Serial Assessment of Mortality in the Neonatal Intensive Care Unit by Algorithm and Intuition: Certainty, Uncertainty, and Informed Consent

William Meadow, Laura Frain, Yaya Ren, Grace Lee, Samir Soneji and John Lantos

*Pediatrics* 2002;109;878-886
DOI: 10.1542/peds.109.5.878

The online version of this article, along with updated information and services, is located on the World Wide Web at:

http://www.pediatrics.org/cgi/content/full/109/5/878
Serial Assessment of Mortality in the Neonatal Intensive Care Unit by Algorithm and Intuition: Certainty, Uncertainty, and Informed Consent

William Meadow, MD, PhD; Laura Frain, MD; Yaya Ren, JD; Grace Lee, BA; Samir Soneji, BA; and John Lantos, MD

ABSTRACT. Objectives. Does predictive power for outcomes of neonatal intensive care unit (NICU) patients get better with time? Or does it get worse? We determined the predictive power of Score for Neonatal Acute Physiology (SNAP) scores and clinical intuitions as a function of day of life (DOL) for newborn infants admitted to our NICU.

Methods. We identified 369 infants admitted to our NICU during 1996–1997 who required mechanical ventilation. We calculated SNAP scores on DOL 1, 3, 4, 5, 7, 10, 14, 21, 28, and weekly thereafter until either death or extubation. We also asked nurses, residents, fellows, and attendings on each day of mechanical ventilation: “Do you think this child is going to live to go home to their family, or die before hospital discharge?”

Results. Two thousand twenty-eight SNAP scores were calculated for 285 infants. On DOL 1, SNAP for nonsurvivors (24 ± 8.7 [standard deviation]) was significantly higher than SNAP for survivors (13 ± 6.1). However, this difference diminished steadily and by DOL 10 was no longer statistically significant (12.7 ± 4.9 vs 10.0 ± 4.8). On each NICU day, at all ranges of SNAP scores, there were at least as many infants who would ultimately survive as would die. Consequently, the positive predictive value of any SNAP value for subsequent mortality was <0.5 on all NICU days. Prediction profiles were obtained for 230 ventilated infants reflecting over 11,000 intubations obtained on 2867 patient days. One hundred fifty-seven (81%) of 192 survivor profiles displayed consistent accurate prediction profiles—at least 90% of their NICU ventilation days were characterized by 100% prediction of survival. Twenty-five (13%) of 192 surviving infants survived somewhat unexpectedly; that is, after at least 1 day characterized by at least 1 estimate of “death.” Thirty-three (60%) of the 55 nonsurvivors died before DOL 10. Eighty-two percent of the prediction profiles for these early dying infants were homogeneous, dismal, and accurate. Twenty-two (40%) of the 55 nonsurvivors died after DOL 10. Seventeen (78%) of these 22 late-dying infants were predicted to live by many observers on many hospital days. Sixty-one (30%) of 230 profiled patients had at least 1 NICU day characterized by at least 1 prediction of death; 26/61 (43%) of these patients were incorrectly predicted; that is, they survived. Seventeen infants who were predicted to die during but survived nonetheless were assessed neurologically at 1 year. Fourteen (82%) of these 17 were not neurologically normal—8 were clearly abnormal, 1 suspicious, and 5 had died.

Conclusions. If absolute certainty about mortality is the only criterion that can justify a decision to withhold or withdraw life-sustaining treatment in the NICU, these data would make such decisions difficult on the first day of life, and increasingly problematic thereafter. However, if we acknowledge that medicine is inevitably an inexact science and that clinical predictions can never be perfect, we can ask the more interesting question of whether good but less-than-perfect predictions of imprecise but ethically relevant clinical outcomes can still be useful. We think that they can—and that they must.


ABBREVIATIONS. NICU, neonatal intensive care unit; ICU, intensive care unit; SNAP; Score for Neonatal Acute Physiology; DOL, day of life; MDI, Mental Development Index; PDI, Physical Development Index; SD, standard deviation; PPV, positive predictive value; ROC, receiver operating characteristic; BW, birth weight; ELBW, extremely low birth weight.
doomed infants from survivors, they could restrict life-sustaining treatments to infants who would benefit, and offer high quality palliative care to those whose death was inevitable. Recognizing this, physicians, ethicists, parents, economists, and policy makers all want to refine prognostic ability and accuracy.

In an attempt to predict the likelihood of survival or nonsurvival for patients admitted to intensive care units (ICUs), illness severity scores were created. These scores (APACHE, later PRISM, Score for Neonatal Acute Physiology [SNAP], and CRIB) for the most part tabulate the physiologic stability of patients in the first 24 hours after ICU admission.12–17 They successfully parse ICU populations into likely survivor and nonsurvivors. In general, the more deranged the physiology, the higher the score, the more likely the patient will die. These prognostications have proven important to counsel patients and their families, to allocate resources, and to level the playing field when attempting to assess the impact of novel ICU innovations, including, as it happens, ICU care itself.18

Illness severity scores have unquestionably succeeded in some of these goals. We, however, were interested in whether they might serve an additional purpose. One dimension of ICU care is often omitted—survivor and nonsurvivors. In general, the more deranged the physiology, the higher the score, the more likely the patient will die. These prognostications have proven important to counsel patients and their families, to allocate resources, and to level the playing field when attempting to assess the impact of novel ICU innovations, including, as it happens, ICU care itself.18

The question naturally arises—does the predictive power of illness severity scores get better with time? Or does it get worse? Perhaps, as initially seems intuitive, patients who will ultimately survive improve over time, whereas doomed patients worsen. If this is the case, then with each passing day assessments of illness severity should possess greater discriminatory power. Alternatively, perhaps many of the physiologic disturbances leading to ICU admission are, at least transiently, correctable for both survivors and doomed patients. If so, then the illness severity scores of the 2 subpopulations would converge, and the ultimate fate of patients would become less clear with the passage of time. Perhaps there is a U-shaped pattern of illness severity scores over time, with transient convergence followed by divergence?

As an alternative to algorithms of physiologic stability, clinical intuitions at the time of admission have also been shown to be predictive of outcomes for ICU patients.12,19–22 The same questions immediately arise for these nonquantitative predictors—are they more and more valid over time, or less so? Do ICU patients “declare” themselves, or “cloak” themselves, with each passing day? Are serial intuitions better, or worse, than serial illness severity scores as predictors for ICU outcomes? Is there an interaction between the 2?

In an attempt to determine the limits of our prognostic ability for newborn infants, and to assess the implications of such limitations for our capacity to inform parents about the potential benefits and burdens of ongoing intensive care, we determined the predictive power of SNAP scores and clinical intuitions as a function of day of life for newborn infants admitted to our NICU.

METHODS

Patient Population

We identified 369 infants admitted to our NICU who required mechanical ventilation during an 18-month interval in 1996–1997. We restricted our study population to infants requiring mechanical ventilation, because infants in the NICU who do not require mechanical ventilation are either highly unlikely to die, or have recognized lethal conditions that have led doctors and parents to agree on a limited treatment plan. The ethical problems raised by such decisions are beyond the scope of the current study. Infants treated with nasal continuous positive airway pressure, but not intubated, were excluded from this analysis. For infants who were reintubated after initial extubation, SNAP and intuition assessments were resumed on the day of reintubation. Each infant contributed only once to an outcome of live/die, no matter how many times the infant was placed on the ventilator. No distinctions were made between conventional intermittent mechanical ventilation and high frequency oscillatory ventilation, both of which were in use in our NICU during the study period.

SNAP Scoring

SNAP was originally developed by Richardson et al.16 to assess derangements in physiology of several organ systems (respiratory, cardiac, renal, and neurologic) in the first 24 hours of life. For 285 ventilated infants, we calculated the SNAP on day of life (DOL) 1, 3, 4, 5, 7, 10, 14, 21, 28, and weekly thereafter until either death or extubation. Extension of the original SNAP score beyond DOL 1 was straightforward. After consultation with Richardson and colleagues (personal communication, May 2000), we concluded that the only significant modification required was in the assessment of PaO₂ for patients who remained on mechanical ventilators but no longer had arterial catheters in place. Rather than ignore that item (which we felt would inappropriately lower the SNAP and underestimate the physiologic derangement), we assessed PaO₂ from capillary blood samples and transcutaneous PaO₂ assessments (which were always available). Assignment to the originally described SNAP categories (PaO₂ >65 torr, 50–65 torr, 30–49 torr, and <30 torr) was straightforward using this modification.

Intuitions Protocol

During 12 consecutive months of our study, we identified 254 mechanically ventilated infants. For each of these patients, on each day one mechanical ventilation protocol was asked intuitions about, and attendings 1 question: “Do you think this child is going to live to go home to their family, or die before hospital discharge?” Each respondent was allowed to answer “live,” “die,” or “uncertain” if she/he could not comfortably answer on that day. Respondents were polled individually and privately, in an attempt to minimize the influence of other respondents on the opinions offered. The investigators spent several hours each day in the NICU, attempting to obtain, at a minimum, the opinions of the primary nurse and primary medical resident, as well as attending and/or fellow for each eligible infant.

Follow-up Neurologic Assessment of Surviving Infants

Eighty-four surviving infants were assessed in our NICU follow-up program until 24 months of age. Infants were categorized as neurologically normal, suspicious, or abnormal using a combination of Bayley Scale of Infant Development (Mental Developmental Index [MDI], Physical Developmental Index [PDI]), and Denver Developmental Screening Examination.23 In an attempt to emphasize ethically relevant neurologic abnormalities, infants were categorized as abnormal only if they had an MDI or PDI score below 75, or their Denver developmental score was <50% of corrected age. Infants with MDI or PDI between 75 and 85, or Denver score between 50% and 75% of corrected age were classified as suspicious.

Statistical Analysis

Data are presented as mean ± standard deviation (SD). Comparisons of SNAP scores between surviving and nonsurviving
populations were performed using a 2-way analysis of variance, with population and DOL as main variables. For individual patients, progression of SNAP versus DOL was analyzed with repeated measures analysis of variance. Other comparisons between surviving and nonsurviving populations were performed with Student t test for normally distributed variables, and Mann-Whitney test for nonnormally distributed variables. Positive predictive value (PPV), negative predictive value, sensitivity, and specificity were calculated in the usual fashion, and receiver-operator characteristic (ROC) curves developed from these calculations. For all evaluations, statistical significance was accepted at a value of \( P < .05 \).

This study was approved by the institutional review board at the University of Chicago.

RESULTS

Patient Demographics

Table 1 presents patient demographics for the 369 ventilated NICU patients who received at least 1 SNAP score or intuition profile during the 18-month study period. Three hundred three (84%) infants survived, and 66 (16%) died before NICU discharge. Not surprisingly, the population of survivors had larger birth weights and longer gestations than nonsurvivors (median 1325 g/31 weeks vs 786 g/26 weeks; both \( P < .001 \) survivors vs nonsurvivors). The median duration of ventilation for nonsurvivors (equivalent to their length of NICU stay) was 4 days, significantly shorter than the median days of mechanical ventilation or length of NICU stay for survivors (6 days and 24 days, respectively; both \( P < .01 \)).

Survival improved with increasing birth weight (BW) for the study population. Forty-one percent of 49 infants with BW <750 g survived, compared with 83% of 72 infants with BW 751 to 1000 g, and 90% of 248 ventilated infants with BW >1000 g.

Serial SNAP Scores as Predictors of Mortality

A total of 2028 SNAP scores were calculated for 248 ventilated infants with BW 750 g survived, compared with 83% of 72 infants with BW 751 to 1000 g, and 90% of 248 ventilated infants with BW >1000 g.

TABLE 1. Demographics of Surviving and Nonsurviving Infants

<table>
<thead>
<tr>
<th></th>
<th>Nonsurvivors (n = 66)</th>
<th>Survivors (n = 303)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>26</td>
<td>31*</td>
</tr>
<tr>
<td>Average</td>
<td>28.6</td>
<td>31.7*</td>
</tr>
<tr>
<td>SD</td>
<td>5.5</td>
<td>4.9</td>
</tr>
<tr>
<td>BW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>786</td>
<td>1325*</td>
</tr>
<tr>
<td>Average</td>
<td>1248</td>
<td>1728*</td>
</tr>
<tr>
<td>SD</td>
<td>866</td>
<td>929</td>
</tr>
<tr>
<td>Days on ventilator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>6*</td>
</tr>
<tr>
<td>Average</td>
<td>13.1</td>
<td>16.1*</td>
</tr>
<tr>
<td>SD</td>
<td>23.4</td>
<td>22.9</td>
</tr>
<tr>
<td>Days in NICU</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>4</td>
<td>24*</td>
</tr>
<tr>
<td>Average</td>
<td>13.1</td>
<td>32.7*</td>
</tr>
<tr>
<td>SD</td>
<td>23.4</td>
<td>29.2</td>
</tr>
</tbody>
</table>

* \( P < .05 \).

SNAP Values vs DOL for Surviving and Non-Surviving Infants

Fig 1. Average SNAP scores as a function of DOL for populations of survivors and nonsurvivors during the first 21 days of life. On DOL 1, SNAP for nonsurvivors (24 \( \pm 6.7 \) [SD]) was significantly higher than SNAP for survivors (13 \( \pm 6.1 \); \( P < .001 \)). This difference diminished steadily over time, as SNAP improved for both groups.

Might the fall in SNAP values for the population of nonsurvivors be an artifact of the death of the children with highest SNAP, and their consequent removal from subsequent analyses? Analysis of serial SNAP values for individual patients suggests that there is more to the phenomenon than this. Nineteen of 45 nonsurvivors died within the first 2 days, and consequently had no trend for their daily SNAP scores. These patients had significantly higher SNAP values than nonsurvivors who died after DOL 2 (27 \( \pm 10 \) vs 21 \( \pm 6 \); \( P < .02 \)). However, of the 26 nonsurvivors who survived to DOL 3 and received at least 2 SNAP values, 22 (88%) had downward trends in their serial SNAP analyses, with final SNAP scores lower than their SNAP on DOL 1. Only 9 (20%) of 45 nonsurvivors had a “U-shaped” course, where SNAP initially fell, then rose again before death. Not surprisingly, for each of 240 survivors, SNAP on the day they were extubated (and consequently exited the study) was lower than SNAP on DOL 1.

Figure 2 elaborates these phenomena by displaying a scattergraph of every SNAP value obtained during the first 10 days of life for 285 ventilated infants. One hundred twenty-five SNAP determinations for the 45 nonsurvivors and 696 SNAP values for 240 surviving infants are presented. The striking aspect of this figure is that on each day, at virtually all ranges of SNAP scores, there are at least as many infants who will ultimately survive as will die. Consequently, the PPV of a SNAP equal to the median SNAP of the nonsurvivors is <.5 at all times (Fig 3). When this analysis is restricted to ELBW infants, where predictive discrimination would be most desirable, this observation remains essentially unchanged (maximum PPV = 0.56).
Predictive Power of SNAP Scores as a Function of Length of NICU Stay

The combination of predictive power and sensitivity for serial SNAP scores is displayed in Fig 4, a family of ROC curves for SNAP values on days between DOL 1 and DOL 10. Sensitivity (varying with true +) is plotted on the Y-axis against 1-Specificity (varying with false +) on the X-axis. The area under the curve for these plots diminishes as a function of DOL, falling from 0.84 on DOL 1 to 0.64 by DOL 10, revealing that SNAP scores become progressively less predictive and sensitive as an index of the subsequent demise.

Serial Intuitions as Predictors of Mortality

Prediction profiles were obtained for 230 of 254 ventilated infants born during the 12-month intuition-collection period. One hundred ninety-two (78%) of these infants survived, while 55 (22%) died. One hundred ninety-two of 199 survivors and 38 of 55 nonsurvivors received prediction profiles. All nonsurvivors not profiled were born, resuscitated, admitted to the NICU, and died before surveyors could obtain any predictions. All survivors not profiled were born, ventilated, and extubated before surveyors could obtain any prognostications. The 230 patient profiles reflect over 11,000 intuitions obtained on 2,867 patient days. The median number of daily intuitions for each ventilated infant was 4. There was no significant difference comparing the number of daily intuitions obtained for nonsurvivors versus survivors.

Prediction Profiles for Survivors

Prediction profiles for survivors reflected 2 distinct hospital courses. Most surviving infants were predicted by all (or almost all) observers to survive on all (or almost all) days of mechanical ventilation. One hundred fifty-seven (82%) of 192 survivor profiles displayed this consistent accurate prediction profile—for 136 (70%) of 192 surviving infants, each and every one of their NICU ventilator-days was characterized by 100% prediction of survival. Twenty-one other survivors had profiles nearly as positive—for these infants, a brief period of uncertainty was followed by increasing confidence in the likelihood of survival, but at least 90% of their NICU ventilation days were characterized by 100% prediction of survival. Figure 5A portrays a representative prediction profile for this group of clear-cut survivors.

Alternatively, for 35/192 (18%) survivors, respondents were less certain about survival during some period of their hospital stay. Twenty-five infants (13% of survivors) were discharged despite at least 1 day characterized by at least 1 estimate of “die.” Indeed, 8 infants survived despite having at least 1 hospital day in which all respondents predicted death. Figure 5B portrays a representative prediction profile.
Prediction Profiles for Nonsurvivors

Death Before DOL 10

Thirty-three (60%) of the 55 nonsurvivors died before DOL 10. All of these infants received prediction profiles. In contrast to the homogeneity that characterized profiles of infants who died before DOL 10, the 22 late-dying infants were a heterogeneous group. Only 5 (22%) of these 22 late-dying infants had the “flat-line” or near “flat-line” profiles that categorized predictions for almost all infants who died before DOL 10. Each of the other 17 late-dying infants was predicted to live by many (if not all) observers on many (if not all) hospital days. Eleven (50%) of these infants suffered, with little warning, a fatal medical catastrophe (necrotizing enterocolitis, sepsis, pneumonia, etc). The rapid and unexpected nature of the demise is emphasized by the observation that for 7 of these late-dying infants, not even 1 day of their hospital stay before the actual day of death was marked by 0% prediction of survival. The prediction profile of 1 of these infants is depicted in Fig 6B.

Six (27%) of 22 late-dying infants had prediction profiles categorized by uncertainty—both within respondents and across days. That is, several hospital days were characterized by “pessimism” (ie, low predictions of survival), alternating with periods of “optimis,” characterized at times by up to 100% prediction of survival. These nonsurvivors often lingered for many weeks before their death. The prediction profiles of these infants resembled those depicted in Fig 5B.

Predictive Power of Intuitions of “Die” as a Function of Length of NICU Stay

Additional insight into predictive power, or lack thereof, of caretaker intuitions of nonsurvival can be derived from determining the PPV of a single day of 100% prediction of demise as a function of the DOL on which the predictions were obtained (Fig 7). During the first 10 days of life, 100% intuitions of nonsurvival retained impressive predictive power, as 80% of patients with at least 1 day of 100% prediction of demise actually died. However, after DOL 14 the PPV of a day of 100% prediction of death fell sharply, and beyond DOL 21 the PPV fell below 50%.

Resources Allocated to Predicted Nonsurvivors

Overall, 61/230 (30%) profiled patients had at least 1 NICU day characterized by at least 1 prediction of death. Twenty-six (43%) of 61 of these patients were incorrectly predicted—that is, they survived. More stringent criteria for prediction of nonsurvival improved predictive power only slightly. Thirty-four (15%) of 230 of patients had at least 1 day characterized by uniform (100%) predictions of death—8 (24%) of these 34 infants were incorrectly predicted; that is, they survived. Indeed, 22/230 (10%) of infants had 3 consecutive days characterized by uniform (100%) prediction of death—4 (18%) of these 22 survived. When the analysis is restricted to extremely low birth weight (ELBW) infants, the results
Concordant, Correct Prediction of Demise

Concordant, Incorrect Prediction of Survival

Fig 6. A, Representative prediction profile for infants who died between before DOL 10. The prediction profiles for these infants were remarkably homogeneous, dismal and accurate. Thirteen of the 16 profiles in this category had 100% prediction of death on every DOL from birth to the day of death. The X signifies that the infant died. B, Representative prediction profile for infants who died unexpectedly, after DOL 10. For 7 of these late-dying infants, not even 1 day of their hospital stay before the actual day of death was marked by 0% prediction of survival. The X signifies that the infant died.

PPV of 100% Prediction of Death vs DOL

Fig 7. PPV of a single day of 100% prediction of death as a function of the DOL on which the predictions were obtained. During the first 10 days of life, 80% of patients with at least 1 day of 100% prediction of death actually died. However, beyond DOL 21 the PPV of a day of 100% prediction of death fell below 50%.

are similar. Seventeen (41%) of 41 ELBW infants predicted on at least 1 day to “die” survived to discharge. Four (18%) 22 ELBW infants with at least 1 day of 100% prediction of death survived to discharge nonetheless.

Viewed from the perspective of resource allocation, 407/2867 (14%) NICU ventilator-days were occupied by infants who eventually died before discharge. However, only 96/2867 (3.3%) NICU ventilator-days were occupied by nonsurvivors who received a day of 100% prediction of death. For ELBW infants, at greatest risk to die, this proportion was even lower 952/1823 (2.9%).

Discordant Predictions of Survival or Death

One hundred eighty-three (80%) 230 infants, representing 2382/2867 (83%) of NICU bed-days had prediction profiles characterized by complete agreement on the part of the caretakers—2242/2867 (78%) days were characterized by complete agreement on survival, 140/2867 (5%) by agreement on nonsurvival. Of the 47/230 (20%) ventilated infants who had at least 1 day of “live/die” disagreement, 21 (45%) died, while 26 (55%) survived. Neither attendings, fellows, residents, nor nurses were significantly more or less optimistic, pessimistic, or correct than the other caretaker groups in their prognostications.

Predictive Power for Residual Neurologic Morbidity

Neurologic status of 84 ventilated infants (median BW/gestational age 1150 g/29 weeks) who survived to NICU discharge was assessed at 1 year of age. Overall, 28 (33%) infants were not normal—15 were clearly neurologically abnormal, 6 had suspicious examinations, and 7 had died since NICU discharge. Table 2 presents the correlation between postdischarge neurologic assessment and caretaker intuitions obtained during NICU. Thirty-two infants (38% of the follow-up cohort) were never predicted either to die or to suffer moderate or severe disability—28 of these 32 infants (88%) were normal at a year. Another 35 infants (42% of the follow-up cohort) were never predicted to die, but had been predicted by at least 1 caretaker on at least 1 day as likely to survive with moderate or severe neurologic sequela—10 (29%) of these 35 were abnormal at a year. Seventeen infants who had been predicted by a care-

TABLE 2. Correlation of NICU Intuitions With Post-NICU Assessments of Neurologic Morbidity

<table>
<thead>
<tr>
<th></th>
<th>Normal outcome</th>
<th>Morbid outcome</th>
<th>Died</th>
</tr>
</thead>
<tbody>
<tr>
<td>No “Die”/No “Moderate/Severe Morbidity” (n = 32)</td>
<td>28 (88%)</td>
<td>4 (12%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Definitely abnormal</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suspicious</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Died</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>No “Die”/Yes “Moderate/Severe Morbidity” (n = 35)</td>
<td>25 (71%)</td>
<td>10 (29%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Definitely abnormal</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suspicious</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Died</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Yes “Die” (n = 17)</td>
<td>3 (18%)</td>
<td>1 (82%)*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Definitely abnormal</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suspicious</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Died</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

* P < .001 “Die” vs No “Die” groups.
taker to die during their NICU stay but had survived to discharge were evaluated. These infants comprised 20% of the follow-up cohort and represented 65% of all survivors predicted to die. Fourteen (82%) of these 17 unexpected survivors were not normal at a year—8 were abnormal, 1 suspicious, and 5 had died after NICU discharge.

DISCUSSION

Intimations of Mortality in the NICU

Prognostications of nonsurvival, whether derived from algorithms or clinical intuitions, have been studied in ICU settings from the elderly to the newborn.12–25 The vast majority of these studies have obtained assessments at a single time point, admission to the ICU, to calibrate the severity of illness of the overall population of ICU patients. These efforts have proved remarkably useful for addressing questions of public policy and quality control across ICUs.

However, as an attempt to inform ethical decision-making for individual patients, they have been less useful. Most patients, especially newborns, are admitted to an ICU for a trial of therapy, tacitly or explicitly agreed to by both the physician caretakers and patient surrogates. After a relatively brief period of intensive care, it is expected that the patients will “declare themselves”—a metaphor for the purported increase in accuracy of prognostic estimates as a function of length of ICU stay. On this view, the ethical appropriateness of continuing NICU support would be revisited periodically, illuminated ever more brightly by ongoing revision of increasingly accurate prognostications.26

In this study, we attempted to validate this intuition by assessing the predictive power of serial assessments of either algorithms or intuitions of mortality. We began by confirming what others have previously recognized—that on DOL 1 it is possible to divide NICU patients into 2 populations, 1 that is likely to survive and 1 that is not. Unfortunately, we have also confirmed that on DOL 1 there is considerable overlap between these populations. Consequently, although serving excellently to “level the playing field” for public policy analysts, both SNAP and intuitions on the day of NICU admission are less useful for individual physicians caring for individual patients.

We then assessed the addition of new dimension—time. We hypothesized that patients for whom NICU care “worked” would get better, have lower SNAP scores, “look” healthier, and be less likely to die, whereas patients for whom NICU care didn’t work would look sicker, have higher SNAP scores, and be correctly intuited as increasingly likely to die. Despite the intuitive appeal of such an hypothesis, our data do not demonstrate it. Why?

The single most important epidemiologic phenomenon in the NICU is that most patients survive—even most sick ventilated patients survive. For the vast majority of these surviving infants, optimistic prognostications are consistent, corroborated, and correct. Moreover, many nonsurvivors, particularly the early dying nonsurvivors, are similarly easy to identify—these infants are moribund virtually their whole (short) lives. However, and here is the counter-intuitive observation, most infants who die after the first few days are not sick their whole life—they improve, at least transiently, and both their illness severity scores and prediction profiles move toward those of the ultimately surviving population. Instead of “declaring themselves,” these infants seem to “cloak themselves”; their ultimate demise becoming less apparent with each succeeding NICU day.

These late-dying, cloaked infants, account for the vast majority of NICU bed-days devoted to nonsurvivors. Futile care, defined as circumstances in which every professional agreed that the infant would not survive, treatment was continued, and the infant did not in fact survive, is apparently very rare in the NICU (3% of overall NICU ventilator-days, in the population studied here). To the extent that protracted treatment was provided to infants who ultimately died, their deaths were unpredictable. Even when these analyses are restricted to ELBW infants, the most ethically relevant subpopulation at greatest risk to die, the phenomena remain robust. These observations underscore our previous findings that little NICU care is spent on nonsurviving infants overall (predicted or not), and stand in sharp contrast to the relatively larger amount of care provided to nonsurviving adults in ICUs.26–32

Are physicians better at predicting death than nurses, or vice versa? The data reveal that neither group was much better than the other—because both are very good. There simply isn’t much prognostic discord in the NICU. That is, there were relatively few patient days where one respondent predicted “live” and another predicted “die before discharge.” The vast majority of survivors were predicted by almost every observer, on almost every day, correctly to survive. In addition, almost half of the nonsurvivors were predicted correctly to die by almost every observer on almost every day. Overall, only 20% of infants had even a single day of disagreement, and disagreement was noted on only 12% of NICU ventilator-days. Moreover, for those few patients where discord was noted, roughly half (21/47) died and half (26/47) lived. These observations suggest that on balance prediction errors were neither overly optimistic nor overly pessimistic. The relative consensus in NICU intuitions stands in marked contrast to comparable data obtained from an adult ICU, where 50% of all patients and 75% of all patient days were marked by discord between one caretaker predicting “live” and another predicting “die before discharge.”33

What of quality of life? Perhaps a lifetime of severe neurologic morbidity is a “worse” outcome than death, at least for some infants and some parents. Extrapolations from our relatively small follow-up population must necessarily be tentative. Nonetheless, some inferences seem appropriate. First, approximately 30% of ventilated infants who survive to NICU discharge will have at least moderate neurologic morbidity. This figure is comparable to those reported in more comprehensive NICU follow-up
studies in both the United States and United Kingdom. Moreover, our current study adds the potential to parse this follow-up population into those infants who are not likely to suffer significant morbidity and those who are. If an infant has what can be considered a relatively “benign” NICU course (ventilated, but never thought likely to either die or survive with significant morbidity), there is roughly a 90% chance that the infant will be neurologically normal at a year. If the infant is sick enough to prompt a prediction of “survive with moderate to severe morbidity” (but never sick enough that anyone predicts death), the glass is partly full (70% of these infants are normal at a year) and partly empty (30% are not). If, however, the infant is sick enough to engender a prediction of death, the likelihood of neurologic abnormality at a year, should the infant survive, exceeds 80%.

Incorporating these morbidity assessments into the predictive power of an intuition of death yields potentially powerful insights. Our data suggest that approximately 40% of infants who are sick enough that a caretaker predicts that they will die in the NICU survive to discharge nonetheless. However, fewer than 20% of these unanticipated survivors were neurologically normal at a year. Consequently, the likelihood of surviving to a year of age with a normal neurologic outcome after a prediction of death in the NICU may be lower than 8 in 100. These observations suggest that if we change the outcome variable from “die” to “die or survive with significant neurologic impairment” counseling after a prediction of death now has an important, almost requisite, feel—and, conversely, parental refusal of prolonged intensive care under these conditions has a newly empowered status.

Several potential methodologic concerns regarding the prediction profiles presented here can be addressed explicitly. First, might our intuition data reflect a self-fulfilling prophecy; that is, once non-survival was predicted, was the balance of NICU care tilted to produce the demise of the infant? We think this unlikely for a number of reasons. First, clinical prognostication of non-survival was just the first step in a sequence of actions that may have led to limitations of treatment. On occasion, these predictions led to discussion with parents and other family members about decisions to withhold or withdraw treatment. However, only rarely did these discussions lead to consensus to withdraw support. The large majority of nonsurvivors in our study were “full codes” up to the day of their death, reflecting the wishes of our parent population.

We cannot rule out more subtle effects of such self-fulfilling prophecies in our infants, analogous to the reduction in therapeutic interventions that has been reported for adults with do-not-resuscitate status. However, we note that any such effects would, by definition, have tended to make prognostication of non-survival more accurate, rather than less so; that is, the self-fulfilling prophecy hypothesis would have made patients who were predicted to die more likely to succumb after their predictions had been made explicit. Thus, the possibility of any such effects biases against the observations noted here, and only serves to emphasize our conclusion that prognostication of non-survival is fraught with uncertainty.

Might our predictions have reflected a “herd” phenomenon? That is, once the opinion of non-survival was articulated (particularly by the attending physician), might others “jump on the bandwagon?” We explicitly tested the possibility of a “herd” bias in a pilot study by comparing predictions of our respondents to predictions of experienced NICU nurses who did not participate in rounds or provide any direct patient care during the study period. In the vast majority of cases, we found agreement between our “blinded” respondents and our study respondents.

Informed Consent: Unreached or Unreachable?

Questions about clinical decision-making at the end-of-life can be addressed under 2 broad moral frameworks. The first focuses on patient autonomy and the belief that patients (or, in the case of children, their parents) are in the best position to determine the course of their treatment. The other broad moral framework focuses on medical futility. By this view, the problem is not that patients are disempowered. Instead, it is that both doctors and patients generally want and choose continued life-sustaining medical treatment unless the treatment is futile, at which point both would choose to forego treatment. Therefore, the challenge is not one of procedural empowerment but of prognostic refinement. If we can learn better how to determine whether a treatment is futile, then doctors and patients will both be willing to forego it.

To be at all useful clinically, futility must be recognized prospectively (that is, before the patient dies as opposed to after the fact), and for individual patients (as opposed to populations). Our data suggest the disturbing possibility that serial assessments of impending death for individual infants, either algorithmic or intuitive, are imperfect to begin with and grow less and less accurate with each passing day. In this regard, our data emphasize and extend the preliminary findings of Richardson et al who also demonstrated the diminished predictive power of SNAP scores obtained during the second week of life. However, if the outcome variable is shifted from “die” to “die or survive with significant neurologic morbidity,” then intuitions in the NICU seem to be powerful predictively.

These observations raise an interesting ethical problem. If absolute certainty about mortality is the only criterion that can justify a decision to withhold or withdraw life-sustaining treatment in the NICU, these data would make such decisions difficult on the first day of life, and increasingly problematic thereafter. However, if we acknowledge that medicine is currently (and inevitably) an inexact science, and that clinical predictions can never be perfect, we can ask the more interesting question of whether good but less-than-perfect predictions of imprecise but ethically relevant clinical outcomes can still be useful. We think that they can—and that they must.
The alternative would be to treat all patients at all times regardless of their current clinical condition or probabilistic prognosis. We think a better use of the data presented in this study is to provide quantitative estimates of the limitations of prognostication. These estimates should be part of the process of informed consent and shared decision-making by which doctors and parents work together to arrive at clinical decisions.

REFERENCES


SERIAL ASSESSMENTS OF MORTALITY IN THE NICU

Downloaded from www.pediatrics.org. Provided by Univ Of Pennsylvania on October 23, 2009
Serial Assessment of Mortality in the Neonatal Intensive Care Unit by Algorithm and Intuition: Certainty, Uncertainty, and Informed Consent
William Meadow, Laura Frain, Yaya Ren, Grace Lee, Samir Soneji and John Lantos
Pediatrics 2002;109;878-886
DOI: 10.1542/peds.109.5.878

<table>
<thead>
<tr>
<th>Updated Information &amp; Services</th>
<th>including high-resolution figures, can be found at: <a href="http://www.pediatrics.org/cgi/content/full/109/5/878">http://www.pediatrics.org/cgi/content/full/109/5/878</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td>This article cites 34 articles, 19 of which you can access for free at: <a href="http://www.pediatrics.org/cgi/content/full/109/5/878#BIBL">http://www.pediatrics.org/cgi/content/full/109/5/878#BIBL</a></td>
</tr>
<tr>
<td>Citations</td>
<td>This article has been cited by 10 HighWire-hosted articles: <a href="http://www.pediatrics.org/cgi/content/full/109/5/878#otherarticles">http://www.pediatrics.org/cgi/content/full/109/5/878#otherarticles</a></td>
</tr>
</tbody>
</table>
| Subspecialty Collections       | This article, along with others on similar topics, appears in the following collection(s):
|                                | Premature & Newborn http://www.pediatrics.org/cgi/collection/premature_and_newborn |
| Permissions & Licensing        | Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.pediatrics.org/misc/Permissions.shtml |
| Reprints                       | Information about ordering reprints can be found online: http://www.pediatrics.org/misc/reprints.shtml |

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN™

Downloaded from www.pediatrics.org. Provided by Univ Of Pennsylvania on October 23, 2009