Predicting outcome in very low birthweight infants using an objective measure of illness severity and cranial ultrasound scanning

P W Fowlie, W O Tarnow-Mordi, C R Gould, D Strang

Abstract

Aim—To investigate the feasibility of developing an objective tool for predicting death and severe disability using routinely available data, including an objective measure of illness severity, in very low birthweight babies.

Method—A cohort study of 297 premature babies surviving the first three days of life was made. Predictive variables considered included birthweight, gestation, 3 day cranial ultrasound appearances and 3 day CRIB (clinical risk index for babies) score. Models were developed using regression techniques and positive predictive values (PPV) and likelihood ratios (LR) were calculated.

Results—On univariate analysis, birthweight, gestation, 3 day CRIB score and 3 day cranial ultrasound appearances were each associated with death. On multivariate analysis, 3 day CRIB score and 3 day cranial ultrasound appearances remained independently associated. A 3 day CRIB score > 4 along with intraventricular haemorrhage (IVH) grade 3 or 4 was associated with a PPV of 64% and an LR of 9.8 (95% confidence limits 3.5, 27.9). Only 3 day CRIB score and 3 day cranial ultrasound appearances were associated with severe disability on univariate analysis. Both remained independently associated on multivariate analysis. A 3 day CRIB score > 4 along with an IVH grade of 3 or 4 was associated with a PPV of 60% and an LR of 24.2 (95% CI 4.4, 133.3).

Conclusion—Incorporating objective measures of illness severity may improve current prediction of death and disability in premature infants.

(*Arch Dis Child Fetal Neonatal Ed* 1998;**78**:F175–F178)

Keywords: severity of illness index; prematurity; CRIB score; positive predictive values; likelihood ratios

Department of Child Health, Ninewells Hospital and Medical School, Dundee, DD1 9SY
P W Fowlie
W O Tarnow-Modi
C R Gould
D Strang

Correspondence to:
Dr Peter Fowlie.

Neonatologists frequently estimate the likely prognosis for very premature infants. However, many neonatal deaths occur within the first two or three days of life usually because the baby succumbs despite continuing intensive care or because continuing treatment is electively withdrawn.[1] As a result, these babies do not tend to present the same prognostic dilemmas as those who survive the immediate newborn period. For those who do survive the first three days, two questions are crucial: what are the chances that this baby will still die; and if this baby survives, what are the chances of severe disability? These questions usually arise when discussing prognosis with parents[2] or when considering withdrawal of intensive treatment.[3–5]

Currently, clinicians amalgamate prognostic data from a variety of sources when predicting the outcome for any given baby.[6–9] Simple objective risk factors such as birthweight and gestational age at birth are often used[10] as well as objective evidence of brain damage obtained from the appearances on cranial ultrasound scans.[11] Over the past decade several models and scoring systems have also been described.[12–16] However, recent evidence suggests that birthweight may not be an accurate guide to prognosis beyond the first few days of life[1] and most of the currently available scoring systems are too complex for use in routine clinical practice. Illness severity can also be an independent predictor of mortality[17] and although neonatologists routinely use their clinical experience to assess subjectively just how sick any given baby might be, little work has been done to examine how objective measures of illness severity such as SNAP (Score for Neonatal Acute Physiology) or CRIB[18 19] might be used in this context. We therefore investigated the possibility of developing a simple predictive model incorporating a measure of illness severity that might provide clinicians with an objective estimate of :(a) the risk of death before discharge; or (b) the risk of severe disability at 18 months in premature babies surviving the first three days of neonatal intensive care. We deliberately explored only the use of data that are *routinely* available and *simple*: birthweight; gestational age at birth; CRIB score on day 3 of life, a simple valid and reliable measure of clinical risk and illness severity[18 20]; and appearances on cranial ultrasound scans on or around day 3 of life.

Method

This was a retrospective cohort study in infants weighing ≤ 1500 g at birth or born at < 31 weeks of gestation between 1988 and 1990 in six Scottish hospitals. Infants born by caesarean section were included but not those born outside the target hospitals and subsequently transferred into one of them. Three of the hospitals were regarded as tertiary referral centres. Data were abstracted from the medical records of each infant: birthweight; gestation; presence of non-lethal congenital malformation; the lowest appropriate concentration of inspired oxygen between 48 and 72 hours; the highest

Table 1   Characteristics of study infants (those infants excluded because of missing data are compared with those with complete data)

	Complete data	n=	Missing data	n=	p value
Male:female	1:1	297	1:1.1	85	0.74*
Median gestation (range)	29 (24–37)	297	31 (24–38)	85	<0.001‡
Median birth weight (range)	1170 (560–1500)	297	1340 (770–1500)	85	<0.001‡
Median 5 minute Apgar score (range)	9 (0–10)	297	9 (3–10)	85	<0.001‡
Median day 3 CRIB score, CRIB72 (range)	2 (0–17)	297	1 (1–5)	64	<0.001‡
Number of infants with grade III or IV IVH on day 3	27 (9%)	297	1 (7%)	14	0.54**
Died	43 (14%)	297	6 (7%)	85	0.10**
Severe disability at 2 years in survivors†	14 (5%)	283	1 (1%)	77	0.21**

*$\star\chi^2$ test, ‡Mann–Whitney U test, **Fisher's exact test.
†See text for definition.

appropriate concentration of inspired oxygen between 48 and 72 hours; and the worst base deficit between 48 and 72 hours. A CRIB score relating to day 3 of life, CRIB72, was then calculated according to the original scoring system.[18] Where they were available, data reporting the cranial ultrasound appearances on or around day 3 of life were also recorded: a variety of ultrasound machines were used across the study centres although all images were reported according to the Papile classification.[21] Data on disability were collected for each surviving infant at a corrected age of 18 months by mailing a questionnaire to the child's health visitor.[22] For the purposes of this study, we have defined a child with severe disability as one who cannot sit unsupported by 18 months and/or is blind, and/or is deaf, requiring bilateral hearing aids, and/or is more than 12 months behind in any other field of development.[23] Specific aspects of fine motor control, communication, or respiratory disease were not considered because of difficulties in case definition. Disability was therefore a dichotomous outcome: a child was either severely disabled at 18 months according to our definition or was not.

Predictive variables considered were birthweight, gestation, appearances on cranial ultrasound scans and CRIB score on day 3 of life. Birthweight, gestation, and CRIB score were regarded as continuous variables. Appearances on cranial ultrasound scan were treated either as ordinal data (grades 0–4) or, for simplicity, converted into a dichotomous outcome (grade 0–2 *vs* grades 3 or 4).

In those infants surviving until the end of day 3, predictive models for (i) subsequent death before hospital discharge and (ii) severe disability at 18 months in survivors were fitted to the data using regression techniques.[24] Each model was assessed by creating receiver operating characteristic (ROC) curves and calculating the area below the curve, Az,[25] as well as by the Hosmer-Lemeshow test of goodness of fit.[26] Models were compared by comparing the areas under the respective ROC curves.[27]

For use in clinical practice, models were transformed into 8×2 tables and positive predictive values and likelihood ratios were calculated along with their 95% confidence intervals.[28] The positive predictive value here represents the number of infants who have a positive "test" and actually experience the outcome of interest (true positives) divided by the total number of infants with a positive "test." The likelihood ratio is defined as the ratio of true positive "tests" to false positive "tests."

Results

Four hundred and twenty eight infants were born in the six units over the study period. Forty six infants died in the first 72 hours of life, leaving 382 babies to contribute data to the analyses. Eighty five of these infants had either missing CRIB data or missing ultrasound data, leaving 297 infants contributing to the models predicting death. Of 254 infants surviving to age 2 with CRIB scores and ultrasound data, 14 were not assessed at 18 months, leaving 240 infants for analysis of models predicting severe disability. The characteristics of babies with and without missing data are shown in table 1.

Forty three babies died after 3 days of life: median age at death day 13, range day 4 to day 308. The common pathologies leading to death included respiratory distress syndrome (n=22), intraventricular haemorrhage (n=15), and pulmonary interstitial emphysema (n=5). Of the 14 severely disabled children included in the analysis, five were blind, four were deaf, seven could not sit unsupported at 18 months and nine had clinically significant delay in another area of development. Most cranial ultrasound scans were performed on day 3 of life: mode day 3 of life, median day 3 of life (quartiles day 2 of life, day 5 of life). Cases were excluded because the notes could not be traced (n=12), the physiological data required to calculate the CRIB score were missing (n=21), no cranial ultrasound scan had been performed (n=77) or follow up data were not available (n=22).

MODEL PREDICTING DEATH BEFORE DISCHARGE FROM THE NEONATAL UNIT
On univariate analysis, the risk of death was significantly associated with birthweight, gestation, CRIB72 and ultrasound appearances; p < 0.05. On multiple logistic regression, CRIB72 (p < 0.0001) and ultrasound appearances (p = 0.0035) remained independently significant

Table 2   Accuracy and validity of models predicting death before discharge

Predictive model	Area under the ROC curve Az, (SE)	Hosmer-Lemeshow χ^2 (p value)
Log odds death $-2.0044 + (0.0034 \times \text{birth weight})$	0.70 (0.05)	9.32 (0.50)
Log odds death $-10.7097 + (0.4423 \times \text{gestation})$	0.74 (0.08)	9.27 (0.51)
Log odds death $4.0334 - (1.6584 \times \text{USS}\star) - (0.4004 \times \text{CRIB72})$	0.89 (0.05)	5.57 (0.85)

*The ultrasound appearances have been treated as dichotomous (IVH grade 0, 1 or 2 *vs* grade 3 or 4) in the model reported for simplicity. The relation holds when the variable is treated as ordinal (IVH grade 0–4).

Table 3   Positive predictive values (PPV) and likelihood ratios (LR) associated with model predicting death before discharge

"Test"	Died	Survived	PPV	LR (95% CI)
CRIB72 > 4 plus IVH grade 3 or 4	9	5	64%	9.8 (3.5–27.9)
CRIB72 = 2–4 plus IVH grade 3 or 4	3	3	50%	5.5 (1.1–26.2)
CRIB72 = 1 plus IVH grade 3 or 4	1	6	14%	0.91 (0.11–7.4)
CRIB72 = 0 plus IVH grade 3 or 4	0	0		
CRIB72 > 4 plus IVH grade 0, 1 or 2	24	44	35%	3.0 (2.0–4.3)
CRIB72 = 2–4 plus IVH grade 0, 1 or 2	5	74	6%	0.37 (0.16–0.86)
CRIB72 = 1 plus IVH grade 0, 1 or 2	2	74	3%	0.15 (0.04–0.58)
CRIB72 = 0 plus IVH grade 0, 1 or 2	0	34	0%	0.08 (0.01–1.3)

Table 4   Positive predictive values (PPV) and likelihood ratios (LR) associated with model predicting severe disability at 2 years

"Test"	Severe disability	No severe disability	PPV	LR (95% CI)
CRIB72 > 4 plus IVH grade 3 or 4	3	2	60%	24.2 (4.4–133.3)
CRIB72 = 2–4 plus IVH grade 3 or 4	0	3	0%	2.2 (0.12–41.2)
CRIB72 = 1 plus IVH grade 3 or 4	0	6	0%	1.2 (0.07–20.3)
CRIB72 = 0 plus IVH grade 3 or 4	0	0		
CRIB72 > 4 plus IVH grade 0, 1 or 2	5	39	11%	2.1 (1.0–4.4)
CRIB72 = 2 or 3 plus IVH grade 0, 1 or 2	3	71	4%	0.70 (0.25–1.90)
CRIB72 = 1 plus IVH grade 0, 1 or 2	1	73	1%	0.20 (0.03–1.50)
CRIB72 = 0 plus IVH grade 0, 1 or 2	2	32	6%	1.00 (0.27–3.8)

while birthweight (p = 0.9345) and gestation (p = 0.1778) were no longer significantly associated with the risk of death. The univariate models using birthweight and gestational age at birth and the multivariate model using CRIB 72 and ultrasound appearances are described in table 2. All three models achieved satisfactory goodness of fit using the Hosmer-Lemeshow χ^2 statistic.[26] In addition, the area under the curve associated with the model combining CRIB72 and ultrasound appearance was significantly greater than the area associated with both the model based on birth weight alone (p = 0.0011) and the model based on gestation alone (p = 0.0436).

Table 3 shows the positive predictive values and likelihood ratios associated with the model based on CRIB72 and ultrasound appearances.

MODEL PREDICTING SEVERE DISABILITY AT 18 MONTHS CORRECTED AGE

On univariate analysis, CRIB72 (p = 0.0051) and ultrasound appearance (p = 0.0193) were both significantly associated with the risk of being severely disabled at 18 months. Neither birthweight (p = 0.0755) nor gestation (p = 0.0719) were associated with this outcome. On multiple regression analysis, both CRIB72 (p = 0.0293) and ultrasound appearance (p = 0.0405) remained independently associated with the risk of severe disability: Log odds severe disability = 3.5871 −(0.1882×CRIB72)−(1.5241×USS); Az(SE) = 0.675411 (0.133221); Hosmer-Lemeshow χ^2 = 5.089515; p = 0.89. The corresponding predictive values and likelihood ratios for this model are shown in table 4.

Discussion

In this cohort 48% of the babies who died before discharge from the neonatal unit did so within the first 72 hours of birth. This is a smaller proportion than was found in the study by Meadow *et al*[1] probably because our population was heavier and more mature, but still justifies our decision to concentrate on developing a model(s) predicting the outcome for very low birthweight or preterm babies *only if they have survived the first three days of life*. We also concentrated on developing models that might help provide answers to highly relevant clinical questions: (1) what are the chances that this baby will survive to be discharged home; and (2) if this baby does survive to go home, what are the chances that s/he will be severely disabled? This approach to predicting outcome has not to our knowledge been reported elsewhere, although others have tried to predict outcome before birth,[12] at or shortly after birth,[14–16] and at discharge home.[13]

We believe we have gone some way towards demonstrating the feasibility of developing simple clinically useful models. We specifically tried to identify those infants we considered to have severe *disability* as it is a concept that clinicians, parents, and lay people readily understand, and it seems to be clinically and practically most relevant when considering prognosis.[29–32] Cranial ultrasonography is a valid and reliable tool for assessing brain damage[33 34] and is an accepted routine investigation in neonatal practice; and the CRIB score is a simple, robust instrument for measuring clinical risk and illness severity which is calculated easily using only six variables, all of which are usually routinely available.[35–37] However, certain methodological issues prevent the immediate application of our models. Although the infants with missing data were similar in their characteristics and outcomes, we cannot be sure what effect their exclusion had on the overall results. Neonatal practice has also changed dramatically over the past 10 years with the introduction of antenatal steroids and surfactant treatment, and although CRIB remains a valid score (G Parry, personal communication), it is not clear what effect this might have on the models presented here. It is also recognised that treatment may sometimes be electively withdrawn and that this practice can be influenced by cranial ultrasound appearances.[38] Without being able to adjust for this practice, therefore, our results might over-estimate the predictive capacity of the models presented. A prospective study that takes these issues into account and validates the models described[39] is therefore needed.

As there is good evidence that neonatal survival has improved,[40] the positive (and negative) predictive values from this study will have changed and there would be advantages to considering the likelihood ratio rather than the predictive value when trying to predict outcome. Assuming the validity of the models is maintained and the current death rate and prevalence of disability for any given population are known, the risk of either outcome can be extrapolated using the likelihood ratio along with a simple nomogram for Bayes's theorem.[28 41] It is worth noting, however, that likelihood ratios between 0.01 and 10 add relatively little to what is known before the results of any "test" are available.[28] Only the likelihood ratio associated with a CRIB score > 4 plus IVH grade 3 or 4, when predicting either death or severe disability, is big enough (9.8 and 24.2

respectively) to alter significantly the post test probability.

Our results suggest that assessment of illness severity permits a more accurate estimate of prognosis than use of birthweight, gestation, or cranial ultrasound scans alone. However, no single piece of evidence can provide a definitive estimate of prognosis for individual babies.[42] Any score is based on population characteristics that may not pertain to the individual baby, relates to outcome data that are by definition already "out of date," and, when derived from relatively small cohorts, will give estimates which lie within wide confidence limits. Clinicians' clinical judgement is a powerful tool in its own right[43] and scoring systems can only provide an additional insight into prognosis. They should never be used in isolation to make clinical decisions, particularly about withdrawal of active treatment, and we would encourage clinicians to continue to use an holistic approach.

1 Meadow W, Reimshisel T, Lantos J. Birth weight specific mortality for extremely low birth weight infants vanishes by four days of life: epidemiology and ethics in the neonatal intensive care unit. *Pediatrics* 1996;**97**:636-43.
2 Perlman NB, Freedman JL, Abramovitch R, Whyte H, Kirplani H, Perlman M. Informational needs of parents of sick infants. *Pediatrics* 1991;**88**:512-18.
3 Doyle L, Wilsher D. Towards guidelines for withholding and withdrawal of life prolonging treatment in neonatal medicine. *Arch Dis Child* 1994;**70**:F66-F70.
4 Lantos JD, Tyson JE, Allen A, *et al*. Withholding and withdrawing life sustaining treatment in neonatal intensive care: issues for the 1990s. *Arch Dis Child* 1994;**71**:F218-23.
5 McHaffie HE, Fowlie PW. *Life, death and decisions: doctors and nurses reflect on neonatal practice*. Hale: Hochland and Hochland, 1996:94-107.
6 Poses MR, Bekes C, Winkler RL, *et al*. Are two (inexperienced) heads better than one (experienced) head? *Arch Intern Med* 1990;**150**:1874-8.
7 Stevens SM, Richardson DK, Gray JE, *et al*. Estimating neonatal mortality risk: an analysis of clinicians' judgements. *Pediatrics* 1994;**93**:945-50.
8 Candee D, Sheehan J, Cook CD, *et al*. Moral reasoning and decisions in dilemmas of neonatal care. *Pediatr Res* 1982;**16**:846-50.
9 Cullen DJ, Chernow B. Predicting outcome in critically ill patients. *Crit Care Med* 1994;**22**:1345.
10 Horbar JD, Onstad L, Wright E; The National Child Health and Human Development Neonatal Research Network. 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**:12-18.
11 Ng PC, Dear PRF. The predictive value of a normal ultrasound scan in the preterm baby - a meta-analysis. *Acta Paediatr Scand* 1990;**72**:286-91.
12 Hobel CJ, Hyvarinen MA, Okada DM, Oh W. Prenatal and intrapartum high-risk neonatal screening, I. Prediction of the high risk-neonate. *Am J Obstet Gynecol* 1973;**117**:1-9.
13 Eckerman CO, Oehler JM, *et al*. Nursery Neurobiologic Risk Score: important factors in predicting outcome in very low birth weight infants. *J Pediatr 1991;118*:783-92.
14 Patterson CC, Halliday HL. Predictions of outcome shortly after delivery for the very low birth weight (less than or equal to 1500g) infant. *Paediatr Perinat Epidemiol* 1988;**2**:221-8.
15 Salamy A, Davis S, Eldredge L, *et al*. Neonatal status: an objective scoring method for identifying infants at risk for poor outcome. *Early Human Dev* 1988;**17**:233-43.
16 Scheiner AP, Sexton ME. Prediction of developmental outcome using a Perinatal Risk Inventory. *Pediatrics* 1991;**88**:1135-43.
17 Richardson DK, Phibbs CS, Gray JE, McCormick MC, Workman-Daniels K, Goldmann DA. Birth weight and ill-

ness severity: independent predictors of neonatal mortality. *Pediatrics* 1993;**91**:969-75.
18 The International Neonatal Network. The CRIB (clinical risk index for babies) score: a tool for assessing the initial neonatal risk and comparing performance of neonatal intensive care units. *Lancet* 1993;**342**:193-8.
19 Richardson DK, Gray JE, McCormick MC, *et al*. Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care. *Pediatrics* 1993;**91**:617-23.
20 Fowlie PW, Tarnow-Mordi, WO, Gould CR, Strang D. Measurement properties of CRIB (Clinical Risk Index for Babies) - reliability, validity beyond the first 12 hours and responsiveness. *Crit Care Med* 1997;**26**:163-80.
21 Papile LA, Burstein J, Burstein R, *et al*. Incidence and evolution of subependymal and intraventricular haemorrhage. A study of infants with birth weight less than 1500g. *J Pediatr* 1978;**92**:529-34.
22 Scottish Neonatal Consultants' Collaborative Group and the International Neonatal Network. CRIB (clinical risk index for babies), mortality, and impairment after neonatal intensive care. *Lancet* 1995;**345**:1020-2.
23 National Perinatal Epidemiology Unit and Oxford Regional Health Authority. *Disability and perinatal care: measurement of health status at two years*. Oxford: National Perinatal Epidemiology Unit, 1994: Appendix 2.
24 Hosmer DW, Lemeshow S. In: *Applied Logistic Regression*. New York: John Wiley & Sons, 1989:106-118.
25 Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29.
26 Hosmer DW, Lemeshow S. In: *Applied Logistic Regression*. New York: John Wiley & Sons, 1989:140-4.
27 Hanley JA, McNeil BJ. A method of comparing the areas under the receiver operating characteristic curves derived from the same cases. *Radiology* 1983;**148**:839-43.
28 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. In: *Clinical Epidemiology: A Basic Science for Clinical Medicine*. 2nd edn. Boston: Little, Brown & Co, 1991:69-152.
29 Tarnow-Mordi WO, Mutch L, Parry G, Cockburn F, McIntosh N, for the Scottish Neonatal Consultants Collaborative Study Group and the International Neonatal Network. CRIB and impairment after neonatal intensive care. *Lancet* 1995;**346**:58-9.
30 World Health Organisation. *International Classification of Impairments, Disabilities and Handicaps*. Geneva: WHO, 1980.
31 Mutch LM, Johnson MA, Morley R. Follow up studies: design, organisation and analysis. *Arch Dis Child* 1989;**64**:1394-402.
32 Johnson A. Disability and perinatal care. *Pediatrics* 1995;**95**:272-4.
33 Hope PL, Gould SJ, Howard S, *et al*. Precision of ultrasound diagnosis of pathologically verified lesions in the brain of very preterm infants. *Dev Med Child Neurol* 1988; **30**:457-71.
34 Pinto-Martin J, Paneth N, Witomski, *et al*. The central New Jersey neonatal brain haemorrhage study: design of the study and reliability of ultrasound diagnosis. *Pediatr Perinat Epidemiol* 1992;**6**:273-84.
35 De Courcy-Wheeler RHB, Wolfe CDA, Fitzgerald A, Spencer M, Goodman JDS, Gamsu HR. Use of CRIB (clinical risk index for babies) score in prediction of neonatal mortality and morbidity. *Arch Dis Child* 1995;**73**:F32-6.
36 Fowlie PW, Gould CR, Parry GJ, Phillips G, Tarnow-Mordi WO. CRIB (clinical risk index for babies) in relation to nosocomial bacteraemia in very low birthweight or preterm infants. *Arch Dis Child* 1996;**75**:F49-F52.
37 Rautonen J, Makela A, Boyd H, Apajasalo M, Pohjavuori M. CRIB and SNAP: assessing the risk of death for preterm neonates. *Lancet* 1994;**343**:1272-3.
38 McHaffie HE, Fowlie PW. Life, death and decisions: doctors and nurses reflect on neonatal practice. Hale: Hochland and Hochland, 1996:58-66.
39 Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules - applications and methodological standards. *N Engl J Med* 1985;**313**:793-9.
40 Escobar GJ, Littenberg B, Petitti DB. Outcome among surviving very low birth weight infants: a meta-analysis. *Arch Dis Child* 1991;**66**:204-11.
41 Fagan TJ. Nomogram for Bayes's Theorem. *N Engl J Med* 1975;**293**:257.
42 Lemeshow S, Klar J, Teres D. Outcome prediction for individual intensive care patients: useful, misused, or abused? *Int Care Med* 1995;**21**:770-6.
43 Knaus WA, Rauss A, Alperovitch A, *et al*. Do objective estimates of chances for survival influence decisions to withhold or withdraw treatment? *Med Decis Making* 1990;**10**:163-71.