Variations in mortality rates among Canadian neonatal intensive care units: interpretation and implications

Jon Tyson, Kathleen Kennedy

ß See related articles pages 173, 193

Differences among hospitals in terms of patient outcomes have attracted intense public and professional attention as indicators of differences in the quality of care provided in those hospitals. However, daunting methodological problems make it difficult to isolate the effects of differences in care from those of other factors affecting outcome. These problems include the need to study large numbers of patients; the need to obtain detailed, accurate and meaningful information; and the need to identify, quantify and adjust appropriately for differences in patient characteristics and severity of illness.[1,2]

In this issue (page 173), Sankaran and colleagues[3] of the Canadian Neonatal Network report a study of exemplary quality that assesses variation in mortality rates among Canadian neonatal intensive care units (NICUs). The study encompassed data for 19 265 infants admitted to 17 NICUs accounting for 75% of the tertiary-level neonatal beds in Canada. Extensive data were collected at the bedside by research personnel using standardized definitions and a customized data entry program with built-in error checking. All patients, including those transferred to other hospitals, were tracked until death or discharge home. The mortality rate for each NICU was adjusted for traditional risk factors (e.g., gestational age, 5-minute Apgar score, congenital anomalies and outborn status [born at another hospital and requiring transfer to the NICU]) and for a relatively new measure of illness severity that is highly correlated with neonatal morbidity and mortality (the Score for Neonatal Acute Physiology, version II, known as SNAP-II).[4] A limitation of this study is that data were collected only for infants admitted to NICUs, so the analysis excludes infants too small or sick to survive transport or too healthy to justify referral to a NICU. The performance of regional NICUs can be gauged and compared on the basis of mortality rates achieved for all infants in particular risk categories within the NICU's referral region. These data do not allow such a comparison.

The unadjusted (crude) mortality rates varied 10-fold among the NICUs (range about 1% to 11%), and for 12 of the 17 NICUs, the crude mortality rate was significantly different from the mean mortality rate. Adjusting the mortality rates for patient risk substantially reduced both the variation among NICUs (range about 2% to 6%) and the number of NICUs that differed significantly from the mean (only 5). For each hospital the 95% confidence interval (indicating the precision with which risk-adjusted mortality could be estimated) overlapped that of multiple other hospitals.

These findings, like those of other studies in neonates and older patients,[5] indicate the importance of careful risk adjustment. Yet even with extensive and accurate data collection and sophisticated risk adjustment, the assessments of mortality rate are likely to be too imprecise and the differences among individual hospitals too small to justify rank ordering these hospitals according to their outcomes. Sankaran and colleagues used state-of-the-art methods, but the differences among hospitals in risk-adjusted mortality rate might be further attenuated if there were better methods to measure and adjust for risk. For example, gestational age clearly has a substantial effect on mortality rate among the infants at highest risk. However, the gestational age of these infants is often uncertain: pediatric estimates of gestational age tend to be greater and often differ substantially from obstetric estimates,[6] and there may be important differences among centres in the manner in which gestational age is assigned. The 5-minute Apgar score and the SNAP-II score are both influenced by postnatal care as well as condition at birth and severity of illness. The risk for infants transferred to a NICU (outborn infants) may differ for different NICUs depending on the care before transfer and whether all, most or only the sickest infants with a particular birth weight, congenital anomaly or diagnosis are transferred.

For these and other reasons, risk adjustment remains an imperfect process for neonates as for older patients.[7] These realities need to be well understood by journalists, consumer advocates, government officials and other groups that publish hospital "report cards" or ratings of patient outcome as measures of quality of care. Moreover, reducing the differences in care among different hospitals is not necessarily a desirable goal. Practices in different NICUs have been shown to differ in a myriad of ways.[8] In comparing different hospitals, there is no way to clearly determine which practice differences cause outcome differences or even to be sure that the outcome differences do not result from unrecognized differences in patient populations. Efforts to standardize care based on observational studies of hospital differences may be difficult, expensive, meddlesome and even harmful.

Why, then, should we bother to estimate and compare the risk-adjusted outcomes for different hospitals? One reason is that this process involves an evaluation of the strength of the relations between various risk factors and outcomes of interest. As in this study, greater knowledge about the biological or social risk factors may lead to a better understanding of the causes of adverse outcomes. Another reason is to determine whether the variation in outcomes is large enough to warrant the extensive effort required to identify and compare practices within the hospitals, to develop specific hypotheses about ways to reduce adverse outcomes and then to test these hypotheses in the most rigorous manner feasible, preferably in a randomized clinical trial. Assessing variation in risk-adjusted outcomes is the first in a series of demanding steps to reduce avoidable illness and death.

In the 17 NICUs studied, risk-adjusted mortality rate — though much less variable than unadjusted morality rate — varied more than 3-fold. The absolute difference between the highest and lowest values for risk-adjusted mortality rate (just over 4%) was comparable to the overall mean value (just over 4%) and equivalent to 1 death per 23 admissions. A range of this magnitude clearly deserves further evaluation. With these data in hand, the investigators in the Canadian Neonatal Network will evaluate practice differences among the NICUs to develop and ultimately test hypotheses about strategies to reduce adverse neonatal outcomes. We applaud their efforts.

Drs. Tyson and Kennedy are with the Center for Clinical Research and Evidence-Based Medicine, University of Texas at Houston Medical School, Houston, Tex.

Competing interests: None declared.

Contributors: Drs. Tyson and Kennedy contributed equally to writing and revising the text.

References

1. Park RE, Brook RH, Kosekoff J, Keesey J, Rubenstein L, Keeler E, et al. Explaining variations in hospital death rates: randomness, severity of illness, quality of care. *JAMA* 1990;264:484-90.
2. Parry GJ, Gould CR, McCabe CJ, Tarnow-Mordi WO, on behalf of the International Neonatal Network and the Scottish Neonatal Consultants and Nurses Collaborative Study Group. Annual league tables of mortality in neonatal intensive care units: longitudinal study. *BMJ* 1998;316:1931-5.
3. Sankaran K, Chien LY, Walker R, Seshia M, Ohlsson A, Lee SK and the Canadian Neonatal Network. Variations in mortality rates among Canadian neonatal intensive care units. *CMAJ* 2002;166(2):173-8. Available: www.cma.ca/cmaj/vol-166/issue-2/0173.html
4. Richardson DK, Corcoran JD, Escobar GJ, Lee SK, the Canadian NICU Network, the Kaiser Permanente Neonatal Minimum Data Set Area Network, and the SNAP-II Study Group. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *J Pediatr* 2000;138:92-100.
5. Jencks SF, Daley J, Draper D, Thomas N, Lenhart G, Walker J. Interpreting hospital mortality data. The role of clinical risk adjustment. *JAMA* 1988;260:3611-6.
6. Donovan EF, Tyson JE, Ehrenkranz RA, Verter J, Wright LL, Korones SB, et al. Inaccuracy of Ballard scores before 28 weeks' gestation. National Institute of Child Health and Human Development Neonatal Research Network. *J Pediatr* 1999;135:147-52.
7. Richardson D, Tarnow-Mardi WO, Lee SK. Risk adjustment for quality improvement. *Pediatrics* 1999;103:255-65.
8. Lee SK, McMillan DD, Ohlsson A, Pendray M, Synnes A, Whyte R, et al. Variations in practice and outcomes in the Canadian NICU Network: 1996–1997. *Pediatrics* 2000;106:1070-9.

Correspondence to: *Dr. Jon E. Tyson, Center for Clinical Research and Evidence-Based Medicine, University of Texas – Houston Medical School, 6431 Fannin St., MSB 2.106, Houston TX 77030; fax 713 500-0519; Jon.E.Tyson@uth.tmc.edu*