Prediction of Neonatal Outcomes in Extremely Preterm Neonates

WHAT’S KNOWN ON THIS SUBJECT: Extremely preterm infants are at high risk of neonatal mortality or morbidities. Existing prediction models focus on mortality, specific morbidities, or composite mortality and morbidity outcomes and ignore differences in outcome severity.

WHAT THIS STUDY ADDS: A simple and practical statistical model was developed that can be applied on the first day after NICU admission to predict outcome severity spanning from no morbidity to mortality. The model is highly discriminative ($C$-statistic = 90%) and internally valid.

abstract

OBJECTIVE: To develop and validate a statistical prediction model spanning the severity range of neonatal outcomes in infants born at ≤30 weeks’ gestation.

METHODS: A national cohort of infants, born at 23 to 30 weeks’ gestation and admitted to level III NICUs in Canada in 2010–2011, was identified from the Canadian Neonatal Network database. A multinomial logistic regression model was developed to predict survival without morbidities, mild morbidities, severe morbidities, or mortality, using maternal, obstetric, and infant characteristics available within the first day of NICU admission. Discrimination and calibration were assessed using a concordance $C$-statistic and the $C_p$ goodness-of-fit test, respectively. Internal validation was performed using a bootstrap approach.

RESULTS: Of 6106 eligible infants, 2280 (37%) survived without morbidities, 1964 (32%) and 1251 (21%) survived with mild and severe morbidities, respectively, and 611 (10%) died. Predictors in the model were gestational age, small ($<$10th percentile) for gestational age, gender, Score for Neonatal Acute Physiology version II $>$20, outborn status, use of antenatal corticosteroids, and receipt of surfactant and mechanical ventilation on the first day of admission. High model discrimination was confirmed by internal bootstrap validation (bias-corrected $C$-statistic = 0.899, 95% confidence interval = 0.894–0.903). Predicted probabilities were consistent with the observed outcomes ($C_p$ $P$ value = .96).

CONCLUSIONS: Neonatal outcomes ranging from mortality to survival without morbidity in extremely preterm infants can be predicted on their first day in the NICU by using a multinomial model with good discrimination and calibration. The prediction model requires additional external validation. Pediatrics 2013;132:e876–e885

AUTHORS: Wen J. Ge, MSc,a,b Lucia Mirea, PhD,a,b Junmin Yang, MSc,a Kate L. Bassil, PhD,a,b Shoo K. Lee, MD, PhD,a,c and Prakesh Kumar S. Shah, MD, MSc,a,c on behalf of the Canadian Neonatal Network

aMaternal–Infant Care (MiCare) Research Centre, Department of Pediatrics, Mount Sinai Hospital, Toronto, Ontario, Canada; and bDepartment of Paediatrics, University of Toronto, Toronto, Ontario, Canada

KEY WORDS infant, prematurity, mortality, morbidities, prediction model, validation

ABBREVIATIONS

BPDP—bronchopulmonary dysplasia
CI—confidence interval
CNN—Canadian Neonatal Network
CPAP—continuous positive airway pressure
GA—gestational age
HFV—high-frequency ventilation
IPPPV—intermittent positive pressure ventilation
IVH—intraventricular hemorrhage
NCT—necrotizing enterocolitis
PVL—periventricular leukomalacia
ROP—retinopathy of prematurity
SGA—small for gestational age
SNAPII—Score for Neonatal Acute Physiology version II

Ms Ge performed the literature review, planned and performed all statistical analyses, provided interpretation and discussion of results, and prepared the manuscript; Dr Mirea conceptualized the idea in conjunction with Dr Shah, planned statistical analyses, interpreted results, and provided critical revisions to the manuscript; Mr Yang contributed to data management and statistical analyses, participated in the interpretation of results, and provided critical revisions to the manuscript; Dr Bassil participated in the interpretation of results and provided critical revisions to the manuscript; Dr Lee was involved in data collection, participated in the interpretation of results, and provided critical revisions to the manuscript; Dr Shah conceptualized the idea in conjunction with Dr Mirea, provided clinical expertise, was involved in data collection, participated in the interpretation of results, and provided critical revisions to the manuscript; and all authors approved the final manuscript as submitted.

(Continued on last page)
Preterm neonates are at elevated risk of neonatal mortality and morbidities including intraventricular hemorrhage (IVH),1 periventricular leukomalacia (PVL),2 retinopathy of prematurity (ROP),3 necrotizing enterocolitis (NEC),4 bronchopulmonary dysplasia (BPD),5 and sepsis during hospitalization in the NICU. Severe morbidities during the neonatal period, commonly characterized by grade 3 or 4 IVH, PVL, stage 3 or higher ROP, stage 3 NEC or severe BPD (need for oxygen and either intermittent positive pressure ventilation [IPPV], high-frequency ventilation [HFV], or continuous positive airway pressure [CPAP] at 36 weeks’ postmenstrual age), are associated with long-term acute and chronic outcomes and neurodevelopmental disability that typically necessitate rehospitalization, ongoing medical care, and family support.6 Mild morbidities including grade 1 or 2 IVH, stage 1 or 2 ROP, stage 2 NEC, and mild BPD (need for oxygen at 36 weeks without positive pressure) are not benign, and they necessitate additional medical resources and follow-up after discharge from the NICU and increase stress for families.

Numerous studies have developed statistical models to predict mortality7,8 or specific morbidities9–14 in infants born very preterm. Investigators have also derived statistical models to predict survival without major morbidities by combining infants who die with those who survive with severe morbidities.15 As physicians and parents face critical decisions about neonatal care with short- and long-term impact on infants’ health and families, it is important to predict with high accuracy the probability of both the devastating outcome of mortality and the desired outcome of survival free of morbidities, as well as the range of possible outcomes including survival with mild or severe morbidities. Among previous prediction models for neonatal mortality in preterm infants, multivariable models predicted mortality better than birth weight or gestational age [GA] alone.8 Furthermore, prediction models examining a combined outcome of mortality and morbidities did not have better performance than models predicting mortality alone.16–19 Therefore, for improved prediction it is important to distinguish between mortality and morbidities and to develop multivariable prediction models that reflect a range of outcomes along the severity spectrum.

The primary objective of this study was to develop a multivariable prediction model for survival free of morbidities, survival with mild morbidities, survival with severe morbidities, or mortality for infants born at ≤30 weeks’ gestation and admitted to level III NICUs in Canada in 2010 and 2011. Our secondary objective was to internally validate the model by applying a bootstrap overfitting or “optimism” correction.

METHODS
Study Population
The Canadian Neonatal Network (CNN) maintains a national standardized database of outcomes and risk factors for infants admitted to level III NICUs in Canada. At each site, data are collected by trained abstractors from patient charts according to the same criteria specified in the CNN Abstractor’s Manual20 and entered electronically into a customized data entry program with built-in error checks. Institutional approval from each site, for data collection and transfer to the CNN coordinating center, is provided either by a local research ethics board or through an institutional quality improvement process.

Infants eligible for this study were born at 23 to 30 weeks’ gestation and admitted to level III NICUs participating in the CNN during a 2-year (January 1, 2010–December 31, 2011) study period. Because of imminent mortality, infants declared moribund (those who on admission were prescribed palliative care and for whom no aggressive treatment was provided) or admitted for palliative care were excluded. In addition, infants were excluded if born with a potentially lethal congenital anomaly, if birth date or gender was missing, or if gender was recorded as ambiguous.

Outcome Definition and Potential Risk Factors
A 4-level outcome was defined indicating survival without morbidities, survival with mild morbidities, survival with severe morbidities, or mortality. Infants who survived without any of the following morbidities were taken as the referent group in all analyses. Mild morbidities included grade 1 or 2 IVH,21 stage 1 or 2 ROP22 in either eye (without need for surgery), stage 2 NEC,23 BPD necessitating only oxygen therapy without positive pressure, or a single episode of infection. Severe morbidities included grade 3 or 4 IVH,21 PVL, stage 3 or higher ROP22 in either eye or surgery for ROP, stage 3 NEC,23 BPD necessitating oxygen and positive pressure in the form of IPPV, HFV, or CPAP, or ≥1 episode of infection. The definitions of mortality and morbidity were consistent across all CNN sites. Mortality (of any cause), PVL, IVH, ROP, and NEC were determined before discharge from the NICU. Because diagnosis of IVH, PVL, ROP, and NEC entails testing, missing data for these morbidities were imputed as negative, assuming testing was not performed because it was not clinically indicated. BPD was defined as the need for oxygen at 36 weeks’ postmenstrual age or at discharge if the infant was discharged before 36 weeks.
A set of 13 covariates of known clinical importance were available for prediction: maternal factors (hypertension, smoking, illicit drug use), infant characteristics (GA, small for gestational age [SGA], gender, Score for Neonatal Acute Physiology version II [SNAPII]) >20, multiple birth, inborn or outborn status, and cesarean delivery), receipt of antenatal corticosteroids, and receipt of surfactant and mechanical ventilation on day 1 in the NICU. In the CNN database, day 1 was defined as the time from admission until midnight of the same day. GA (complete weeks) was defined as the best estimate based on the date of intrafertilization, early ultrasound, last menstrual period, or obstetric estimate, followed by pediatric estimate, in that hierarchical order. Birth weight was available but not used in analyses because of its strong correlation with GA (Pearson correlation coefficient = .75). SGA was derived as weight <10th percentile for GA according to a gender-specific Canadian reference for birth weight and GA. Chorioamnionitis was excluded as a potential predictor because of the large percentage (26%) of missing data. All potential predictors considered had <5% missing data, as recommended for complete case data analyses.

**Statistical Analyses**

The distribution of each covariate was compared between the 4 outcome groups by using the Pearson $\chi^2$ test. The Cochrane–Armitage trend test for binary covariates and the Mantel–Haensel $\chi^2$ test for multilevel factors assessed risk patterns with increasing outcome severity. A generalized logistic regression model was developed to predict survival free of morbidities, mild morbidities, severe morbidities, or mortality using automatic stepwise variable selection among the set of 13 covariates available for model development. Because the main goal was to develop a good prediction model rather than the best-fitting model for the available data, a more liberal significance cutoff, $P$ value <.1 rather than the typical $P$ value <.05, was applied for inclusion or exclusion of covariates.

A number of measures were computed by using predicted probabilities from the final multivariable model to assess apparent predictive ability. For each outcome group, calibration plots examined agreement between the observed data and predicted probabilities by using a loess smoothing algorithm (smoothing parameter of 0.5). Measures of predictive accuracy, positive and negative predictive value, sensitivity, and specificity were evaluated considering the highest probability estimate as the predicted outcome and a range of cutoff points. Goodness of fit for the final model was tested by using the $C_p$ statistic, which corresponds to an extension of the Hosmer–Lemeshow test to multinomial logistic regression. A Brier skill score compared the generalized Brier score of the final multivariable model with that from the model including only the intercept. The ability of the final multivariable model to discriminate between outcome groups was evaluated by using a concordance $C$-statistic and corresponding 95% confidence interval (CI). For a binary outcome, the $C$-statistic corresponds to the area-under-the-curve measure. For multilevel outcomes, the $C$-statistic assesses concordance and discordance within pairs of subjects from different outcome groups (details provided in the Supplemental Information).

Measures of model performance computed by using predicted probabilities from the final multivariable model developed using the complete data are subject to overfitting and therefore are optimistic. For the $C$-statistic, internal validation was performed using a bootstrap optimism correction computed based on 200 bootstrap samples. The bootstrap is a statistical sample reuse process that can be applied to obtain nearly unbiased estimates of future model performance without holding back data when selecting predictors and obtaining final estimates of model parameters. Bootstrap samples were generated by random sampling a total number of subjects ($N$), with replacement, from the original data. For each bootstrap sample, a prediction model was developed, and 2 $C$-statistics were estimated using the bootstrap data and the original data. The bootstrap optimism correction was computed as the mean of the difference between these 2 $C$-statistics across all bootstrap samples. The bootstrap-corrected $C$-statistic was obtained by subtracting the optimism correction from the resubstitution $C$-statistic computed by using predicted probabilities from the final multivariable model developed by using the original data. All analyses were performed by using SAS version 9.2 (SAS Institute Inc, Cary, NC), including macros %BVAL and %XVAL revised for generalized logistic regression, and R version 2.13.2 software (R Project for Statistical Computing), including the nonbinROC package. Statistical significance was assessed by using 2-sided $P$ values at the 5% testing level.

**RESULTS**

A total of 6424 infants born between 23 and 30 weeks’ GA with known birth date were admitted to NICUs participating in the CNN in 2010 and 2011. Of these, 318 infants were excluded; details about excluded infants are provided in Fig 1. Among the final 6106 (95% of 6424) infants eligible for this study, 2280 (37.3%) survived without morbidities.
1964 (32.2%) had mild morbidities, 1251 (20.5%) developed severe morbidities, and 611 (10.0%) died. Table 1 presents the joint distribution of morbidities used to derive the outcome groups. Statistically significant differences between outcome groups were detected for all factors, except for gender (borderline significance), multiple birth, maternal smoking, and maternal drug use (Table 2). Significant trends in outcome severity were detected for all factors, except for gender (borderline significance), maternal smoking, and maternal drug use (Table 2).

Table 3 presents odds ratios and 95% CIs for the 8 factors (GA, SGA, gender, SNAP II >20, outborn status, no antenatal corticosteroids, surfactant on day 1, and IPPV or HFV on day 1) selected in the final multivariable model developed using the complete data. Compared with that of survival free of morbidities, the odds of each mortality, severe morbidities and mild morbidities were significantly higher for lower GA and for infants with SGA, SNAP II >20, or mechanical ventilation using IPPV or HFV on day 1 in the NICU. Moreover, the odds ratio estimate for GA, SGA, SNAP II >20, and mechanical ventilation was larger as outcome severity increased, indicating stronger effects. Mortality and severe morbidities were higher for infants whose mothers did not receive antenatal corticosteroids. The odds of mortality were significantly higher in male infants, and the odds of severe morbidities were higher for infants treated with surfactant on day 1.

Predicted probabilities were generally close to the diagonal line (perfect prediction), except for mortality, which was underestimated by probabilities >.80 (Supplemental Fig 2). Measures summarizing the distribution of predicted probabilities and quantifying prediction ability are presented in Table 4. Prediction accuracy was highest for mortality (91%) and lowest for survival with mild morbidities (63%). Additional results using a range of cutoff values are presented in Supplemental Table 6. No statistically significant differences were detected between predicted probabilities and observed outcomes ($C_p$ test $P = .96$), indicating good model fit. The Brier score = .60 from the multivariable model was closer to 0 (perfect prediction) than the Brier score = .70 from the model without any covariates, yielding a Brier skill score of 0.15, which indicates that covariates in the multivariable model explain 15% of variability in the data. The bootstrap optimism correction of 0.0036 was negligible, reducing the resubstitution $C$-statistic = 0.902 (95% CI = 0.897–0.906) to a bias-corrected $C$-statistic = 0.899 (95% CI = 0.894–0.903).

The complete data prediction model was developed with the following characteristics as baseline: birth at 30 weeks’ GA, not SGA, female gender, SNAP II ≤20, received antenatal corticosteroids, inborn status, no surfactant administration, and no IPPV or HFV on day 1 in the NICU. For an infant with these baseline characteristics, the predicted probability of survival without morbidities, mild morbidities, severe morbidities, and mortality is equal to .76, .20, .04, and .004, respectively (Table 5, first row). To facilitate computation of predicted probabilities for infants with characteristics different from baseline, Table 5 shows the change in the probability of each outcome level associated with a change from baseline for each risk factor. For example, if an infant has the same baseline characteristics except for SNAP II >20, then the probability of survival free of morbidities decreases by .11, whereas the probability of mild morbidities, severe morbidities, and mortality increases by .04, .05, and .02, respectively (Table 5). For an infant outborn at 28 weeks’ GA who is SGA, male, with SNAP II >20 and treated with antenatal corticosteroids, with surfactant on day 1, and with IPPV or HFV on day 1, the predicted probability of survival free of morbidities decreases by .11, whereas the probability of mild morbidities, severe morbidities, and mortality increases by .04, .05, and .02, respectively (Table 5). For an infant outborn at 28 weeks’ GA who is SGA, male, with SNAP II >20 and treated with antenatal corticosteroids, with surfactant on day 1, and with IPPV or HFV on day 1, the predicted probability of survival free of morbidities decreases by .11, whereas the probability of mild morbidities, severe morbidities, and mortality increases by .04, .05, and .02, respectively (Table 5).
The probability of no morbidities on day 1 is .14, calculated as .76 (baseline). The values listed in Table 5 can be used similarly to compute the probability of mild morbidity as .57, severe morbidity as .25, and mortality as .05.

**DISCUSSION**

The multivariable model derived in this population-based cohort of extremely preterm infants has high (90%) discrimination to predict, on the first day of admission to the NICU, survival without morbidities, survival with mild morbidities, survival with severe morbidities, or mortality during NICU hospitalization. The 8 predictors include routinely collected data on GA, SGA, gender, inborn or outborn status, antenatal corticosteroid use, SNAPII score, surfactant use on day 1, and receipt of IPPV or HFV or CPAP on day 1. The values listed in Table 5 can be used similarly to compute the probability of mild morbidity as .57, severe morbidity as .25, and mortality as .05.

**TABLE 1**

<table>
<thead>
<tr>
<th>Morbidities (N Missing)</th>
<th>IVH or PVL</th>
<th>ROP</th>
<th>BPD</th>
<th>NEC</th>
<th>Sepsis</th>
</tr>
</thead>
<tbody>
<tr>
<td>IVH or PVL (638 missing)</td>
<td>None</td>
<td>Mild</td>
<td>Severe</td>
<td>None</td>
<td>Mild</td>
</tr>
<tr>
<td>None</td>
<td>4157</td>
<td>0</td>
<td>0</td>
<td>3252</td>
<td>653</td>
</tr>
<tr>
<td>Mild (grade 1 or 2 IVH)</td>
<td>0</td>
<td>1175</td>
<td>0</td>
<td>789</td>
<td>280</td>
</tr>
<tr>
<td>Severe (grade 3 or 4 IVH or PVL)</td>
<td>0</td>
<td>0</td>
<td>794</td>
<td>539</td>
<td>178</td>
</tr>
<tr>
<td>ROP (2787 missing)</td>
<td>None</td>
<td>3325</td>
<td>789</td>
<td>559</td>
<td>4653</td>
</tr>
<tr>
<td>Mild (stage 1 or 2 ROP)</td>
<td>633</td>
<td>280</td>
<td>178</td>
<td>0</td>
<td>1091</td>
</tr>
<tr>
<td>Severe (stage ≥3 ROP or surgery)</td>
<td>179</td>
<td>106</td>
<td>77</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BPD (573 missing)</td>
<td>None</td>
<td>3243</td>
<td>814</td>
<td>554</td>
<td>3976</td>
</tr>
<tr>
<td>Mild (oxygen without IPPV, HFV, or CPAP)</td>
<td>625</td>
<td>238</td>
<td>150</td>
<td>561</td>
<td>318</td>
</tr>
<tr>
<td>Severe (oxygen with IPPV, HFV, or CPAP)</td>
<td>269</td>
<td>123</td>
<td>90</td>
<td>216</td>
<td>148</td>
</tr>
<tr>
<td>NEC (5754 missing)</td>
<td>None</td>
<td>3950</td>
<td>1091</td>
<td>713</td>
<td>4443</td>
</tr>
<tr>
<td>Mild (stage 2 NEC)</td>
<td>137</td>
<td>47</td>
<td>39</td>
<td>126</td>
<td>58</td>
</tr>
<tr>
<td>Severe (stage 3 NEC)</td>
<td>50</td>
<td>37</td>
<td>42</td>
<td>84</td>
<td>27</td>
</tr>
<tr>
<td>Sepsis (0 missing)</td>
<td>None</td>
<td>3487</td>
<td>883</td>
<td>549</td>
<td>3999</td>
</tr>
<tr>
<td>Mild (single episode of infection)</td>
<td>515</td>
<td>237</td>
<td>173</td>
<td>545</td>
<td>272</td>
</tr>
<tr>
<td>Severe (multiple episodes of infection)</td>
<td>135</td>
<td>55</td>
<td>72</td>
<td>108</td>
<td>91</td>
</tr>
</tbody>
</table>
severe. Mild morbidities were defined to include all other stages or degrees of morbidities, such that the desired outcome was survival free of any morbidity. Additional refinement to differentiate between type and number of severe and mild morbidities is possible, but requires larger data samples.

To facilitate prediction, investigators have developed graphical tools, but these are limited to prediction of binary outcomes (mortality versus survival, mortality or severe morbidity versus survival free of severe morbidity) and consider the range of only 2 predictors, typically GA and birth weight. In contrast, our prediction model discriminates across a spectrum of outcome severity levels by using 8 factors with variable impact on outcome levels. Furthermore, distinguishing outcome severity levels improves prediction; the $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.

Factors in our model include GA, SGA, gender, inborn or outborn status, and antenatal corticosteroid use, which are well-known predictors of survival in preterm infants. As expected, GA was the strongest predictor; GA measured in completed weeks was modeled as a continuous measure requiring linear and quadratic terms that are difficult to include all other stages or degrees of morbidities, such that the desired outcome was survival free of any morbidity. Additional refinement to differentiate between type and number of severe and mild morbidities is possible, but requires larger data samples.

To facilitate prediction, investigators have developed graphical tools, but these are limited to prediction of binary outcomes (mortality versus survival, mortality or severe morbidity versus survival free of severe morbidity) and consider the range of only 2 predictors, typically GA and birth weight. In contrast, our prediction model discriminates across a spectrum of outcome severity levels by using 8 factors with variable impact on outcome levels. Furthermore, distinguishing outcome severity levels improves prediction; the $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.

Factors in our model include GA, SGA, gender, inborn or outborn status, and antenatal corticosteroid use, which are well-known predictors of survival in preterm infants. As expected, GA was the strongest predictor; GA measured in completed weeks was modeled as a continuous variable rather than as a categorical variable with fixed categories of severity. The $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.

Factors in our model include GA, SGA, gender, inborn or outborn status, and antenatal corticosteroid use, which are well-known predictors of survival in preterm infants. As expected, GA was the strongest predictor; GA measured in completed weeks was modeled as a continuous variable rather than as a categorical variable with fixed categories of severity. The $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.

Factors in our model include GA, SGA, gender, inborn or outborn status, and antenatal corticosteroid use, which are well-known predictors of survival in preterm infants. As expected, GA was the strongest predictor; GA measured in completed weeks was modeled as a continuous variable rather than as a categorical variable with fixed categories of severity. The $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.

Factors in our model include GA, SGA, gender, inborn or outborn status, and antenatal corticosteroid use, which are well-known predictors of survival in preterm infants. As expected, GA was the strongest predictor; GA measured in completed weeks was modeled as a continuous variable rather than as a categorical variable with fixed categories of severity. The $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.

Factors in our model include GA, SGA, gender, inborn or outborn status, and antenatal corticosteroid use, which are well-known predictors of survival in preterm infants. As expected, GA was the strongest predictor; GA measured in completed weeks was modeled as a continuous variable rather than as a categorical variable with fixed categories of severity. The $C$-statistic = 0.902 from the multinomial model exceeds the $C$-statistic = 0.817 obtained from a logistic model with the same predictors but using a binary indicator outcome for survival without severe morbidity, derived by pooling infants who survived with none or mild morbidities together and pooling those who survived with severe morbidities, together with those who died.
Our model also included receipt of surfactant and mechanical ventilation on day 1 of admission, which were generally predictive of poorer outcomes but provided minimal increase in the C-statistic and may be excluded for a more parsimonious model. Furthermore, if SNAP-II is replaced by the Apgar or Clinical Risk Index for Babies score, a predictive model can be developed that would be applicable at the time of admission to the NICU. Site was considered as a potential confounder, but because no changes >10% in parameter estimates associated with site were detected, site was excluded as a possible predictor to develop a prediction model that would be generally useful for infants admitted to sites outside the CNN.

Our model was well calibrated for survival without morbidity, survival with mild morbidity, and survival with severe morbidity but tended to underpredict mortality in high-risk infants. The discriminative ability of our model was internally validated by using a bootstrap approach to estimate and correct bias in the C-statistic due to overfitting. The advantage of bootstrap validation is use of the complete data for model development and nearly unbiased estimates of prediction ability.30 We also considered alternative approaches for internal validation, including split-sample, leave-one-out, and k-fold cross-validation, with comparable estimates of predictive ability (results not shown). Split-sample validation entails random splitting of the data into independent training and validation subsets for model development and assessment, respectively. Although split-sample validation provides unbiased estimates of predictive ability, model building using a smaller training subset may exclude important risk factors that do not reach the required statistical significance threshold because of reduced power. Validation by using a reduced sample will increase the variability of the C-statistic estimate. Furthermore, a single random split can produce substantial imbalances in the distribution of the outcome and predictors, between the training and validation data subsets, resulting in poor model performance.38

As an alternative to the split-sample approach, sample-reuse methods, including leave-one-out and k-fold cross-validation, use the complete data for both training and validation. In k-fold cross-validation the complete data are randomly divided into k equal parts (typically 5–10), and each subset is in turn retained for validation while the other k – 1 subsets are used to train a predictive model. Leave-one-out is a special case of k-fold cross-validation with k = total sample size. The disadvantage of k-fold and leave-one-out cross-validation is that multiple models are validated, and it is difficult to interpret.15,34 Because of the high correlation between GA and birth weight, we derived and used SGA (<10th percentile birth weight for GA by gender) as a predictor instead of birth weight. Although the majority (>85%) of infants in this recent (2010–2011) Canadian cohort received antenatal corticosteroids, infants who were not treated with antenatal corticosteroids were at significant disadvantage. Whereas previous studies examined Apgar score at 1 or 5 minutes or the Clinical Risk Index for Babies score to quantify illness severity, we used SNAP-II score >20 to provide an objective assessment of infants’ health independent of GA and birth weight. The SNAP-II score is computed using 5 physiologic measures obtained in the first 12 hours of NICU admission and is a proven predictor of neonatal mortality.37

### TABLE 4 Distribution of Predicted Probabilities and Measures of Predictive Ability From the Complete Data Multivariable Model Using 6106 Infants Born at ≤30 Weeks’ Gestation

<table>
<thead>
<tr>
<th>Measure</th>
<th>No Morbidity</th>
<th>Mild Morbidity</th>
<th>Severe Morbidity</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infants 23–30 wk GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>.004</td>
<td>.026</td>
<td>.042</td>
<td>.005</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>(.142–.606)</td>
<td>(.237–.411)</td>
<td>(.089–.319)</td>
<td>(.012–.106)</td>
</tr>
<tr>
<td>Mean</td>
<td>.370</td>
<td>.329</td>
<td>.208</td>
<td>.093</td>
</tr>
<tr>
<td>Median</td>
<td>.346</td>
<td>.346</td>
<td>.184</td>
<td>.033</td>
</tr>
<tr>
<td>Maximum</td>
<td>.762</td>
<td>.592</td>
<td>.533</td>
<td>.741</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.725</td>
<td>.628</td>
<td>.777</td>
<td>.910</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>.606</td>
<td>.439</td>
<td>.447</td>
<td>.529</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>.927</td>
<td>.730</td>
<td>.833</td>
<td>.832</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>.733</td>
<td>.470</td>
<td>.515</td>
<td>.305</td>
</tr>
<tr>
<td>Specificity</td>
<td>.520</td>
<td>.705</td>
<td>.988</td>
<td>.972</td>
</tr>
<tr>
<td>Brier score (Brier skill score)</td>
<td>.595 (.153)</td>
<td>.902 (.897–.906)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infants 23–28 wk GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum</td>
<td>.004</td>
<td>.026</td>
<td>.102</td>
<td>.013</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>(.081–.334)</td>
<td>(.283–.451)</td>
<td>(.187–.373)</td>
<td>(.051–.207)</td>
</tr>
<tr>
<td>Mean</td>
<td>.205</td>
<td>.364</td>
<td>.288</td>
<td>.143</td>
</tr>
<tr>
<td>Median</td>
<td>.154</td>
<td>.391</td>
<td>.283</td>
<td>.076</td>
</tr>
<tr>
<td>Maximum</td>
<td>.514</td>
<td>.552</td>
<td>.535</td>
<td>.741</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.789</td>
<td>.574</td>
<td>.687</td>
<td>.863</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>.466</td>
<td>.445</td>
<td>.448</td>
<td>.533</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>.829</td>
<td>.736</td>
<td>.763</td>
<td>.895</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>.256</td>
<td>.679</td>
<td>.378</td>
<td>.351</td>
</tr>
<tr>
<td>Specificity</td>
<td>.925</td>
<td>.515</td>
<td>.811</td>
<td>.952</td>
</tr>
<tr>
<td>Brier score (Brier skill score)</td>
<td>.648 (.148)</td>
<td>.887 (.881–.885)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The outcome with the highest probability was considered as the outcome predicted by the model.
recommend a final predictive model for future use.

Strengths of our study are the use of a large contemporary cohort representing a population-based sample from across Canada, and meticulous data collection within the network. Data from all NICUs participating in the CNN were collected according to established protocol standards, with data processing undergoing a number of error-checking steps. The 13 potential risk factors were selected to represent maternal and infant factors of known clinical importance that are commonly recorded. To quantify the impact of predictors on the probability of each outcome group, we have provided the change in outcome probability associated with changes in each risk factor, such that the predicted probabilities of any infant can be computed with simple arithmetic.

Our study was limited by incomplete data on race or ethnicity, and chorioamnionitis, both of which have been reported to be associated with neonatal outcomes. 8,39 Although multiple imputation methods can be applied to include covariates with missing data, the resulting model would not be applicable to future infants for whom these data are missing. We were unable to perform external validation because a comparable independent data set was not readily available; however, we are establishing collaborations with several other national neonatal networks to externally validate our prediction model in future work.

CONCLUSIONS

We have developed a multivariable model to predict survival without morbidity, survival with mild morbidity, survival with severe morbidity, or mortality for extremely preterm infants admitted to level III NICUs in Canada. The model predicted outcome severity levels with high discrimination and was internally validated. The information about the 8 predictors included in the model is available on the first day of admission to the NICU, thereby facilitating reliable prognostic information for physicians and families early in the neonatal period.

ACKNOWLEDGMENTS

Site investigators of the CNN were as follows: Prakeshkumar S. Shah, director, CNN; Wayne Andrews, Janeway Children’s Health and Rehabilitation Centre (St John’s, NL); Keith Barrington, Sainte Justine Hospital (Montreal, QC); Wendy Yee, Foothills Medical Centre (Calgary, AB); Barbara Bullied, Everett Chalmers Hospital (Fredericton, NB); Rody Canning, Moncton Hospital (Moncton, NB); Ruben Alvaro, St Boniface General Hospital (Winnipeg, MB); Orlando da Silva, London Health Sciences Centre (London, ON); Kimberly Dow, Kingston General Hospital (Kingston, ON); Michael Dunn, Sunnybrook Health Sciences Centre (Toronto, ON); Adele Harrison, Victoria General Hospital (Victoria, BC); Kyong-Soon Lee, The Hospital for Sick Children (Toronto, ON); Zarin Kalapesi, Regina General Hospital (Regina, SK); Lajos Kovacs, Jewish General Hospital (Montreal, QC); Douglas D. McMillan, IWK Health Centre (Halifax, NS); Prakeshkumar S. Shah, Mount Sinai Hospital (Toronto, ON); Cecil Ojah, St John Regional Hospital (St John, NB); Abraham Peliowski and Khalid Aziz, Royal Alexandra Hospital (Edmonton, AB); Bruno Piedboeuf, Centre Hospitalier Universitaire de Québec (Sainte Foy, QC); Patricia Riley, Montreal Children’s Hospital (Montreal, QC); Daniel Faucher, Royal Victoria Hospital (Montreal, QC); Nicole Rouvinez-Bouali, Children’s Hospital of Eastern Ontario (Ottawa, ON); Koravangattu Sankaran, Royal University Hospital (Saskatoon, SK); Mary Seshia, Health Sciences Centre (Winnipeg, MB); Sandesh Shivananda, Hamilton Health Sciences Centre (Hamilton, ON); Zenon Cieslak, Royal Columbian Hospital (New Westminster, BC); Anne Synnes, Children’s and Women’s Health Centre of British Columbia (Vancouver, BC); Valerie Bertelle, Centre Hospitalier Universitaire de Sherbrooke (Fleuri Mont, QC); Andrzej Kajetanowicz, Cape Breton Regional Hospital (Cape Breton, NS);
REFERENCES


(Continued from first page)

www.pediatrics.org/cgi/doi/10.1542/peds.2013-0702
doi:10.1542/peds.2013-0702
Accepted for publication Jul 23, 2013

Address correspondence to Lucia Mirea, PhD, Maternal–Infant Care (MiCare) Research Centre, Mount Sinai Hospital, 700 University Ave, Suite 8-500, Toronto, Ontario M5G 1X8. E-mail: lmirea@mtsain.on.ca

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).
Copyright © 2013 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Organizational support to the Canadian Neonatal Network was provided by the Canadian Institutes of Health Research Team in Maternal–Infant Care (MiCare). The Maternal–Infant Care Research Centre was supported by the Ministry of Health and Long-term Care, Ontario, Canada.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.
Prediction of Neonatal Outcomes in Extremely Preterm Neonates

Wen J. Ge, Lucia Mirea, Junmin Yang, Kate L. Bassil, Shoo K. Lee and Prakeshkumar S. Shah

Pediatrics 2013;132;e876; originally published online September 23, 2013; DOI: 10.1542/peds.2013-0702
Prediction of Neonatal Outcomes in Extremely Preterm Neonates
Wen J. Ge, Lucia Mirea, Junmin Yang, Kate L. Bassil, Shoo K. Lee and Prakeshkumar S. Shah

Pediatrics 2013;132;e876; originally published online September 23, 2013;
DOI: 10.1542/peds.2013-0702

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/132/4/e876.full.html