NATIONAL QUALITY FORUM

Measure Submission and Evaluation Worksheet 5.0

This form contains the information submitted by measure developers/stewards, organized according to NQF's measure evaluation criteria and process. The evaluation criteria, evaluation guidance documents, and a blank online submission form are available on the submitting standards web page.

NQF #: 0334   **NQF Project:** Pulmonary Project
(for Endorsement Maintenance Review) **Original Endorsement Date:** May 15, 2008 **Most Recent Endorsement Date:** May 15, 2008 **Last Updated Date:** Jan 20, 2012
BRIEF MEASURE INFORMATION
De.1 Measure Title: PICU Severity-adjusted Length of Stay
Co.1.1 Measure Steward: Virtual PICU Systems, LLC
De.2 Brief Description of Measure: The number of days between PICU admission and PICU discharge.
2a1.1 Numerator Statement: Number of PICU days, PICU days = Number of days between PICU admission and PICU discharge
2a1.4 Denominator Statement: Discharges from the PICU (including tranfers to other units) during the time period being reported
2a1.8 Denominator Exclusions: Patients => 18 years of age
1.1 Measure Type: Outcome **2a1. 25-26 Data Source:** Administrative claims, Electronic Clinical Data : Registry, Paper Records **2a1.33 Level of Analysis:** Facility **1.2-1.4 Is this measure paired with another measure?** No **De.3 If included in a composite, please identify the composite measure** (*title and NQF number if endorsed*): Recommend use in conjunction with 0334 (PICU Severity Adjusted Length of Stay) and 0335 (PICU Unplanned Readmission Rate) as balancing measures.

STAFF NOTES (*issues or questions regarding any criteria*)
Comments on Conditions for Consideration:
Is the measure untested?   Yes☐ No☐   If untested, explain how it meets criteria for consideration for time-limited endorsement:
1a. Specific national health goal/priority identified by DHHS or NPP addressed by the measure (*check De.5*): 5. Similar/related endorsed or submitted measures (*check 5.1*): Other Criteria:
Staff Reviewer Name(s):

1. IMPACT, OPPORTUNITY, EVIDENCE - IMPORTANCE TO MEASURE AND REPORT
Importance to Measure and Report is a threshold criterion that must be met in order to recommend a measure for endorsement. All three subcriteria must be met to pass this criterion. See guidance on evidence. *Measures must be judged to be important to measure and report in order to be evaluated against the remaining criteria.* (evaluation criteria)
1a. High Impact: H☐ M☐ L☐ I☐ (*The measure directly addresses a specific national health goal/priority identified by DHHS or NPP, or some other high impact*

aspect of healthcare.)

De.4 **Subject/Topic Areas** *(Check all the areas that apply)*:   Pulmonary/Critical Care, Pulmonary/Critical Care : Critical Care
De.5 **Cross Cutting Areas** (*Check all the areas that apply*):

1a.1 Demonstrated High Impact Aspect of Healthcare: High resource use, Patient/societal consequences of poor quality, Severity of illness

1a.2 If "Other," please describe:

1a.3 Summary of Evidence of High Impact *(Provide epidemiologic or resource use data)*:
ICUs are a source of signficant health care cost (1,2,5) PICUs have been shown to have varying degrees of efficiency with consumption of resources that could have been provided elsewhere (3,4). This, coupled with the potential for hospital acquired infections, supports caring only for those patients in the ICU who require ICU-level care.

1a.4 Citations for Evidence of High Impact cited in 1a.3: 1. Russell LB: The role of technology assessment in cost control, in McNeil BJ, Cravalho EG (eds): Critical Issues in Medical Technology. Boston, Auburn House, 1980, pp 129-138.
2. Knaus WA, Draper EA, Wagner DP: The use of intensive care: New research initiatives and their implications for national health policy. Milbank Q
1983;61:561-583
3. Pollack MM, Ruttimann UE, Glass NL, et al: Monitoring patients in pediatric intensive care. Pediatrics 1985;76:719-724.
4. Pollack MM. Getston PR, Ruttimann UE, et al. Efficiency of intensive care. A compartive analysis of eight pediatric intensive care units. JAMA 1987; 358:1481-1486
5. Brilli RJ, Spvetz A, Branson, RD, et al. Critical care delivery in the intensive care unit: Defining clinical roles and the best practice model. Critical Care Medicine; 2001: 29 (10), 2007-2019.

1b. Opportunity for Improvement: H☐ M☐ L☐ I☐
(*There is a demonstrated performance gap - variability or overall less than optimal performance*)

1b.1 Briefly explain the benefits (improvements in quality) envisioned by use of this measure:
Use of this measure, coupled with 0335 and 0336 as balancing measures, allows for evaluation of appropriateness of resource utilization.

1b.2 Summary of Data Demonstrating Performance Gap *(Variation or overall less than optimal performance across providers)*:
[*For <u>Maintenance</u> – Descriptive statistics for performance results <u>for this measure</u> - distribution of scores for measured entities by quartile/decile, mean, median, SD, min, max, etc.*]
Adult data has found 24% of admissions to adult ICUs were for observation only in one study (1), while 77% of admissions to an adult ICU were for monitoring alone in another study (2)

Analysis of data PICUs submitting data in Q3 2011 to the VPS system revealed a of range of severity adjusted LOS from 1.71 to 4.02 days
Pediatric critical care studies found that 27% of admissions to one PICU received no benefit beyond what could be provided elsewhere (3), while an analysis of eight PICUs demonstrated varying efficiency with a potential saving of 5.1 to 17.2% of ICU days of care through earlier discharge (4).

1b.3 Citations for Data on Performance Gap: [*For <u>Maintenance</u> – Description of the data or sample for measure results reported in 1b.2 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]
1.Knaus WA, Draper EA, Wagner DP: The use of intensive care: New research initiatives and their implications for national health policy. Milbank Q
1983;61:561-583
2. Thibault GE, Mulley AG, Barnett CO, et al: Medical intensive care: Indications, interventions, and outcome. N Engl J Med 1980;302:938-942.
3. Pollack MM, Ruttimann UE, Glass NL, et al: Monitoring patients in pediatric intensive care. Pediatrics 1985;76:719-724.
4. Pollack MM. Getston PR, Ruttimann UE, et al. Efficiency of intensive care. A compartive analysis of eight pediatric intensive care units. JAMA 1987; 358:1481-1486

1b.4 Summary of Data on Disparities by Population Group: [*For Maintenance –Descriptive statistics for performance results for this measure by population group*]

Population differences have not been found to be variable in pediatric intensive
care therapies. A study examined whether medical resources and outcomes for children admitted to pediatric intensive care units differed according to race, gender, or insurance status. After adjustment for differences in illness severity, standardized mortality ratios and overall resource use were similar with regard to race, gender, and insurance status, but uninsured children had significantly shorter lengths of stay in the pediatric intensive care unit. Uninsured children also had significantly greater physiologic derangement on admission (mortality probability, 8.1%; 95% confidence interval [CI], 6.2-10.0) than did publicly insured (3.6%; 95% CI, 3.2-4.0) and commercially insured patients (3.7%; 95% CI, 3.3-4.1). Consistent with greater physiologic derangement, hospital mortality was higher among uninsured children than insured children.

1b.5 Citations for Data on Disparities Cited in 1b.4: [*For Maintenance – Description of the data or sample for measure results reported in 1b.4 including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included*]

Lopez A, Tilford J, Anand K, et al. Variation in pediatric intensive care therapies and outcomes by race, gender and insurance status. Ped Crit Care Med 2006; 7(1). 2-6.

1c. Evidence (*Measure focus is a health outcome OR meets the criteria for quantity, quality, consistency of the body of evidence.*)
Is the measure focus a health outcome?   Yes☐ No☐   <u>If not a health outcome</u>, rate the body of evidence.

Quantity: H☐ M☐ L☐ I☐   Quality: H☐ M☐ L☐ I☐   Consistency: H☐ M☐ L☐ I☐

Quantity	Quality	Consistency	Does the measure pass subcriterion1c?
M-H	M-H	M-H	Yes☐
L	M-H	M	Yes☐ IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise **No**☐
M-H	L	M-H	Yes☐ IF potential benefits to patients clearly outweigh potential harms: otherwise **No**☐
L-M-H	L-M-H	L	No ☐

Health outcome – rationale supports relationship to at least one healthcare structure, process, intervention, or service	Does the measure pass subcriterion1c? Yes☐ IF rationale supports relationship

1c.1 Structure-Process-Outcome Relationship *(Briefly state the measure focus, e.g., health outcome, intermediate clinical outcome, process, structure; then identify the appropriate links, e.g., structure-process-health outcome; process- health outcome; intermediate clinical outcome-health outcome)*:
This measure indirectly measures process (decision making related to PICU discharge) while directly measuring PICU resource utilization.

1c.2-3 Type of Evidence *(Check all that apply)*:
Selected individual studies (rather than entire body of evidence)

1c.4 Directness of Evidence to the Specified Measure *(State the central topic, population, and outcomes addressed in the body of evidence and identify any differences from the measure focus and measure target population)*:
Measurement of ICU and PICU length of stay (LOS) is was included as a measure or focus of study in over 9000 publications in a recent Pubmed search. Risk-adjustment of LOS is an established and accepted methodology.

1c.5 Quantity of Studies in the Body of Evidence *(Total number of studies, not articles)*:

1c.6 Quality of Body of Evidence (*Summarize the certainty or confidence in the estimates of benefits and harms to patients across studies in the body of evidence resulting from study factors. Please address: a) study design/flaws; b) directness/indirectness of the evidence to this measure (e.g., interventions, comparisons, outcomes assessed, population included in the evidence); and c) imprecision/wide confidence intervals due to few patients or events*): Length of stay may vary significantly based on the severity of illness of the patient at the time of admission. Failure to adjust for patient-level severity of illness may

result in inappropriate comparison.

1c.7 Consistency of Results across Studies *(Summarize the consistency of the magnitude and direction of the effect):* There are two categories of studies measuring LOS: those that use a risk-adjustment method and those that don´t. For those studies that compare the two approaches, risk adjustment is found to be an important factor. However, short of the studies that compare methodologies, there is no way to otherwise make comparisons

1c.8 Net Benefit *(Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms)*:
Risk-adjustment of LOS accounts for otherwise unexplained variation within and between centers which may result in flawed interpretations of performance.

1c.9 Grading of Strength/Quality of the Body of Evidence. Has the body of evidence been graded? No

1c.10 If body of evidence graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.11 System Used for Grading the Body of Evidence: Other

1c.12 If other, identify and describe the grading scale with definitions: per 1c.9 above no grading has been done.

1c.13 Grade Assigned to the Body of Evidence:

1c.14 Summary of Controversy/Contradictory Evidence: None

1c.15 Citations for Evidence other than Guidelines*(Guidelines addressed below)*:
Ruttimann UE, Pollack MM. Variability in duration of stay in pediatric intensive care units: A multiinstitutional study. The Journal of Pediatrics;1996:128(1), 35-43.

Straney L, Clements A, Slater A. Quantifying variation of paediatric length fo stay among intensive care units in Australia and New Zealand. Qual Saf Health Care. 2010. 19 1-5

Starney LD, Clement A, Alexander J, Slater A. Measuring efficiency in Australian and New Zealand paediatric intensive care units. Intensive Care Med. 2010; 36(8):1410-16

Niskanen M, Reinikainen M, Pettila V. Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. Intensive Care Med 2009; 35(6). 1060-7

1c.16 Quote verbatim, <u>the specific guideline recommendation</u> *(Including guideline # and/or page #)*:
N/A

1c.17 Clinical Practice Guideline Citation:

1c.18 National Guideline Clearinghouse or other URL: n/A

1c.19 Grading of Strength of Guideline Recommendation. Has the recommendation been graded? No

1c.20 If guideline recommendation graded, identify the entity that graded the evidence including balance of representation and any disclosures regarding bias:

1c.21 System Used for Grading the Strength of Guideline Recommendation: Other

1c.22 If other, identify and describe the grading scale with definitions: per 1c.19 above no grading has been done.

1c.23 Grade Assigned to the Recommendation:
1c.24 Rationale for Using this Guideline Over Others:
Based on the NQF descriptions for rating the evidence, what was the <u>developer's assessment</u> of the quantity, quality, and consistency of the body of evidence? **1c.25** Quantity: Moderate **1c.26** Quality: High**1c.27** Consistency: High
Was the threshold criterion, *Importance to Measure and Report*, met? (*1a & 1b must be rated moderate or high and 1c yes*) Yes☐ No☐ Provide rationale based on specific subcriteria:
For a new measure if the Committee votes NO, then STOP. For a measure undergoing endorsement maintenance, if the Committee votes NO because of 1b. (no opportunity for improvement), it may be considered for continued endorsement and all criteria need to be evaluated.

2. RELIABILITY & VALIDITY - SCIENTIFIC ACCEPTABILITY OF MEASURE PROPERTIES

Extent to which the measure, <u>as specified</u>, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. (<u>evaluation criteria</u>)

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate field. Supplemental materials may be referenced or attached in item 2.1. See <u>guidance on measure testing</u>.

S.1 **Measure Web Page** *(In the future, NQF will require measure stewards to provide a URL link to a web page where current detailed specifications can be obtained)*. Do you have a web page where current detailed specifications for <u>this</u> measure can be obtained? Yes

S.2 **If yes, provide web page URL:** https://portal.myvps.org/document/NQFMeasures.pdf

2a. RELIABILITY. Precise Specifications and Reliability Testing: H☐ M☐ L☐ I ☐

2a1. Precise Measure Specifications. (*The measure specifications precise and unambiguous.*)

2a1.1 **Numerator Statement** *(Brief, narrative description of the measure focus or what is being measured about the target population, e.g., cases from the target population with the target process, condition, event, or outcome)*:
Number of PICU days, PICU days = Number of days between PICU admission and PICU discharge

2a1.2 **Numerator Time Window** *(The time period in which the target process, condition, event, or outcome is eligible for inclusion)*:
Submitted quarterly for all discharges during that time period

2a1.3 **Numerator Details** *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
All patients < 18 years of age

2a1.4 **Denominator Statement** *(Brief, narrative description of the target population being measured)*:
Discharges from the PICU (including tranfers to other units) during the time period being reported

2a1.5 **Target Population Category** *(Check all the populations for which the measure is specified and tested if any)*: Children's Health

2a1.6 **Denominator Time Window** *(The time period in which cases are eligible for inclusion)*:
Submitted quarterly for all discharges during that time period

2a1.7 **Denominator Details** (*All information required to identify and calculate the target population/denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
Patient age, Date of discharge

2a1.8 Denominator Exclusions *(Brief narrative description of exclusions from the target population)*:
Patients => 18 years of age

2a1.9 Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, codes with descriptors, and/or specific data collection items/responses)*:
Patient age

2a1.10 Stratification Details/Variables *(All information required to stratify the measure results including the stratification variables, codes with descriptors, definitions, and/or specific data collection items/responses)*:
Risk-adjustment using approved severity of illness tool.

2a1.11 Risk Adjustment Type *(Select type. Provide specifications for risk stratification in 2a1.10 and for statistical model in 2a1.13)*: Statistical risk model   **2a1.12 If "Other," please describe:**

2a1.13 Statistical Risk Model and Variables *(Name the statistical method - e.g., logistic regression and list all the risk factor variables. Note - risk model development should be addressed in 2b4.)*:
Selection criteria for risk adjustment tool for pediatric ICU's:
- Tool must allow quality assessment and comparison between intensive care units, and must be widely used
- Tool must be valid and reliable for severity adjustment and measurement of quality of care provided
- Computation of mortality risk must be in the public domain (i.e. free ofcharge)
- Algorithms must receive ongoing validation and recalibration

The PRISM 3 model meets these criteria.

1. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996;24:743-52.

2a1.14-16 Detailed Risk Model Available at Web page URL (or attachment). Include coefficients, equations, codes with descriptors, definitions, and/or specific data collection items/responses. Attach documents only if they are not available on a webpage and keep attached file to 5 MB or less. NQF strongly prefers you make documents available at a Web page URL. Please supply login/password if needed:
URL
https://portal.myvps.org/document/NQFMeasures.pdf

2a1.17-18. Type of Score: Rate/proportion

2a1.19 Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*: Better quality = Score within a defined interval

2a1.20 Calculation Algorithm/Measure Logic *(Describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; aggregating data; risk adjustment; etc.)*:
Numerator of number of days between PICU admission and PICU discharge is determined.
All discharges including transfer from PICU are counted for same time period to serve as denominator.

Risk stratification addressed using PRISM 3 methodology.

PRISM 3 is a valid, realiable and internationally accepted risk measurement tool. The methodology and measure specifications have been published(1) and are available at https://portal.myvps.org/document/NQFMeasures.pdf
1. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996;24:743-52.

2a1.21-23 Calculation Algorithm/Measure Logic Diagram URL or attachment:

2a1.24 Sampling (Survey) Methodology. If measure is based on a sample (or survey), provide instructions for obtaining the sample, conducting the survey and guidance on minimum sample size (response rate):
N/A. All patients are included.

2a1.25 Data Source *(Check all the sources for which the measure is specified and tested)*. If other, please describe:
 Administrative claims, Electronic Clinical Data : Registry, Paper Records

2a1.26 Data Source/Data Collection Instrument (*Identify the specific data source/data collection instrument, e.g. name of database, clinical registry, collection instrument, etc.*): No mandatory data source or collection instrument for PICU community. Potential resources include PICU-specific databases or the VPS database (myvps.org).

Thus, 2a1.27 and 2a1.30 are not applicable

2a1.27-29 Data Source/data Collection Instrument Reference Web Page URL or Attachment:

2a1.30-32 Data Dictionary/Code Table Web Page URL or Attachment:

2a1.33 Level of Analysis (*Check the levels of analysis for which the measure is specified and tested)*:   Facility

2a1.34-35 Care Setting (*Check all the settings for which the measure is specified and tested)*: Hospital/Acute Care Facility

2a2. Reliability Testing. (*Reliability testing was conducted with appropriate method, scope, and adequate demonstration of reliability.*)

2a2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Severity adjusting of LOS is an established method which has been demonstrated to be both valid and reliable, in addition to superior to unadjusted (raw) LOS. Further measure testing is not indicated.

2a2.2 Analytic Method *(Describe method of reliability testing & rationale)*:

2a2.3 Testing Results *(Reliability statistics, assessment of adequacy in the context of norms for the test conducted)*:

2b. VALIDITY. Validity, Testing, <u>including all Threats to Validity</u>:   H☐ M☐ L☐ I☐

2b1.1 Describe how the measure specifications *(measure focus, target population, and exclusions)* **are consistent with the evidence cited in support of the measure focus (***criterion 1c)* **and identify any differences from the evidence:**
Exclusion criteria for the severity adjustment tool (PRISM 3) assures accuracy of severity adjusted LOS calculation.

2b2. Validity Testing. (*Validity testing was conducted with appropriate method, scope, and adequate demonstration of validity.)*

2b2.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Severity adjusting of LOS is an established method which has been demonstrated to be both valid and reliable, in addition to superior to unadjusted (raw) LOS. Further measure testing is not indicated.

2b2.2 Analytic Method *(Describe method of validity testing and rationale; if face validity, describe systematic assessment)*:

2b2.3 Testing Results *(Statistical results, assessment of adequacy in the context of norms for the test conducted; if face validity, describe results of systematic assessment)*:

POTENTIAL THREATS TO VALIDITY. (*All potential threats to validity were appropriately tested with adequate results.*)

2b3. Measure Exclusions. (*Exclusions were supported by the clinical evidence in 1c or appropriately tested with results demonstrating the need to specify them.*)

2b3.1 Data/Sample for analysis of exclusions *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
The endorsed severity adjustment methodology used for calculating severity adjusted LOS excludes:
• PICU patients >=18 years of age
• PICU patients under the age of 18 years with a stay < 2 hours in the PICU or < 2 consecutive sets of vital signs consistent with life
• Patients admitted to PICU for palliative care
• Preterm infants post-gestational age 36 weeks

Palliative cases are excluded because the intention is that the patient will likely die in the ICU, skewing SMR calculations if the patients are included.

The other exclusions are consistent with the use of the PRISM 3 instrument for severity adjustment. The tool has not been validated in patients <36 weeks gestation, > or equal to 18 years, or if not in the PICU at least two hours/for two vital signs to be taken.

Further validation of these exclusions is not indicated.

2b3.2 Analytic Method *(Describe type of analysis and rationale for examining exclusions, including exclusion related to patient preference)*:

2b3.3 Results *(Provide statistical results for analysis of exclusions, e.g., frequency, variability, sensitivity analyses):*

2b4. Risk Adjustment Strategy. (*For outcome measures, adjustment for differences in case mix (severity) across measured entities was appropriately tested with adequate results.*)

2b4.1 Data/Sample *(Description of the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
PRISM 3 is a valid, realiable and internationally accepted risk measurement tool. The methodology and measure specifications have been published(1) and are available at https://portal.myvps.org/document/NQFMeasures.pdf

Calibration reassessment has been performed with plans for future model enhancement

2b4.2 Analytic Method (*Describe methods and rationale for development and testing of risk model or risk stratification including selection of factors/variables)*:

2b4.3 Testing Results *(Statistical risk model: Provide quantitative assessment of relative contribution of model risk factors; risk model performance metrics including cross-validation discrimination and calibration statistics, calibration curve and risk decile plot, and assessment of adequacy in the context of norms for risk models. Risk stratification: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata)*:

2b4.4 If outcome or resource use measure is not risk adjusted, provide rationale and analyses to justify lack of adjustment:

2b5. Identification of Meaningful Differences in Performance. (*The performance measure scores were appropriately analyzed*

and discriminated meaningful differences in quality.)

2b5.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
No sampling was done. However, the data available from the VPS system reveals that the severity adjusted length of stay among 80 participating PICUs ranged from 1.71 to 4.02 days in the third quarter of 2011. This indicates that there is unit specific variance. As numerators, denominators and all definitions are standardized with an IRR >96%, this variation reflects differences in care and not the measurement itself.

2b5.2 Analytic Method *(Describe methods and rationale to identify statistically significant and practically/meaningfully differences in performance)*:

2b5.3 Results *(Provide measure performance results/scores, e.g., distribution by quartile, mean, median, SD, etc.; identification of statistically significant and meaningfully differences in performance)*:

2b6. Comparability of Multiple Data Sources/Methods. (*If specified for more than one data source, the various approaches result in comparable scores.*)

2b6.1 Data/Sample *(Describe the data or sample including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included)*:
Existing literature identifies the shortcomings of not severity adjusting LOS. Thus no attempt to report unadjusted LOS has been made.

2b6.2 Analytic Method *(Describe methods and rationale for testing comparability of scores produced by the different data sources specified in the measure)*:

2b6.3 Testing Results *(Provide statistical results, e.g., correlation statistics, comparison of rankings; assessment of adequacy in the context of norms for the test conducted)*:

2c. Disparities in Care: H☐ M☐ L☐ I ☐ NA☐ (*If applicable, the measure specifications allow identification of disparities.*)

2c.1 If measure is stratified for disparities, provide stratified results *(Scores by stratified categories/cohorts)*: N/A. This is consistent with published literature.

2c.2 If disparities have been reported/identified (e.g., in 1b), but measure is not specified to detect disparities, please explain:

2.1-2.3 Supplemental Testing Methodology Information:

Steering Committee: Overall, was the criterion, *Scientific Acceptability of Measure Properties*, met?
(*Reliability and Validity must be rated moderate or high*) Yes☐ No☐
Provide rationale based on specific subcriteria:

If the Committee votes No, STOP

3. USABILITY

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making. (**evaluation criteria**)

C.1 **Intended Purpose/ Use** *(Check all the purposes and/or uses for which the measure is intended)*:   Public Reporting, Quality Improvement (Internal to the specific organization)

3.1 **Current Use** *(Check all that apply; for any that are checked, provide the specific program information in the following questions)*:   Public Reporting, Quality Improvement with Benchmarking (external benchmarking to multiple organizations), Quality Improvement (Internal to the specific organization)

3a. Usefulness for Public Reporting: H⬜ M⬜ L⬜ I⬜
(*The measure is meaningful, understandable and useful for public reporting.*)

3a.1. **Use in Public Reporting - disclosure of performance results to the public at large** *(If used in a public reporting program, provide name of program(s), locations, Web page URL(s)).* If not publicly reported in a national or community program, state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [*For Maintenance – If not publicly reported, describe progress made toward achieving disclosure of performance results to the public at large and expected date for public reporting; provide rationale why continued endorsement should be considered.*]
We are aware of no instances of public reporting of this measure at present. There is no barrier beyond interest of national or community programs for this data.

3a.2. **Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting**. If usefulness was demonstrated (e.g., focus group, cognitive testing), describe the data, method, and results: There is potential to enhance reporting of LOS data by eliminating misleading variation caused by not accounting for patient-level severity of illness. While not immediately intuitive, the concept of risk adjustment is widely used in other settings,e.g., standardized mortality ratio, and thus there exists no reason why this measure should or could not be used for public reporting.

3.2 **Use for other Accountability Functions (payment, certification, accreditation).** If used in a public accountability program, provide name of program(s), locations, Web page URL(s): None that we are aware of.

3b. Usefulness for Quality Improvement: H⬜ M⬜ L⬜ I⬜
(*The measure is meaningful, understandable and useful for quality improvement.*)

3b.1. **Use in QI.** If used in quality improvement program, provide name of program(s), locations, Web page URL(s):
[*For Maintenance – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement*].
This approach is used internally by multiple PICUs participating in the VPS clinical database. For instance, the 72-bed PICU at the Children´s Hospital of Wisconsin trends severity-adjusted LOS as part of its improvement program

3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated *(e.g., QI initiative)*, describe the data, method and results:
Use of severity adjusted LOS for QI purposes is superior to using an unadjusted LOS because of the elimination of misleading variation caused not accounting for by patient-level severity of illness.

Overall, to what extent was the criterion, *Usability*, **met? H⬜ M⬜ L⬜ I⬜**
Provide rationale based on specific subcriteria:

4. FEASIBILITY

Extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement. **(evaluation criteria)**

4a. Data Generated as a Byproduct of Care Processes: H⬜ M⬜ L⬜ I⬜

4a.1-2 **How are the data elements needed to compute measure scores generated?** *(Check all that apply).*
Data used in the measure are:
generated by and used by healthcare personnel during the provision of care, e.g., blood pressure, lab value, medical condition, Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

4b. Electronic Sources: H☐ M☐ L☐ I☐

4b.1 Are the data elements needed for the measure as specified available electronically *(Elements that are needed to compute measure scores are in defined, computer-readable fields)*: Some data elements are in electronic sources

4b.2 If ALL data elements are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources: All necessary data may be available electronically if an organization has implemented an EHR. In the absence of an EHR, manual data collection would be required.

4c. Susceptibility to Inaccuracies, Errors, or Unintended Consequences: H☐ M☐ L☐ I☐

4c.1 Identify susceptibility to inaccuracies, errors, or unintended consequences of the measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results:
Manual data abstraction with entry into a multi-institutional clinical PICU database (the VPS (myvps.org) has been completed for the variables used in this measure since 2002. Currently, 99 hospitals and 117 PICUs are abstracting and entering data with an aggregate interrater reliability of 96.78%

4d. Data Collection Strategy/Implementation: H☐ M☐ L☐ I☐

A.2 Please check if either of the following apply (*regarding proprietary measures*):
4d.1 Describe what you have learned/modified as a result of testing and/or operational use of the measure regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues *(e.g., fees for use of proprietary measures)*:
The data necessary for capturing length of stay in relatively simple and not burdensome. Data collection for the severity adjustment component is not significant but quite feasible. For instance, the group of 99 hospitals and 117 PICUs using the VPS database have collected these elements for more than 470,000 patient encounters between 2002 and 3rd Quarter of 2011

Finally, the elements needed for determining the SMR denominator are also used in NQF measure 0343

Overall, to what extent was the criterion, *Feasibility*, met? H☐ M☐ L☐ I☐
Provide rationale based on specific subcriteria:

OVERALL SUITABILITY FOR ENDORSEMENT

Does the measure meet all the NQF criteria for endorsement? Yes☐ No☐
Rationale:

If the Committee votes No, STOP.
If the Committee votes Yes, the final recommendation is contingent on comparison to related and competing measures.

5. COMPARISON TO RELATED AND COMPETING MEASURES

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure before a final recommendation is made.

5.1 If there are related measures *(either same measure focus or target population)* or competing measures *(both the same measure focus and same target population)*, list the NQF # and title of all related and/or competing measures:
0702 : Intensive Care Unit (ICU) Length-of-Stay (LOS)

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications completely harmonized? Yes

5a.2 If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden:

5b. Competing Measure(s)

5b.1 If this measure has both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures *(e.g., a more valid or efficient way to measure quality);* OR provide a rationale for the additive value of endorsing an additional measure. *(Provide analyses when possible)*:
The measures have complementary target populations with population-specific (and appropriate) risk adjustment tools. The requirement of population-specific tools precludes use of only one measure.

CONTACT INFORMATION

Co.1 Measure Steward (Intellectual Property Owner): Virtual PICU Systems, LLC, 4470 W Sunset Blvd, Suite 440, Los Angeles, California, 90027

Co.2 Point of Contact: Christine, Gall, cgall@myvps.org, 262-439-9640-

Co.3 Measure Developer if different from Measure Steward: NACHRI (Pedi-QS), 401 Wythe Street, Alexandria, Virginia, 22314

Co.4 Point of Contact: Ellen, Schwalenstocker, PhD, eschwalenstocker@nachri.org, 703-797-6045-

Co.5 Submitter: Christine, Gall, cgall@myvps.org, 262-439-9640-, Virtual PICU Systems, LLC

Co.6 Additional organizations that sponsored/participated in measure development:
National Association of Children´s Hospitals and Related Institutions, Child Health Corporation of America, Medical Management Planning, VPS

Co.7 Public Contact: Christine, Gall, cgall@myvps.org, 262-439-9640-, Virtual PICU Systems, LLC

ADDITIONAL INFORMATION

Workgroup/Expert Panel involved in measure development
Ad.1 Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

Ad.2 If adapted, provide title of original measure, NQF # if endorsed, and measure steward. Briefly describe the reasons for adapting the original measure and any work with the original measure steward:

Measure Developer/Steward Updates and Ongoing Maintenance
Ad.3 Year the measure was first released: 2008
Ad.4 Month and Year of most recent revision:
Ad.5 What is your frequency for review/update of this measure? 3 years
Ad.6 When is the next scheduled review/update for this measure? 01, 2012

Ad.7 Copyright statement:

Ad.8 Disclaimers:

Ad.9 Additional Information/Comments:

Date of Submission (*MM/DD/YY*): 10/18/2011

EXHIBIT C
FEES OR CHARGES TO OTHER ORGANIZATIONS

VPS Participant Fee Schedule

Total Annual Unit Admissions	Annual Unit VPS Participation Fee
<500	$15,625
500-999	$18,750
1,000-1,499	$21,875
1,500-2,000	$25,000
2,000	$31,250

VPS participation fees are based upon a sliding scale of unit admissions for a calendar year, and a one-time unlimited license fee of $13,000.

For each additional unit:
- The one-time license fee for a second unit will be adjusted dependent upon the continuity of the data collection staff across both participating units:
    - Same data collection teams: $13,000 (1st unit) + $11,700 (2nd unit, 10% adjustment) = $24,700 one-time fee
    - Different data collection teams: $13,000 (1st unit) + $12,350 (2nd unit, 5% adjustment) = $25,350 one-time fee
- The ongoing annual participation fee for a second unit (whichever is smaller) will be discounted 15%

PRISM 3 data collection instructions (taken from the VPS definitions manual 1/28/12).

Timeframe for Data Collection
- The first 12 hours of admission to the ICU**.**
- If the patient does not have an ICU length of stay of 12 hours, a minimum of two hours in the ICU is required. There must be a minimum of at least 2 hours of vital signs compatible with life (e.g. the patient should not have been in a continuous state of resuscitation). See below.
- For patients that have a length of stay in the ICU of less than two hours, do not complete a PRISM 3 score.
- Only data obtained while the patient is in the ICU is utilized.

Approved use of data from the pre-ICU care time:
Data from the pre-ICU care time is not admissible with the exception of neurologic status and admission ICU labs in the following circumstances:
- If the patient was iatrogenically sedated or paralyzed during the entire ICU timeframe for data collection, use the most recent, accurate mental status assessment **_PRIOR TO_** the admission to the ICU (i.e. the coma status in the Emergency Department).
- Lab work performed in the Emergency Department just prior to ICU admission may be used.
  Lab work from the Emergency Department should not be used if the lab results were used to determine appropriate patient placement (critical care versus acute care).

Inclusion Criteria

- All ICU admissions that are included in the VPS (excludes overflow or boarder patients) with a minimum length of stay of two hours in the ICU and without the exclusion criteria below.

Exclusion Criteria

- ICU patients with lengths of stay less than two hours or who never achieve vital signs consistent with life.
- Non-ICU patients admitted to the ICU. This includes:
  - boarder/overflow patients
  - patients using the ICU as a post-operative or post-procedure recovery room
  - patients in the ICU for dialysis only
  - step-down (intermediate) patients admitted to the ICU if there is a specific institutional designation for such patients
  - patients who change to step-down (intermediate) status during their ICU stay should cease data collection when this change to intermediate status is made
- Data should not be obtained from the pre-terminal period.
  - Determining the pre-terminal period is only necessary if the patient dies during the first 12 hours of the ICU stay.
  - Often, a cardiac arrest will initiate a pre-terminal period.
  - If it is difficult to determine when the pre-terminal period begins, delete the last two to four hours of data. If this results in less than two hours of data available for evaluation, assess the information available for the first two hours of care. If this **_DOES NOT_** include an obvious pre-terminal period, use this data for data collection (i.e. delete less than the last two hours of data in order to achieve the minimum two hour timeframe for data collection) and include the patient in PRISM 3.
- Non-survivors of the ICU require at least two cardiopulmonary vital signs compatible with survival and collected during consecutive hours (in the timeframe for data collection) in order to be included in PRISM 3.*

The following criteria should be used for **exclusion of neurologic data** in the generation of a PRISM 3 score but will <u>not</u> exclude a patient from receiving a PRISM 3 score altogether:

- Patients with chronic "coma" or chronically altered mental status are excluded from collection of neurologic data in PRISM 3. Patients should only have neurologic data entered if altered mental status occurs in the ICU or due to the illness requiring ICU admission and if drugs are not a causative factor*.

* For the purposes of PRISM 3 only, collection of PRISM 3 variables will stop if the patient is downgraded to intermediate or floor status. PRISM 3 data should only be collected for patients who meet critical care criteria.

Missing Data

If data are not available for one or more variables during the timeframe of data collection, check the box to the right of the field and leave the data entry field blank.

How to Select Data for PRISM 3
- For each PRISM 3 variable select the **most abnormal values** during the timeframe of data collection.
- Cardiovascular data are collected from the cardio-respiratory vital sign flow sheet.
- Blood gases data should be collected from official laboratory reports or from respiratory flow sheets.
- Lab data (chemistry and hematology) must be collected from official laboratory reports or the laboratory records.
- Neurologic data should be collected from the neurologic vital sign sheets, relevant sections of the cardio-respiratory vital sign sheets, and/or nurse's notes.

<p align="center">Variables used to calculate the PRISM 3 risk of mortality</p>

Definitions
- **PRISM 3 12 hour score**
  The method of calculating the PRISM 3 12 hour score has been published and is available publicly. The reference is: Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996;24:743-52.

- **PRISM 3 12 hours score squared**
  The method of calculating the PRISM 3 12 hour score has been published and is available publicly. The reference is: Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996;24:743-52.

For the following questions, a response of yes = 1; no = 0 (and therefore, no responses are removed from the calculation).
- **Pre-ICU care area**
  Admitted from an inpatient care area?
  A yes response indicates that the patient came from an in inpatient area. The response to this question was designed to identify inpatients that require an upgrade in care related to deterioration. All patients coming from the operation room or the recovery room (PACU) with a prior stay on the inpatient unit should answer this question "NO". For patients transported to your facility from another inpatient facility, this includes an inpatient unit at that facility.

- **Operative status**

Is the patient post operative?
The patient is post operative if they had surgery within 24 hours before or after admission to the ICU. According to the operative status definition, patients that had a heart catherization (either interventional or diagnostic) are not considered post operative, unless they had another operative procedure.

- **Acute diagnosis of diabetes**
  Does the patient have an acute diagnosis of diabetes?
  A yes response requires that the patient was admitted for treatment of an acute phase of diabetes, namely diabetic ketoacidosis. This does not include patients with a diagnosis of diabetes that are admitted to the ICU for other reasons.

- **Pre-ICU cardiac massage**
  Was there closed or open chest cardiac massage (meaning cardiac compressions) immediately prior to this ICU admission?

- **Age**
  Was the patient a neonate?
  Neonate is defined as patients that are 28 days or less at the time of the ICU admission.

PRISM 3 Risk of Mortality calculation

Calculation of the Logit (r)
The current PRISM 3 logit algorithm used to calculate the risk of mortality includes the above variables and the y-intercept. The current coefficients as well as their positivity or negativity will be posted on the VPS website https://portal.myvps.org/document/NQFMeasures.pdf.

Calculation of the Risk of Mortality (Probability of Death)

PRISM 3 ROM (POD) = exp r / (1 + exp[r]) = $e^r / (1+e^r)$

PRISM III: An updated Pediatric Risk of Mortality score

Murray M. Pollack, MD, FCCM; Kantilal M. Patel, PhD; Urs E. Ruttimann, PhD

Objectives: The relationship between physiologic status and mortality risk should be reevaluated as new treatment protocols, therapeutic interventions, and monitoring strategies are introduced, and as patient populations change. We developed and validated a third-generation pediatric physiology-based score for mortality risk, Pediatric Risk of Mortality III (PRISM III).

Design: Prospective cohort.

Setting: There were 32 pediatric intensive care units (ICUs): 16 pediatric ICUs were randomly chosen and 16 volunteered.

Patients: Consecutive admissions at each site were included until at least 11 deaths per site occurred.

Measurements and Main Results: Physiologic data included the most abnormal values from the first 12 and the second 12 hrs of ICU stay. Outcomes and descriptive data were also collected. Physiologic variables where normal values change with age were stratified by age (neonate, infant, child, adolescent). The database was randomly split into development (90%) and validation (10%) sets. Variables and their ranges were chosen by computing the risk of death (odds ratios) relative to the midrange of survivors for each physiologic variable. Univariate and multivariate statistical procedures, including multiple logistic regression analysis, were used to develop the PRISM III score and mortality risk predictors.

Data were collected on 11,165 admissions (543 deaths). The PRISM III score has 17 physiologic variables subdivided into 26 ranges. The variables most predictive of mortality were minimum systolic blood pressure, abnormal pupillary reflexes, and stupor/coma. Other risk factors, including two acute and two chronic diagnoses, and four additional risk factors, were used in the final predictors. The PRISM III score and the additional risk factors were applied to the first 12 hrs of stay (PRISM III-12) and the first 24 hrs of stay (PRISM III-24). The Hosmer-Lemeshow chi-square goodness-of-fit evaluations demonstrated absence of significant calibration errors (p values: PRISM III-12 development = .2496; PRISM III-24 development = .1374; PRISM III-12 validation = .4168; PRISM III-24 validation = .5504). The area under the receiver operating curve and Flora's z-statistic indicated excellent discrimination and accuracy (area under the receiver operating curve – PRISM III-12 development 947 ± 0.007; PRISM III-24 development 0.958 ± 0.006; PRISM III-12 validation 0.941 ± 0.021; PRISM III-24 validation 0.944 ± 0.021; Flora's z-statistic – PRISM III-12 validation = .7479; PRISM III-24 validation = .9225), although generally, the PRISM III-24 performed better than the PRISM III-12 models. Excellent goodness-of-fit was also found for patient groups stratified by age (significance levels: PRISM III-12 = .1622; PRISM III-24 = .4137), and by diagnosis (significance levels: PRISM III-12 = .5992; PRISM III-24 = .7939).

Conclusions: PRISM III resulted in several improvements over the original PRISM. Reassessment of physiologic variables and their ranges, better age adjustment for selected variables, and additional risk factors resulted in a mortality risk model that is more accurate and discriminates better. The large number of diverse ICUs in the database indicates PRISM III is more likely to be representative of United States units. (Crit Care Med 1996; 24:743–752)

KEY WORDS: severity of illness index; mortality prediction; pediatrics; critical illness; patient outcome assessment; intensive care unit, pediatric

Severity of illness assessment has been crucial for a wide range of pediatric, neonatal, and adult intensive care unit (ICU) uses, including quality assessments, controlling for severity of illness in clinical studies, and studies of ICU resource utilization and management (1–6). Although severity of illness is a familiar medical concept, it is sometimes difficult to define. In the context of intensive care, a rational and objective way to define and quantify severity of illness is through the development of probability models predicting mortality risk (7). Such predictive models have been developed for all age groups (8–13). Future uses of outcome probabilities may even include decision-making for individual patients, if predictors achieve a sufficient level of accuracy and validity (14).

The relationship between physiologic status and mortality risk may change as new treatment protocols, therapeutic interventions, and monitoring strategies are introduced. Patient populations may also change as new therapies ameliorate the requirement for ICU care, and new patient groups may emerge, often as a result of other medical advances. Predictive models evolve as databases become larger and additional patient

The physiologic variables and their ranges, as well as diagnostic and other risk variables reflective of mortality risk, were reevaluated to update and improve the performance of the score.

characteristics can be integrated into the predictive algorithms.

The Pediatric Risk of Mortality (PRISM) is a second-generation, physiology-based predictor for pediatric ICU patients. PRISM was initially derived from the Physiologic Stability Index (8, 15). The goal of the present study was the development and validation of PRISM III, a third-generation score, based on a sample of 11,165 admissions to 32 pediatric ICUs, representing a wide diversity of organizational and structural characteristics. Specifically, the physiologic variables and their ranges, as well as diagnostic and other risk variables reflective of mortality risk, were reevaluated to update and improve the performance of the score. In addition, since minimizing the time period for assessing mortality risk is advantageous for evaluating pediatric ICU quality, we developed a 12-hr prediction model as well as a 24-hr prediction model. Concepts that guided this effort included the following: a) maximizing the predictive performance while keeping the number of variables and their ranges to a minimum, using variables that are readily available and clearly definable while maintaining the assumptions inherent in the Physiologic Stability Index and PRISM that unmeasured variables are assumed to be normal; and b) avoidance of therapeutic variables that may be unduly influenced by practice patterns.

MATERIALS AND METHODS

Study Sites. There were 32 study sites. The selection process for the first 16 units has been previously reported (1, 16). A stratified sample of pediatric

ICUs representing a broad range of organizations and structures was randomly selected based on size, unit coordination, presence or absence of a pediatric intensivist, and teaching status of the hospital. In addition, a data set from 18 volunteer units were collected in 1993 and 1994, although two units were excluded because they did not meet criteria for data reliability. The characteristics of these units are shown in Table 1.

Patients. Consecutive admissions to each pediatric ICU were included, unless they met the criteria for exclusion specified below. Readmissions to the pediatric ICU during the same hospitalization were analyzed as separate patients because each admission presented a separate opportunity for a pediatric ICU outcome. Excluded from the study were: a) admissions for recovery from procedures normally cared for in other hospital locations; b) patients staying in the ICU <2 hrs; c) patients transferred from the study pediatric ICU to another ICU because their outcome could not be clearly credited to either ICU; and d) patients admitted in a state of continuous cardiopulmonary resuscitation who never achieved stable vital signs for at least 2 hrs. If deaths occurred in the operating room, the patients were included if the operation occurred during the pediatric ICU stay and was a therapy for the illness requiring pediatric ICU care. Terminally ill patients who were transferred from the pediatric ICU for "comfort care" after discontinuation of a pediatric ICU

Table 1. Study site and patient characteristics

PICUs (n = 32)	
Beds (n)	4–40
Volume (patients/month)	14–151
No. with intensivists	25
No. with pediatric critical care training programs	8
Mortality rates (%)	2.2–16.4
Hospital Characteristics	
Pediatric beds (n)	20–325
Patient Characteristics by PICU	
Sample size (n)	110–674
Deaths (n)	11–28
Age in months (mean)	21–75
Emergency admissions (%)	43–91
Postoperative admissions (%)	16–58
Admissions from inpatient units (%)	25–72

PICU, pediatric intensive care unit.

technology (e.g., mechanical ventilation) were included as pediatric ICU patients for the 24 hrs after pediatric ICU discharge because 24 hrs is a routine observational time after technology is discontinued. Terminally ill patients transferred from the pediatric ICU for comfort care while technological support was maintained were included as pediatric ICU patients until 24 hrs after the technological support was discontinued. Patients transferred out of the pediatric ICU with technological support who were not considered terminal (e.g., chronic mechanical ventilation) were classified as survivors.

All institutions collected information on all admissions. When the last death in each pediatric ICU's sample occurred, all patients admitted before that death remained in the study. All pediatric ICUs submitted patient logs. These logs were assessed to ensure that no deaths were left out, data on at least 97% of patients were included, and none of the patients who lacked data died.

Data. In the first 16 pediatric ICUs, data collection methods were taught at site visits. In the volunteer pediatric ICUs, a video tape teaching program was used. For both groups, a detailed protocol manual was supplied. Patient data included the following information: age; gender; pediatric ICU and hospital outcomes (survival, death); admission and discharge diagnoses classified by system and etiology of disease; elective/emergency status; operative status; clinical service of primary responsibility; admission source (same hospital nursing unit, referral hospital nursing unit, home, physician office/clinic); transportation to hospital by an organized transport system (helicopter, fixed wing, ambulance, none); previous pediatric ICU admission during the current hospitalization; cardiac massage before the pediatric ICU or hospital admission; and selected critical care modalities used in the first 24 hrs of the pediatric ICU stay. In addition to the diagnostic classification using system and etiology of disease, we also investigated a more traditional diagnostic system, using the common diagnoses (asthma, pneumonia, meningitis, seizures, head trauma, other trauma, human immunodeficiency virus status, congenital heart disease, diabetes, sepsis, and

bronchopulmonary dysplasia). Diagnoses were determined from admission-day information.

Physiologic data included the most abnormal values from the first 12 hrs and the second 12 hrs of pediatric ICU stay. In the first 16 units, data collection involved obtaining photocopies of the vital sign and laboratory records, and the appropriate items were extracted at the data center. In the second group of units, the data were collected at the sites. The data consisted of the following: systolic and diastolic blood pressures; heart rate; respiratory rate; temperature (oral, axillary, or core); coma status; pupillary reactions; pupillary size and equality; concentrations of sodium, potassium, total CO_2, bicarbonate, total and direct bilirubin, total and ionized calcium, glucose, blood urea nitrogen, creatinine, and albumin; hemoglobin; white blood cell count; platelet count; prothrombin and partial thromboplastin times; pH and Pco_2 (arterial, venous, or capillary); and Pao_2 with a simultaneous Fio_2. Whole blood as well as serum/plasma measurements of sodium, potassium, and glucose were also collected. For variables where both high and low abnormalities may reflect increased mortality risk, we collected both the high and the low values. Thus, both high and low values of the same physiologic variable could contribute to severity of illness. Heart rate, respiratory rate, and blood pressure were not included at times when crying or iatrogenic agitation was noted. Physiologic data accumulated during the preterminal period in patients dying within the first 24 hrs of pediatric ICU care were not included in the study when death was obvious (usually, the last 2 to 4 hrs of life).

Since altered mental status can be influenced by a variety of iatrogenic interventions, we only considered mental status for children with known acute central nervous system disease, or where acute central nervous system disease secondary to an acute, systemic event (e.g., hypoxia, hypotension) was a possibility. In addition, we did not include mental status assessments for the 2 hrs after sedatives, paralyzing drugs, or anesthetic agents. If patients were sedated or paralyzed during the entire assessment period, the mental status assessment most proximate to pediatric

ICU admission without sedation, paralysis, or anesthesia was used (usually in the emergency department). Altered mental status was defined as a Glasgow Coma Scale score of <8, or stupor or coma.

Physiologic variables, where normal physiologic values are age dependent, were stratified into the following age groups: neonates (<1 month); infants ($\geq$1 to 12 months); children ($\geq$12 to 144 months); and adolescents (>144 months). Age-adjusted variables included the following: systolic blood pressure; diastolic blood pressure; heart rate; respiratory rate; concentrations of blood urea nitrogen, creatinine, albumin, and bilirubin; hemoglobin, prothrombin time, partial thromboplastin time, and Pao_2.

When several variables overlapped significantly in the assessment of physiologic dysfunction, we attempted to combine them into a composite variable. This approach was most pertinent for acidosis variables and clotting variables. For example, we combined pH and total CO_2 into a variable representing acidosis.

The reliability of the data collection, entry, and verification processes were formally checked by reabstracting a random selection of at least 23 cases from each institution after completion of the initial data collection. The reabstractions were subjected to the identical processes of data entry and verification, and PRISM scores were recalculated. Institutions were included if the intraclass correlation coefficient of reliability (17) for their abstraction/reabstraction of PRISM scores was >0.80, resulting in the exclusion of two of the volunteer pediatric ICUs.

Variable and Range Selection. Initially, we developed separate prediction models for two time periods, one for the first 12 hrs and one for the first 24 hrs of pediatric ICU stay. Our approach to this portion of PRISM III development assumed that deviations of physiologic variables from the midrange (40th to 60th percentiles) of survivors positively contributed to mortality risk, with larger deviations reflecting higher mortality risks. Appropriate variable ranges that significantly contributed to mortality prediction were investigated initially using univariate logistic regression analysis. The risk of death (odds ratios)

relative to the midrange of survivors was computed for each physiologic variable. Continuous physiologic variables were initially subdivided into ranges based on percentiles of survivors (5%, 10%, 20%, 30%, 40% to 60%, 70%, 80%, 90%, 95%). In some instances, the resulting cutoff points were modified based on clinical judgment. The variable ranges were absorbed into the midrange under the following conditions: a) the logistic regression coefficients of the variable ranges were not significant ($p \geq .25$) and they bordered the midrange or a range that had been combined with the midrange; or b) none of the deaths had variable values in midrange. Such a variable range was then combined with the range displaying the most similar regression coefficient. When the regression coefficients of two or more adjacent ranges were within the standard errors, the ranges were also combined.

These univariate procedures yielded 21 physiologic variables with 78 ranges for inclusion into the multivariate logistic model. Table 2 illustrates the ranges for one of the variables, systolic blood pressure. The logistic model utilized a stepwise variable inclusion procedure (18). The ranges of the predictor variables were included in the logistic regression model, one at a time, as long as the Akaike Information Criterion decreased. Subsequently, to obtain the best subset of ranges for each variable and, at the same time guard against overfitting, we employed a cross-validation by a repeated training-testing method (19). For that purpose, ten random validation samples, each consisting of 10% of the total sample, were generated without replacement. The ten complementary 90% samples served as training samples to develop the "best" model for that sample, based on the minimum Akaike Information Criterion.

Table 2. Systolic blood pressure ranges (mm Hg) selected by the univariate methods to be included in the multivariate logistic regression model

	Low			High
Neonate	<40	40–50	51–55	>125
Infant	<45	45–55	56–65	>135
Child	<55	55–65	66–75	>150
Adolescent	<65	65–75	76–85	>190

PRISM III

CARDIOVASCULAR/NEUROLOGIC VITAL SIGNS (1-6)

Systolic Blood Pressure (mm Hg)
Measurement ___

	Score=3	Score=7
Neonate	40-55	<40
Infant	45-65	<45
Child	55-75	<55
Adolescent	65-85	<65

Temperature
Measurement ___

	Score=3
All Ages	<33 °C (91.4 °F) or >40.0 °C (104.0 °F)

Mental Status
Measurement ___

	Score=5
All Ages	Stupor/Coma (GCS <8)

Heart Rate (beats per minute)
Measurement ___

	Score=3	Score=4
Neonate	215-225	>225
Infant	215-225	>225
Child	185-205	>205
Adolescent	145-155	>155

Pupillary Reflexes
Measurement ___

	Score=7	Score=11
All Ages	One fixed,	Both fixed
		one reactive

ACID-BASE/BLOOD GASES (1,2,7,8)

Acidosis (Total CO_2 (mmol/L) or pH)
Measurement ___

	Score=2	Score=6
All Ages	pH 7.0-7.28 or total CO_2 5-16.9	pH <7.0 or total CO_2 <5

pH
Measurement ___

	Score=2	Score=3
All Ages	7.48-7.55	>7.55

PCO_2 (mm Hg)
Measurement ___

	Score=1	Score=3
All Ages	50.0-75.0	>75.0

Total CO_2 (mmol/L)
Measurement ___

	Score=4
All Ages	>34.0

PaO_2 (mm Hg)
Measurement ___

	Score=3	Score=6
All Ages	42.0-49.9	<42.0

CHEMISTRY TESTS (1,2,9)

Glucose
Measurement ___

	Score=2
All ages	>200 mg/dL or >11.0 mmol/L

Potassium (mmol/L)
Measurement ___

	Score=3
All ages	>6.9

continued

PRISM III (continued)

Creatinine
Measurement ___

	Score=2
Neonate	>0.85 mg/dL or >75 μmol/L
Infant	>0.90 mg/dL or >80 μmol/L
Child	>0.90 mg/dL or >80 μmol/L
Adolescent	>1.30 mg/dL or >115 μmol/L

Blood Urea Nitrogen (BUN)
Measurement ___

	Score=3
Neonate	>11.9 mg/dL or >4.3 mmol/L
All Other Ages	>14.9 mg/dL or >5.4 mmol/L

HEMATOLOGY TESTS (1,2)

White Blood Cell Count (cells/mm³)
Measurement ___

	Score=4
All ages	<3,000

Prothrombin Time (PT) or Partial Thromboplastin Time (PTT) (seconds)
Measurement ___

	Score=3
Neonate	PT >22.0 or PTT >85.0
All Other Ages	PT >22.0 or PTT >57.0

Platelet Count (cells/mm³)
Measurement ___

	Score=2	Score=4	Score=5
All ages	100,000-200,00	50,000-99,999	<50,000

TOTAL PRISM III SCORE ___

OTHER FACTORS (10)

□ nonoperative CV disease □ chromosomal anomaly □ cancer □ previous PICU admission □ Pre-ICU CPR
□ post-operative □ acute diabetes (eg DKA) □ admission from inpatient unit (exclude post-operative patients)

Notes:
1. PRISM III mortality risk equations are available for the first 12 hours and the first 24 hours of PICU care.
2. General: Use the highest and/or the lowest values for scoring. When there are both low and high ranges, PRISM III points may be assigned for the low and the high ranges. Readmissions are included as separate patients. Exclude admissions routinely cared for in other hospital locations, staying in the PICU (< 2 hours; and those admitted in continuous CPR who do not achieve stable vital signs for ≥ 2 hours. Deaths occurring in the OR are included only if the operation occurred during the PICU stay and was a therapy for the illness requiring PICU care. Terminally ill patients transferred from the PICU for "comfort care" are included as PICU patients for the 24 hours following PICU discharge or, if receiving technologic support, until 24 hours after the technologic support is discontinued. Ages: Neonate = 0 - <1 month; Infant = ≥1 month - 12 months; Child = ≥12 months - 144 months; Adolescent >144 months.
3. Heart Rate: Do not assess during crying or iatrogenic agitation.
4. Temperature: Use rectal, oral, blood, or axillary temperatures.
5. Pupillary Reflexes: Nonreactive pupils must be >3 mm. Do not assess after iatrogenic pupillary dilatation.
6. Mental Status: Include only patients with known or suspected, acute CNS disease. Do not assess within 2 hours of sedation, paralysis, or anesthesia. If there is constant paralysis and/or sedation, use the time period without sedation, paralysis, or anesthesia closest to the PICU admission for scoring. Stupor/coma is defined as GCS score < 8 or stupor/coma using other mental status scales.
7. Acid-Base: Use calculated bicarbonate values from blood gases only if total CO_2 is not measured routinely. pH and PCO_2 may be measured from arterial, capillary, or venous sites.
8. PaO_2: Use arterial measurements only.
9. Whole Blood Corrections: Whole blood measurements should be increased as follows: glucose - 10%; sodium - 3 mmol/L; potassium - 0.4 mmol/L. (Pediatric Reference Ranges, Soldin SJ, Hicks JM eds, AACC Press, Washington, D.C., 1995).
10. Nonoperative CV disease includes acute cardiac and vascular conditions as the primary reasons for admission. Cancer and chromosomal anomalies are acute or chronic. Previous PICU admission and pre-PICU CPR refer to the current hospital admission. CPR requires cardiac massage. Post-operative is the initial 24 hours following an OR surgical procedure. Catheterizations are not post-operative. Acute diabetes includes acute manifestation of diabetes (e.g. DKA) as the primary reason for PICU admission. Admission from routine care area includes all inpatient locations except the operating or recovery rooms.

© Children's National Medical Center, May 1995

Figure 1. The Pediatric Risk of Mortality III (PRISM III) score. Numbers in parentheses refer to Notes. CV, cardiovascular; PICU, pediatric intensive care unit; ICU, intensive care unit; CPR, cardiopulmonary resuscitation; DKA, diabetic ketoacidosis; CNS, central nervous system; GCS, Glasgow Coma Scale; OR, operating room.

The associated 10% sample was used as the test sample for validating the corresponding model. For each of the ten resulting "best" models, model fit was assessed by computing the mean square deviance in both the training and the validation samples (20). The final model for the physiologic variables was selected from the ten "best" candidates as the one that displayed the maximum difference between the mean square deviance values in the associated training and validation samples, while the Hosmer-Lemeshow goodness-of-fit test was not significant at a level of $p > .10$ (20). This process maximally separated the training and validation samples with respect to the prediction performance of the model, enabling the testing of the most deviant validation sample for goodness-of-fit. After the selection of the final physiologic model, the logistic regression coefficients were scaled to yield integer scores for the individual variable ranges. The sum of these scores constitutes PRISM III.

After the development of the physiologic portion of the PRISM score, diagnostic and other risk variables were tested for effect on mortality prediction. The association of these risk variables with outcome was assessed by multivariate logistic regression analyses in the previously selected ten training samples, with PRISM III as a covariate in the model. Variable inclusion was based on minimizing the Akaike Information Criterion value in each sample. The final model was chosen from the best training sample models that included only variables selected in the majority of the training samples and yielded the highest prediction accuracy, while maintaining the Hosmer-Lemeshow goodness-of-fit test with $p > .10$ in both the training and validation samples. The goodness-of-fit test assessed model calibration, while prediction accuracy was measured by the area under the receiver operating characteristic curve (21). Model fit in the validation sample was also assessed by Flora's method (22).

Finally, the performance of the previous version of the PRISM physiology score was compared directly with PRISM III by using the variables (the physiologic score, age, and operative status) and observation period (24 hrs) as specified by the previous version of

PRISM. Improvements in the Akaike Information Criterion and the log-likelihood ratio were compared using the percentage improvement in the training set. The area under the receiver operating curve was compared in both the training and validation sets using Hanley's method (23).

Table 3. Model fit and performance measures in the training sample (n = 9,997; 483 deaths)

	χ^2 (df)	AIC	AUC (SEM)	Hosmer-Lemeshow χ^2 12 df	p Value
PRISM III-12					
PRISM III-12	1902.404 (1)	1970.878	.929 (.008)	35.877	.0003
PRISM III-12 + (PRISM III-12)^2	1910.861 (2)	1964.421	.929 (.008)	17.683	.1257
PRISM III-12 with additional variables	2023.520 (10)	1867.762	.946 (.007)	14.854	.2496
PRISM III-24					
PRISM III-24	2045.877 (1)	1827.405	.947 (.007)	39.300	.0001
PRISM III-24 + (PRISM III-24)^2	2060.464 (2)	1814.818	.947 (.007)	19.966	.0677
PRISM III-24 with additional variables	2167.509 (10)	1723.773	.958 (.006)	17.335	.1374

df, degrees of freedom; AIC, Akaike Information Criterion; AUC, area under the receiver operating characteristic curve; SEM, standard error of the mean; PRISM III-12, Pediatric Risk of Mortality III score applied to the first 12 hrs of pediatric intensive care unit stay; PRISM III-24, Pediatric Risk of Mortality III score applied to the first 24 hrs of pediatric intensive care unit stay.

Additional variables include the following: a (PRISM III)^2 term; admission for treatment of acute complications of diabetes; nonoperative cardiovascular conditions (e.g., congenital heart disease, cardiomyopathies, myocarditis, heart failure, dysrhythmias, cardiac complications of drugs, cardiogenic shock from any etiology, systemic hypertension, pulmonary hypertension, vasculitis); chromosomal anomalies; oncologic disease (acute or chronic); admission from an inpatient care area (excluding operating room or recovery room); postoperative status; previous pediatric intensive care unit admission; and preintensive care unit cardiac massage.

The Hosmer-Lemeshow chi-square values are for the 14 risk intervals given in Table 4.

Table 4. Hosmer-Lemeshow goodness-of-fit test in the training sample (n = 9,997; 483 deaths) for the model using the Pediatric Risk of Mortality III score applied to the first 12 hrs of pediatric intensive care unit stay with additional variables

Probability of Death (%)	Survivors		Deaths		Standardized Mortality Ratio (SEM)
	Expected	Observed	Expected	Observed	
0–<1	6,451.35	6,458	27.65	21	0.759 (.190)
1–<2.5	1,545.12	1,536	24.88	34	1.366 (.199)
2.5–<5	670.14	668	24.86	27	1.086 (.197)
5–<10	353.51	358	27.49	23	0.837 (.183)
10–<15	164.69	165	23.31	23	0.987 (.194)
15–<25	144.08	148	34.92	31	0.888 (.151)
25–<35	60.57	63	26.43	24	0.908 (.162)
35–<45	46.24	42	30.76	35	1.138 (.139)
45–<55	24.29	18	24.71	31	1.255 (.141)
55–<65	19.80	18	29.20	31	1.062 (.117)
65–<75	16.39	16	37.61	38	1.010 (.090)
75–<85	10.43	12	42.57	41	0.963 (.068)
85–<95	5.81	9	60.19	57	0.947 (.038)
≥95	1.59	3	68.41	67	0.979 (.018)

Standardized Mortality Ratio, observed mortality rate/expected mortality rate; SEM, standard error of the mean; χ^2, 12 degrees of freedom = 14.854, p = .2496.

The additional variables are given in Table 3.

RESULTS

Data were collected on 11,165 admissions (543 deaths). In the first data set of 16 pediatric ICUs, there were 5,415 admissions; in the second data set of 16 volunteer pediatric ICUs, there were 5,750 admissions. Population characteristics are summarized in Table 1.

Multivariate logistic regression modeling resulted in a PRISM III score based on the first 12 hrs of care, consisting of 17 physiologic variables subdivided into 26 ranges, and a PRISM III score based on the first 24 hrs of care, consisting of 17 physiologic variables subdivided into 26 ranges. The variables selected for the first 12 hrs and first 24 hrs were identical, with the exception of potassium concentration, which was included in the first 12-hr score but not in the first 24-hr score, and respiratory rate, which was included in the first 24-hr score but not in the first 12-hr score. The ranges and their relative contributions to risk prediction scores were almost identical. Subsequent analysis demonstrated similar performance between the PRISM III score specifically determined from the first 24 hrs applied to the first 12 hrs, and the PRISM III score determined from the first 12 hrs applied to the first 24 hrs. Therefore, only a single set of physiologic variables and ranges derived from the first 12 hrs was used for determining physiologic status in both the first 12 hrs and first 24 hrs of care (PRISM III, Fig. 1). The PRISM III score, when obtained from the first 12 hrs, is denoted as PRISM III-12. The PRISM III score, when obtained from the first 24 hrs, is denoted as PRISM III-24. Five physiologic variables are age adjusted. For some variables (e.g., systolic blood pressure), different physiologic ranges are used for each age group, while for other variables (e.g., partial thromboplastin time), several of the age groups share the same physiologic ranges. When both high and low ranges are included for a physiologic variable (e.g., pH), PRISM points may be assigned for both the high and the low range if abnormalities in both ranges occur. The variables representing acidosis and coagulation are composite variables, combining in an "either/or" format the most extreme deviation of either variable. This structure worked as well as more complicated variable combination schemes and was simpler to use. Data collection rules have been provided in the *Notes* at the end of Figure 1.

The variables that were most predictive of mortality, as indicated by the highest PRISM III scores, were minimum systolic blood pressure, abnormal pupillary reflexes, and stupor/coma. Variables in the original PRISM that are not included in PRISM III are diastolic blood pressure, respiratory rate, Pao_2/Fio_2, and bilirubin and calcium concentrations. Variables that are included in PRISM III but not in PRISM are temperature, pH, Pao_2, creatinine concentration, blood urea nitrogen concentration, white blood cell count, and platelet count.

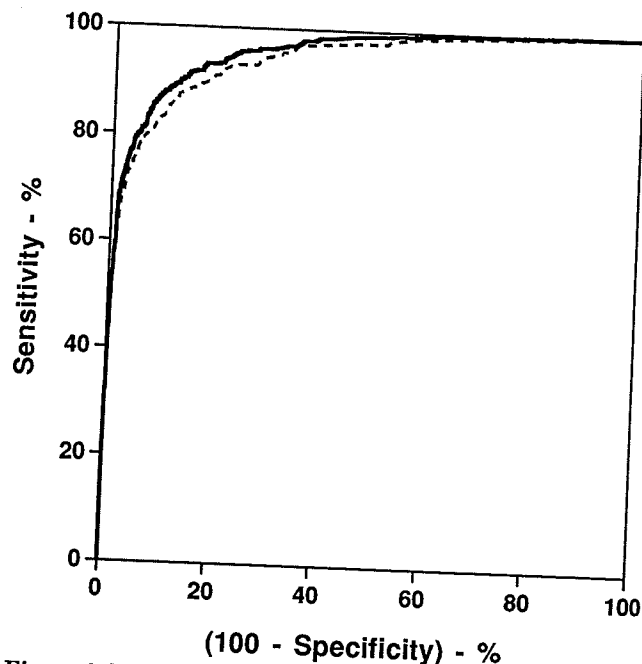After selection of the PRISM III physiologic variables and their ranges, additional predictive factors were tested for their effects on mortality prediction by building logistic regression models with either PRISM III-12 or PRISM III-24 as a covariate. This approach resulted in the

Figure 2. Observed and predicted mortality rates in risk intervals for the training sample (n = 9,997; deaths = 483). *PRISM III*, Pediatric Risk of Mortality III.

Figure 3. Receiver operating curves for the training sample (n = 9,997; deaths = 483). *Solid line*, Pediatric Risk of Mortality III score applied to the first 24 hrs of pediatric intensive care unit stay with all risk factors (area under the receiver operating curve = 0.959 ± 0.006); *dashed line*, Pediatric Risk of Mortality III score applied to the first 12 hrs of pediatric intensive care unit stay with all risk factors (area under the receiver operating curve = 0.946 ± 0.007).

748

inclusion of a PRISM III squared term (24), two acute diagnoses (diabetes and nonoperative cardiovascular disease), two diagnoses reflecting acute and chronic health status (chromosomal anomalies, oncologic disease), and four additional risk variables reflecting pre-ICU risk factors (operative status, pre-ICU care area, pre-ICU cardiac massage, and previous ICU admissions) (Table 3). Table 4 illustrates the goodness-of-fit data for the PRISM III-12 model, with all significant risk variables. Figure 2 illustrates the observed and expected mortality rates for all training models. Overall, the additional risk variables contributed 5% to the variance explained by the models, while PRISM III contributed 95%. Figure 3 shows the receiver operating characteristic curves.

The performance of the predictors in the validation sample is shown in Table 5. For all models, the Hosmer-Lemeshow chi-square and Flora's z-statistic indicated excellent fit in this independent sample, although generally, the PRISM III-24 model performed better than the PRISM III-12 model. Table 6 shows the goodness-of-fit to the validation data for the best performing model: PRISM III-24 with the additional diagnostic and risk factors. Figure 4 illustrates the observed and expected mortality rates for all of the validation models.

Two additional goodness-of-fit evaluations were done, using the total sample to assess model calibration for different patient groups. First, patients were stratified by the major diagnostic categories causing death, and the full PRISM III-12 and PRISM III-24 models were tested. In both cases, the fit was excellent (PRISM III-12: chi-square, 6 degrees of freedom = 4.576, p = .5992; PRISM III-24: chi-square, 6 degrees of freedom = 3.118, p = .7939). Table 7 shows the data for the PRISM III-12 model. The performance in the age groups was similarly tested and both full models performed well (PRISM III-12: chi-square, 4 degrees of freedom = 6.541, p = .1622; PRISM III-24: chi-square, 4 degrees of freedom = 3.944, p = .4137). The performance of the PRISM III-24 model for the different age groups is shown in Table 8.

Finally, PRISM III-24 was compared with the original PRISM, as described in the Materials and Methods section. In the training set, the Akaike Information Criterion improved by 18.4% (PRISM 2214.23; PRISM III-24 1807.749), the −2ln(likelihood ratio) improved by 24.4% (PRISM 1663.051; PRISM III 2069.533), and area under the receiver operating curve improved by 3.9% (PRISM 0.914; PRISM III 0.950, p < .0001). In the validation set, the area under the receiver operating curve was also significantly improved by 9.0% (PRISM 0.831; PRISM III 0.906, p < .0005).

Table 5. Performance measures in the validation sample (n = 1,168; 60 deaths)

	Hosmer-Lemeshow χ^2		Expected Deaths (no.)	Flora's Method		AUC (SEM)
	5 df	p		z	p	
PRISM III-12						
PRISM III-12	7.072	.2154	57.08	.517	.6048	.889 (.028)
PRISM III-12 + (PRISM III-12)^2	5.920	.3141	57.73	.396	.6919	.889 (.028)
PRISM III-12 with additional variables	4.992	.4168	58.22	.321	.7479	.941 (.021)
PRISM III-24						
PRISM III-24	3.328	.6496	59.01	.183	.8550	.897 (.027)
PRISM III-24 + (PRISM III-24)^2	3.032	.6951	59.53	.085	.9321	.897 (.027)
PRISM III-24 with additional variables	3.993	.5504	59.48	.097	.9225	.944 (.021)

df, degrees of freedom; AUC, area under the receiver operating characteristic curve; SEM, standard error of the mean; PRISM III-12, Pediatric Risk of Mortality III score applied to the first 12 hrs of pediatric intensive care unit stay; PRISM III-24, Pediatric Risk of Mortality III score applied to the first 24 hrs of pediatric intensive care unit stay.

Hosmer-Lemeshow chi-square (5 degrees of freedom) for the risk intervals: 0–<3%; ≥3%–10%; ≥10%–25%; ≥25%–45%; ≥45%–65%; and ≥65%.

Additional variables are given in Table 3.

Table 6. Goodness-of-fit test for the validation sample (n = 1,168; 60 deaths) for the model using the Pediatric Risk of Mortality III score applied to the first 24 hrs of pediatric intensive care unit stay with the additional risk variables

Probability of Death (%)	Survivors		Deaths		Standardized Mortality Ratio (SEM)
	Expected	Observed	Expected	Observed	
0.00–<0.03	942.43	939	5.57	9	1.616 (.421)
0.03–<0.10	101.17	103	5.83	4	0.686 (.401)
0.10–<0.25	41.76	41	8.24	9	1.093 (.316)
0.25–<0.45	13.30	15	7.70	6	0.779 (.285)
0.45–<0.65	6.87	6	9.13	10	1.096 (.216)
0.65–<1.00	2.99	4	23.01	22	0.956 (.067)
Total	1,108.52	1,108	59.48	60	1.009 (.090)

Standardized Mortality Ratio, observed mortality rate/expected mortality rate; SEM, standard error of the mean.

Chi-square = 3.993, degrees of freedom = 5, p = .5504. The Flora z = 0.097, p = .9225.

The additional variables are given in Table 3.

DISCUSSION

The development of PRISM III resulted in several improvements over the original PRISM. First, the physiologic variables and their ranges were reevaluated. The variables and the ranges in PRISM had been originally selected based on the subjective opinions of physicians who developed the Physiologic Stability Index. When the PRISM score was developed from these variables, objectivity was added, but a

Figure 4. Observed and predicted mortality rates in risk intervals for the validation sample (n = 1,168; deaths = 60). *PRISM III*, Pediatric Risk of Mortality III.

reevaluation of the original ranges was not undertaken. In this study, we objectively reassessed the predictive power of the physiologic variables and their ranges, eliminating some ranges that did not contribute significantly to mortality risk (e.g., high systolic blood pressure), and revising the ranges of the retained physiologic variables. Some physiologic variables have been eliminated and others—including temperature, pH, Pao_2, creatinine concentration, blood urea nitrogen concentration, white blood cell count, and platelet count—have been added. Although these are important changes, the variables with the greatest importance in outcome prediction are the same in both scores: low systolic blood pressure, altered mental status, and abnormal pupillary reflexes.

Second, age issues, clear data collection instructions, precise variable definitions, and strict rules for patient inclusions and exclusions were addressed at the outset of this study. While age was included as an explicit variable in the original PRISM score, it is included in the PRISM III score in a logically and clinically more convincing form by using appropriate age-adjusted physiologic variable ranges. Subsequent model fit evaluations demonstrated the success of these adjustments. A formal operational method for assessing mental status also was established to account for the frequent use of sedation and paralysis. Other variables included in the prediction model are better defined, making the score less vulnerable to "gaming." Two diagnostic entities, chromosomal abnormalities and oncologic disease, reflect underlying health status as well as acute disease status. Two acute

Table 7. Hosmer-Lemeshow goodness-of-fit assessment in the total sample, stratified by diagnoses, for the model using the Pediatric Risk of Mortality III score applied to the first 12 hrs of pediatric intensive care unit stay with the additional variables

Diagnosis	Survivors		Deaths		Standardized Mortality Ratio (SEM)
	Expected	Observed	Expected	Observed	
Cardiac—nonoperative	563.07	561	69.93	72	1.030 (.080)
Cardiac—operative	1,016.34	1,016	52.66	53	1.006 (.117)
CNS infections	341.40	340	13.60	15	1.103 (.214)
HIE	60.56	61	19.44	19	0.977 (.106)
Head trauma	1,058.49	1,050	110.51	119	1.077 (.052)
Respiratory infections	984.68	981	28.32	32	1.130 (.155)
Sepsis	202.19	196	45.81	52	1.135 (.092)
Miscellaneous	6,396.75	6,417	201.25	181	0.899 (.053)
Total	10,623.49	10,622	541.51	543	1.003 (.030)

Standardized Mortality Ratio, observed mortality rate/expected mortality rate; SEM, standard error of the mean; CNS, central nervous system; HIE, hypoxic-ischemic encephalopathy.

The total chi-square = 4.576 with 6 degrees of freedom (p = .5992) and Flora's z = 0.091 (p = .9274).

diagnoses include nonoperative cardiovascular disease and acute diabetes (primarily diabetic ketoacidosis). Other risk factors include operative status, pre-ICU care area, pre-ICU cardiac massage, and previous ICU admission.

Third, the relationship between physiologic status, as measured by PRISM III, and outcomes has been calibrated to a contemporary, well-defined, large reference sample. The set of 32 pediatric ICUs represents about 10% of all pediatric ICUs in the

United States. These units encompass a wide diversity of organizational structure and patient mixes. This diversity makes the sample sufficiently representative for most units, enabling PRISM III to be used in the comparative assessment of pediatric ICU outcomes in essentially all pediatric ICUs.

Our method of developing the PRISM III models continued the evolution toward a parsimonious predictor. The Physiologic Stability Index incorporated 102 discrete physiologic ranges of 34 physiologic variables

Table 8. Hosmer-Lemeshow goodness-of-fit assessment in the total sample, stratified by age, for the model using the Pediatric Risk of Mortality III score applied to the first 24 hrs of pediatric intensive care unit stay with the additional variables

Age Group (months)	Survivors		Deaths		Standardized Mortality Ratio (SEM)
	Expected	Observed	Expected	Observed	
0–<1	547.94	551	55.06	52	0.944 (.096)
1–<12	2,628.24	2,614	157.76	172	1.090 (.052)
12–<36	2,195.05	2,203	101.95	94	0.922 (.070)
36–<72	1,490.08	1,489	67.92	69	1.016 (.082)
72–≤144	1,921.01	1,915	78.99	85	1.076 (.076)
>144	1,840.16	1,850	80.84	71	0.878 (.078)

Standardized Mortality Ratio, observed mortality rate/expected mortality rate; SEM, standard error of the mean.

The total chi-square with 4 degrees of freedom = 3.944 (p = .4137) and Flora's z = 0.030 (p = .9758).

The additional variables are given in Table 3.

selected by physicians for their clinical importance. PRISM reduced the number of physiologic variables to 14 and their ranges to 34. While PRISM III added several new variables, the total number of ranges was reduced. Differences in the frequency of measuring variables associated with individual pediatric ICUs are unlikely to influence the reliability or accuracy of PRISM III (25). An alternative approach of including more physiologic ranges could have been accomplished by applying less strict statistical criteria for variable and range inclusion. However, this approach may have increased the variability of the predictor, decreasing the power of detecting truly existing differences from the expected mortality rates. More importantly, it could produce a biased ("overfitted") model that might perform very well in the training sample but poorly in an independent sample by incorporating idiosyncrasies of the training sample, and thus, may be biased. The excellent performance in the training sample may generate an unjustified confidence in the predictor's prediction accuracy.

Overall, all PRISM III prediction models were accurately calibrated and achieved good discrimination. The PRISM III-24 model with the diagnostic and other risk variables performed best. This result was expected, since PRISM III-24 incorporates the most information over the longest time period. However, the other models also performed very well and are suitable for quality assessment. We recommend using the PRISM III models with the additional variables since these models may increase the applicability to a wider variety of case-mix samples. The use of the PRISM III-12 model is appealing for quality assessments since, by shortening data acquisition time, it better separates the observation from the treatment period, while the PRISM III-24 model is more accurate for individual patient mortality risk assessments.

As expected, PRISM III performed better than PRISM, even when limited to the variables originally included in PRISM. The improvement in the area under the receiver operating curve was similar to the improvement seen with more recent versions of adult severity scores compared with their previous versions (26). Newer versions of severity of illness scores, such as PRISM III, will need revisions and recalibrations to maintain their relevance to contemporary patient populations.

REFERENCES

1. Pollack MM, Cuerdon TC, Patel KM, et al: Impact of quality-of-care factors on pediatric intensive care unit mortality. *JAMA* 1994; 272:941–946
2. Pollack MM, Getson PR, Ruttimann UE, et al: Efficiency intensive care. A comparative analysis of eight pediatric intensive care units. *JAMA* 1987; 258:1481–1486

3. Knaus WA, Wagner DP, Zimmerman JE, et al: Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993; 118:753–761

4. Shortell SM, Zimmerman JE, Rousseau D, et al: The performance of intensive care units: Does management make a difference? *Med Care* 1994; 32:508–525

5. Rapoport J, Teres D, Lemeshow S, et al: A method for assessing the clinical performance and cost-effectiveness of intensive care units: A multicenter inception cohort study. *Crit Care Med* 1994; 22:1385–1391

6. Pollack MM, Alexander SR, Clarke N, et al: Improved outcomes from tertiary center, pediatric intensive care: A statewide comparison of tertiary and non-tertiary care facilities. *Crit Care Med* 1991; 19:150–159

7. Lemeshow S, Le Gall JR: Modelling the severity of illness of ICU patients. *JAMA* 1994; 272:1049–1055

8. Pollack MM, Ruttimann UE, Getson PR: The Pediatric Risk of Mortality (PRISM) Score. *Crit Care Med* 1988; 16:1110–1116

9. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993; 270:2957–2963

10. Lemeshow S, Teres D, Klar J, et al: Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993; 270:2478–2486

11. Knaus WA, Wagner DP, Draper EA, et al: The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; 100:1619–1636

12. Richardson DK, Gray JE, McCormick MC, et al: Score for neonatal acute physiology: A physiologic severity index for neonatal intensive care. *Pediatrics* 1993; 91:617–623

13. Tarnow-Mordi W, Ogston S, Wilkinson AR, et al: Predicting death from initial disease severity in very low birth weight infants: A method for comparing the performance of neonatal units. *BMJ* 1990; 300:1611–1614

14. Knaus WA, Wagner DP, Lynn J: Short-term mortality predictions for critically ill hospitalized adults: Science and ethics. *Science* 1991; 254:389–394

15. Pollack MM, Ruttimann UE, Getson PR, et al: Accurate prediction of pediatric intensive care outcome: A new quantitative method. *N Engl J Med* 1987; 316:134–139

16. Pollack MM, Cuerdon TC, Getson PR: Pediatric intensive care units: Results of a national study. *Crit Care Med* 1993; 21:607–614

17. Fleiss JL: The Design and Analysis of Clinical Experiments. New York, John Wiley & Sons, 1986, pp 2–3

18. SAS/STAT User's Guide, Version 6. Fourth Edition. Volume 2. Cary, NC, SAS Institute, pp 1071–1126

19. Zhang P: Model selection via multifold cross validation. *Ann Statistics* 1993; 21:299–313

20. Hosmer DW, Lemeshow S: Applied Logistic Regression. New York, John Wiley & Sons, 1989

21. Ruttimann UE: Statistical approaches to development and validation of predictive instruments. *Crit Care Clin* 1994; 10:19–35

22. Flora JD: A method for comparing survival of burn patients to a standard survival curve. *J Trauma* 1978; 18:701–705

23. Hanley JA, McNeil BJ: A method of comparing the areas under the receiver operating characteristic curves derived from the same cases. *Radiology* 1983; 148:839–843

24. Kay R, Little S: Transformations of the explanatory variables in the logistic regression model for binary data. *Biometrika* 1987; 74:495–501

25. Pollack MM, Patel KM, Ruttimann UE, et al: Frequency of variable measurement in 16 pediatric intensive care units: Influence on accuracy and potential for bias in severity of illness assessment. *Crit Care Med* 1996; 24:74–77

26. Xavier C, Artigas A, Bion J, et al: A comparison of severity of illness scoring systems for intensive care unit patients: Results of a multicenter, multinational study. *Crit Care Med* 1995; 23:1327–1335

Length of stay and efficiency in pediatric intensive care units

Urs E. Ruttimann, PhD,[†] Kantilal M. Patel, PhD, and Murray M. Pollack, MD

Objective: Assessment of pediatric intensive care unit (PICU) efficiency with a length of stay prediction model and validation of this assessment by an efficiency measure based on daily use of intensive care unit-specific therapies.

Design: Inception cohort study of data acquired between 1989 and 1994.

Setting: Thirty-two PICUs, 16 selected randomly and 16 volunteering.

Subjects: Consecutive admissions of 10,658 patients (466 deaths) who stayed at least 2 hours and up to 12 days in the PICU.

Measurements: Length of stay and its prediction from a model with admission day data (PRISM III-24, diagnostic factors, mechanical ventilation). For validation, 11 PICUs recorded each patient's "efficient" days, that is, days when at least one PICU-specific therapy was given. PICU efficiency was computed as either the ratio of the observed efficient days or the days accounted for by the predictor variables to the total care days, and the agreement was assessed by Spearman's rank correlation analysis.

Results: The total care days provided by each PICU (n = 32) were well predicted by the length of stay model (r = 0.946). The agreement in 11 validation PICUs between therapy-based efficiency (range 0.30 to 0.67) and predictor-based efficiency (range 0.31 to 0.63) was excellent (rank correlation r = 0.936, p < 0.0001).

Conclusion: PICU efficiency comparisons with either method are nearly equivalent. Predictor-based efficiency has the advantage that it can be computed from admission day data only. (J Pediatr 1998;133:79-85)

Tracking resource use of patients in intensive care units is an important aspect of ICU and hospital management. Two methods are currently in use for assessing the appropriateness of resource use. The first is an efficiency evaluation based on the criterion whether on each day a patient used at least one therapy that is best delivered in the ICU.[1] By this criterion each patient's care day was categorized as "necessary" or "unnecessary." Unit efficiency is defined as the ratio of the total number of care days ICU-specific therapy was given divided by the total number of patient care days provided. The second method for comparing resource use among hospitals is by length of ICU stay, with appropriate adjustments made for diagnoses, severity of illness, and other case-mix variables. Studies of adult and pediatric ICUs found large interinstitutional differences of case-mix adjusted length of stay.[2,3]

LOS	Length of stay
(P)ICU	(Pediatric) intensive care unit
PRISM III-24	Pediatric risk of mortality, version III, 24 hour assessment

There are advantages and drawbacks associated with the use of either efficiency evaluation method. The therapy-based measure has the advantage of face validity and conceptual appeal, because criteria for classifying patients requiring intensive care are clear and in most cases uncontroversial. The main disadvantage is that it requires daily collection of data on therapies provided for each patient. Also, it could be misused by implying that patients classified as "inefficient consumers" did not benefit from intensive care. The main advantage of the case-mix adjusted LOS approach is that only data from the admission day are required to adjust for patient-specific needs, making it relatively easy and inexpensive to use. However, its shortcoming is that it has never been validated against other measures of efficient resource use. Vali-

Table I. Institution characteristics

	Validation PICUs (n = 11)		Other PICUs (n = 21)	
	Median	Range	Median	Range
Sample size	342	180-616	270	104-667
Mortality rate (%)	4.6	2.1-6.7	5.0	2.3-17.3
Mean care days/patient	2.60	1.91-2.95	2.43	1.52-3.34
Emergency admissions (%)	60	47-81	73	42-91
Postoperative admissions (%)	38	26-57	36	16-60
From in-hospital sites (%)	60	40-74	53	24-89

Table II. Generalized linear regression model (inverse Gaussian) for length of stay (n = 9558)

Variable	Length of stay ratio*	95% Confidence interval	p Value†
PRISM III-24	‡	‡	0.0001
(PRISM III-24)**2	‡	‡	0.0001
Primary diagnoses			
CNS infections	1.41	1.28-1.56	0.0001
Neoplastic diseases	1.22	1.13-1.31	0.0001
Asthma	0.91	0.85-0.96	0.0045
Pneumonia	1.50	1.40-1.61	0.0001
Drug overdoses	0.74	0.70-0.79	0.0001
CV nonoperative	1.22	1.14-1.32	0.0001
CV operative	0.89	0.83-0.95	0.0006
Diabetes	0.74	0.67-0.81	0.0001
Admission specifications			
Postoperative	0.92	0.88-0.96	0.0004
Inpatient	1.17	1.13-1.22	0.0001
Previous ICU admission	1.26	1.15-1.38	0.0001
Therapy			
Mechanical ventilation	1.68	1.60-1.77	0.0001
Model intercept (± SEM) = 1.423 ± 0.021 days			

CNS, Central nervous system; CV, cardiovascular system.
*Effect of the variable after adjusting for the effects of all other variables in the model.
†Log-likelihood ratio compared with the chi-squared distribution with 1 degree of freedom.
‡See Fig. 2.
Model fit: Scaled deviance = 9558 (chi-square with 9543 degrees of freedom, $p > 0.45$). Observed versus predicted length of stay, mean (± SEM) in: training sample (n = 9,558): 2.351 (± 0.032) versus 2.360 (± 0.011), $p > 0.64$; test sample (n = 1,100): 2.461 (± 0.069) versus 2.419 (± 0.035), $p > 0.49$.

dation is especially important, because a large percentage of the variability in LOS is not explained by severity-adjustment models.

This study investigated the agreement of the two resource use assessment methods in a sample of pediatric ICUs where data were available to compute efficiency measures by both methods. Hence, the hypothesis was that identical rankings of pediatric ICU efficiencies are obtained when assessed by either a case-mix adjusted prediction model with admission-day observations only or by the daily monitoring of the use of PICU-specific therapies.

METHODS

Study Sites

Initially, data from 34 institutions (16 randomly selected, 18 volunteering) were obtained.[4-6] Two of the volunteering institutions were not included in this sample because their data did not meet preset standards of reliability (see Data

section in following text). The randomly selected and the volunteering units were comparable with respect to nursing and physician staffs and PICU and hospital characteristics.[4] Eleven of the PICUs were used to validate sample the case-mix predicted with the observed efficient care days. A summary of characteristics relevant to this investigation are shown in Table I.

Patients

All consecutively admitted patients were included, unless they met the following exclusion criteria: (1) admissions for recovery from procedures normally cared for in other hospital locations, (2) patients staying in the PICU <2 hours, (3) patients transferred from the study's PICU to another PICU, and (4) patients admitted in a state of continuous cardiopulmonary resuscitation who never achieved stable vital signs for at least 2 hours.

Data

Patient information consisted of dates and times of PICU admission and discharge, age, sex, PICU and hospital outcomes (survival, death), diagnosis based on data available within the first 24 hours, elective/emergency status, operative status, clinical service of primary responsibility, preadmission care area, mode of transportation to the hospital, previous PICU admission during the current hospitalization, cardiac massage before hospital admission, and the use of selected critical care modalities including mechanical ventilation and vasoactive agent infusion during the first 24 hours of care. Physiologic data collected were those required for the PRISM III score[6] and included the most abnormal values observed during the first 24 hours of the PICU stay. In addition to these admission-day data, 11 institutions collected daily data on PICU therapies given during each patient's stay in the unit.

All sites were required to reabstract data from 23 to 30 randomly selected patients so that their PRISM scores could be recomputed. Only institutions for which the intraclass coefficient of reliability[7] of their abstracted versus reabstracted PRISM scores was >0.80 were

used in this study, leaving 32 institutions for further analysis.

Admission diagnoses were classified by the physiologic system of primary dysfunction (18 classes; e.g., neurologic, respiratory) and by the cause of primary dysfunction (18 classes; e.g., congenital, neoplastic).[3,8] Only combinations of the system and cause classifications with >1% of the patients were retained as separate entities. Small categories sharing some similarities were combined for the purpose of meeting this minimum size, and the remaining small-size categories were combined into a miscellaneous group, resulting in 21 diagnostic groups. Common clinical names were assigned to these diagnostic groups; for example, the combinations of respiratory system diseases with the cause groups infection or immunology/allergy were labeled pneumonia or asthma, respectively.

Efficiency

PICU days were defined efficient if a patient required at least one therapy that is best given in an ICU setting. Mechanical ventilation and vasoactive agent infusions were used in >95% of the efficient care days in a previous study,[1] and thus these therapies were considered ICU-specific in this study. The total number of efficient days divided by the total number of care days provided by a unit served as a therapy-based measure of PICU efficiency.[1] As an alternative, an efficiency assessment based on an LOS prediction model was defined. Such a model projects for each patient an expected LOS from admission-day data. Only the part of this predicted LOS accounted for by patient-specific predictor variables was considered efficient. Hence, model-based efficiency was computed for each unit by dividing the sum of the predicted efficient PICU days by the total number of care days provided. Agreement between the two efficiency assessment methods was evaluated by applying each to the same set of patients in 11 PICUs.

Statistical Analyses

LENGTH OF STAY PREDICTION. The availability of a newly updated pediatric
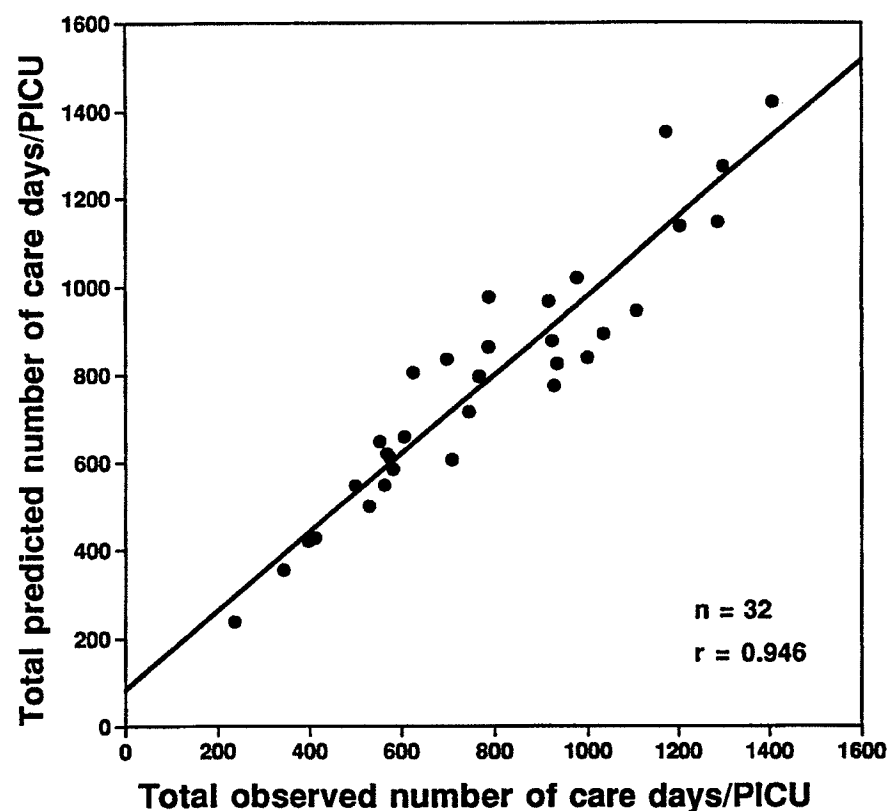
Fig. 1. Comparison of total predicted versus observed number of care days for each PICU. Predicted length of stay was computed for each patient by generalized linear regression model in Table II.

severity-of-illness assessment system (PRISM III-24) necessitated the development of a new LOS prediction model. In this new model LOS is assumed to be inversely Gaussian distributed, rendering it better suited to compare mean LOS among institutions than a previously developed predictor.[3] As in the previous model,[3] patients staying longer than the upper 95th percentile of the LOS distribution (>12 days) were excluded from the model building process to prevent a few patients with atypically long stays from exerting undue influence on the predictor coefficients.

The LOS predictor was developed by fitting a generalized linear regression model[9] to the observed LOS data with the log link function.[10] Because potential predictors were considered PRISM III-24, the 21 primary admission diagnoses, preadmission sites and preadmission care (pre-ICU cardiopulmonary resuscitation, previous PICU admission during the current hospitalization), age, sex,

and ICU-specific therapies required during the first 24 hours were used. The total database was randomly split into a 90% training sample for model development and a 10% validation sample. Predictor variables were included in the model if they improved the fit with $p < 0.05$ (log-likelihood ratio test). Goodness of fit of the final model was assessed by testing the scaled deviance from the training data (chi-squared test) and paired t tests comparing the predicted with the observed LOS for each patient (in both split samples separately). Lack of fit was considered statistically significant for $p < 0.05$.

COMPARISON OF OBSERVED EFFICIENT WITH PREDICTED EFFICIENT CARE DAYS. This comparison was accomplished by investigating the agreement between the predicted efficiency of each unit (with the LOS model) and the therapy-based efficiency of the same unit. The association of the mean predicted
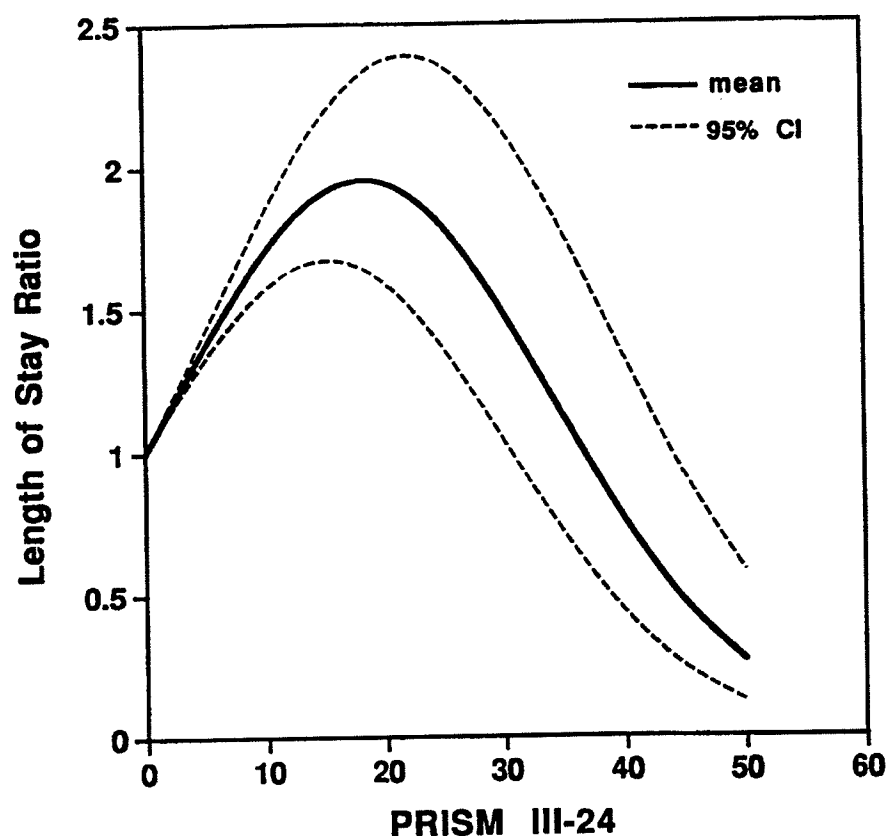
81

Fig. 2. Effect of severity of illness assessed by PRISM III-24 scores on length of stay relative to mean (± SEM) stay of 1.63 (± 0.02) days at score 0. *Broken lines* indicate 95% confidence limits.

each PICU the total number of care days is shown in Fig. 1, demonstrating a very high correlation (r = 0.946) with the observed number of care days.

Fig. 2 illustrates the effect of severity of illness in terms of PRISM III scores on LOS. PRISM III-24 has a biphasic effect on the LOS ratio, leading to an increase in the expected LOS up to a score of 18 and, thereafter, a decrease caused by a progressively larger number of early deaths as disease severity increases further. Patients with PRISM III-24 scores between 10 and 30 stayed between 60% to 94% longer compared with the mean (± SEM) stay of 1.63 (± 0.02) days of patients with a score of 0.

The effects of other patient-related factors on LOS are summarized in Table II in terms of their adjusted odds ratios. The adjusted odds ratio for each factor provides the amount by which the LOS is to be multiplied if the factor is present while all other factors remain fixed. Hence, it quantifies the relative influence of each factor on LOS in the presence of the simultaneous effects of all other factors. The intercept of the regression was 1.42 ± 0.02 days; that is, patients had an average LOS of approximately 1.4 days that was not accounted for by PRISM III-24 or the case-mix factors.

LOS with the mean efficient days in each unit was tested by linear regression analysis and nonparametrically by Spearman's rank correlation. A value of p < 0.05 was considered to indicate a significant association.

RESULTS

The 32 institutions yielded a database of 11,165 patients, of whom 10,658 (466 deaths) stayed less or equal to 12 days (≤95th percentile) in intensive care. This set broke down into 4080 patients with 167 (4.1%) deaths in the 11 validation units and 6578 patients with 299 (4.5%) in the other 21 PICUs. Table I shows the population characteristics of the individual PICUs. There were no significant group differences with respect to the average population characteristics among the 11 validation units and the 21 other units. However, there was a substantial variation (p < 0.001)

of mortality rates, emergency admissions, postoperative admissions, and admissions from other inpatient units among the individual hospitals within each group.

Predicted Length of Stay

A regression model with PRISM III-24 and case-mix factors to predict LOS for each patient was fitted to the data of 9558 patients in a 90% training sample, and its performance was evaluated in the remaining 1100 patients from the 10% test sample. The variables and factors found to be significantly associated with LOS are listed in Table II. Model performance criteria shown at the bottom of Table II indicate good data fit (scaled deviance, p > 0.45) and unbiased prediction of the mean LOS in both the training (p > 0.64) and the test samples (p > 0.49). The application of the predictor model to the total sample achieved a reduction in the variance of LOS by r^2 = 0.22. Performance of the model in predicting for

Qualitatively, similar LOS increases or decreases were predicted for comparable diagnostic groups by this new predictor and by a previous model.[3] For example, the LOS ratios predicted by the new versus the old model were for neoplastic diseases 1.22 versus 1.26, asthma 0.91 versus 0.89, pneumonia 1.50 versus 1.26, and drug overdoses 0.74 versus 0.84. However, a rigorous quantitative comparison cannot be made, because the previous predictions were made from a now outdated severity scores, which cannot be recomputed from the present PRISM III-24 scores. The new predictor is simpler in that the number of significant diagnostic groups has been reduced from 10 to 8, and interaction terms of certain diagnoses with the PRISM score are no longer required. Among the primary diagnoses, pneumonia, central nervous system infections, nonoperative cardiovascular diagnoses, and neoplastic diseases were all associated with increased

(22% to 50%) expected LOS. Patients with drug overdoses, diabetes, asthma, and operative cardiovascular conditions stayed for a significantly shorter time (–9% to –26%). Some general admission conditions (admission from an inpatient unit and previous PICU admissions) had longer expected stays (17% and 26%), whereas a postoperative status was associated with shorter (–8%) stays. Patients requiring mechanical ventilation at any time during their first 24 hours in the PICU stayed on average 68% longer than patients who did not undergo ventilation.

Efficiency

The total number of care days provided by each of the 11 PICUs in the validation sample ranged between 539 and 1407 days (median 818 days), whereas the total predicted care days ranged between 508 and 1425 days (median 870 days), well in agreement. The corresponding mean predicted care days (total care days/number of patients) provided by each PICU varied between 2.15 and 2.87 days. The intercept in the prediction model indicated that each PICU provided a mean of 1.42 care days that could not be accounted for by severity, diagnoses, and other case-mix factors. Subtraction of this unexplained (i.e., inefficient) stay from the mean predicted LOS resulted in the mean predicted efficient stay for each PICU. This mean predicted efficient stay/PICU ranged from 0.73 to 1.45 days, shown along the vertical axis in Fig. 3. Addressing the efficiency assessment based on therapy use, the total number of efficient days in these 11 units ranged between 285 and 610 days (median 378 days). The corresponding mean observed efficient days/PICU (total number of efficient days/number of patients) varied between 0.55 and 1.71 days, which are plotted along the horizontal axis. Linear regression analysis (regression line shown in the graph) yielded a highly significant ($p < 0.0001$) correlation with a coefficient of $r = 0.959$. Hence, there was a very strong association between mean predicted efficient days with the LOS model and mean observed efficient care days with therapy-based criteria. However, because a measure of efficiency on an absolute scale
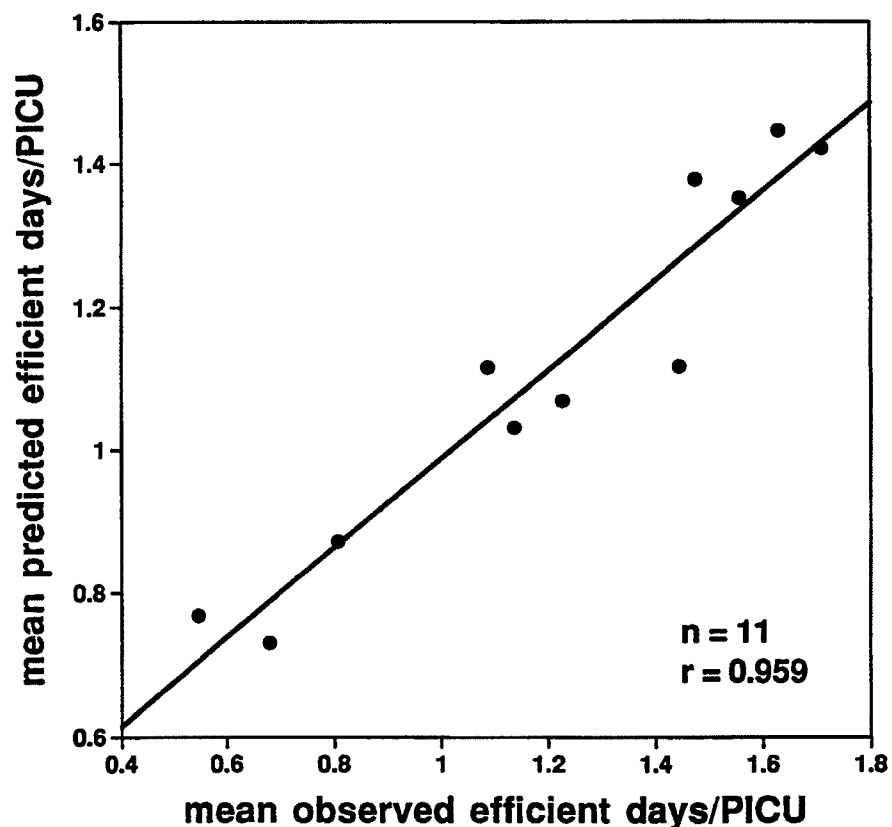
Fig. 3. Relation between mean efficient care days predicted by patient-mix factors in length-of-stay regression model and mean efficient care days when ICU-specific therapies were given. Shown are data from 11 validation PICUs and linear least-square regression line.

does not exist, interhospital comparisons should be performed only on a relative basis; that is, only a ranking of the units in terms of their efficiency assessment should be used. Therefore it is important that the efficiency ranking obtained by either method is maintained. Addressing this point, the ranking orders of the PICUs in terms of mean efficient days were well replicated by the two methods (Spearman's rank correlation coefficient of $r_s = 0.945, p < 0.0001$). There was only one disagreement in the ranks by a difference of two, whereas all other rankings either agreed or were off by only one.

Finally, therapy-based efficiency has been defined as the ratio of the total observed efficient care days relative to the total provided care days.[1] This is equivalent to the ratio of (mean observed efficient days)/(mean observed LOS). In the same way model-predicted efficiency can be defined as the ratio of total pre-

dicted efficient care days divided by total observed care days, or equivalently, as (mean predicted efficient days)/(mean observed LOS). Therapy-based efficiency ranged between 0.30 and 0.67 compared with the model-based efficiency ranging from 0.31 to 0.63. Fig. 4 demonstrates the high correlation ($r = 0.927, p < 0.0001$) between the two assessments. Of more important practical relevance, Spearman's rank correlation coefficient of $r_s = 0.936$ ($p < 0.0001$) indicates an excellent agreement between the two efficiency rankings; that is, they may indeed be used interchangeably.

DISCUSSION

Traditionally, medicine has focused on quality without regard to cost. As the growth of health care expenditures has risen relative to the gross domestic prod-

83
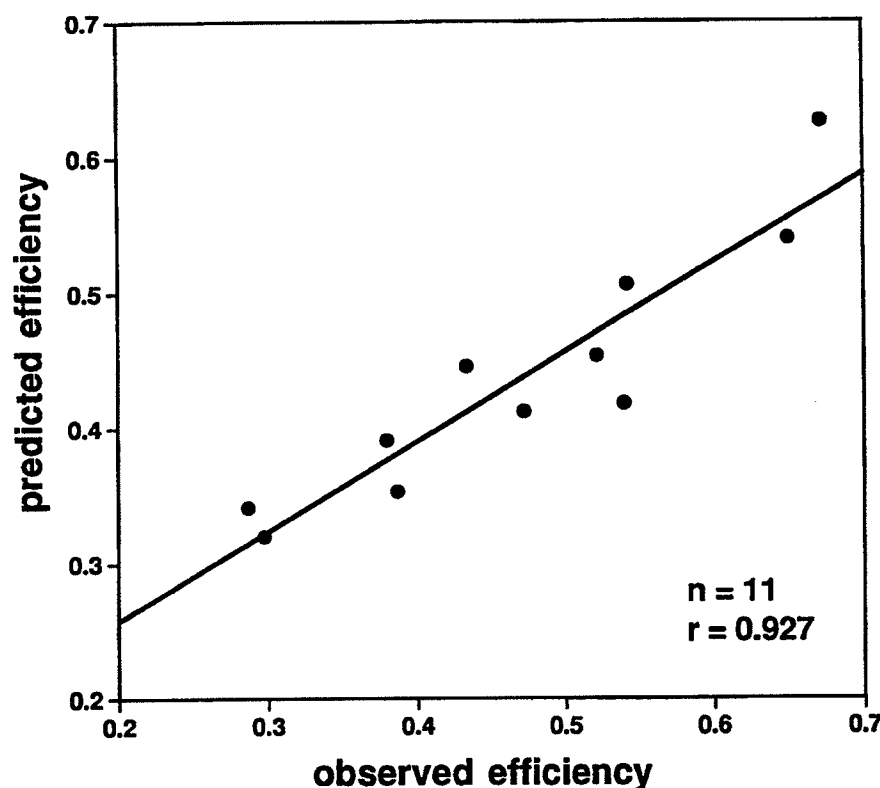
Fig. 4. Comparison of predicted (model based) efficiency with observed (based on therapy use) efficiency in 11 validation PICUs. *Line* indicates linear least-square regression.

uct, efforts have focused on finding areas that are suitable for cost reduction without reducing quality. The ICU has been one area that has received attention because of its relatively high costs; although ICU beds are <10% of the inpatient beds, they account for approximately 22% of the hospital costs.[11]

There are reasons to believe that there are too many PICUs and PICU beds, and this could result in overuse or inefficient use of these beds. Despite the declining number of pediatric inpatient facilities and beds, the number of PICUs and the number of PICU beds increased in the 1980s by 59% and 76%, respectively.[12] Groeger et al.[13] found that the use of PICUs was lowest among all types of ICUs. When there are too many ICU beds, the severity of admitted patients may decline.[14]

The goal of this analysis was the validation of an efficiency assessment based on case-mix adjusted LOS with efficiency based on daily use of PICU specific

therapy. There was excellent agreement between the two methods; both the number of efficient days predicted by the model and the predicted efficiency correlated highly with the respective computations based on daily therapy use. The LOS projected by the model has two main components; the first is accounted for by the regression coefficients of the explanatory variables, and the second is the model intercept, which cannot be accounted for by the predictor variables. Hence, the first component "explained" the effects of severity of illness and specific patient factors, whereas the unaccounted component (1.4 days per patient) was considered inefficient stay. The interpretation of the accounted part of the stay as efficient was validated by comparing it with the number of days PICU-specific therapies were given in 11 units of the sample. Very strong associations were observed by linear and, more importantly, rank correlation analyses, and therefore the methods were demon-

strated to be largely equivalent. However, the advantage of the predictor-based efficiency assessment is that it can be derived from admission-day data only, obviating the need for daily data collection over each patient's stay. This advantage enables completion of the data acquisition over the same period as required for the determination of the PRISM III-24 score, yielding a concurrent assessment of ICU quality in terms of severity-adjusted mortality rates and ICU efficiency.

The selection of the relevant diagnoses and other case-mix factors was consistent between an earlier[3] and the updated model, and their effects on LOS were comparable in magnitude. Increasing the stay most notably were mechanical ventilation on the admission day, pneumonia, central nervous system infections, and ICU readmission. Factors associated with shorter stays included drug overdoses, diabetes, recovery from cardiovascular system surgery, and asthma. Also, the fraction of the observed LOS variability accounted for by the current model (r^2 = 0.22) was similar to the performance of the earlier predictor (r^2 = 0.24). These explanatory fractions are comparable to those attained by LOS predictors for adult ICUs (r^2 = 0.23, r^2 = 0.10).[15] Because of these relatively low explanatory powers, current LOS predictors can be reliably applied only to groups of patients, because they occur in interhospital comparisons. The high correlation between observed and predicted care days for individual institutions shown in Fig. 1 and the small errors between observed and predicted mean LOS shown in Table II demonstrate that good prediction accuracy can be achieved in such applications.

Patient care in PICUs involves both the provision of care to patients using PICU therapies ("efficient consumers") and to those being monitored because they might need such care in an acute manner ("inefficient consumers"). Both classifications routinely apply to the same patients. Some monitored patients eventually require therapies, and most patients receiving PICU therapy progress to a monitor-only phase during recovery. The ideal mix of PICU therapy

days and monitored days (e.g., efficiency) is not known. The efficient use of PICU beds also depends on other institutional resources such as other monitoring areas and bed availability in other hospital locations.[16] Because there is no accepted ideal efficiency rate, PICUs may gain insight into their practice patterns by comparing their efficiency with those of other PICUs. Toward this application, we found that the efficiency rankings obtained by the two assessment methods are highly correlated and can be used interchangeably.

Current trends make monitoring of both quality and efficiency relevant. Both are important to maintain excellence in care and to support continuous quality improvement activities. Monitoring will enable PICUs to compare themselves with a national database to ensure that their internal quality and efficiency standards remain current. The recent emphasis on benchmarking including the initiative by the Joint Commission an Accreditation of Healthcare Organizations (ORYX Initiative) focuses on risk-adjusted methods that can be used for comparison with other institutions. We have presented methods enabling such comparisons in terms of relative efficiency rankings based on PICU bed use and have demonstrated their applications in a sample of 11 PICUs.

REFERENCES

1. Pollack MM, Getson PR, Ruttimann UE, Steinhart CM, Kanter RK, Katz RW, et al. Efficiency of intensive care: a comparative analysis of eight pediatric intensive care units. JAMA 1987;258: 1481-6.
2. Knaus WA, Wagner DP, Zimmerman JE, Draper EA. Variations of mortality and length of stay in intensive care units. Ann Intern Med 1993;118:753-61.
3. Ruttimann UE, Pollack MM. Variability in duration of stay in pediatric intensive care units: a multi-institutional study. J Pediatr 1996;128:35-44.
4. Pollack MM, Cuerdon TC, Getson PR. Pediatric intensive care units: results of a national study. Crit Care Med 1993;21: 607-14.
5. Pollack MM, Cuerdon TC, Patel KM, Ruttimann UE, Getson PR, Levetown M. Impact of quality of care factors on pediatric ICU mortality. JAMA 1994; 272:941-6.
6. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. Crit Care Med 1996; 24:743-52.
7. Fleiss JL. The design and analysis of experiments. New York: John Wiley & Sons; 1986. p. 2-3.
8. Gonnella JS. Disease staging: clinical criteria. 3rd ed. Santa Barbara (CA): SysteMetrics/McGraw-Hill; 1986.
9. SAS Technical Report P-24, SAS/STAT Software: The GENMOD Procedure, Release 6.09.
10. McCullagh P, Nelder JA. Generalized linear models. London: Chapman and Hall; 1983. p. 15-24.
11. Halpern NA. Federal and nationwide intensive care units and health care costs. Crit Care Med 1994;22:2001-7.
12. Pollack MM. Cost containment. The pediatric perspective. New Horizons 1994; 2:305-31.
13. Groeger JS, Strosberg MA, Halpern NA, Raphaeli RC, Kaye WE, Guntupalli KK, et al. Descriptive analysis of critical care units in the United States. Crit Care Med 1992;20:846-63.
14. Marshall MF, Schwenzer KJ, Fletcher JC, Durbin CG Jr. Influence of political power, medical provincialism, and economic incentives on the rationing of surgical intensive care units beds. Crit Care Med 1992;20:387-94.
15. Shortell SM, Zimmerman JE, Rousseau DM, Gillies RR, Wagner DP, Draper EA, et al. The performance of intensive care units: Does good management make a difference? Med Care 1994;32:508-25.
16. Stambouly JJ, Pollack MM, Ruttimann UE. An objective method to evaluate rationing of pediatric intensive care beds. Intensive Care Med 1991;17:154-8.