

## ORIGINAL ARTICLES

## Comparison of mortality risk: a score for very low birthweight infants

R F Maier, M Rey, B C Metze, M Obladen

### Abstract

**Aim**—To develop and evaluate a score which quantifies mortality risk in very low birthweight (VLBW) infants (birthweight below 1500 g) at admission to the neonatal intensive care unit.

**Methods**—Five hundred and seventy two VLBW infants admitted from 1978 to 1987 were randomly assigned to a cohort (n = 396) for score development and a cohort (n = 176) for score validation. Two hundred and ninety four VLBW infants admitted from 1988 to 1991 were used to compare risk adjusted mortality between the two eras.

**Results**—Using multiple regression analysis, birthweight, Apgar score at 5 minutes, base excess at admission, severity of respiratory distress syndrome, and artificial ventilation were predictive of death in the development cohort. According to regression coefficients, a score ranging from 3 to 40 was developed. At a cutoff of 21, it predicted death in the validation cohort with a sensitivity of 0.85, a specificity of 0.73, and a correct classification rate of 0.76. The area under the receiver operating characteristic curve was 0.86. There was no significant difference in risk severity and in risk adjusted mortality between the eras 1978–87 and 1988–91.

**Conclusion**—The present score is robust, easily obtainable at admission, and permits early randomisation based on mortality risk.

(Arch Dis Child 1997;76:F146-F151)

Keywords: mortality risk; scoring system; very low birthweight

Department of  
Neonatology,  
Virchow-Klinikum,  
Humboldt-Universität,  
Augustenburger  
Platz 1, D-13353  
Berlin, Germany

R F Maier  
M Rey  
B C Metze  
M Obladen

Correspondence to:  
Dr Rolf Maier.

Accepted 19 November 1996

Infant mortality is one of the leading public health problems in developed countries. Very low birthweight (VLBW) is a major cause of death in the first year of life.<sup>1</sup> Most VLBW infants (birthweight below 1500 g) receive intensive care and neonatal mortality has diminished considerably during the past four decades.<sup>2</sup> New therapeutic methods have been investigated in numerous clinical trials. But are the results comparable? In some randomised, placebo controlled surfactant trials results

differed considerably between collaborating centres.<sup>3,4</sup> The assumption that populations defined by birthweight and gestational age are comparable in all hospitals participating in a multicentre trial or during different time periods is questionable.

An approach towards quantifying mortality risk is to develop predictive models using logistic regression. These models sum up patient data related to anamnestic and clinical risk factors and to severity of disease, to assess the probability of a patient's survival. Such models have been validated in adult (APACHE, TRISS) and in paediatric intensive care (PRISM, PeRF).<sup>5-9</sup> Attempts have also been made to predict the death of inpatient VLBW infants from their initial disease severity.<sup>10-12</sup> A score for neonatal acute physiology (SNAP) was developed by Richardson *et al*, using the severest physiological impairments in each organ system during the first 24 hours.<sup>13</sup> Tarnow-Mordi *et al* published another scoring system, the clinical risk index for babies (CRIB), which takes into account birthweight, gestational age, maximum and minimum fraction of inspired oxygen (FIO<sub>2</sub>) and maximum base deficit during the first 12 hours, as well as presence of congenital malformations.<sup>14</sup> The last two scores estimate the risk of dying in hospital with high accuracy, but cannot be assessed before 12–24 hours of age, during which time therapeutic interventions, such as surfactant substitution, or assignment to specific study groups, must already have taken place. The aim of this study was (1) to establish a score which quantifies mortality risk of VLBW infants on admission to the neonatal intensive care unit (NICU) and (2) to assess period specific changes in risk adjusted mortality in our NICU.

### Methods

From 1978 to 1991, 978 VLBW infants with a median (5th; 95th percentiles) birthweight of 1150 (680; 1470) g and a median (5th; 95th percentiles) gestational age of 29.3 (25.0; 35.0) weeks were admitted to our NICU. Six hundred and fifty five of these infants were cared for from 1 January 1978 to 31 December 1987 and 323 from 1 January 1988 to 31 December 1991. Charts of infants admitted before 1 January 1987 were reviewed retro-

Table 1 Mortality risk score for very low birthweight infants obtainable at admission to neonatal intensive care

Variable	Range/grade	Category	Score
Birthweight (g)	1250–1499	4	3
	1000–1249	3	6
	750–999	2	9
	< 750	1	12
Grade of respiratory distress syndrome	0	0	0
	I	1	2
	II	2	4
	III	3	6
	IV	4	8
Apgar score at 5 minutes	> 8	5	0
	7 to 8	4	2
	5 to 6	3	4
	3 to 4	2	6
Artificial ventilation	< 3	1	8
	No	0	0
Base excess at admission (mmol/l)	Yes	4	8
	≥ -2.0	1	0
Score maximum	-2.1 to -5.0	2	1
	-5.1 to -8.0	3	2
	-8.1 to -10.0	4	3
	< -10	5	4
			40

Table 2 Stepwise forward multiple regression analysis showing effect of five independent variables on mortality in very low birthweight infants and resulting equation to predict mortality

Variable	Logistic $\beta$ slopes	Standard error	Regression coefficient	P value
Birthweight	-0.8387	0.1768	-0.2231	<0.0001
Respiratory distress syndrome	0.7051	0.1384	0.2411	<0.0001
Apgar score at 5 minutes	-0.5530	0.1613	-0.1539	0.0006
Artificial ventilation	0.5349	0.1930	-0.1174	0.0056
Base excess at admission	0.3684	0.1321	0.1185	0.0053
Constant	-0.3690	1.1202		0.7419
Equation to predict mortality:				
Pr(mortality) = 1 / (1 + exp (-(-0.37 - 0.84 BW + 0.71 RDS - 0.55 APGAR + 0.53 VENT + 0.37 BE)))				

Pr = probability; BW = birthweight category; RDS = respiratory distress syndrome category; VENT = ventilatory support category; BE = base excess category (as defined in table 1).

Table 3 Characteristics of very low birthweight infants at admission to intensive care unit in two time periods. (If not indicated otherwise, values are given as median (5th; 95th percentiles))

	Group 1 (n=655)	Group 2 (n=323)	P value
Date of birth	1/1978 - 12/1987	1/1988 - 12/1991	
Inborns n (%)	220 (33.6)	233 (72.1)	< 0.01
Infants resuscitated by a neonatologist n (%)	87 (13.3)	202 (62.5)	< 0.01
Boys n (%)	333 (50.8)	177 (54.8)	0.24
Birthweight (g)	1160 (700; 1470)	1130 (660; 1470)	0.19
Gestational age (weeks)	29.8 (25.5; 35.0)	29.1 (25.0; 34.2)	0.02
Apgar score at 1 minute	5 (1; 9)	4 (1; 8)	0.10
Apgar score at 5 minutes	8 (3; 10)	7 (4; 10)	0.75
pH	7.31 (7.08; 7.50)	7.33 (7.13; 7.52)	<0.01
Base excess (mmol/l)	-5 (-14; -1)	-3 (-12; +3)	< 0.01
PO <sub>2</sub> (torr)	47 (25; 79)	45 (19; 77)	0.26
PCO <sub>2</sub> (torr)	44 (25; 72)	43 (26; 66)	0.58
Blood glucose (mg/dl)	90 (25; 175)	70 (27; 160)	0.05
Rectal temperature (°C)	36.0 (34.0; 37.5)	37.0 (35.7; 38.4)	< 0.01
Infants with RDS n (%)	347 (53.0)	124 (38.4)	< 0.01
Ventilated infants n (%)	438 (66.9)	249 (77.1)	< 0.01

RDS=respiratory distress syndrome.

spectively; thereafter data were collected prospectively. One hundred and twelve infants (83 in the first and 29 in the second period) were excluded from this evaluation either because more than 10% of the required data were missing or because the infants' gestational ages exceeded 33 completed weeks and the problems attendant on very immature babies were unlikely. Conditions incompatible with long term survival, such as severe malformations or inborn errors of metabolism, were not excluded. Infants born outside our perinatal

centre and transported after birth were called outborns. The outcome variable "death in hospital" was defined as death during the first stay in hospital (range 410 days) following preterm birth. Gestational age was estimated from early ultrasound derived fetal morphometric data or from the given date of last menstruation or, if this information was not available or unreliable, from clinical criteria according to Finnström.<sup>15</sup> Infants' Apgar scores were assessed at 1, 5, and 10 minutes after birth.<sup>16</sup> Immediately after admission to the NICU, rectal temperature was measured and arterial or capillary blood was drawn for gas analysis and determination of base excess and glucose. Respiratory distress syndrome (RDS) was diagnosed according to Hjalmarsen and was graded according to Giedion *et al.*<sup>17 18</sup> In none of the infants had surfactant been administered before admission. This study was approved by the local ethical committee.

#### SCORE DEVELOPMENT

To develop and validate a score for mortality risk we divided the 572 VLBW infants included between 1 January 1978 and 31 December 1987 into two cohorts using the random number generator of a personal computer, with a random split of about 70:30. This led to a cohort for score development of 396 infants and a cohort for score validation of 176 infants. Using forward stepwise multiple logistic regression analysis, with death in hospital as the dependent variable, the prognostic variables for the score were selected from the following variables obtained on admission to the NICU: inborn *vs* outborn; delivery room resuscitation by a neonatologist; Apgar score; gender; gestational age; birthweight; rectal temperature; pH; PO<sub>2</sub>; PCO<sub>2</sub>; base excess; blood glucose; artificial ventilation; grade of respiratory distress syndrome. For logistic regression analysis, continuous variables were transformed into five categories each. In order to keep weighting of the variables balanced, artificial ventilation was categorised as 4, the birthweight class above 1499 g as 0 (table 1). The regression coefficients for categories of those variables which were significantly (P < 0.05) associated with death in hospital were converted into integers in a logistic model.<sup>19</sup> A bivariate analysis was calculated to look for collinearity.

#### SCORE VALIDATION

The score was validated in the other cohort of 176 VLBW infants born in the same period using the receiver operating characteristic curve.<sup>20</sup> This curve represents the ability of a test to discriminate between populations (survivors *vs* non-survivors). It plots the tradeoff between sensitivity and false positive rate (or 1 minus specificity) over the full range of cutoff points. A good test would have a high sensitivity (true positive rate) and include few false positives and the resulting area under the curve would approach 1.0.

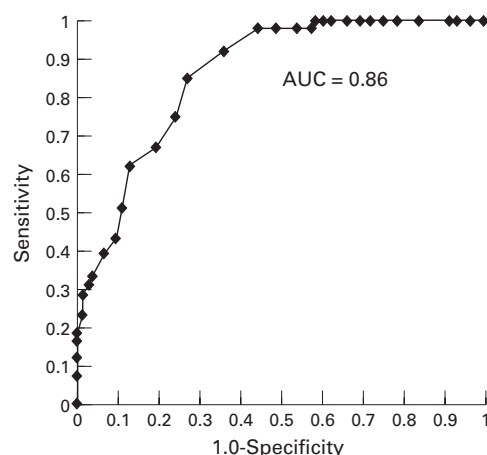


Figure 1 ROC curve for score to predict death in VLBW infants; AUC is area under curve.

#### STATISTICS

To establish the validity of the logistic regression and to check that the observed and predicted values were similar the Hosmer Lemeshow goodness of fit test was performed. This indicates where the score differs significantly from what is observed. A Mantel Haenszel stratified  $\chi^2$  test was used to capture the common odds ratio between eras. Ranks of variables were compared using the Mann-Whitney U test and proportions using  $\chi^2$  with Yates' correction. The stratified Mantel Haenszel test was performed using the BMDP Statistical Software (University of California, USA). All other calculations were done using the software package SPSS-PC+ 3.1 (SPSS, Inc., Chicago, Illinois, USA).

#### Results

Table 2 shows the logistic  $\beta$  slopes, the standard errors, the regression coefficients, and the P values of the five factors found by stepwise logistic regression analysis to be predictive of death during first stay in hospital. The equation to predict mortality is also given in table 2. Although gestational age and mortality were significantly associated in a bivariate analysis, gestational age was not predictive of death when using multiple logistic regression analysis.

Taking into account the regression coefficients obtained for the five predictors in the logistic regression analysis, we established a disease severity score up to a maximum of 40 points (table 1).

Figure 1 reveals the receiver operating characteristic curve resulting from the score's sensitivity and specificity to predict mortality in the validation cohort at cutoff points from 3 to 40. The area under the curve, representing the score's quality, was 0.86. The prevalence of death in hospital after preterm birth was 0.31 in our validation group. At a cutoff of 21 points we observed a sensitivity of 0.85, a specificity of 0.73, a negative predictive value of 92%, and a positive predictive value of 55%. At this cutoff, 76% of the patients were classified correctly. At a cutoff of 18 points, the positive predictive value was 44% and the negative predictive

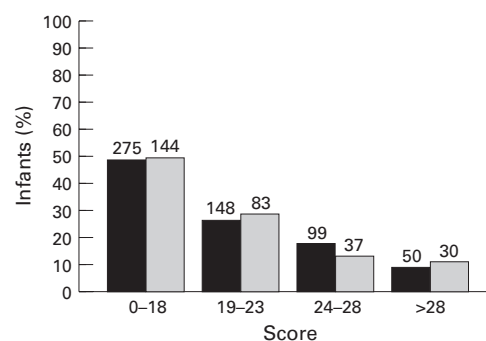


Figure 2 Distribution of LBW infants to the ascending ranges of the score for mortality risks:  $\leq 18$  = low risk; 19-23 moderate risk; 24-28 high risk; and  $> 28$  extreme risk. Absolute numbers of infants admitted are shown above the columns. Solid columns indicate the period 1978-87 ( $n = 572$ ) and dashed columns the period 1988-91 ( $n = 294$ ).

value 98%; at a cutoff of 29 points the positive predictive value was 80% and the negative predictive value 78%.

For score application we defined four mortality risk groups: a score up to 18 points was classified as "mild risk" and a score from 19 to 23 points as "moderate risk." Infants with 24 to 28 points were assessed to be at "high risk," and with more than 28 points to be at "extreme risk" of death in hospital.

Characteristics of VLBW infants studied from 1978 to 1987 and from 1988 to 1991 are shown in table 3. Significant differences between the two periods were found for (1) the percentage of inborns; (2) qualification of the resuscitating physician; (3) gestational age; (4) pH, base deficit, and temperature at admission; (5) incidence of RDS; and (6) frequency of artificial ventilation.

The distribution of infants to the ascending ranges of score in both periods did not differ significantly (fig 2), although there was a significant increase of infants with birthweight below 750 g in the later period ( $P < 0.01$ ). In all four risk groups mortality was lower in the more recent period (fig 3). However, this difference did not reach significance (table 4).

The Hosmer Lemeshow goodness of fit test showed that there were no significant deviations between observed and expected values for the development cohort ( $P = 0.56$ ), the validation cohort ( $P = 0.99$ ), and the most recent cohort of 1988-91 ( $P = 0.71$ ).

#### Discussion

Crude comparisons of hospital or period specific mortality can be highly misleading. Reliable assessment of neonatal outcome is impossible without correcting for major risk factors, particularly birthweight and initial disease severity. Maddock *et al* looked at the outcome of mechanically ventilated newborn infants in two geographically defined regions.<sup>21</sup> They found a clear difference in neonatal mortality after adjusting for birthweight. This could have been due to different ventilatory managements. However, higher mortality in one population may also be explained by lower disease severity in the other or by different numbers of extremely immature infants.

Our aim was to develop and validate a score which permits classification of VLBW infants according to their mortality risk as soon as possible after admission to the NICU. Using multiple regression analysis, five variables, all of which are easy to obtain on admission, were found to be highly predictive for death in hospital: birthweight; Apgar score at 5 minutes; base excess at admission; severity of RDS; and artificial ventilation. Surprisingly, gestational age was not significantly associated with death. We had expected that extremely immature babies would be more likely to die in hospital than more mature infants. However, other authors also found only a weak association between gestational age and death in hospital.<sup>12 14</sup> As expected, the logistic regression analysis showed a collinearity between our predictive variables “grade of RDS” and “artificial ventilation.” However, if we abandoned the variable “artificial ventilation” in our calculation, the prognostic model had an inferior quality: the area under the ROC curve decreased from 0.86 to 0.81.

This scoring system permits the identification of VLBW infants with similar disease severity and similar mortality risk and the comparison of neonatal intensive care among different hospitals or time periods. In our study the cohort from 1978 to 1987 did not differ from that of 1988 to 1991 with respect to the initial mortality risk. The rationale for defining these two periods were major changes in perinatal and neonatal conditions, such as a regionalisation programme for high risk pregnancies and premature births, resuscitation by neonatologists, and surfactant therapy for respiratory distress syndrome. In this context we are aware that interpreting NICU mortality over 14 years is difficult due to increasing knowledge, developing techniques, and changing staff.

In each risk group mortality was lower in the period 1988–91 than in 1978–87. However, this difference did not reach significance. This may be due to too small numbers increasing the risk of type II error.

Like other authors, we did not achieve our goal of predicting death with an accuracy of 100%. The prevalence of death during first stay in hospital was 0.31 in our validation group. The positive predictive value was 55% at a cut-off of 21 points. This increased to 80% with a cutoff of 29 points. Within this score range, prediction was difficult although the infants were at high mortality risk. Below this range

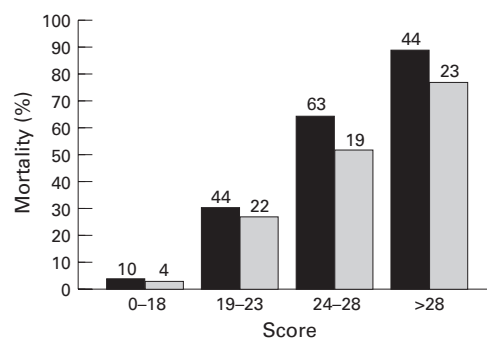


Figure 3 Mortality of VLBW infants in relation to ascending ranges of the score for mortality risk:  $\leq 18$  = low risk; 19–23 moderate risk; 24–28 high risk; and  $> 28$  extreme risk. Absolute numbers of infants who died are shown above the columns. Solid columns indicate the period 1978–87 (n = 572) and dashed columns indicate the period 1988–91 (n = 294).

( $\leq 18$ ) prediction of survival was accurate (negative predictive value 98%); a high score ( $> 28$ ) implied death in hospital in about 80%. The area under the ROC curve, a quality criterion of the test model, was 0.86 for our score compared with 0.90 for the APACHE III and CRIB scores and 0.92 for the PRISM and SNAP-PE scores.<sup>6 8 14 22</sup>

Reports by Horbar *et al*, Maddock *et al*, Patterson and Halliday, Tarnow-Mordi *et al*, the International Neonatal Network, and Richardson *et al*, as well as our own study, strongly suggest that predictive models for outcome in VLBW infants can be developed that are robust and easily obtainable.<sup>10-14 21 22</sup> The advantage of our score system is its very early evaluation of risk mortality, especially if infants have to be assigned to different study groups in the first hours after birth. In contrast to our score obtained on admission to the NICU, Tarnow-Mordi *et al*, the International Neonatal Network, and Richardson *et al* scored after 12 and 24 hours, respectively, during which time considerable therapeutic interventions in the NICU had already taken place.<sup>11 13 14</sup> The therapeutic management in the NICU does not influence our score. However, it must be acknowledged that four components of our score—Apgar, grade of RDS, artificial ventilation, and base excess—are susceptible to therapeutic bias due to prior neonatal intervention, such as policies for resuscitation, elective intubation, surfactant therapy, administration of alkali in the delivery room, etc.

Our score may become an efficient tool for settling some problems affecting controlled clinical trials, especially if infants have to be assigned to different study groups soon after admission to the NICU. It may allow early randomisation based on risk of mortality. It also may help to differentiate treatment effects and variations in mortality risk over time. This score was not intended to limit care of the individual infant. Because of the obvious ethical issues involved, we want to emphasise that our score system is not sufficiently accurate to identify those patients who cannot be saved.

Table 4 Comparison of risk adjusted mortality between 1978-87 and 1988-91

Score	1978-87			1988-91			Odds ratio	95% confidence interval	P value
	Infants admitted (n)	Infants died (n)	Mortality (%)	Infants admitted (n)	Infants died (n)	Mortality (%)			
$\leq 18$	275	10	3.6	144	4	2.8	1.31	0.42-4.10	0.67*
19-23	148	44	29.7	83	22	26.5	1.12	0.72-1.73	0.64*
24-28	99	63	63.6	37	19	51.4	1.26	0.90-1.81	0.19*
$> 28$	50	44	88.0	30	23	76.7	1.15	0.92-1.43	0.23*
Overall	572	161	28.1	294	68	23.1	1.19	0.96-1.45	0.12#

\* $\chi^2$  Mantel-Haenszel test to compare mortality of infants with this score in both periods.  
#stratified Mantel-Haenszel test to compare risk adjusted mortality in both periods.

We thank in particular W Tarnow-Mordi, G Parry, P Stevens, and H Segerer for valuable suggestions, and W Hopfenmüller and A Mohnhaupt for statistical advice.

Supported by BMFT, Public Health 07PHF01, No B3.

- 1 Lee KS, Paneth N, Gartner LM, Pearlman M. The very-low-birth-weight rate: Principal predictor of neonatal mortality in industrialized populations. *J Pediatr* 1980;97:759-64.
- 2 Stewart AL, Reynolds EO, Lipscomb AP. Outcome for infants of very low birthweight: Survey of world literature. *Lancet* 1981;i:1038-40.
- 3 Collaborative European Multicenter Study Group. Surfactant replacement therapy for severe neonatal respiratory distress syndrome: An international randomized clinical trial. *Pediatrics* 1988;82:683-91.
- 4 Merritt TA, Hallman M, Berry C, Pohjavuori M, Edwards DK, Jaaskelainen J, et al. Randomized, placebo-controlled trial of human surfactant given at birth versus rescue administration in very low birthweight infants with lung immaturity. *J Pediatr* 1991;118:581-94.
- 5 Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. *Crit Care Med* 1985;13:818-29.
- 6 Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991;100:1619-36.
- 7 Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: The TRISS method. Trauma score and the injury severity score. *J Trauma* 1987;27:370-8.
- 8 Pollack MM, Ruttimann UR, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med* 1988;16:1110-6.
- 9 Timmons OD, Havens PL, Fackler JC. Predicting death in pediatric patients with acute respiratory failure. *Chest* 1995;108:789-97.
- 10 Patterson CC, Halliday HL. Prediction of outcome shortly after delivery for the very low birthweight ( $\leq 1500$  g) infant. *Paediatr Perinat Epidemiol* 1988;2:221-8.
- 11 Tarnow-Mordi W, Ogston S, Wilkinson AR, Reid E, Gregory J, Saeed M, Wilkie R. Predicting death from initial disease severity in very low birthweight infants: A method for comparing the performance of neonatal units. *BMJ* 1990;300:1611-4.
- 12 Horbar JD, Onstad L, Wright E. Predicting mortality risk for infants weighing 501 to 1500 grams at birth. A National Institutes of Health Neonatal Research Network report. *Crit Care Med* 1993;21:12-8.
- 13 Richardson DK, Gray JE, McCormick MC, Workman K, Goldmann DA. Score for neonatal acute physiology: A physiologic severity index for neonatal intensive care. *Pediatrics* 1993;91:617-23.
- 14 The International Neonatal Network. The CRIB (clinical risk index for babies) score: A tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;342:193-8.
- 15 Finnström O. Studies on maturity in newborn infants. IX. Further observations on the use of external characteristics in estimating gestational age. *Acta Paediatr Scand* 1977;66:601-4.
- 16 Apgar V. A proposal for a new method of evaluation of the newborn infant. *Curr Res Anesth Analg* 1953;32:260-7.
- 17 Hjalmarson O. Epidemiology and classification of acute, neonatal respiratory disorders. A prospective study. *Acta Paediatr Scand* 1981;70:773-83.
- 18 Giedion A, Haefliger H, Dangel P. Acute pulmonary X-ray changes in hyaline membrane disease treated with artificial ventilation and positive end-expiratory pressure (PEP). *Pediatr Radiol* 1973;1:145-52.
- 19 Cole TJ. Scaling and rounding regression coefficients into integers. *Appl Statist* 1993;42:261-8.
- 20 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
- 21 Maddock CR, Carpenter RG, Gardner A. Mechanical ventilation for the newborn. *Lancet* 1988;i:707.
- 22 Richardson DK, Phibbs CS, Gray JE, McCormick MC, Workman DK, Goldmann DA. Birth weight and illness severity: Independent predictors of neonatal mortality. *Pediatrics* 1993;91:969-75.

### Commentary

The authors of the "Berlin score" must be congratulated on developing a very promising tool for stratifying infants in individual hospitals in randomised controlled trials according to risk of mortality. However, it would be unwise to use it for making risk adjusted comparisons of mortality between hospitals or over time. The inclusion of ventilatory support before admission would only be valid in the score if all hospitals had similar policies. In fact, policies for artificial ventilation vary widely. Similarly, clinical assessments of respiratory distress syndrome and Apgar scores may vary with local practice. As these three items account for up to 60% of the maximum risk, their inclusion in the score could lead to significant bias when

adjusting for risk of mortality between hospitals or over time. This would be of little importance in randomised controlled trials if infants were stratified by risk within individual hospitals.

The ideal scoring system for neonatal risk requires items which are prognostically accurate, objective, routinely available, measured at the moment of birth and independent of the quality of prior care. Unfortunately, this holy grail is never likely to be achieved. Measures of risk based on physiological derangement or illness severity can be standardised, thus reducing bias due to interobserver variation. However, it is impossible to disentangle whether measures of illness severity reflect intrinsic disease, the effects of antenatal or early neonatal care, or a combination of all three. Even demographic measures of risk like gestation and birthweight are not inured from this problem as they also reflect prior quality of care. For example, variations in rates of very preterm delivery may reflect differences in obstetric management. Furthermore, variations in use of antenatal steroids<sup>1</sup> may lead to substantial variations in gestation specific severity of illness. Unless the use or neglect of antenatal steroids is accurately accounted for, variations in gestation specific mortality could wrongly be attributed to differences in neonatal care. Scores which include illness severity measured shortly after birth can account for variations in use of antenatal steroids<sup>2</sup> and predict mortality more accurately than demographic variables alone.<sup>3-5</sup> They can therefore adjust more reliably for differences in prior risk, treatment, and illness severity than demographic variables alone, allowing more reliable inferences about differences in subsequent quality of care to be drawn. However, these inferences can only be made about quality of care given *after* the point when risk has been measured.

Discrepancies in mortality after adjusting for risk at different points in the continuum of perinatal care can offer fascinating insights. When adjustment for risk is made using only gestation or birthweight, comparisons of outcome take no account of information after birth. More accurate adjustment for risk can be made using CRIB,<sup>3,5</sup> which takes account of events up to 12 hours after birth. In unpublished data, risk adjusted hospital mortality was compared in very low birthweight or preterm infants admitted to neonatal units in the UK and a developing country. Mortality was twice as high in the developing country after adjusting for risk using birthweight, but similar after adjusting for risk using CRIB. One explanation is that provision of antenatal or immediate neonatal care was worse in the developing country, but that neonatal care from up to 12 hours after birth was equally effective. (Tarnow-Mordi WO, Gould C, Parry GJ, et al, for International Neonatal Network. Risk adjusted league tables in neonatal intensive care. Proceedings of the First International Conference on the Scientific Basis of Health Services, London 1995).

It is interesting to speculate why the authors of the Berlin score found no fall in mortality

after adjusting for risk on admission between 1978–87 and 1988–91, despite presumed advances in neonatal care. There was an improvement in acid base status and rectal temperature and in the rate of respiratory distress syndrome on admission over those periods. This could reflect improved antenatal care, more skilled resuscitation by trained neonatologists, or both, particularly if gestation or birthweight specific mortality fell in the second period. Otherwise the lack of improvement in mortality could imply that the samples were too small to show a difference, that there was no significant improvement in any aspect of management before or after admission, or that the score was not valid in adjusting for changes in risk over time. Table 3 shows that the decrease in respiratory distress syndrome accompanied an increase in artificial ventilation at admission between the two periods. Clearly, a more aggressive approach to starting ventilatory support before admission was adopted in the second period. An intriguing possibility is that the score *underestimated* the true level of risk associated with this aggressive approach. This may have masked a real improvement over time in risk adjusted outcome after admission.

What advances can we expect in the development of new neonatal scoring systems? Prognostically accurate measurements of illness severity as soon as possible after birth remain a priority, as the authors of the Berlin score have emphasised. One exciting new

measure of illness severity in the newborn which merits serious consideration is blood lactate, which seems a better indicator of tissue perfusion and prognosis than either pH or base excess.<sup>6</sup> The more neonatal units which invest in equipment to measure blood lactate routinely, the sooner it will be feasible to test its prognostic value in new neonatal scoring systems.

I thank Martin Ward Platt, Michael Weindling and members of the UK Neonatal Staffing Study Steering Group for helpful suggestions.

WILLIAM TARNOW-MORDI  
International Neonatal Network  
Department of Child Health  
University of Dundee  
Ninewells Hospital and Medical School  
Dundee DD1 9SY

- 1 Scottish Neonatal Consultants' Collaborative Study Group, International Neonatal Network. Trends and variations in use of antenatal corticosteroids to prevent neonatal respiratory distress syndrome: recommendations for national and international comparative audit. *Br J Obstet Gynaecol* 1996;**103**:534–40.
- 2 Tarnow-Mordi W, Cooke RWI, Parry G, Ogston S. The CRIB score. *Lancet* 1993;**342**:612–13.
- 3 International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool for assessing initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;**342**:193–8.
- 4 Gray JE, Richardson DK, McCormick MC, *et al.* Birthweight and illness severity: independent predictors of NICU mortality. *Pediatrics* 1993;**91**:969–75.
- 5 DeCourcy Wheeler RHB, Wolfe CDA, Fitzgerald A, *et al.* Use of the CRIB (clinical risk index for babies) score in prediction of neonatal mortality and morbidity. *Arch Dis Child* 1995;**73**:F32–6.
- 6 Deshpande SA, Ward Platt MP. Association between blood lactate and acid base status and mortality in ventilated infants. *Arch Dis Child* 1997;**76**:F15–F20.