CRIB, son of Apgar, brother to APACHE

Neonatologists spend most of their time dealing with a captive population of preterm infants, whose prospects for survival are very closely related to gestation and birthweight. Survival rates at different gestations or in different birthweight categories may seem therefore to be good ‘performance indicators’, and are often used to compare the results from different units, regions, or countries.

There are two major problems with such comparisons. Firstly, mortality cannot be regarded as the only adverse outcome of neonatal intensive care. Major disability in survivors is not uncommon, and mortality rates and major morbidity rates may not be positively correlated, indeed they may have a reciprocal relation. Secondly, while birthweight and gestational age are undoubtedly major prognostic indicators, there are other factors like the severity of respiratory disease and the presence of congenital abnormalities which have a significant influence on outcome. These factors may vary with time in a single unit, confounding the study of trends. They are also affected by social and geographical variations which limit the validity of regional and international comparisons. The complex effects on mortality of referral patterns between district general hospitals and referral centres may also make more parochial comparisons inappropriate and such comparisons have certainly led to vigorous controversy in the past.\(^1\)\(^2\) Fresh life has been breathed into these old arguments in the new United Kingdom health market. Mortality rates, even if corrected for birthweight, are crude and inappropriate performance indicators.

**Scoring systems in intensive care units**

One of the earliest neonatal scoring systems was the Apgar score, designed in the 1950s to quantify the need for resuscitation of the newborn infant and widely used since then for a variety of other, often inappropriate, purposes. Several different indices of neonatal disease severity have subsequently been used for research, many of which have been based on respiratory status and ventilatory requirements. More sophisticated scoring systems for patients in intensive care have been developed over the past decade or so and are now widely used. These include the various incarnations of the APACHE (acute physiology and chronic health evaluation) score in adult intensive care, the PRISM (pediatric risk of mortality) score in paediatric intensive care, and the SNAP (score for neonatal acute physiology) in the neonatal unit. Details of these scoring systems, their history, and relative performance as predictors of outcome are discussed in detail in a review by Richardson and Tarnow-Mordi.\(^3\)

Severity of illness may be based on diagnostic categories, risk factors, therapeutic requirements, physiological measurements or combinations of all of these. The information derived by stratifying patients in this way can be used to ensure adequate randomisation in clinical trials, to compare quality of outcome between different centres, or to ensure fair allocation of resources between centres with different patient populations.\(^3\) These scores may also be used as a prognostic index to determine management of individual patients. It is this latter use which is a major concern, and in this respect it is at least slightly comforting that the International Neonatal Network has chosen for their scoring system\(^4\) the pastoral acronym CRIB (clinical risk index for babies) rather than an equivalent of the macho APACHE so fondly wielded by anaesthetists.

**Derivation of the CRIB score**

Tarnow-Mordi\(^5\) and the four and a half column inches of collaborators that make up the International Neonatal Network reported the details of development of the CRIB score in the *Lancet*.\(^4\) The score was designed by retrospective analysis of the records of 812 very low birthweight (<1500 g) or very preterm (<31 weeks’ gestation) infants without lethal congenital abnormality, admitted to four United Kingdom tertiary referral centres. Using death as the dependent variable, logistic regression was used to determine the prognostic value of 40 perinatal variables. Because initial disease severity was so likely to have an important influence on outcome, physiological measurements in the first 12 hours of life were included as prognostic variables, although it was realised that these are, to some extent, outcome variables rather than independent predictors.

The CRIB score was derived from the six most predictive variables which were stratified to give a simply derived total score out of 23. Birthweight (divided into four ranges), gestation (above 24 weeks’ or not), and the presence or absence of congenital abnormalities (life threatening or otherwise) were the three intrinsic variables. The maximum and minimum ‘appropriate’ FiO\(_2\) and the maximum base deficit in the first 12 hours were the most predictive early indices of disease severity, where ‘appropriate’ FiO\(_2\) was that associated with a satisfactory Po\(_2\) or oxygen saturation.
Validating and using CRIB

Using receiver operator characteristic curves (ROC), the CRIB score was a considerably better predictor of mortality than a coin tossing model (encouraging), but also more predictive than birthweight alone. The score was then applied prospectively to a different population of very low birthweight neonates for use in teaching hospital neonatal units and it performed equally well in predicting death on this 'validation cohort'. Results compared favourably with other clinical scoring systems in adult, paediatric, and neonatal intensive care.3 The data from the original ‘development’ cohort were then added to those from the ‘validation’ cohort and further data from one other tertiary referral centre and four district general hospitals. Together with their description of the derivation of the CRIB score, the group were therefore able to publish a comparison of the outcomes for these infants cared for in nine tertiary referral centres and four district general hospitals. Raw mortality data showed little difference between tertiary referral centres and district general hospitals, but the tertiary referral centres tended to have lower mortality rates when these were adjusted for birthweight. When the mortality rates were adjusted for risk of adverse outcome using CRIB scoring, all the tertiary referral centres had lower mortality rates than the district general hospitals. The odds ratio for hospital death in the district general hospitals compared with that at the tertiary referral centres was 1.45 (95% confidence interval 1.01–2.11) adjusted for birthweight and 2.12 (95% CI 1.39–3.24) adjusting by the CRIB score. These differences were not explained by postnatal transfer practices because the odds ratios increased when only inborn infants were considered.

Differences in populations or practices?

It appeared, therefore, that the case mix in the tertiary referral centres comprised smaller and sicker infants than the district general hospitals, and that these factors were responsible for the superficially similar performance of the two types of unit. It is not surprising that regional units should have smaller babies, but it is perhaps surprising that the condition of the babies in the first 12 hours of life (which contributes so heavily to the CRIB adjustments) must have been, on the whole, worse. It is unlikely that obstetric management or initial resuscitation were more effective in district general hospitals, where staffing levels are less adequate and paediatric senior house officers may be distracted by general paediatric duties. It is feasible that blood gas analysis is carried out earlier or more frequently in tertiary referral centres, in which case the highest base deficit will tend to be higher and a similar population will appear to be in worse condition, but this potential bias is unlikely to have a large effect. There is a slight residual doubt, however, that a score which is derived from a tertiary referral centre population and validated on a tertiary referral centre population, is not necessarily applicable to a district general hospital population if the relation between CRIB score and true severity of illness is different in that setting. Of course, attitudes to withdrawal of intensive care may also have major effects on the relation between CRIB score and mortality in such a system. The advantage of a scoring system that is partly dependent on early management, as the authors admit, is that it may fail to distinguish the baby with severe disease, who is ideally managed, from the infant who is less sick but also less optimally treated.5 6 However, unless the tertiary referral centres provided uniformly less satisfactory care in the first 12 hours of life this effect could not explain the apparently greater disease severity in their very low birthweight babies.

Implications for changing patterns of care

Having looked for the methodological issues which may confound the fine detail, an unsurprising conclusion remains. Neonatal units in tertiary referral centres, staffed with specialist neonatologists, with more medical staff, selected nursing staff and usually more expensive equipment, have corrected survival rates greater than district general hospitals. So does this evidence suggest that we should be aiming at an increasingly centralised neonatal service in this country, with regional units expanding and district general hospitals (increasingly threatened by the legislation on junior doctors' hours) ceasing to provide neonatal intensive care? Such a move would be ironic at a time when public and political pressure is tending to devolve rather than centralise childbirth. Taking the CRIB data at face value, a centralist policy may prevent some babies dying in the neonatal period. The cost would be a significant increase in parental inconvenience and decrease in parent-infant contact for a very large number of families, problems which in themselves may be unacceptable or lead to substantial later morbidity.

Following the publication of the CRIB data one correspondent remarked on the very large differences in CRIB adjusted mortality between the individual tertiary referral centres.7 Small numbers and large confidence intervals notwithstanding, a very different political message would have emerged if the CRIB study had highlighted the two-fold difference in adjusted mortality between different tertiary referral centres. If the response to these data is to recommend delivery of all very low birthweight infants in tertiary referral centres it would be equally 'logical' and of course equally unacceptable to deliver all infants in those tertiary referral centres with the lowest CRIB adjusted mortality. Indeed, the reductio ad absurdum would be to consider transferring all high risk mothers to Columbia University in New York, where 28 day survival without chronic lung disease seems more likely than elsewhere.8 9

The compromise, of course, is to provide good quality neonatal intensive care near to home for most low birthweight infants in their local district general hospitals, and to use tertiary referral centres to their full potential by ensuring that the smallest and sickest infants are transferred there (preferably in utero). It's a difficult balance that most regions, or former regions, were working towards. Whether the philosophy will withstand the onslaught of market forces remains to be seen.

Uses and abuses of scoring systems

If some form of scoring system helps to compare different systems of care on a more rational basis than before, then that is obviously a good thing. There are obviously advantages to a simple system like the CRIB score which is quick and dirty, or 'robust' as the epidemiologists say. One could argue whether it should be further refined or made more robust, but it makes sense when planning services to seek to improve the comparability of survival and morbidity statistics. The CRIB score may be a useful tool when applied to populations, but it is a blunt instrument selected in a way which may not be extended to individuals. Of course gestation, severity of respiratory disease, and the presence of congenital abnormalities will be considered in management decisions about very low birthweight infants. However, the assignment of a number to an individual is a dangerous practice, especially if men in grey suits begin to influence clinical decisions. As the paper from the International Neonatal Network points out: 'if in the next few years advances in management improve survival of severely ill infants, any previously accurate cut-off value of
CRIB may overpredict death. We must be wary that CRIB does not become a tool for clinical risk assessment by purchasers.

PETER HOPE

Department of Paediatrics,
John Radcliffe Hospital,
Headington,
Oxford OX3 9DU

5 Grant JM. The CRIB score. Lancet 1993; 342: 612.
6 Hughes-Davies TH. The CRIB Score. Lancet 1993; 342: 938.