

## SNAP-II Predicts Severe Intraventricular Hemorrhage and Chronic Lung Disease in the Neonatal Intensive Care Unit

Li-Yin Chien, MPH, ScD  
Robin Whyte, MBBS, FRCPC, MSc  
Paul Thiessen, MD, FRCPC  
Robin Walker, MBBS, FRCPC  
David Brabyn, MBChB, FRACP, FRCPC  
Shoo K. Lee, MBBS, FRCPC, PhD  
The Canadian Neonatal Network

### OBJECTIVE:

To determine whether the Score for Neonatal Acute Physiology, Version II (SNAP-II), improved prediction of severe ( $\geq$ grade III) intraventricular hemorrhage (IVH) and chronic lung disease (CLD) when compared to models using gestational age (GA) and traditional risk factors (e.g., Apgar score, small-for-gestational-age, sex, outborn status).

### STUDY DESIGN:

We examined 4226 infants  $\leq$ 32 weeks' GA admitted to 17 Canadian neonatal intensive care units between 1996 and 1997. We compared prediction models for severe IVH and CLD, with and without SNAP-II.

### RESULTS:

SNAP-II was a significant and independent predictor of severe IVH and CLD. Addition of SNAP-II to models using GA and traditional risk variables significantly ( $p < 0.05$ ) improved model prediction (AUC 0.8 for severe IVH; 0.83 for CLD). Models were well calibrated ( $p > 0.05$  for Hosmer-Lemeshow goodness of fit test).

### CONCLUSION:

Addition of SNAP-II to models using GA and traditional risk factors significantly improves prediction of severe IVH and CLD.

*Journal of Perinatology* (2002) 22, 26–30 DOI: 10.1038/sj/jp/7210585

### BACKGROUND

With increasing survival of preterm infants, reducing the incidence of morbidity (e.g., chronic lung disease [CLD] and severe [ $\geq$ grade III] intraventricular hemorrhage [IVH]) is a priority in the neonatal intensive care unit (NICU).<sup>1</sup> Wennberg et al.<sup>2</sup> advocated studying small area variations as a method for improving quality of care. Small area variations represent the practice patterns of small groups of physicians. Differences in practice may have different outcomes, and if these differences can be identified, strategies can be designed for improving quality of care.<sup>2–4</sup> However, valid comparisons of outcomes can only be made if adjustments are made for risks that are not under the practitioner's control. These differences in patient risks are sometimes referred to as "case mix"<sup>3</sup> or "initial risk".<sup>5,6</sup>

The Score for Neonatal Acute Physiology, Version II (SNAP-II),<sup>7</sup> is a simplified neonatal illness severity score that measures six empirically weighted, physiology-based items during a 12-hour period, including lowest blood pressure, lowest temperature, PO<sub>2</sub>/FiO<sub>2</sub> ratio, lowest serum pH, seizures, and urine output. SNAP-II has a score range of 0 (low severity) to 115 (high severity). SNAP-II has been validated for mortality risk prediction in three different populations.<sup>7</sup> Addition of SNAP-II to prediction models using gestational age (GA) and other traditional risk factors (e.g., sex, birth weight, Apgar score, small-for-gestational-age [SGA], outborn status) significantly improved prediction of NICU mortality.<sup>7</sup> The objectives of the current study were to examine whether SNAP-II independently predicts severe IVH and CLD, and whether addition of SNAP-II to models using GA and other traditional risk factors improves prediction of severe IVH and CLD.

### METHODS

#### Study Population

The Canadian Neonatal Network<sup>8</sup> comprises 17 NICUs throughout Canada, representing 75% of all tertiary level NICU beds in Canada, and serving a population of about 22 million people. In 1996,

*Department of Pediatrics (L.Y.C., P.T., S.K.L.), University of British Columbia, Vancouver, BC, Canada; Centre for Community Health and Health Evaluation Research (L.Y.C., S.K.L.), Vancouver, BC, Canada; Department of Pediatrics (R.W.), Dalhousie University, Halifax, NS, Canada; Department of Pediatrics (R.W.), University of Ottawa, Ottawa, ON, Canada; and Department of Pediatrics (D.B.), Royal Columbian Hospital, New Westminster, BC, Canada.*

*This study was supported by grants 40503 and 00152 from the Medical Research Council of Canada. Additional funding was provided by the B.C. Children's Hospital Foundation; Calgary Regional Health Authority; Division of Neonatology, Children's Hospital of Eastern Ontario; Child Health Program, Health Care of St. John's; The Neonatology Program, Hospital for Sick Children; Lawson Research Institute; Midland Walwyn Capital; Division of Neonatology, McMaster Health Sciences Centre; Mount Sinai Hospital; North York General Hospital Foundation; Saint Joseph's Health Centre; University of Western Ontario; Women's College Hospital.*

*Address correspondence and reprint requests to Shoo K. Lee, MBBS, FRCPC, PhD, Canadian Neonatal Network, Canadian Neonatal Network Coordinating Centre, 4480 Oak Street, Room E-414, Vancouver, BC, Canada V6H 3V4.*

Canada had a population of 30 million<sup>9</sup> with about 357,000 annual births.<sup>10</sup> The 17 NICUs range in size from 18 to 97 beds, from 133 to 1129 admissions, and from 0 to 7500 births annually. All, but one, are regional tertiary perinatal centers. Transports from community hospitals account for about 26% of admissions. During the study period from January 8, 1996 to October 31, 1997, 19,507 patients were admitted to the Canadian Neonatal Network.<sup>8</sup> In this study, we excluded infants who: (1) were released to the regular nursery in <24 hours; (2) were admitted to the NICU at >48 hours of age or after having been discharged home; (3) were moribund on admission, i.e., where an explicit physician decision not to provide life support was made at the time of NICU admission; (4) died in the delivery room; and (5) were >32 weeks' GA, because CLD and IVH were most relevant to infants  $\leq$ 32 weeks' GA. There were 4907 infants who met the inclusion criteria. Of these, 334 (6%) infants who did not have complete information were excluded. For CLD validation, we further excluded 347 (7%) infants who died before 36 weeks' corrected GA. The remaining 4226 infants were included in the analysis of CLD. For analysis of IVH, we included only infants ( $n=3778$  infants or 77%) who had cranial ultrasound examinations during the first 2 weeks of life. Characteristics of the study population are shown in Table 1. When compared to infants without cranial ultrasound, infants with cranial ultrasound had lower mean birth weight (1227 vs 1730 g), lower mean GA (29 vs 31 weeks), and higher mean admission SNAP-II (15 vs 7).

### Data Abstraction

Trained research assistants abstracted patient information from the mothers' and infants' charts at each participating hospital on a daily basis.<sup>8</sup> Data were directly entered into laptop computers at the bedside using a customized data entry program with built-in error checking and a standard manual of operations and definitions. Data were electronically transmitted to the Canadian Neonatal Network Coordinating Centre, at the British Columbia Research Institute for Children's and Women's Health, for verification. Potential data errors were re-checked by site research assistants. Data management was conducted by the Canadian Neonatal Network Coordinating Centre, in concert with a Steering Committee comprising experienced researchers and neonatologists representing each of the five geographic regions (British Columbia, Prairie provinces,

Ontario, Quebec, Atlantic provinces) in Canada, and with site investigators representing each of the 17 participating hospitals. Patient information was collected until death or discharge from the NICU. Patients transferred to another hospital were tracked until death or discharge home and outcome information collected. Data analysis was performed using each infant rather than each admission. The study was approved by institutional review boards at all 17 hospitals.

### Patient Information

Patient information collected included demographic information, status of infant at birth, admission illness severity (SNAP<sup>11</sup> and SNAP-II<sup>7</sup>), use of technology and resources, and selected patient outcomes. Study variables were defined according to the Canadian Neonatal Network SNAP Project Abstractor Manual. GA was defined as the best obstetric estimate based on early prenatal ultrasound, obstetric examination, and obstetric history, unless the postnatal pediatric estimate of gestation differed from the obstetric estimate by more than 2 weeks. In that case, the pediatric estimate of GA based on the Ballard Score<sup>12</sup> or the best estimate of the attending neonatologist was used instead. An infant was defined as SGA if the birth weight was less than the third percentile for GA according to the British Columbia provincial growth charts established by Whitfield<sup>13</sup> in 1992. Severe IVH was defined as  $\geq$ grade III IVH according to the criteria of Papile et al.<sup>14</sup> from head ultrasound performed before 14 days of life. CLD was defined as oxygen dependency at 36 weeks' corrected GA for an infant who was born at  $\leq$ 32 weeks' gestation.<sup>15</sup>

### Data Analysis

**Criterion validity.** We compared SNAP-II against SNAP<sup>11</sup> because the former was the simplified version of the latter and SNAP has been widely used in clinical research.<sup>16–18</sup> We used the area under the Receiver Operator Characteristic (ROC) curve<sup>19,20</sup> to compare discrimination and the multiple regression model  $\chi$ -squared analysis to compare explanatory power of the models. A higher  $\chi$ -squared value indicates better explanatory power.

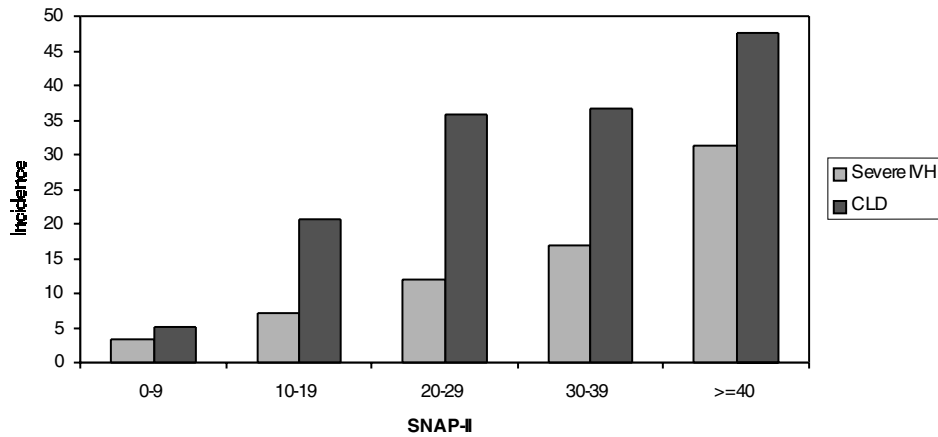
**Theoretical consistency.** Mean scores were compared for infants with and without severe IVH and CLD. SNAP-II was plotted against incidence of severe IVH and CLD.

**Discrimination.** We assessed performance of SNAP-II and other risk variables for prediction of severe IVH and CLD using the area under the ROC curve technique of Hanley and McNeil.<sup>19,20</sup> The area under the ROC curve (AUC) is a measure of discrimination, which gives the probability of any patient with the disease having a higher score than any patient without the disease.<sup>21,22</sup> An AUC of 1.0 indicates perfect discrimination, whereas an area of 0.5 would be completely random.

**Predictive performance and calibration.** Multiple logistic regression<sup>23</sup> was used to model IVH and CLD. In each model, SNAP-II was tested singly and in combination with other risk variables.

**Table 1** Characteristics of the Study Cohort

	IVH sample ( $n=3778$ )	CLD sample ( $n=4226$ )
Male (%)	56.2	56.2
Mean birth weight $\pm$ SD (g)	1190 $\pm$ 426	1390 $\pm$ 457
Mean GA $\pm$ SD (weeks)	29 $\pm$ 2	29 $\pm$ 2
Apgar at 5 minutes <7 (%)	10.9	7.4
Outborn (%)	17.3	14.4
SGA (%)	5.2	4.0
Mean admission SNAP-II $\pm$ SD	14 $\pm$ 13	11 $\pm$ 11



**Figure 1.** Incidence of severe ( $\geq$ grade III) IVH and CLD rises with increasing admission SNAP-II.

Likelihood ratio statistics were used to examine if addition of SNAP-II significantly ( $p < 0.05$ ) improved model prediction. The relative contribution of each variable (SNAP-II, GA, and other traditional risk factors) to model predictive power was determined by examining the relative decrease in likelihood ratio when each variable was removed from the full model. We used the Hosmer-Lemeshow method<sup>24</sup> to test goodness of fit and assess deviations between the observed and expected number of severe IVH and CLD cases. A  $p$  value of  $> 0.05$  indicates no significant difference between the observed and expected values, and therefore, the fit of the model is acceptable.

**RESULTS**

**Criterion Validity**

SNAP-II compared well with SNAP in predicting severe IVH and CLD. For prediction of severe IVH, AUC were  $0.73 \pm 0.01$  for SNAP, and  $0.73 \pm 0.02$  for SNAP-II; model  $\chi^2$  were 222 for SNAP, and 219 for SNAP-II. For prediction of CLD, AUC were  $0.74 \pm 0.01$  for SNAP, and  $0.78 \pm 0.01$  for SNAP-II; model  $\chi^2$  were 381 for SNAP, and 470 for SNAP-II.

**Theoretical Consistency**

Mean SNAP-II was significantly ( $p < 0.001$ ) higher among infants with severe IVH than among those without severe IVH (mean SNAP-II = 26.4 vs 13.6, respectively), and higher ( $p < 0.001$ ) among infants with CLD than among those without CLD (mean SNAP-II = 20.5 vs 9.5, respectively). The incidence of severe IVH and CLD increased with increasing SNAP-II (Figure 1).

**Discrimination**

When used alone, SNAP-II discriminated presence from absence of severe IVH and CLD with AUC of 0.73 and 0.78, respectively. For comparison, the AUC for using GA alone was 0.75 for severe IVH and 0.77 for CLD, which were not significantly different from the AUC using SNAP-II alone ( $p = 0.13$  for severe IVH;  $p = 0.19$  for CLD). Models using GA plus SGA (AUC = 0.75 for IVH; 0.78 for CLD) were not significantly different from models using GA alone and SNAP-II

alone. Addition of SNAP-II to models using GA plus SGA significantly ( $p < 0.05$ ) improved AUC (0.78 for IVH; 0.82 for CLD). The full prediction models using SNAP-II plus GA and variables (outborn status and low Apgar score for severe IVH; SGA, sex, low Apgar, and outborn status for CLD) were more discriminative than models using SNAP-II alone, GA alone, and GA plus variables in predicting severe IVH and CLD (Table 2).

**Predictive Performance**

SNAP-II was a significant predictor of severe IVH and CLD (Table 3), even after adjustment for GA and other perinatal risks. Addition of SNAP-II to models using GA plus variables significantly improved model prediction (likelihood ratio statistics = 62.4 for severe IVH and 134.3 for CLD; both  $p < 0.0001$ ). Each point increment in SNAP-II score increased the odds of severe IVH and CLD by 4% and 5%, respectively. Each week increment in GA decreased the odds of severe IVH and CLD by 22% and 28%, respectively. Low Apgar score at 5 minutes and outborn status were associated with higher risk of severe IVH and CLD. SGA and male sex were associated with

Model	GA	SNAP-II	GA + variables*	SNAP-II + GA + variables*
<i>IVH</i>				
AUC	0.75	0.73	0.78	0.80†
Hosmer-Lemeshow $p$ value	0.38	0.09	0.34	0.16
<i>CLD</i>				
AUC	0.77	0.78	0.80	0.83†
Hosmer-Lemeshow $p$ value	0.98	0.01	0.87	0.51

\*Variables are outborn status and Apgar at 5 minutes  $< 7$  for IVH; SGA, male, Apgar at 5 minutes  $< 7$ , and outborn status for CLD.  
 †Indicates AUC significantly ( $p < 0.05$ ) different from models using GA, SNAP-II, and GA + variables for each outcome.

**Table 3** Prediction Models for Severe ( $\geq$ Grade III) IVH and CLD

Risk factors	Odds ratio	95% Confidence interval	Percent contribution to predictive power
<i>Model: severe (<math>\geq</math>grade III) IVH</i>			
GA	0.78	0.74, 0.82	41
SNAP-II	1.04	1.03, 1.05	30
Outborn status	2.77	2.08, 3.66	23
Apgar at 5 minutes $<7$	1.77	1.20, 2.42	6
<i>Likelihood ratio reduction = 17.5%</i>			
<i>Hosmer-Lemeshow goodness of fit statistic, <math>p=0.16</math></i>			
<i>Model: CLD</i>			
GA	0.72	0.69, 0.75	54
SNAP-II	1.05	1.04, 1.06	30
SGA	3.92	2.66, 5.78	9
Male	1.43	1.18, 1.75	3
Apgar at 5 minutes $<7$	1.55	1.16, 2.08	2
Outborn status	1.48	1.15, 1.91	2
<i>Likelihood ratio reduction = 21.7%</i>			
<i>Hosmer-Lemeshow goodness of fit statistic, <math>p=0.51</math></i>			

increased risk of CLD. Removal of SNAP-II and other variables from the full regression models resulted in a likelihood ratio reduction of 17.5% for severe IVH and 21.7% for CLD. We calculated relative contributions of SNAP-II and other variables to the model predictive power. SNAP-II contributed a significant proportion to model prediction (30% for severe IVH and CLD), next only to GA (Table 3). SNAP-II and GA together accounted for most of model prediction (71% for severe IVH and 84% for CLD).

### Calibration

Because models for prediction of severe IVH and CLD should include SNAP-II plus all other significant variables (Table 2), the relevant consideration is the calibration of the full model and not the calibration of single variables alone in the model. In this regard, the models using SNAP-II plus GA plus other variables were well calibrated ( $p > 0.05$  for Hosmer-Lemeshow goodness of fit test), indicating no significant difference between the observed and expected number of IVH and CLD cases, respectively (Table 2).

### DISCUSSION

Our results show that SNAP-II is an independent and significant predictor of severe IVH and CLD. Addition of SNAP-II to traditional risk factors, such as GA and perinatal risks, increases the predictive performance and discriminative power of the models for prediction of severe IVH and CLD without loss of calibration. SNAP-II and GA together account for more than 70% of the model prediction of severe IVH and CLD. Failure to control for differences in admission illness severity may result in erroneous conclusions about the true incidence of severe IVH and CLD when

comparing outcomes between institutions. In order to provide fair and valid comparisons between institutions, outcome comparisons should utilize the best tools available. Our results show that SNAP-II improved the quantification of risk for prediction of severe IVH and CLD, and should be used in addition to traditional risk factors to more appropriately represent the case mix or initial risk for severe IVH and CLD.

We found that SNAP-II performed as well as SNAP in predicting severe IVH and CLD. Richardson et al.<sup>7</sup> previously showed that SNAP-II performed as well as SNAP for prediction of mortality. Because SNAP-II (six items) is quicker and simpler to use than SNAP (34 items), SNAP-II should be used in preference to SNAP as a measure of illness severity for most purposes.

Richardson et al.<sup>7</sup> separately described SNAP-II Perinatal Extension (SNAP-II-PE) as a mortality risk score. SNAP-II-PE incorporated three additional perinatal risk variables (SGA, BW, Apgar score at 5 minutes) into the score, which improved prediction of mortality. We found that independently of SNAP-II, different risk variables were predictive of IVH (GA, Apgar score at 5 minutes, outborn status) and CLD (GA, SGA, Apgar score at 5 minutes, sex, outborn status) in the multiple regression models (Tables 2 and 3). Outcomes other than IVH and CLD may utilize yet other different combinations of risk variables in multiple regression analysis. Therefore, SNAP-II should be used in preference to SNAP-II-PE as a measure of illness severity. The underlying reason is that SNAP-II is a pure measure of physiologic derangement, whereas SNAP-II-PE incorporates nonphysiologic mortality risk variables that may not be relevant to other outcomes. Consequently, each outcome should be modeled separately using SNAP-II and relevant risk variables. The Clinical Risk Index for Babies (CRIB)<sup>25</sup> has the same drawback as SNAP-II-PE because CRIB also incorporates perinatal risk variables designed to optimize prediction of mortality risk but not necessarily other outcomes.

SNAP-II is a simple, objective, physiology-based measure of illness severity that is relatively inexpensive to obtain. The six items in SNAP-II are readily available in neonatal medical records and a score can be assigned in 4 to 6 minutes. An easily calculated illness severity score can be used by clinicians to predict patient outcomes and resource use. SNAP-II has the potential to serve as an adjunct to existing clinical and administrative databases. Routine collection of illness severity may enhance studies of practice variation and facilitate quality improvement efforts.

### LIMITATIONS

The main limitation of SNAP-II is that it requires data collection over 12 hours. Consequently, SNAP-II may measure not only admission illness severity but also interventions during the first 12 hours of NICU admission. This limitation is intrinsic to all physiologic scores such as SNAP and CRIB.<sup>26</sup> Studies in adult and pediatric intensive care suggest that reduction of the time interval

for data collection reduces the prediction of the scores significantly.<sup>27</sup> Given the clinical significance of illness severity scores, the trade-off between purity and accuracy of measurement is decidedly favorable.

## CONCLUSION

SNAP-II predicts severe IVH and CLD independently of GA and other perinatal risk factors. Addition of SNAP-II to GA and other traditional risk factors increases the model prediction of severe IVH and CLD. SNAP-II should be used in combination with other risk variables for quantifying the case mix or initial risk in outcome comparisons.

*Members of the Canadian Neonatal Network and Canadian Neonatal Coordinating Center.*

## Acknowledgment

The authors thank Douglas Richardson, MD, MBA, for his valuable advice in drafting the manuscript.

## References

- Kliegman RM. Neonatal technology, perinatal survival, social consequences, and the perinatal paradox. *Am J Public Health* 1995;85:909–13.
- Wennberg JE, Blowers L, Parker R, Gittelsohn AM. Changes in tonsillectomy rates associated with feedback and review. *Pediatrics* 1977;59(6):821–6.
- Richardson DK, Tamow-Mordi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics* 1999;103:255–65.
- Gould JB. Vital records for quality improvement. *Pediatrics* 1999;103:278–90.
- Bard H. Assessing neonatal risk: CRIB vs. SNAP. *Lancet* 1993;342:449–50.
- Fowlie PW, Gould CR, Tamow-Mordi WO, Strang D. Measurement properties of the clinical risk index for babies — reliability, validity beyond the first 12 hours, and responsiveness over 7 days. *Crit Care Med* 1998;26:163–8.
- Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: simplified newborn illness severity and mortality risk scores. *J Pediatr* 2001;138:92–100.
- Lee SK, McMillan D, Ohlsson A, et al. Variations in practice and outcomes of the Canadian NICU Network 1996–7. *Pediatrics* 2000;106:1070–9.
- Statistics Canada. Population [cited February 23, 2001] [six screens]. Available from URL: <http://www.statcan.ca/english/Pgdb/People/population/demo.02.htm>
- Statistics Canada. Population [cited February 23, 2001] [six screens]. Available from URL: <http://www.statcan.ca/english/Pgdb/People/population/demo.04.htm>
- Richardson DK, Gray JE, McCormick MC, Workman-Daniels K, Goldmann D. Score for Neonatal Acute Physiology (SNAP): a physiologic severity index for neonatal intensive care. *Pediatrics* 1993;91:617–23.
- Ballard JL, Novak KK, Driver M. A simplified assessment of fetal maturation of newly born infants. *J Pediatr* 1979;95:769–74.
- Whitfield M. British Columbia provincial growth charts. B.C. Children's Hospital, 1992.
- Papile L-A, Burstein J, Burstein R, Koffler H. Incidence and evolution of subependymal and intraventricular hemorrhage: a study of infants with birth weights less than 1500 gm. *J Pediatr* 1978;92:529–34.
- Shennan AT, Dunn MS, Ohlsson A, Lennox K, Hoskins EM. Abnormal pulmonary outcomes in premature infants: prediction from oxygen requirement in the neonatal period. *Pediatrics* 1988;82:527–32.
- Richardson DK, Phibbs CS, Gray JE, McCormick MC, Workman-Daniels K, Goldmann DA. Birth weight and illness severity: independent predictors of neonatal mortality. *Pediatrics* 1993;91:969–75.
- Petridou E, Richardson DK, Dessypris N, et al. Outcome prediction in Greek neonatal intensive care units using a Score for Neonatal Acute Physiology (SNAP). *Pediatrics* 1998;101:1037–44.
- Escobar GJ, Fischer A, Li DK, Kremers R, Armstrong, MA. Score for neonatal acute physiology: validation in three Keiser Permanente neonatal intensive care units. *Pediatrics* 1995;96:918–22.
- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29–36.
- Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristics curves derived from the same cases. *Radiology* 1983;148:839–43.
- Ruttimann UE. Statistical approaches to development and validation of predictive instruments. *Crit Care Clin* 1994;10:19–35.
- Lemonshow S, Le Gall JR. Modelling the severity of illness of ICU patients. A systems update. *JAMA, J Am Med Assoc* 1994;272:1049–55.
- Hosmer DW, Lemeshow S. Model-building strategies and methods for logistic regression. *Applied Logistic Regression*. Chapter 4. New York, NY: John Wiley and Sons; 1989. p. 82–133.
- Hosmer DW, Hosmer T, le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997; 16(9):965–80.
- International Neonatal Network. The CRIB (Clinical Risk Index for Infants Score) score — a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;342:193–8.
- Richardson DK, Tamow-Mordi WO, Escobar GJ. Neonatal risk scoring systems: can they predict mortality and morbidity? *Clin Perinatol* 1998;25:591–611.
- Pollack MM, Patel KM, Ruttimann U, Cuerdon T. Frequency of variable measurement in 16 pediatric intensive care units: influence on accuracy and potential for bias in severity of illness assessment. *Crit Care Med* 1996;24:74–7.