Variations in Definitions of Mortality Have Little Influence on Neonatal Intensive Care Unit Performance Ratings

Jochen Profit, MD, MPH[1,2], Jeffrey B. Gould, MD, MPH[3,4], David Draper, PhD[5], John A. F. Zupancic, MD, ScD[6,7], Marc A. Kowalkowski, MS[2,8], LeChauncy Woodard, MD, MPH[2,8], Kenneth Pietz, PhD[2,8], and Laura A. Petersen, MD, MPH[2,8]

Objective To measure the influence of varying mortality time frames on performance rankings among regional neonatal intensive care units (NICUs) in a large state.

Study design We performed a cross-sectional data analysis of very low birth weight infants receiving care at 24 level 3 NICUs. We tested the effect of 4 definitions of mortality: (1) death between admission and end of birth hospitalization or up to 366 days; (2) death between 12 hours of age and the end of birth hospitalization or up to 366 days; (3) death between admission and 28 days; and (4) death between 12 hours of age and 28 days. NICUs were ranked by quantifying their deviation from risk-adjusted expected mortality and dividing them into 3 tiers: top 6, bottom 6, and in between.

Results There was wide interinstitutional variation in risk-adjusted mortality for each definition (observed minus expected *z*-score range, −6.08 to 3.75). However, mortality-based NICU rankings and classification into performance tiers were very similar for all institutions in each of our time frames. Among all 4 definitions, NICU rank correlations were high (>0.91). Few NICUs changed relative to a neighboring tier with changes in definitions, and none changed by more than one tier.

Conclusion The time frame used to ascertain mortality had little effect on comparative NICU performance. *(J Pediatr 2013;162:50-5)*.

Similar to outcomes of patients in other areas of medicine, preterm infants receiving care in neonatal intensive care units (NICUs) experience variations in quality of health care and clinical outcomes that cannot readily be explained by differences in underlying clinical risk.[1-7] For example, risk-adjusted mortality rates among NICUs participating in the Vermont Oxford Network show up to 3-fold differences.[6] Health policy makers and health care payers are promoting comparative performance assessments to improve the value of health care.[8,9] Many clinicians are wary of this competitive model of quality-improvement; one concern is whether the data will allow for valid comparisons among providers.[10]

Studies assessing adult inpatient quality of care have shown that standardized mortality rates for congestive heart failure based on in-hospital and 30-day mortality may be relatively similar despite differing discharge practices and rates.[11] Another study did not yield markedly different results when using data from 30 days versus data from 180 days postadmission; moreover, according to the 180-day data, the 30-day data were good predictors of the best and worst quintiles of hospitals.[12] However, still another study found widely varying general in-hospital mortality rates according to the evaluation method used.[13]

In the NICU setting, the time frame used to ascertain mortality varies, depending on expert input and the properties of available data repositories. Public stakeholders and neonatal quality of care collaboratives have applied differing definitions of mortality for measurement of comparative performance; for example, the Joint Commission's neonatal mortality measure focuses on neonates who expire before 28 days of age; all liveborn neonates are included, as are transfers-in (with no limit set on the day of transfer).[14] Transfers-out are excluded from this definition. The Agency for Healthcare Research and Quality definition includes in-hospital deaths of infants born at that institution (inborn) and those transferred from another hospital (outborn).[15] Quality collaboratives have traditionally focused on death during the birth hospitalization.

Differences in these definitions may influence provider performance. For example, inclusion of deaths after day of life 28 but before discharge, although uncommon, may highlight opportunities for improvement in chronic respiratory care or health care maintenance, including avoidance of late infections. The effect of certain definitions on comparisons of provider performance must

CPQCC	California Perinatal Quality Care Collaborative
NICU	Neonatal intensive care unit

From the [1]Department of Pediatrics, Baylor College of Medicine, Texas Children's Hospital; [2]Section of Health Services Research, Department of Medicine, Baylor College of Medicine, Houston, TX; [3]California Perinatal Quality Care Collaborative; [4]Division of Neonatology, Perinatal Epidemiology and Health Outcomes Research Unit, Stanford University, Palo Alto, CA; [5]Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, Santa Cruz, CA; [6]Department of Neonatology, Beth Israel Deaconess Medical Center; [7]Division of Newborn Medicine, Harvard Medical School, Boston, MA; and [8]Houston VA Health Services Research and Development Center of Excellence, Health Policy and Quality Program, Michael E. DeBakey VA Medical Center, Houston, TX

Supported by the Veterans Administration Center (VA HSR&D CoE HFP90-20). J.P. is supported by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (grant K23 HD056298), and L.P. is the recipient of an American Heart Association Established Investigator Award (0540043N). The views expressed in this report are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. The authors declare no conflicts of interest.

 http://dx.doi.org/10.1016/j.jpeds.2012.06.002

be made explicit to draw appropriate and relevant conclusions supporting care improvement. In the present study we examined the effects of changes in measurement and sample definitions of mortality (ie, the time frame used to ascertain mortality) on NICU performance ratings.

Methods

This study used data from the California Perinatal Quality Care Collaborative (CPQCC), which comprises more than 100 member hospitals, including all of the state's 24 regional centers. The CPQCC maintains a real-time, risk-adjusted perinatal data system. Measurement bias is addressed through standardized data collection procedures. Data are abstracted from the medical record locally, using data abstraction protocols. The data are then transmitted to the CPQCC, where they are subjected to logic and completeness checks. In the event of inconsistencies, the data are reconciled with the medical record.

We attempted to minimize selection bias by using exclusion criteria designed to minimize exclusion of patients from the database, minimize systematic bias from discretionary care decisions at the border of viability, maximize comparability across NICUs by defining a clinically homogenous patient population, and ensure that clinical outcomes of care were attributed to the NICU under investigation. The exclusion criteria are listed in **Table I** (available at www.jpeds.com).

Our study sample included infants with birth weight <1500 g and gestational age of ≥25 weeks. The upper-limit weight cutoff is congruent with inclusion criteria into the CPQCC small infant database and is in harmony with other measures of quality endorsed by the National Quality Forum.[16] With regard to including infants at the border of viability, we based our criteria on a study reporting that 84% of surveyed neonatologists considered treatment clearly beneficial for infants born at >25 weeks gestational age. Only 41% of the neonatologists considered treatment beneficial at 24 1/7-6/7 weeks gestation, highlighting the potential bias introduced by inclusion of this group in our analyses.[17] Using standard CPQCC definitions, we excluded patients with major congenital anomalies (those associated with an increased mortality risk), because we were interested in examining mortality in a "healthy" preterm population. We also excluded infants who remained in the hospital beyond 1 year of life, because data collection is truncated at that age and survival is uncertain.

Patient Transfer

Data from the CPQCC database used for this study were collected at the patient level but not at the hospital level. This made it difficult to confidently attribute responsibility for certain outcomes to individual institutions for transferred patients. We considered 2 main time periods for transfers: early transfer-in and late transfer-out. The timing of early transfer-in cases depended on whether the transferring facility intended immediate postpartum neonatal transfer owing to unavailability of adequate neonatal intensive care support or whether infants received a trial of therapy in the hope of retaining care locally. We excluded infants admitted beyond 3 days of age (day 1 being the day of birth), to avoid penalizing (or crediting) institutions for the quality of care delivered at the transferring institution.

Among transferred infants, the location of a clinical event was not recorded in the database. Transfer-out of moribund infants might introduce bias, but was minimized in this analysis through our focus on regional centers. Thus, we included infants transferred out for convalescent care (85% of those transferred out). Deaths in this group were rare, and arguably the regional centers to whom these deaths were attributed assumed some degree of responsibility by sending an infant to a partner institution for convalescent care.

In infants who were transferred for other reasons (predominantly for medical services, diagnostic imaging, or surgery), confidently attributing responsibility for death to the receiving or transferring institution can be difficult. Thus, we omitted these infants from our analysis.

Sample

A total of 5289 very low birth weight infants receiving care at all of California's 24 level 3 regional centers between 2004 and 2007 met our inclusion criteria. Fifteen of these 24 centers are designated as level 3D, because they perform open-heart surgery and/or extracorporeal membrane oxygenation for medical conditions. The remaining 9 centers are classified as level 3C (subspecialty NICUs with extensive capabilities not including open-heart surgery or extracorporeal membrane oxygenation).[18] Five regional centers did not have birthing facilities onsite. We used multiyear analysis because there a small number of very low birth weight infants received care at some of the institutions. **Figure 1** (available at www.jpeds.com) shows the number of excluded infants by criterion.

Dependent Variables: Mortality

We used 4 competing time frames to ascertain mortality, presented in order of decreasing inclusivity (**Figure 2**). All of the time frames refer to the birth hospitalization and include the time after transfer. The 4 time frames are defined as: (1) death between admission and end of birth hospitalization or up to 366 days; (2) death between 12 hours of age and end of birth hospitalization or up to 366 days; (3) death between admission and 28 days; and (4) death between 12 hours of age and 28 days.

We set the early cutoff for inclusion into this study to NICU admission or 12 hours of life. This time difference represents potential bias owing to admission of moribund patients who are admitted for comfort care but expected to die imminently.

We excluded delivery room deaths from mortality definitions because of concerns regarding validity and reliability; validity because neonatal providers might not be responsible for the delivery room death of a moribund infant, and reliability because there may be local practice variation in