

# PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

## **Assessment and Optimization of Mortality Prediction Tools for Admissions to Pediatric Intensive Care in the United Kingdom**

Anthony R. Brady, David Harrison, Stephanie Black, Sam Jones, Kathy Rowan, Gale Pearson, Jane Ratcliffe, Gareth J. Parry and on behalf of the UK PICOS Study Group

*Pediatrics* 2006;117:e733-e742; originally published online Mar 1, 2006;

DOI: 10.1542/peds.2005-1853

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

<http://www.pediatrics.org/cgi/content/full/117/4/e733>

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2006 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



# Assessment and Optimization of Mortality Prediction Tools for Admissions to Pediatric Intensive Care in the United Kingdom

Anthony R. Brady, MSc<sup>a</sup>, David Harrison, PhD<sup>a</sup>, Stephanie Black, MSc<sup>a</sup>, Sam Jones, BSc<sup>b</sup>, Kathy Rowan, DPhil<sup>a</sup>, Gale Pearson, MB, ChB<sup>c</sup>, Jane Ratcliffe, MB, ChB<sup>d</sup>, Gareth J. Parry, PhD<sup>b</sup>, on behalf of the UK PICOS Study Group

<sup>a</sup>Intensive Care National Audit & Research Centre, London, United Kingdom; <sup>b</sup>Health Services Research, School of Health and Related Research, University of Sheffield, Sheffield, United Kingdom; <sup>c</sup>Birmingham Children's NHS Trust, Birmingham, United Kingdom; <sup>d</sup>Royal Liverpool Children's NHS Trust, Liverpool, United Kingdom

The authors have indicated they have no financial relationships relevant to this article to disclose.

## ABSTRACT

**OBJECTIVE.** To assess the Pediatric Risk of Mortality (PRISM, PRISM III-12, and PRISM III-24) systems and the Pediatric Index of Mortality (PIM and PIM2) systems for use in comparing the risk-adjusted mortality of children after admission for pediatric intensive care in the United Kingdom.

**METHODS.** All PICUs in the United Kingdom were invited to participate. Predicted probability of PICU mortality was calculated using the published algorithms for PIM, PIM2, and PRISM and compared with observed mortality. These scores, along with PRISM III-12 and PRISM III-24, whose algorithms are not published, were optimized for the United Kingdom.

**RESULTS.** Of 26 PICUs in the United Kingdom, 22 (85%) were recruited, and sufficient prospective data were collected from 18 (69%) units on 10 197 (98%) of 10 385 admissions between March 2001 and February 2002. All published tools were found to have poor calibration but provided good discriminatory power. After estimation of UK-specific coefficients, only PIM2, PRISM III-12, and PRISM III-24 had satisfactory calibration. All models provided good discriminatory power. Funnel plots for all of the recalibrated models indicated that the risk-adjusted mortality for all units was consistent with random variation.

**CONCLUSIONS.** PIM2, PRISM III-12, and PRISM III-24 all were found to be suitable for use in a UK PICU setting. All tools provided similar conclusions in assessing the distribution of risk-adjusted mortality in UK PICUs. It now is important that these tools be used to monitor outcome and improve the quality of pediatric intensive care within the United Kingdom.

[www.pediatrics.org/cgi/doi/10.1542/peds.2005-1853](http://www.pediatrics.org/cgi/doi/10.1542/peds.2005-1853)

doi:10.1542/peds.2005-1853

### Key Words

clinical score, intensive care, mortality prediction, outcome assessment, quality of care

### Abbreviations

PRISM—Pediatric Risk of Mortality  
PIM—Pediatric Index of Mortality  
UK PICOS—United Kingdom Pediatric Intensive Care Outcome Study

Accepted for publication Oct 13, 2005

Address correspondence to Gareth J. Parry, PhD, Department of Medicine, Children's Hospital Boston, 300 Longwood Ave, Boston, MA 02115. E-mail: [gareth.parry@childrens.harvard.edu](mailto:gareth.parry@childrens.harvard.edu)

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275). Copyright © 2006 by the American Academy of Pediatrics

THE PROVISION OF intensive care to infants, children, and adults increased steadily during the latter half of the 20th century, with particular, rapid expansion during the 1970s and 1980s.<sup>1</sup> In pediatric intensive care, the growth in activity has followed somewhat later and has been accompanied by little objective evaluation of quality of care provided or the effectiveness of specific treatments. Furthermore, evidence to support planned national infrastructure is limited when compared with the significant resources expended on pediatric intensive care.<sup>2,3</sup>

Within the United Kingdom, the equitable provision of pediatric intensive care, service structure, and rigorous evaluation of new and existing treatments all require methods to account for variation in case mix. Rigorously conducted randomized, controlled trials of sufficient statistical power are the "gold standard" research design to compare different aspects of medical practice. However, there are a number of situations in which random allocation is not possible for either ethical or logistic reasons.<sup>4</sup> When randomization is impossible, the optimal research design consists of prospective studies that adjust outcomes for variation in risk factors between hospitals according to prespecified comparators.<sup>5,6</sup> The term "risk adjustment" has been coined to describe the process of adjusting for such risk factors when comparing outcomes after intensive care.

Risk-adjustment tools that predict death in PICUs have become established in the past 20 years. The Physiologic Stability Index, published in 1984,<sup>7</sup> has been updated twice, using data from North American PICUs, and renamed, first as the Pediatric Risk of Mortality (PRISM) and more recently as PRISM III.<sup>8,9</sup> PRISM III currently provides the risk-adjustment tool for the United States based Pediatric Intensive Care Unit Evaluations (PICUEs) system, which provides comparative reports to participating units under a licensing arrangement.<sup>10</sup> Various versions of the PRISM family of risk-adjustment tools have been used extensively in the United States and to some extent in the Netherlands to inform policy and organizational decisions.<sup>5,11</sup>

The alternative to the PRISM family of tools is the Pediatric Index of Mortality (PIM), which was updated recently to PIM2.<sup>12,13</sup> PIM was published in 1997, based on admissions to PICUs in Australia and 1 in the United Kingdom. PIM2 was published in 2003 and included 15 191 admissions from 10 PICUs in Australia and New Zealand and 4 PICUs in the United Kingdom. Although PIM was tested in 4 UK PICUs, none of the tools has been externally validated for general use in UK PICUs.<sup>14</sup>

Prediction tools must discriminate well between deaths and survivors and be well calibrated before they can be applied usefully to assess or standardize comparisons of PICUs or to correct for case-mix differences between groups in observational studies. Single-center studies in the United Kingdom and studies in other countries have generated conflicting evidence for the

accuracy of predictions from both the PRISM and the PIM families of tools.<sup>11,15-18</sup> Such single-center studies, often occurring several years after the publication of the tools, are likely to be unreliable because it is to be expected that risk-adjusted mortality will improve over time and that the scoring system hence will become decalibrated.<sup>19</sup> For risk-adjusted mortality rates to be meaningful, epidemiologic studies must either use a tool that is derived from their own data or use the most recent recalibration of the chosen tool. Consequently, a joint working group of the Department of Health and the Medical Research Council concluded in 1997 that no risk-adjustment method had been validated sufficiently for general use in PICUs in the United Kingdom and stated that "work on patient risk adjustment is a high priority, as without better methods of risk adjustment, research on the effectiveness of IC/HDC (Intensive Care/High Dependency Care) will be severely limited."<sup>20</sup> The United Kingdom Pediatric Intensive Care Outcome Study (UK PICOS), reported here, aimed to assess and optimize objectively the competing prediction tools in a large, prospective, and representative sample of admissions to UK PICUs.

## METHODS

All PICUs in the United Kingdom were invited to participate in UK PICOS. Staff at participating units were asked to collect prospectively standardized information using dedicated software on consecutive admissions to their unit between March 5, 2001, and March 4, 2002. Data collection staff were trained to enhance standardization of data collection between units. Data were also validated both locally, using checks built into the data collection software, and at the central processing center, where extensive data validation and query resolution were conducted. At the end of the data collection period, all units were asked to provide an independent count, using departmental databases or admission books, of the number of admissions and deaths during the study period. Units that collected data on <80% of admissions were excluded in whole or were excluded in part when it was possible to identify clearly periods of satisfactory ( $\geq 80\%$ ) coverage.

The information that was collected on each admission included demographic details, the presence in the medical history of any of 15 significant events or premorbid conditions, the location before admission, the reason for admission to the PICU, measurements of the required physiologic parameters up to 24 hours after admission, outcome at discharge from the ICU, and eventual hospital discharge (status and length of stay).<sup>21</sup> Patients who were already dead or had clinical signs of brain death on admission to the PICU were excluded from the study.

The mortality prediction tools under consideration were PIM, PIM2, PRISM, and PRISM III,<sup>8,9,12</sup> all of which aim to predict mortality in the admitting PICU. PIM and PIM2 use information on 7 and 10 factors, respectively,

to calculate the probability of death during the intensive care admission. These are gathered between first contact with a PICU doctor and up to 1 hour after admission to the PICU. PRISM uses measurements on up to 14 physiologic values during the first 24 hours after admission to the PICU. The PRISM III system can be broken down further into PRISM III-12, using measurements from up to 17 physiologic values during the first 12 hours after admission to a PICU, and PRISM III-24, using measurements from up to 17 physiologic values during the first 24 hours. In both PRISM and PRISM III, the physiologic values are used to form scores, which when combined with 2 (PRISM) or 8 (PRISM III) other factors, are used to generate predicted probability of mortality in the PICU.

### Data Completeness

All tools assume that score components that are not recorded are normal. Data were collected to allow us to predefine “not applicable” to indicate when a tool component was appropriately not recorded, for example, no blood gases were taken and “missing” to indicate all other not recorded tool components. The rate of not applicable and missing tool components then was examined for each tool by PICU.

### Statistical Analysis

#### *Published Models*

Models were applied to the data using coefficients and instructions that were published by the original authors of the tools and by personal communication with these authors for any necessary clarification. Access to the coefficients that were necessary to calculate a probability for PRISM III was refused by the main author; therefore, it could not be included in the initial stages of this evaluation (M.M. Pollack, Children’s National Medical Center, Washington, DC, personal communication, 2003). These published models were assessed by investigating their calibration and discrimination for PICU mortality.

#### *Discrimination*

Predictions from each model were assessed using the c index (area under the receiver operating characteristic curve) for discrimination. The c index is the probability of concordance between outcomes and predictions. In this study, it represents the probability that a randomly chosen patient who died will have a higher predicted probability of mortality than a randomly chosen patient who survived. Published c-index criteria suggest that  $\geq 0.7$  is acceptable,  $\geq 0.8$  is good, and  $\geq 0.9$  is excellent.<sup>22</sup>

#### *Calibration*

How well the predicted probabilities of mortality that were generated by the risk-adjustment models compared with the observed mortality was assessed using the

Hosmer-Lemeshow  $\chi^2$  test and Cox’s calibration regression.<sup>23–25</sup> For each risk-adjustment model, for the Hosmer-Lemeshow goodness-of-fit test, patients were categorized into 10 groups according to quintiles of their associated predicted probability of mortality, and the observed and expected outcomes were compared using a  $\chi^2$  statistic. For the Cox regression model, the probability of mortality that was generated by each risk-adjustment model was transformed into a log odds of mortality. This then was fitted as an explanatory variable in a logistic regression model to predict mortality. Perfect calibration would be indicated by a model with a constant term of 0 and a slope term of 1. Significant differences from these values give a quantifiable indication of where the calibration of the models has failed.

#### *Optimization of Current Models*

The tools were optimized for the United Kingdom by reestimation of coefficients using logistic regression. In the case of PRISM and PRISM III, the score was entered into the regression rather than the individual component variables of the tool. Optimized models were assessed by random allocation of PICUs (stratified by annual admissions) into development and validation samples in a 2:1 ratio. Reestimation of coefficients was conducted in the development sample, and model predictions were assessed in the external validation sample using the discrimination and calibration measures described above. This approach avoids the bias that is inherent when the same data are used both to fit the model and to emulate the application of the tool at new PICUs.<sup>26,27</sup> The process was repeated 5000 times to avoid basing conclusions on a single random split of the data that may have favored 1 of the models by chance. Summaries of the 5000 resulting discrimination and calibration statistics were calculated to guide the success of the optimization.

#### *Funnel Plots*

To illustrate the assessment of the published models and optimization of models, we produced a series of funnel plots of PICU mortality ratios by number of admissions along with corresponding 99.8% control limits.<sup>28</sup> Funnel plots are a variation on control charts that identify units with unexpectedly high (above upper control limit) or low (below the lower control limit) mortality.<sup>29,30</sup> They are becoming established as a superior alternative to league tables and ranks for the presentation of performance measures. Units with mortality ratios within the control limits are defined to have mortality consistent with random variation.

Mortality ratios were calculated by dividing the observed PICU mortality by an expected mortality. A funnel plot of crude mortality ratios first was produced using the mortality of the whole population of admissions as an expected mortality. Then 2 types of funnel plot were produced for each risk-adjustment method,

first using the published coefficients to generate the expected mortality and second using the optimized coefficients to generate the expected mortality. Ethical approval was granted for this study by the Trent Multi-Research Ethics Committee (MREC/99/4/046).

## RESULTS

Twenty-six PICUs were invited to participate in UK PICOS, and 22 (85%) agreed. The 4 units that did not take part all are based in London and admitted a median of 342 admissions in 2001 compared with a median of 452 in units that participated. Four units were excluded after the data collection period because of low or unverified coverage. In the remaining 18 units, data were collected on 10 197 (98%) of 10 385 admissions during the study period. The median age of included admissions was 1.4 years (interquartile range: 0.2–6.6 years), and 5772 (57%) were male (Table 1). Just more than one third (3813; 37%) were admitted to the PICU directly from the operating theater, a minority (622; 16%) of which were unplanned. The 3 most common reasons for admission directly from the operating theater were complex congenital cardiac abnormality (330; 8.7%), congenital ventricular septal defect (262; 6.9%), and kyphoscoliosis (259; 6.8%). In contrast, most of those who were admitted to the PICU other than directly from the operating theater were unplanned (5727 of 6384; 90%). The 3 most common reasons for these admissions were bronchiolitis (548; 8.6%), status epilepticus or uncontrolled seizure (548; 8.6%), and primary brain injury (278; 4.4%).

Of the 10 197 admissions included in the analysis, 633 (6.2%) patients died in the admitting PICU. An additional 301 (3.0%) died before discharge from the hospital and 516 (5.4%) were lost to follow-up, making total hospital mortality 9.2%. There was wide variation

between PICUs in both the characteristics and the outcome of patients who were admitted (Table 1).

## Data Completeness

PIM2 and PRISM III excluded the most admissions from mortality prediction (7.1% and 8.3%, respectively), whereas PIM and PRISM have no explicit exclusion criteria (Table 2). Of eligible admissions, 42% (range: 13%–85% between PICUs) for PIM, 55% (range: 29%–97%) for PRISM, and 46% (range: 25%–86%) for PRISM III-12 and -24 required at least 1 assumption of normality as a result of a nonrecorded tool component's being not applicable (Table 2). Also, 10% (range: 1%–48% between PICUs) for PIM, 80% (range: 29%–99%) for PRISM, and 80% (range: 28%–99%) for PRISM III-24 required at least 1 assumption of normality as a result of a nonrecorded tool component's being missing (Table 2). However, the bulk of the missing tool components relating to PRISM and PRISM III were attributed to the Glasgow Coma Score component's being missing in 77% of eligible admissions. Excluding the Glasgow Coma Score, 24% (range: 5%–56%) for PRISM and 23% (range: 1%–56%) for PRISM III-12 and -24 required at least 1 assumption of normality as a result of a nonrecorded tool component's being missing (Table 2).

## Published Models

### Discrimination

The discrimination of the 3 published tools is shown in Table 2. All tools provided "good" (*c* index >0.8) discriminatory power.<sup>22</sup> PIM2 (*c* index = 0.84) had significantly (*P* < .001) better discrimination than PIM (*c* index = 0.81) but not significantly (*P* = .105) better discrimination than PRISM (*c* index = 0.82).

TABLE 1 Number of Admissions, Age, Gender, Source of Admission, and Unit Outcome by PICU

PICU	No. of Admissions	Age, y, median	Male, %	Direct From Theatre, %	Unplanned, %	Unit Mortality, %	Unit Length of Stay, h, median
A	317	1.7	57	27	82	6.9	31.5
B	432	3.1	53	30	64	3.2	32.0
C	674	1.4	61	39	63	4.9	59.5
D	1116	0.6	56	19	76	10.1	59.2
E	226	2.5	52	22	76	4.4	36.9
F	520	1.9	55	21	87	7.5	64.6
G	985	1.0	57	52	46	4.8	44.4
H	335	5.1	53	40	72	5.7	47.5
I	1110	1.0	60	49	58	7.1	43.3
J	781	0.8	57	41	39	4.1	92.2
K	447	1.2	53	38	65	9.0	57.4
L	1026	1.5	52	46	50	5.0	23.2
M	279	2.2	62	35	70	6.8	37.8
N	278	1.5	60	21	81	3.6	46.6
O	420	2.7	57	33	77	6.4	38.0
P	340	2.5	57	40	64	5.9	34.6
Q	607	0.7	58	39	57	6.6	34.5
R	304	1.0	58	49	54	5.9	52.3
Total	10 197	1.4	57	37	62	6.2	46.2



**TABLE 2** Exclusion Criteria and Discrimination and Calibration Characteristics of the PIM, PIM2, PRISM, PRISM III-12, and PRISM III-24 Systems Using Published and UK PICOS Recalibrated Models

	PIM	PIM2	PRISM	PRISM III-12	PRISM III-24
Admissions excluded from models					
Age > 15 y	188	188	0	0	0
Transferred to another ICU	0	543	0	543	543
In unit < 2 h	0	0	0	313	313
Admissions included in model assessment					
N	10 009	9475	10 197	9352	9352
%	98.2	92.9	100	91.7	91.7
Admissions with at least 1 not recorded tool component being not applicable					
%	42	42	55	50	46
PICU range	13–85	13–85	29–97	33–86	25–86
Admissions with at least 1 not recorded tool component being missing					
%	10	10	80	80	80
PICU range	1–48	1–48	29–99	28–99	28–99
Admissions with at least 1 not recorded tool component being missing, excluding Glasgow Coma Score					
%			24	23	23
PICU range			5–56	1–56	1–56
Published calibration models					
ROC area, c index (95% CI)	0.81 (0.79–0.83)	0.84 (0.82–0.86)	0.82 (0.80–0.84)		
Hosmer-Lemeshow $\chi^2$	86.7	39.8	541.9		
P	<.001	<.001	<.001		
Cox's method					
Slope (95% CI)	0.83 (0.77–0.89)	0.87 (0.81–0.93)	0.66 (0.61–0.70)		
Intercept (95% CI)	–0.54 (–0.69 to –0.39)	–0.15 (–0.31 to 0.02)	–1.33 (–1.44 to –1.23)		
UK PICOS recalibration models					
ROC area, c index, median (quartiles)	0.81 (0.79–0.83)	0.84 (0.83–0.85)	0.85 (0.84–0.86)	0.88 (0.87–0.89)	0.88 (0.87–0.90)
Hosmer-Lemeshow $\chi^2$ , median (quartiles)	13.8 (9.0–26.3)	14.5 (10.6–20.0)	15.1 (10.9–23.4)	14.9 (10.7–20.0)	13.2 (9.7–18.2)
P < .05, n (%)	1766 (35)	1595 (32)	1898 (38)	1602 (32)	1232 (25)
Cox's method					
Slope, median (quartiles)	0.97 (0.92–1.03)	0.97 (0.92–1.03)	0.98 (0.94–1.04)	0.95 (0.89–1.03)	0.96 (0.89–1.03)
Intercept, median (quartiles)	–0.07 (–0.17 to 0.06)	–0.05 (–0.30 to 0.14)	–0.02 (–0.13 to 0.10)	–0.11 (–0.25 to 0.10)	–0.08 (–0.25 to 0.12)

CI indicates confidence interval; ROC, receiver operating characteristic.

### Calibration

All models displayed significant lack of fit and therefore poor calibration according to the Hosmer-Lemeshow test. The Cox calibration regression found that PIM and PRISM tended to overestimate risk in high-risk admissions, and PIM2 tended to underestimate risk in low- and medium-risk admissions (Fig 1 and Table 2). PRISM also underestimated risk in very-low-risk (<1%) admissions. The slope estimate from Cox's calibration regression was less than the ideal value of 1 for all tools, suggesting that predictions generally were exaggerated compared with observed risk.

### Optimization of Current Models

The mean sample size for reestimation of model coefficients varied according to exclusion criteria, ranging from 6185 for PRISM III to 6751 for PRISM. The mean sample size for validating the models ranged between

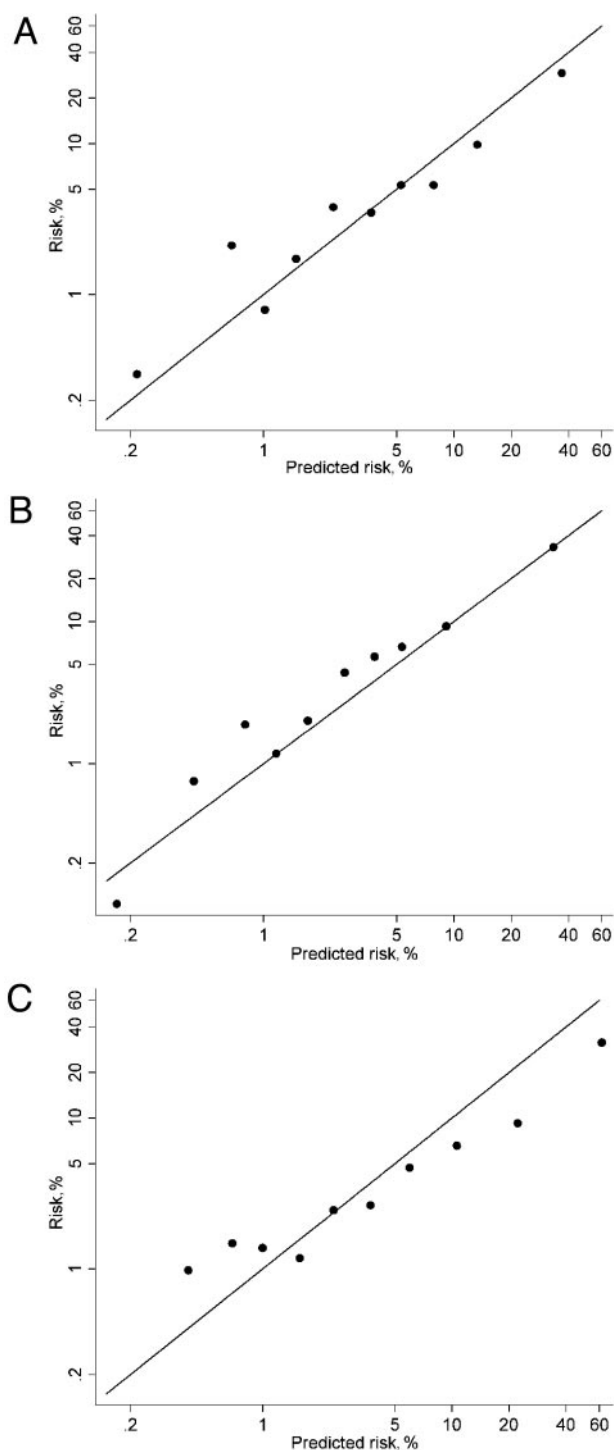
3164 and 3443. The median discrimination and calibration statistics from the 5000 validation samples are shown in Table 2.

### Discrimination

On average, all tools provided “good” (c index >0.8) discriminatory power.<sup>22</sup> In the whole sample, the discrimination for both PRISM III-12 and PRISM III-24 (median c index = 0.88) was found to be significantly ( $P < .001$  in all cases) higher than for PRISM (median c index = 0.85), PIM (median c index = 0.81), and PIM2 (median c index = 0.84).

### Calibration

Using the Hosmer-Lemeshow test, we found a lack of calibration to be significant at the 5% level in only 25% of samples for PRISM III-24 compared with 32% of



**FIGURE 1**  
Calibration plots for PIM, PIM2, and PRISM. Plots show the observed risk by expected risk in deciles of expected risk using a log scale.

samples for PIM2 ( $P < .001$ ). In the whole sample, the Hosmer-Lemeshow test indicated significant lack of fit for PIM ( $P = .03$ ) and PRISM ( $P = .01$ ) and adequate fit for PIM2 ( $P = .32$ ), PRISM III-12 ( $P = .13$ ), and PRISM III-24 ( $P = .37$ ). All models displayed a slight tendency toward overfitting, indicated by the average slope esti-

mate's being  $<1$  from Cox's recalibration regression in validation data. Calibration plots for models that were reestimated and assessed in the whole sample are shown in Fig 2.

### Funnel Plots

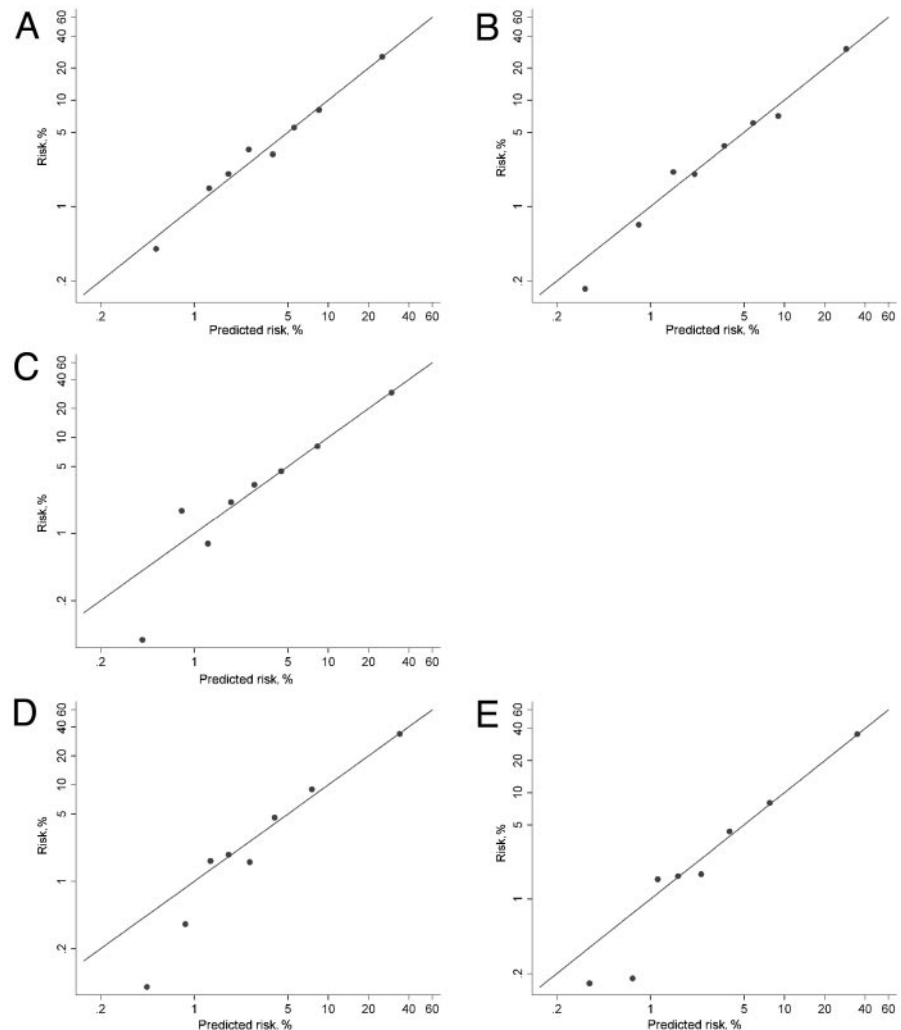
Figure 3 contains a funnel plot of the crude mortality ratio by number of admissions for each unit, which indicates that 1 unit seemed to have an unadjusted mortality ratio that was higher than expected. The funnel plots that were produced using the published models indicate that 1 unit had adjusted mortality that was lower than expected for PIM, 2 units had adjusted mortality that was higher than expected for PIM2, and 7 units had adjusted mortality that was lower than expected for PRISM, including 1 unit for which PIM had identified as having mortality that was higher than expected. The funnel plots for the recalibrated models indicate that the adjusted mortality for all units is consistent with random variation for all models.

### DISCUSSION

Admissions to PICUs in the United Kingdom vary considerably in their characteristics, making adjustment for case mix imperative when comparing outcomes between different units. We compared all major published tools for predicting PICU mortality using a large, representative sample of consecutive admissions to 18 UK PICUs between March 2001 and February 2002. All published models (PIM, PIM2, and PRISM) displayed good discrimination but poor calibration. Including all models and after recalibration on UK data, only PIM2, PRISM III-12, and PRISM III-24 displayed adequate fit. In addition, both PRISM III-12 and PRISM III-24 demonstrated significantly higher discrimination than the other tools.

The data completeness seems to be higher for PIM and PIM2 than for PRISM, PRISM III-12, and PRISM III-24. However, this was found to be attributable primarily to low completeness of the Glasgow Coma Score associated with the various PRISM scores and does not seem to have affected the overall performance of these scores.

The overall calibration of PIM and PIM2 was better than PRISM when published coefficients were used. This is illustrated further by the corresponding funnel plots, in which PRISM identified 7 units as having adjusted mortality outside the control limits whereas PIM identified 1 and PIM2 identified 2. This is not surprising as PRISM is much older (published in 1988) and based on a relatively small sample ( $N = 2642$ ) of patients in North America.<sup>8</sup> If PRISM had shown perfect calibration, then this would have suggested that risk-adjusted mortality after admission to UK PICUs in 2001 was similar to that of North American PICUs in 1988. Miscalibration of PRISM in validation samples has been noted several



**FIGURE 2**

Calibration plots for recalibrated PIM, PIM2, PRISM, PRISM III-12, and PRISM III-24. Plots show the observed risk by expected risk in deciles of expected risk using a log scale. Forming deciles lead to 0 deaths in some groups; therefore, octiles were used in preference.

times before.<sup>15,17,18</sup> PIM was published in 1997, and PIM2 was published in 2003. This group recently performed a large study that compared PIM, PIM2, PRISM, and PRISM III in PICUs in Australia and New Zealand and concluded that PIM2 was the most accurate.<sup>31</sup> For a tool to calibrate well in a validation sample, all factors that influence outcome must either be included in the model or have the same distribution in the validation sample as the sample used to develop the model. Differences between countries and over time make this second condition less likely. PICUs have been found to vary considerably in their structure, organization, and staffing across European countries and still can be anticipated to do so between the United Kingdom and Australia.<sup>2,32</sup>

This study reinforces the need to use a representative and large number of units when assessing such risk-adjustment methods in another country or health care system. A small or unrepresentative selection of units could lead to confusion between unit or overall system performance and the need to recalibrate the score. Even including all units may result in poor calibration, but this

does not necessarily invalidate the scores as a simple recalibration of these scores can lead to their being useful and valid tools for that country or health care system.

The performance of PIM2, PRISM III-12, and PRISM III-24 using coefficients estimated in UK data were acceptable, although PRISM III-12 and PRISM III-24 displayed an advantage in terms of discrimination. The funnel plots of adjusted mortality ratios for all of the recalibrated models suggested that mortality in all of the units in this study was consistent with random variation and illustrates further the importance of recalibrating the models.

The use of physiologic readings up to 24 hours after admission has been criticized by the authors of PIM. They argued that such readings may be recognizing death, rather than predicting it, in the substantial proportion of children who die within the first 24 hours (29% of deaths in our data).<sup>12</sup> In addition, quality-of-care differences may be masked because PICUs that mismanage care in the first 24 hours will seem to have a worse case mix.<sup>12,33</sup> The development of the PRISM



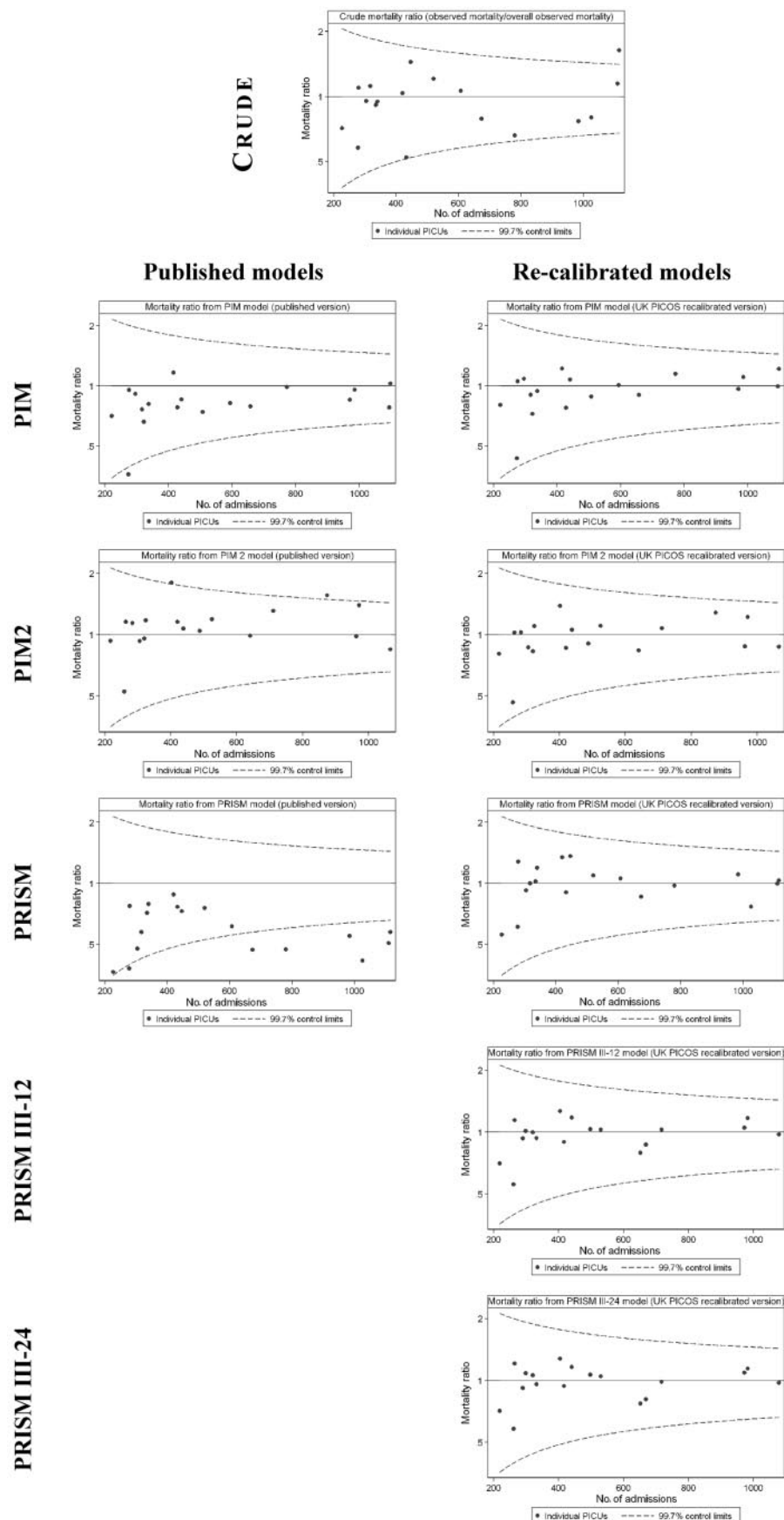


FIGURE 3

Funnel plot of mortality ratio by number of admissions for each unit. The mortality ratio is calculated for each PICU by obtaining the expected number of deaths according to each scoring system and dividing the observed number of deaths by the expected number. The mortality ratio then is plotted on the y axis against the number of admissions to the PICU on the x axis. To satisfy the condition that if the distribution of the mortality ratios is random there exists an  $\sim 5\%$  chance of a unit's falling outside the control limits, then the upper and lower control limits must represent not 95% confidence intervals but 99.7% (confidence intervals around a mortality ratio of 1 by number of admissions<sup>29</sup>). For the 18 PICUs considered here, each has a probability of falling outside the confidence control limits of  $1 - .9971 = .0029$ . Therefore, overall, there is a probability of at least 1 unit's falling outside the control limits by chance of  $1 - .9971^{18} = .05$ . In other words, to fulfill the criteria that there be a 5% probability of a unit's falling outside the control limits by chance alone, the limits must be set at 99.7%.

III-12 tool was motivated by these concerns.<sup>9</sup> However, others have argued that by using admission values only, as in PIM and PIM2, there is an increased risk for observing nontypical point estimate values that misrepresent the true initial health status of the patient.<sup>34</sup> The results of this study show that good discrimination and calibration are achievable without using information that is obtained after the first hour in the PICU. The amount of data to be collected is an important consideration in the design of any study; therefore, PIM2 may be valuable for most quality improvement and audit studies that require the minimum of data collection necessary to undertake the study. Larger scale research projects with funding to collect additional data could reasonably choose to use PRISM III variables with the provisos mentioned above.<sup>27,35</sup> Whichever risk-adjustment method is chosen, units that wish to participate in comparative research studies or collaborative quality improvement efforts must ensure that participating centers not only collect the data for the risk-adjustment tool in a standardized manner but also that standardized reporting of outcomes are fed back to the units.

Any variation in risk-adjusted PICU mortality needs to be tempered by an assessment of variation in longer term outcome. A reduction in mortality is of dubious benefit if it is at the expense of increased severe morbidity. Standardized and reliable methods of measuring health status (or at least clinical status) are required for use in the United Kingdom.<sup>36</sup> Currently, only the PRISM family of tools has been shown to have any relationship to longer term outcomes in the United States.<sup>37</sup>

## CONCLUSION

This study confirms that risk-adjustment methods that are developed primarily in other countries require validation before being used to provide risk-adjusted outcomes of PICU mortality for units within a new health care setting. It is also important that the calibration of these tools be reassessed periodically to ensure their continued validity. Identifying suitable risk-adjustment tools is only a first step. It now is important that they be applied effectively to monitor outcome and improve the quality of pediatric intensive care within the United Kingdom.<sup>30</sup>

## ACKNOWLEDGMENTS

This study was funded by the UK Medical Research Council (G9900013). The study sponsor played no part in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

UK PICOS Steering Group: Stephanie Black, Anthony Brady, Caroline Goldfrad, and Kathy Rowan, Intensive Care National Audit & Research Centre; Brigitte Colwell, John Brazier, Sam Jones, Chris McCabe, Jon Nicholl, Gareth Parry, and Khadija Rantell, Katherine Stevens,

University of Sheffield; Neil Bennett and Charles Stack, Sheffield Children's Hospital; David Hallworth, Royal Hospital for Sick Children, Glasgow; Philip Holland, General Infirmary at Leeds; Ian Murdoch, Guy's & St Thomas Hospital, London; Gale Pearson, The Birmingham Children's Hospital NHS Trust; Jane Ratcliffe, Royal Liverpool Children's Hospital.

UK PICOS Study Group: Participating PICUs were based at the following NHS Trusts: Addenbrookes NHS Trust, Cambridge; Birmingham Children's Hospital, Birmingham; Bristol Royal Hospital for Children, Bristol; Leeds General Infirmary, Leeds; Royal Liverpool Children's NHS Trust (Alder Hey) Liverpool; Sheffield Children's Hospital, Sheffield; Guy's & St Thomas Hospital Trust, London; Royal Hospital for Sick Children, Glasgow; Royal Hospital for Sick Children, Edinburgh; Newcastle Upon Tyne Hospitals NHS Trust, Newcastle; Royal Manchester Children's Hospital, Manchester; The John Radcliffe Hospital, Oxford; Great Ormond Street Hospital for Children, London; Royal Hospital for Sick Children, Belfast; Northern Ireland, Queen's Medical Centre, Nottingham; Southampton General Hospital, Southampton; North Staffordshire Hospital NHS Trust, Stoke-on-Trent; University Hospitals of Leicester, Leicester; and University Hospital of Wales, Cardiff.

We thank the staff of participating units for their support and time in contributing to UK PICOS.

## REFERENCES

1. Gemke RBJ. *Outcome Assessment of Pediatric Intensive Care: Principles and Applications* [doctoral thesis]. Utrecht, Netherlands: University of Utrecht; 1994
2. Pearson G, Shann F, Barry P, et al. Should paediatric intensive care be centralised? Trent versus Victoria. *Lancet*. 1997;349:1213-1217
3. National Co-ordinating Group on Paediatric Intensive Care. *Paediatric Intensive Care: A Framework for the Future*. Leeds, United Kingdom: NHS Executive Leeds; 1997
4. Black N. Why do we need observational studies to evaluate the effectiveness of health care? *Br Med J*. 1996;312:1215-1218
5. Pollack MM, Cuerdon TT, Patel KM, Ruttimann UE, Getson PR, Levetown M. Impact of quality of care factors on pediatric intensive care unit mortality. *JAMA*. 1994;272:941-946
6. UK Neonatal Staffing Study Group. A prospective evaluation of patient volume, staffing and workload in relation to risk-adjusted outcomes in a random, stratified sample of all UK neonatal intensive care units. *Lancet*. 2002;359:99-107
7. Yeh TS, Pollack MM, Ruttimann UE, Holbrook PR, Fields AI. Validation of a physiologic stability index for use in critically ill infants and children. *Pediatr Res*. 1984;2:171-179
8. Pollack MM, Ruttimann UE, Getson PR. Pediatric risk of mortality (PRISM) score. *Crit Care Med*. 1988;16:1110-1116
9. Pollack MM, Patel KM, Ruttimann UE. PRISM III: an updated pediatric risk of mortality score. *Crit Care Med*. 1996;24:743-752
10. Children's National Medical Center. PICUES: The International Standard for PICU Evaluations of Quality and Efficiency of Care. Available at: [www.dcchildrens.com/picues/](http://www.dcchildrens.com/picues/). Accessed February 9, 2006
11. Gemke RBJ, Bonsel GJ. Comparative assessment of pediatric intensive care: a national multicenter study. *Crit Care Med*. 1995;23:238-245

12. Shann F, Pearson G, Slater A, Wilkinson K. Paediatric index of mortality (PIM): a mortality prediction model for children in intensive care. *Intensive Care Med.* 1997;23:201–207
13. Slater A, Shann F, Pearson G. PIM2: a revised version of the Paediatric Index of Mortality. *Intensive Care Med.* 2003;29:278–285
14. Pearson GA, Stickley J, Shann F. Calibration of the paediatric index of mortality in UK paediatric intensive care units. *Arch Dis Child.* 2001;84:125–128
15. Bertolini G, Ripamonti D, Cattaneo A, Apolone G. Pediatric risk of mortality: an assessment of its performance in a sample of 26 Italian intensive care units. *Crit Care Med.* 1998;26:1427–1432
16. Gemke RBBJ, van Vught AJ. Scoring systems in pediatric intensive care: PRISM III versus PIM. *Intensive Care Med.* 2002;28:204–207
17. Wells M, Riera-Fanego JF, Luyt DK, Dance M, Lipman J. Poor discriminatory performance of the Pediatric Risk of Mortality (PRISM) score in a South African intensive care unit. *Crit Care Med.* 1996;24:1507–1513
18. Goddard JM. Pediatric risk of mortality scoring overestimates severity of illness in infants. *Crit Care Med.* 1992;20:1662–1665
19. Parry GJ, Tucker JS, Tarnow-Mordi WO. CRIB II: an update of the clinical risk index for babies score. *Lancet.* 2003;361:1789–1791
20. MRC/DOH Working Party on Intensive Care. *The Research Needs and Opportunities Relevant to the NHS.* London, United Kingdom: Medical Research Council; 1997
21. Young JD, Goldfrad C, Rowan K. Development and testing of a hierarchical method to code the reason for admission to intensive care units: the ICNARC Coding Method. *Br. J Anaesth.* 2001;87:543–548
22. Murphy-Filkins R, Teres D, Lemeshow S, Hosmer DW. Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model: how to distinguish a general from a specialty intensive care unit. *Crit Care Med.* 1996;24:1968–1973
23. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* 1982;247:2543–2546
24. Hosmer DW, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. *Commun Stat.* 1980;A9:1043–1069
25. Cox DR. Two further applications of a model for binary regression. *Biometrika.* 1958;45:562–565
26. Efron B. How biased is the apparent error rate of a prediction rule? *J Am Stat Assoc.* 1986;81:461–470
27. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med.* 2000;19:453–473
28. Spiegelhalter D. Funnel plots for institutional comparison. *Qual Saf Health Care.* 2002;11:390–391
29. Simpson JM, Evans N, Gibberd RW, Heuchan AM, Henderson-Smart DJ. Analysing differences in clinical outcomes between hospitals. *Qual Saf Health Care.* 2003;12:257–262
30. Paediatric Intensive Care Audit Network. National Report 2003–2004. Universities of Leeds, Leicester, and Sheffield. Sheffield, United Kingdom: PICANet; 2005
31. Slater A, Shann F, for the ANZICS Paediatric Study Group. The suitability of the Paediatric Index of Mortality (PIM), PIM2, the Paediatric Risk of Mortality (PRISM), and PRISM III for monitoring the quality of pediatric intensive care in Australia and New Zealand. *Pediatr Crit Care Med.* 2004;5:447–454
32. Nipshagen MD, Polderman KH, DeVicor D, Gemke RJ. Pediatric intensive care: result of a European survey. *Intensive Care Med.* 2002;28:1797–1803
33. Randolph AG. Paediatric index of mortality (PIM): do we need another paediatric mortality prediction score? *Intensive Care Med.* 1997;23:141–142
34. Tarnow-Mordi WO, Tucker J, Parry GJ. Should paediatric intensive care be centralised? *Lancet.* 1997;350:66–67
35. Kuhlthau K, Ferris TGG, Iezzoni LI. Risk adjustment for pediatric quality indicators. *Pediatrics.* 2004;113:210–216
36. Keizer NF, Bonsel GJ, Gemke RBBJ. Health status prediction in critically ill children: a pilot study introducing standardized health ratios. *Qual Life Res.* 1997;6:192–199
37. Ruttimann UE, Pollack MM, Fiser DH. Prediction of three outcome states from pediatric intensive care. *Crit Care Med.* 1996;24:78–85

## Assessment and Optimization of Mortality Prediction Tools for Admissions to Pediatric Intensive Care in the United Kingdom

Anthony R. Brady, David Harrison, Stephanie Black, Sam Jones, Kathy Rowan, Gale Pearson, Jane Ratcliffe, Gareth J. Parry and on behalf of the UK PICOS Study Group

*Pediatrics* 2006;117:e733-e742; originally published online Mar 1, 2006;

DOI: 10.1542/peds.2005-1853

<b>Updated Information &amp; Services</b>	including high-resolution figures, can be found at: <a href="http://www.pediatrics.org/cgi/content/full/117/4/e733">http://www.pediatrics.org/cgi/content/full/117/4/e733</a>
<b>References</b>	This article cites 32 articles, 9 of which you can access for free at: <a href="http://www.pediatrics.org/cgi/content/full/117/4/e733#BIBL">http://www.pediatrics.org/cgi/content/full/117/4/e733#BIBL</a>
<b>Citations</b>	This article has been cited by 2 HighWire-hosted articles: <a href="http://www.pediatrics.org/cgi/content/full/117/4/e733#otherarticles">http://www.pediatrics.org/cgi/content/full/117/4/e733#otherarticles</a>
<b>Subspecialty Collections</b>	This article, along with others on similar topics, appears in the following collection(s): <b>Office Practice</b> <a href="http://www.pediatrics.org/cgi/collection/office_practice">http://www.pediatrics.org/cgi/collection/office_practice</a>
<b>Permissions &amp; Licensing</b>	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: <a href="http://www.pediatrics.org/misc/Permissions.shtml">http://www.pediatrics.org/misc/Permissions.shtml</a>
<b>Reprints</b>	Information about ordering reprints can be found online: <a href="http://www.pediatrics.org/misc/reprints.shtml">http://www.pediatrics.org/misc/reprints.shtml</a>

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

