The statistical model developed in this analysis includes an indicator variable for birth at 23 weeks’ gestational age because the decision-making process for admission is different from that at ≥24 weeks. In the United Kingdom, for births at 23 weeks’ gestational age, health professionals discuss with parents the provision of active intervention given the individual circumstances; at 24 weeks, the broad expectation is that active intervention and intensive care will be started unless an infant is in very poor condition at birth.41 Internationally, there seems to be consensus on the level of intervention at ≥25 weeks but the situation is less clear for those infants born at 23 and 24 weeks.42

Other mortality prediction models have been proposed of varying complexity,3 but the aim of the current article was to update a simple prediction model with a small number of highly predictive variables. We believe that the use of weight and gestational age at birth, together with the infant’s gender, allows the estimation of survival rates that are clinically useful without relying on drug administration or physiologic or pharmacologic tests that may not always be performed. The final model used to estimate the survival probabilities is relatively straightforward and contains no interactions between the predictors. The inclusion of birth weight in predictive models for

![Grid of predicted survival according to gestational age, birth weight, and gender.](image)

FIGURE 1
Grid of predicted survival according to gestational age, birth weight, and gender. The main number in each cell is the predicted survival at the midpoint of the cell (week of gestational age plus 4 days and birth weight midpoint). The numbers in parentheses are the lowest and highest predicted survival within that cell. A, Female. B, Male.
preterm infants has been undertaken previously in various ways; for example, observed birth weight, ratio to gestation-specific mean birth weight, difference from gestation-specific mean birth weight, and percentile. In practice, there is little difference in predicted values obtained by using these various approaches; indeed, they are all approximations of the same underlying model. In this analysis, observed birth weight was included because this is the most straightforward variable to interpret and does not require the choice or estimation of gestation/gender-specific mean birth weight. These data showed no evidence for increased mortality at higher gestation-specific birth weights. Previous evidence for this relationship has been equivocal, with some articles reporting an increased risk of mortality for infants with large birth weights for gestational age whereas others have not. Although some of this inconsistency may be explained by variations in the type of infants included in the studies, it may also be due to the statistical methods used. Fractional polynomials allow more flexibility in the modeling of the relationship between birth weight and mortality than conventional polynomials.

The fit of the model to the observed data was investigated for the whole data set and also for clinical subgroups defined a priori. Although any model may seem to predict well overall, this may be the result of underestimating the outcome probability for some groups and overestimating it for others. It is important, therefore, that subgroups of the data also be examined. The model derived here demonstrated good goodness-of-fit and calibration both overall and for all subgroups. The discriminatory performance of the models was less clear. Although the discriminatory performance was good overall and for the subgroups defined according to gestational age, it was poor for the subgroups defined according to gestational age. This finding is not unexpected because gestational age is a strong predictor of survival, and its effect will not be seen for subgroups defined over a narrow range of gestational ages. However, the important characteristic of this model is its calibration, which was shown to be good.

In this article, the survival estimates were derived for singletons of white ethnic origin because the data set was too small to be able to confidently obtain estimates of sufficient precision for infants from multiple births or of non-white ethnic origin. However, the final model showed good overall predictive performance for singletons of non-white ethnic origin and for all multiple births, although there was insufficient data to confirm whether this held for subgroups of infants.

Previous estimates of survival obtained by using TNS data have also included survival estimates for all fetuses known to be alive at the onset of labor by including data from the national Confidential Inquiry into Maternal and

FIGURE 2
Contour plot of predicted survival according to gestational age, birth weight, and gender. The contour lines join combinations of gestational age and birth weight of equal estimated probability of survival. Birth weight percentiles are shown for information. A, Female. B, Male.
Child Health. Currently, data on stillbirths and infants who died before admission to an NICU are not available for England, and it was only possible therefore to produce these charts for infants admitted to neonatal intensive care. Because the charts in this article report survival for infants who survived to admission, they will overestimate survival rates for all births. When the additional data become available, the charts for all births will be updated.

CONCLUSIONS

These survival charts have been updated and will be of use to clinicians, parents, and managers.

APPENDIX: FINAL MODEL DETAILS

The final logistic model to predict survival is given by:

\[
\text{Logit} (\text{Survival}) = \beta_0 + \beta_1 \times \text{gestation} + \beta_2 \times \text{birth weight} + \beta_3 \times \text{gender} + \beta_4 \times \text{gest23}
\]

Where:

- \(\text{gestation}\) is the infant’s gestational age in completed weeks and days as a decimal (i.e., 28.5 weeks takes the value 28.429);
- \(\text{birth weight}\) is the infant’s birth weight in kilograms;
- \(\text{gender}\) is the infant’s gender: for males, \(\text{gender} = 0\); for females, \(\text{gender} = 1\);
- \(\text{gest23}\) is an indicator for birth at <24\(^{+0}\) weeks’ gestational age:

\[
\text{gest23} = 1 \text{ if gestation at birth <24\(^{+0}\} \text{ weeks}; \text{gest23} = 0 \text{ otherwise.}
\]

The estimates for the model parameters are:

- \(\beta_0 = -10.28 \text{ (SE 1.36)}\)
- \(\beta_1 = 0.46 \text{ (SE 0.05)}\)
- \(\beta_2 = -0.47 \text{ (SE 0.13)}\)
- \(\beta_3 = 0.45 \text{ (SE 0.16)}\)
- \(\beta_4 = -0.28 \text{ (SE 0.39)}\)

ACKNOWLEDGMENTS

We acknowledge the continuing help and collaboration of the hospitals delivering neonatal care in the study area. We also thank Dr. Oliver Kuss for the use of the SAS macro %GOFLOGIT to calculate Farrington’s test.

REFERENCES

12. Draper ES, Manktelow B, Field DJ, James D. Tables for predicting survival for preterm births are updated. BMJ. 2003;327(7419):872
42. Pignotti MS, Donzelli G. Perinatal care at the threshold of viability: an international comparison of practical guidelines for the treatment of extremely preterm births. Pediatrics. 2008;121(1). Available at: www.pediatrics.org/cgi/content/full/121/1/e193
49. Manktelow BN, Draper ES. Modelling neonatal mortality by gestational age, birth weight and sex among very preterm babies using fractional polynomials. In: Poster presented at 27th Annual Conference of the International Society for Clinical Biostatistics; August 30, 2006; Geneva, Switzerland
Population-Based Estimates of In-Unit Survival for Very Preterm Infants
Bradley N. Manktelow, Sarah E. Seaton, David J. Field and Elizabeth S. Draper

*Pediatrics* 2013;131:e425; originally published online January 14, 2013;
DOI: 10.1542/peds.2012-2189

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/131/2/e425.full.html