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 <u>submitting standards web page</u>.

NQF #: 0343 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 30, 2012

BRIEF MEASURE INFORMATION

De.1 Measure Title: 0343 PICU Standardized Mortality Ratio

Co.1.1 Measure Steward: Virtual PICU Systems, LLC

De.2 Brief Description of Measure: The ratio of actual deaths over predicted deaths for PICU patients.

2a1.1 Numerator Statement: Actual number of deaths occurring in PICU.

2a1.4 Denominator Statement: Predicted mortality, "Predicted mortality" = Number of deaths expected based on assessed physiologic risk of mortality

Include all PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment

2a1.8 Denominator Exclusions: Preterm infants and/or adults who are admitted to the PICU in addition to patients admitted solely for palliative care

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*): N/A

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 <u>endorsed</u> or submitted measures (*check 5.1*):

Other Criteria:

Staff Reviewer Name(s):

1. IMPACT, OPPORTUITY, 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 <u>guidance on evidence</u>. *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

(*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*):

PICUs are a source of significant healthcare cost(1,2)

Risk adjusted mortality has been identified as a major focus of quality assessment in ICUs (6) and standardized mortality ratio (SMR) is recognized as a valid and reliable tool for comparing outcome between ICUs (3,4) Quality of care in the PICU may be independent of indicators of overall hospital care. (5)

SMR has been seen

1a.4 Citations for Evidence of High Impact cited in 1a.3: 1. Chalom R, Raphaely R, Costarino A. Hospital costs of pediatric intensive care. Crit Care Med; 1999. 27(10). 2079-2085.

2. Kelm S, Pollack M, Geston P. Cost, resuource utilation and severity of illness in intensive care. J Peds. 1990; 116(2). 231-237.

3. Knaus W, Wagner D, Zimmerman JE, Draper E. VAriations in moratlity and length of stay in intensive care units. Ann of Int Med. 1993;118(10).753-761

4.Selker H. Systems for comparing actual and predicted mortality rates: characteristics to promote cooperation in improving hospital care. Ann Int Med. 1993; 118228(10). 820-822

5. Pollack MM, Cuerdon TT, Patel, KM, et.al. Impact of quality-of-care factors on pediatric intensive care unit mortality. Journal of the American Medical Association;1994:272 (12). 941-946.

6. Glance, L.G., Osler TM, Dick A. Rating the quality of intensive care units: Is it a function of the intensive care unit scoring system? Critical Care Medicine; 2002.30(9):1976-1982.

1b. Opportunity for Improvement: H M L

(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 SMR allows for comparing care in different PICUs while allowing for accounting for variations in the severity of patients' illness. While there is no "right" mortality rate for a PICU, comparisons of SMR can guide quality improvement efforts.

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.]*

SMR varies widely in a cohort of PICUs sharing data through the VPS system. According to the latest data through September 2011 available in the VPS, the average SMR ranged from 0.00 to 1.76.

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. Pollack MM, Cuerdon TC, Getson PR, et al. Pediatric intensive care units: Results of a national survey. Crit Care Med 1993; 21:607–614.)

1b.4 Summary of Data on Disparities by Population Group: [*For <u>Maintenance</u> – Descriptive statistics for performance results <u>for this measure</u> 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 <u>Maintenance</u> – 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	М	Yes IF additional research unlikely to change conclusion that benefits to patients outweigh harms: otherwise No						
M-H	L	M-H	Yes IF potential benefit	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 specifically relates to a health outcome- mortality.

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):

Standardized mortality ratio is a well established methodology for comparing two or more care settings while adjusting for patient level case-mix data. Consequently, there is no perceived need in the PICU community for SMR re-evaluation aside from periodic recalibration of the tool used for risk adjustment. Hospital SMR has been

1c.5 Quantity of Studies in the Body of Evidence (*Total number of studies, not articles*): A search of Pubmed returned 155 articles specific to using SMR in an intensive care unit setting.

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): The standardized mortality ratio is simply a ratio. It is accepted as an appropriate quality measure for ICU settings. The literature is consistent on the value of using SMR with three caveats. First, use of a calibrated tool for severity adjustment has been identified as important. Second, a recent publication by Brinkman in Critical Care Medicine (2012; 40(2)373-378) identified that use of a physiology-based tool when calculating SMR is superior than to using adminstrative data. Finally, it is important to note that premature transfer of patients from

an ICU can lower the SMR, creating a potential for "gaming" measures (Kahn Chest. 2007;131(1):68-75). Use of measure 343 in conjunction with measure 334, 335 and 336 address the potentil for gaming.

1c.7 Consistency of Results across Studies (Summarize the consistency of the magnitude and direction of the effect): SMR has been found to be a reliable and valid tool for measuring ICU quality. Individual ICUs have demonstrated a decrease in SMR as a result of targeted improvement efforts.

1c.8 Net Benefit (Provide estimates of effect for benefit/outcome; identify harms addressed and estimates of effect; and net benefit - benefit over harms):

SMR is beneficial in comparing outcomes of otherwise dissimilar PICUs while allowing individual organizations to trend their own 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: N/A

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: N/A

1c.14 Summary of Controversy/Contradictory Evidence:

1c.15 Citations for Evidence other than Guidelines *(Guidelines addressed below)*: Brinkman S, Abu-Hanna A van der Veen A. A comparison of the performan of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ration of intensive care units.Critical Care Medicine (2012; 40(2)373-378

Kahn JM, Kramer AA, Rubenfield GD. Transferring critically ill patients out of hospital improves the standardized mortality ratio: a simulation study.

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: http://qualitymeasures.ahrq.gov/content.aspx?id=33605&search=standardized+mortality+ratio

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 developer's assessment of the quantity, quality, and

See Guidance for Definitions of Rating Scale: H=High; M=Moderate; L=Low; I=Insufficient; NA=Not Applicable

NQF #0343 0343 PICU Standardized Mortality Ratio, Last Updated Date: Jan 30, 2012
consistency of the body of evidence? 1c.25 Quantity: Moderate 1c.26 Quality: Moderate1c.27 Consistency: High
Was the threshold criterion, <i>Importance to Measure and Report</i> , met? (<i>1a & 1b must be rated moderate or high and 1c yes</i>) 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. (evaluation criteria) 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 this 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): Actual number of deaths occurring in PICU.
2a1.2 Numerator Time Window (<i>The time period in which the target process, condition, event, or outcome is eligible for inclusion</i>): All PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment. Data submission quarterly with reporting on annual basis.
 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: Exclusions: 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
2a1.4 Denominator Statement (Brief, narrative description of the target population being measured): Predicted mortality, "Predicted mortality" = Number of deaths expected based on assessed physiologic risk of mortality
Include all PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment
2a1.5 Target Population Category (Check all the populations for which the measure is specified and tested if any): Children's

2a1.6 Denominator Time Window (*The time period in which cases are eligible for inclusion*): All PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment. Data submission quarterly with reporting on annual basis. 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): Inclusions: • All PICU patients < 18 year of age admitted to the PICU for greater than 2 hours or with at least two consecutive sets of vital signs consistent with life with risk of mortality assessment 2a1.8 Denominator Exclusions (Brief narrative description of exclusions from the target population): Preterm infants and/or adults who are admitted to the PICU in addition to patients admitted solely for palliative care 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): All PICU patients >= 18 years of age, PICU patients with a stay < 2 hours or < 2 consecutive sets of vital signs consistent with life, Deaths occurring outside the PICU, Preterm infants post-gestational age < 36 weeks, Patients admitted to PICU for palliative care: AAP Committee on Bioethics 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): No additional stratification occurs beyond the risk adjustment inherent to this measure. That is, the expected mortality that serves as the denominator in this measure specifically accounts for the severity of illness of patients included in the measure. No further stratification is appropriate based on current literature. 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 guality 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. The risk model was developed using forward stepping logistic regression. Final variables were selected using a significance level p<0.05. The risk factor variables used in the version of PRISM 3 currently in use in the VPS dataset include: PRISM 3 12-hour score PRISM 3 12-hour score squared Pre-ICU care area **Operative status** Acute diagnosis of diabetes Pre-ICU cardiac massage Age 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: Attachment NQF response Risk of Mortality PRISM 3.docx

2a1.17-18. Type of Score: Ratio

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 = Lower score

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.):

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):

SMR is an established method which has been demonstrated to be reliable and valid. Further 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, including all Threats to Validity: H M L

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 SMR 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*):

SMR is an established method which has been demonstrated to be reliable and valid. Further 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 SMR 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

Arial Narrow

Calibration reassessment has been performed with plans for future model enhancementPRISM 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

See 2b4.2 for additional information on analytic method.

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*):

Methods

Training sample

The training sample used 7375 admissions including 213 deaths. The model predicted 213.002 deaths (SMR = 1.00, z = 0.000, p = 1.0).

Discrimination

The area under the ROC curve, used to assess discrimination, was excellent with c = 0.930 + 0.009. Discrimination was further assessed using the misclassification rate for deaths, with a cutpoint for classifying deaths predicted to live assigned arbitrarily at p = 0.5. Using this method, the misclassification rate for deaths was 0.657.

Calibration

The Hosmer-Lemeshow statistic, used with seven strata of mortality (0.00-0.01, >0.01-0.035, >0.035-0.075, >0.075-0.150, >0.150-0.250, >0.250-0.500, and >0.500), selected to ensure sufficient data in each stratum, demonstrated excellent calibration (H-L statistic = 5.4179, p = 0.36703).

Test sample

3208 admissions, including 93 deaths, were included in the test sample. Overall, 97.150 deaths were predicted (SMR = 0.957, z = -0.542, p = 0.583).

Discrimination

Discrimination as assessed by the area under the ROC curve was excellent (c=0.944 +/- 0.013). The misclassification rate for deaths was 0.591 using the same method as used in the training sample.

Calibration

The Hosmer-Lemeshow test, using the same mortality strata as in the training sample was excellent (HL = 7.205, p = 0.206).

2b4.3 Testing Results (<u>Statistical risk model</u>: 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. <u>Risk stratification</u>: Provide quantitative assessment of relationship of risk factors to the outcome and differences in outcomes among the strata): See 2b4.2 for additional information on analytic method including calibration and discrimination results.

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

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):

A population of 117 PICUs using the VPS database found SMR varies widely. According to the latest data through September 2011 available in the VPS, the average SMR ranged from 0.00 to 1.76.

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

This was merely a calcuation of SMR at individual unit level with comparison.

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):

Published literature indicates use of administrative data is inappropriate for determination of SMR when clinical data is available. Thus, this is a clinical measure.

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

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), Quality Improvement with Benchmarking (external benchmarking to multiple organizations)

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 (<i>If used in a public reporting program, provide name of program(s), locations, Web page URL(s)</i>). <u>If not publicly reported in a national or community program</u> , state the reason AND plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement: [<i>For Maintenance</i> – <i>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.</i>] Publication of SMR with peer comparison is used by various hospitals nationally. One example can be found on the Children's Hospital of Wisconsin Quality webpage (http://www.chw.org/display/PPF/DocID/46554/Nav/1/router.asp). Similarly, SMR is
publically available for the Cincinnati Children's (http://www.cincinnatichildrens.org/about/quality-measures/system-level- measures/standardized-picu-mortality-ratio/)
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: Mortality (particularly the death of a child) is readily understandable and meaningful. Focus group tesing of families in anticipation of publication of the aforementioned hospital SMR data found this to be an important measure to families.
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):
3b. Usefulness for Quality Improvement: H M H L I 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 <u>Maintenance</u> – If not used for QI, indicate the reasons and describe progress toward using performance results for improvement]. While mortality is a relatively infrequent event (occuring in ~3% of PICU admissions), when adjusted for severity of illness using a methodology like SMR, it allows for longitudinal trending of PICU performance. Additionally, PICUs who apply the same risk
adjustment methodology can compare outcomes and benchmark.
3b.2. Provide rationale for why the measure performance results are meaningful, understandable, and useful for quality improvement. If usefulness was demonstrated (<i>e.g.</i> , <i>Ql initiative</i>), describe the data, method and results: Confidential reports by specific PICUs participating in the VPS clinical database have demonstrated single site improvments in SMR as a result of sustained QI interventions.
Overall, to what extent was the criterion, <i>Usability</i> , met? H M L I
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? (<i>Check all that apply</i>). 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 (<i>regarding proprietary measures</i>): 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 (<i>e.g., fees for use of proprietary measures</i>): The work required in collecting those elements necessary for calculating expected mortality (SMR denominator) is not insignificant 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 0334						
Overall, to what extent was the criterion, <i>Feasibility</i> , met? H M L I						
Does the measure meet all the NOF 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: 0229 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following heart failure (HF) hospitalization for patients 18 and older

0230 : Hospital 30-day, all-cause, risk-standardized mortality rate (RSMR) following acute myocardial infarction (AMI) hospitalization for patients 18 and older

5a. Harmonization

5a.1 If this measure has EITHER the same measure focus OR the same target population as <u>NQF-endorsed measure(s)</u>: Are the measure specifications completely harmonized? No

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

Measures 0229 and 0230 both lack "a traditional numerator and denominator". Neither apply to the population of the measure being resubmitted. They share the same measure focus in attempting to measure mortality; however measures 0229 and 0230 utilize claims data without evidence for comparability, much less ability to be harmonized with PICU SMR. Instead, these measures appear complementary. Measures 0229 and 0230 both lack "a traditional numerator and denominator". Neither apply to the population of the measure being resubmitted. They share the same measure focus in attempting to measure mortality; however measures 0229 and 0230 both lack "a traditional numerator and denominator". Neither apply to the population of the measure being resubmitted. They share the same measure focus in attempting to measure mortality; however measures 0229 and 0230 utilize claims data without evidence for comparability, much less ability to be harmonized with PICU SMR. Instead, these measures 0229 and 0230 utilize claims data without evidence for comparability, much less ability to be harmonized with PICU SMR. Instead, these measures appear complementary with PEDS-HC-10..

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*): N/A

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, 22042

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-, VPS 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-, VPS 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

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 <u>PRIOR TO</u> 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 <u>DOES NOT</u> 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.

<u>Missing Data</u>

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 cardiorespiratory vital sign sheets, and/or nurse's notes.

Variables used to calculate the PRISM 3 risk of mortality

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 yintercept. The current coefficients as well as their positivity or negativity will be posted on the VPS website <u>https://portal.myvps.org/document/NQFMeasures.pdf</u>.

Calculation of the Risk of Mortality (Probability of Death)

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

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 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

From the Departments of Anesthesiology (Dr. Pollack) and Pediatrics (Drs. Pollack, Patel, and Ruttimann), George Washington University School of Medicine; Children's National Medical Center (Drs. Pollack and Patel), Center for Health Services and Clinical Research, Children's Research Institute (Drs. Pollack and Patel); National Institute on Alcohol Abuse and Alcoholism (Dr. Ruttimann), the National Institutes of Health, Washington, DC.

Supported, in part, by grant MCH-110584 from the Maternal and Child Health Bureau (Title V, Social Security Act), Health Resources and Services Administration, Department of Health and Human Services.

See page 752 for information regarding PRISM III and updated PRISM algorithms.

Address requests for reprints to: Murray M. Pollack, MD, Children's National Medical Center, 111 Michigan Avenue, NW, Washington, DC 20010.

Copyright © 1996 by Williams & Wilkins

Crit Care Med 1996 Vol. 24, No. 5

everity 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 guantify 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 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.

he physiologic

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_o, 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, with a simultaneous Fio,. 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 Pao₂.

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 $(p \ge .25)$ 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 Infant Child Adolescent	<40 <45 <55 <65	40–50 45–55 55–65 65–75	51–55 56–65 66–75 76–85	>125 >135 >150 >190

PRISM III (continued)	Creatinine Blood Urea Nitrogen (BUN) Measurement Measurement	$\frac{\text{Score}=2}{\text{Neonate}} = \frac{\text{Score}=2}{\text{Neonate}} = \frac{\text{Score}=3}{\text{Neonate}}$ Neonate >0.85 mg/dL or >75 μ mol/L Neonate >11.9 mg/dL or >4.3 mmol/L Infant >0.90 mg/dL or >5.4 mmol/L All Other Ages >14.9 mg/dL or >5.4 mmol/L Infant >0.90	Child > 0.50 mg/dL or > 115 µmo/L Adolescent > 1.30 mg/dL or > 115 µmo/L HEMATOLOGY TESTS (1,2)		White Blood Cell Count (cells/mur') Prothrombin Time (PT) or Partial Thromboplastin Time (PTT) (seconds)	MeasurementMeasurementScore=4Score=3Score=4NeonateAll ages<3,000All Other AgesPT > 22.0 or PTT > 57.0	Platelet Count (cells/mm*)	Measurement $\underline{\qquad}$ Measurement $\underline{\qquad}$ Score=2 $\underline{\qquad}$ Score=2 $\underline{\qquad}$ Score=5	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 thichest and/or the lowest values for sochine. 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 hasping locations, susying in the PICU > 2 hours, and house admitted in conditions TeV who do not achieve table vial agais for 2 hours. Denotes The Recurring in the OR are included only if the operation occurred during the PICU sity and was therapy for the functions requiring PICU area. Terminally ill patients transferred from the PICU for somotic acte are included PICI I rations for the 24 hours. Death for the 24 hours and hour a three predicts and not a three predicts are active reheated as PICI rations for the 24 hours. Death for the 24 hours and hour a bruck of the predict acted to the predict of the predict acted by the predict of the predict acted by the predict of the predic	support is discontinued. Ages: Neonate = $0 \cdot <1 \mod 1$; anoth: 12 months; Child = $\ge 12 \mod 1$; Adolescent > 144 months. Adolescent > 144 months.	 Heart Rate: Do not assess during crying or iatrogenic agilation. Temperature: Use rectal, yail, blood, or adilary temperatures. Dubilitor Reflexes: Nonreaction multis multi be 5 nm not access sher iatroenic multilare dilatation. 	6. Mental Satura function of particular with known or suspected, acute CNS data water. Do not assess within 2 hours of sedation, paralysis, or paralysis, factured entry particular subjects and/or sedation, use the time period without sedation, paralysis, or anesthesia closest to the PICU admission for scoring. Supportoma is defined as GCS score < 8 or supportoma using other restrictions of the provided provided action of the provided period without sedation.	neutral values 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 interlait, capitalizary, or venous sites.	 PaO₂: Use arterial measurements only. Whole blood measurements should be increased as follows: glucose - 10%; sodium - 3 mmol/L; potassium -0.4 mmol/L. (Pediatic Reference Ranges, Soldin SI, Hicks IM eds, AACC Press, Washington, D.C., 1995). 	10. Nonoperative CV discase 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. Cath- cterizations 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
		I	Score = 4 > 225 > 225 > 205	> 155		Score=11 Both fixed one reactive					1			2 Score=6 6 <42.0						1		
	VITAL SIGNS (1-6)	rt Rate (beats per minute) asurement	Score=3 onate 215-225 ant 215-225	dolescent 145-155	upillary Reflexes	feasurement <u>Score=7</u> VII Ages One fixed			ASES (1,2,7,8)	Total CO ₂ (mmol/L)	Measurement	Score=4 All Ages >34.0	PaO ₁ (mm Hg)	Measurement <u>Score=3</u> All Ages 42.0-49.9				TS (1,2,9)	Potassium (mmol/L)	Measurement	Score=3 All ages >6.9	
PRISM III	CARDIOVASCULAR/NEUROLOGIC	l Pressure (mm Hg) Hea	Score=3 Score=1 40-55 < 40 Ne 45-65 < 45 Int	65-85 <65 Au	Pi	nt N <u>Scorte=3</u> <33 °C (91.4 °F) or >40.0 °C (104.0 °F)	rtus	ent	ACID-BASE/BLOOD G	fotal CO ₂ (mmol/L) or pH)	cnt 1	Score=2 Score=6 pH 7.0-7.28 pH < 7.0 or total CO ₂ 5-16.9 or total CO ₂ < 5		ent. <u>Score=2</u> , <u>Score=3</u> 7.48-7.55 > 7.55	n Hg)	cnt	Score=1 Score=3 50.0-75.0 > 75.0	CHEMISTRY TES		ent	> 200 mg/dL or > 11.0 mmol/L continued	's National Medical Center, May 1995
		Systolic Bloo Measurement	Neonate Infant	Child Adolescent	Temperatı	Measurem All Ages	Mental St	Measuren All Ages		Acidosis (Measuren	All Ages	Hq	Measurer All Ages	PCO ₁ (II	Measuren	All Ages		Glucose	Measurer	All ages	Childre

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; CU, 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-offit. 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 goodnessof-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 loglikelihood 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)

				Hosı Lemes	mer- how χ²
	χ^2 (df)	AIC	AUC (sem)	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	Surv	ivors	Dea	aths	Standardized				
(%)	Expected	Observed	Expected	Observed	Ratio (SEM)				
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.

5

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



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.

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, Pao₂/Fio₂, 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,



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).

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

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 zstatistic 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-offit 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.

ple

00

1 =

III

nit

ve

 \mathbf{re}

;ay

3 =

re

ty

.ic M

ri-

he

), 5

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

	$\begin{array}{c} Hosmer-\\ Lemeshow\chi^2 \end{array}$		Expected	Flora's Method			
	5 df	р	Deaths (no.)	z	р	AUC (SEM)	
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: $0 \ll 3\%$; $\geq 3\% - 10\%$; $\geq 10\% - 25\%$; $\geq 25\% - 45\%$; $\geq 45\% - 65\%$; and $\geq 65\%$.

Additional variables are given in Table 3.

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

	Survi	Survivors		Deaths	
Probability of Death (%)	Expected	Observed	Expected	Observed	Mortality Ratio (SEM)
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.

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).

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 he relationship between physio logic status, as measured by PRISM III, and outcomes has been calibrated to a contemporary, well defined, large reference sample.

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,, 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 ageadjusted 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



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.

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

	Survivors		Dea	Standardized		
Diagnosis	Expected	Observed	Expected	Observed	Mortality Ratio (SEM)	
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, welldefined, 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)	Surv	Survivors		Deaths	
	Expected	Observed	Expected	Observed	Ratio (SEM)
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.

ACKNOWLEDGMENTS

This study was, in part, an independent effort of members of the National Association of Children's Hospitals and Related Institutions. The following study sites and institutional coordinators participated: Mark E. Swanson, MD, Arnold Palmer Hospital for Children and Women, Orlando, FL; Jacob Hen, Jr, MD, Bridgeport Hospital, Bridgeport, CT; Bob Lynch, MD, Cardinal Glennon Children's Hospital, Saint Louis, MO; James Fackler, MD, Children's Hospital, Boston, MA; Barbara A. Jackson, MD, and Stephen Levine, MD, Children's Hospital, New Orleans, LA; Mary W. Lieh-Lai, MD,

Children's Hospital of Michigan, Detroit, MI; Frank Allman, MD, Children's Hospital Medical Center of Akron, Akron, OH; William A. Spohn, MD, The Children's Medical Center, Dayton, OH; James D. Wilkinson, MD, Children's National Medical Center, Washington, DC; Tara Snellgrove, MD, and Linda Lai, RN, Cook Children's Medical Center, Fort Worth, TX; Peter Quint, MD, Emanuel Children's Hospital and Health Center, Portland, OR; Suresh Havalad, MD, Lutheran General Hospital, Park Ridge, IL; David M. Habib, MD, Medical University of South Carolina, Charleston, SC; Kathleen Winder, RN, Memorial Hospital, Colorado Springs, CO; Linda Marzano, MD, Miami Children's Hospital, Miami, FL; Suzanne Sander, RN, and Stephen C. Kurachek, MD, Minneapolis Children's Medical Center, Minneapolis, MN; Richard M. Ruddy, MD, New York Medical Center, Valhalla, NY; Rex Northup, MD, Sacred Heart Hospital, Pensacola, FL; Maggie Halley, MD, Saint Joseph's Hospital and Medical Center, Phoenix, AZ; David L. Peterman, MD, Saint Luke's Regional Medical Center, Boise, ID; James H. Jose, MD, Scottish Rite Children's Medical Center, Atlanta, GA; Gary A. Neidich, MD, Sioux Valley Hospital, University of South Dakota School of Medicine, Sioux Falls, SD; Marjory K. Waterman, RN, MN, Southwest Texas Methodist Hospital, San Antonio, TX; Maria Cox, RN, Tampa Children's Hospital, Tampa, FL; Wallace W. Marsh, MD, and John Cochran, MD, Texas Tech University Health Science Center and University Medical Center Hospital, Lubbock, TX; Arno Zaritsky, MD, University of North Carolina, Chapel Hill, NC.

The authors also wish to acknowledge the individual efforts of Timothy Cuerdon, PhD, Marcia Levetown, MD, Walter Clark, BA, Kia Banks, BA, and Joyce Sheppard, RN.

REFERENCES

- 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
- 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
- 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
- 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 nontertiary 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

PRISM III and updated PRISM algorithms are copyrighted and may be the subject of one or more patents held by Children's Research Institute. The equations are available without charge for research uses including the independent verification of their accuracy and reliability. Children's National Medical Center may receive compensation resulting from nonresearch uses of PRISM III and PRISM algorithms. new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA 1993; 270:2957-2963

- 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
- 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
- Richardson DK, Gray JE, McCormick MC, et al: Score for neonatal acute physiology: A physiologic severity index for neonatal intensive care. *Pediat*rics 1993; 91:617-623
- 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
- Knaus WA, Wagner DP, Lynn J: Shortterm mortality predictions for critically ill hospitalized adults: Science and ethics. *Science* 1991; 254:389–394
- 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
- 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

- SAS/STAT User's Guide, Version 6. Fourth Edition. Volume 2. Cary, NC, SAS Institute, pp 1071-1126
- Zhang P: Model selection via multifold cross validation. Ann Statistics 1993; 21:299-313
- Hosmer DW, Lemeshow S: Applied Logistic Regression. New York, John Wiley & Sons, 1989
- 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
- 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
- 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