

oshpd

Office of Statewide Health Planning and Development



# California Intensive Care Outcomes Project (CALICO)

May 2007

**Arnold Schwarzenegger**

Governor  
State of California

**Kimberly Belshé**

Secretary  
Health and Human Services Agency

**David M. Carlisle, M.D., Ph.D.**

Director  
Office of Statewide Health Planning and Development

**Final Report**

**CALIFORNIA INTENSIVE  
CARE OUTCOMES  
(CALICO) PROJECT**

**May 2007**

# **CALIFORNIA INTENSIVE CARE OUTCOMES (CALICO) PROJECT**

## **Office of Statewide Health Planning and Development**

Joseph Parker, Ph.D.  
*Director, Healthcare Outcomes Center*

### **Study Consultants**

#### **University of California, San Francisco**

R. Adams Dudley, M.D., M.B.A.  
*Principal Investigator*

Michael Kuzniewicz, M.D., M.P.H.  
*Co-investigator*

Mitzi Dean, M.S., M.H.A.  
*Project Director*

Rondall Lane, M.D., M.P.H.  
*Co-investigator*

Deborah Rennie  
*Programmer/Analyst*

Peter Bacchetti, Ph.D.  
*Statistician*

Ted Clay, M.S.  
*Programmer/Analyst*

Stephen Crane, M.B.A.  
*Research Associate*

Hal Luft, Ph.D.  
*Co-investigator*

## California Health Policy and Data Advisory Commission

Vito J. Genna, Chair  
*Representing Group Long-Term Care*

Corinne Sanchez, Esq.  
*General Member*

Marjorie B. Fine, M.D., F.A.C.S.  
*General Member*

Josh Valdez, D.B.A.  
*Representing Health Insurance Industry*

Sol Lizerbram, D.O.  
*General Member*

William Brien, M.D.  
*Representing Physicians and Surgeons*

Vacant  
*Representing Disproportionate Share Hospitals*

Adama L. Iwu  
*Representing Group Prepayment Health Services Plan*

Sonia Moseley, R.N.  
*Representing Labor Health Coalitions*

Janet Greenfield, R.N.  
*Representing Freestanding Ambulatory Surgery Centers*

Jerry Royer, M.D., M.B.A.  
*Representing Hospitals*

Howard L. Harris, Ph.D.  
*General Member*

Kenneth M. Tiratira, M.P.A.  
*Representing Business Health Coalitions*

Kathleen Maestas  
*Acting Executive Director*

### California Health Policy and Data Advisory Commission Technical Advisory Committee

Jerry Royer, M.D., M.B.A., Chair  
*Representing Hospitals*

Douglas Bagley  
*Representing Hospitals*

Kathy McCaffrey, M.A., R.H.I.A.  
*Representing Medical Records Practitioner*

Marilyn Chow, R.N., D.N.Sc.  
*Representing Nurses*

Laura B. Gardner, M.D., M.P.H., Ph.D.  
*Representing Research Community*

Mark Hlatky, M.D.  
*Representing Research Community*

David E. Hayes-Bautista, Ph.D.  
*Representing Research Community*

Diana Petitti, M.D.  
*Representing Research Community*

Robert H. Brook, M.D., Sc.D.  
*Representing Research Community*

Laurie Sobel, J.D.  
*Representing Consumers Union*

Nancy Donaldson, R.N., D.N.Sc.  
*Representing California Nurses Association*

Vacant  
*Representing Physicians*

William S. Weil, M.D.  
*Representing Physicians*

Elizabeth Carolyn Abbott  
*Representing Health Access*

# Index

	<i>Page(s)</i>
1. Executive Summary.....	1-1 to 1-5
2. Literature Summary.....	2-1 to 2-9
3. Hospital and Patient Sampling.....	3-1
4. Patient Inclusion and Exclusion Criteria .....	4-1 to 4-3
5. Hospital Selection and Patient Population.....	5-1 to 5-3
6. Data Collection.....	6-1 to 6-2
7. Data Quality.....	7-1 to 7-5
8. Model Descriptions and Development.....	8-1 to 8-8
9. Customization of the Models.....	9-1 to 9-2
10. Patient Discharge Database Risk-adjustment Model.....	10-1 to 10-2
11. APACHE III System Model.....	11-1 to 11-2
12. Results I: Performance of the Original Models.....	12-1 to 12-4
13. Results II: Performance of the Customized Models.....	13-1 to 13-2
14. Results III: Performance of the PDD Models.....	14-1 to 14-3
15. Results IV: Performance of the APACHE III System Model.....	15-1 to 15-2
16. Results V: Comparative Performance of the Hospitals.....	16-1 to 16-4
17. Summary and Conclusions.....	17-1 to 17-3
Appendix.....	A-1 to A-18

# Tables and Figures:

	<i>Page(s)</i>
Table 2.1 Performance of the models in the original datasets.....	2-4
Table 2.2 Variables collected by each model.....	2-5
Table 2.3 Summary of studies assessing performance of the models.....	2-6
Table 5.1 Characteristics of hospitals included in the study.....	5-2
Table 5.2 Patients excluded from the CALICO database.....	5-2
Table 5.3 Patients excluded from the comparative analysis of the models.....	5-3
Table 5.4 Summary of patient population characteristics.....	5-3
Table 6.1 Listing of additional variables in the CALICO data collection.....	6-2
Table 7.1 Inter-rater reliability of variables collected.....	7-3 to 7-4
Table 8.1 MPM <sub>0</sub> II variables and coefficients.....	8-2
Table 8.2 SAPS II variables and point assignments.....	8-3
Table 8.3 The APACHE II severity of disease classification system.....	8-5
Table 8.4 The APACHE II weights for age and chronic health.....	8-5
Table 8.5 APACHE II diagnoses and coefficients.....	8-6
Table 11.1 Crude mortality by system and medical/surgical status.....	11-2
Table 11.2 Inter-rater reliability of the APACHE III models.....	11-2
Table 12.1 Summary of discrimination – Original models.....	12-1
Figure 12.1 Receiver operator curves for the four models.....	12-1
Table 12.2 Summary of calibration – Original models.....	12-2
Figure 12.2 Calibration curves – Original models.....	12-3
Table 13.1 Performance of the re-estimated models in the validation sample.....	13-1
Figure 13.1 Calibration curves – Re-estimated models.....	13-2
Table 14.1 Discrimination of the PDD and re-estimated models.....	14-1
Table 14.2 Calibration of the PDD models.....	14-2
Figure 14.1 Calibration curves of the PDD models.....	14-3

# Tables and Figures:

	<i>Page(s)</i>
Table 15.1 Discrimination of the APACHE III system model.....	15-1
Table 15.2 Calibration of the APACHE III system model.....	15-1
Figure 15.1 Calibration curves of the APACHE III models.....	15-2
Figure 16.1 SMR and 95% confidence interval of the hospitals.....	16-1
Table 16.1 Statistically significant high and low hospital outliers.....	16-3
Table 16.2 Rank order of hospitals' SMRs.....	16-4
Table 16.3 Spearman rank correlation coefficients.....	16-4
Table A.1 Hosmer-Lemeshow Goodness-of-Fit (C test) – Original models.....	A-1 to A-2
Table A.2 Hosmer-Lemeshow Goodness-of-Fit (H test) – Original models.....	A-3 to A-4
Table A.3 MPM <sub>0</sub> II variables and re-estimated coefficients .....	A-5
Table A.4 SAPS II variables and re-estimated coefficients .....	A-5
Table A.5 APACHE II variables and re-estimated coefficients .....	A-6
Table A.6 APACHE III variables and re-estimated coefficients .....	A-7
Table A.7 MPM <sub>0</sub> II re-estimated coefficients and odds ratios (100% Sample).....	A-8
Table A.8 SAPS II re-estimated coefficients and odds ratios (100% Sample).....	A-8
Table A.9 APACHE II re-estimated coefficients and odds ratios (100% Sample)...	A-9
Table A.10 APACHE III re-estimated coefficients and odds ratios (100% Sample)	A-10 to A-11
Table A.11 Combined APACHE II diagnostic categories.....	A-11
Table A.12 Combined APACHE III diagnostic categories.....	A-12
Table A.13 PDD model estimated coefficients and odds ratios (estimation).....	A-13 to A-15
Table A.14 PDD model estimated coefficients and odds ratios (100% sample)...	A-16 to A-17
Table A.15 Hospital Outliers by SMR method.....	A-18

# Executive Summary

## Background

The California Hospital Outcomes Project (CHOP) is an initiative mandated by the State of California and conducted by the Office of Statewide Health Planning and Development (OSHPD) to produce public reports comparing hospital outcomes for patients treated in California hospitals for selected conditions, procedures, and units. Intensive care unit (ICU) mortality was selected as a potential subject for outcome reporting by the OSHPD staff and the California Health Policy and Data Advisory Commission Technical Advisory Committee because of the high mortality rate in ICUs, data showing large variations in ICU performance as measured by risk-adjusted mortality rates, evidence that intervention could effect beneficial change in mortality, and the high cost of care in these specialized units.

There are four ICU risk adjustment models that are widely used, the Mortality Probability Model II at admission or “zero hours” (MPM<sub>0</sub>II), the Simplified Acute Physiology Score II (SAPS II), and the Acute Physiology and Chronic Health Evaluation, versions II and III (APACHE II and APACHE III). The utility of these four models in assessing the performance of modern ICUs is unknown, as they were developed from the mid-1980s to the early 1990s. Furthermore, no attempt has been made to compare these models, all of which use data obtained from chart abstractions, to models using only data already available in discharge abstracts. Since the existing mortality models vary significantly in the data burden they place on hospitals, studies are needed to determine whether models that place greater data burden on hospitals offer sufficiently greater predictive accuracy and/or data reliability to justify the additional burden.

## Objectives

The goals of the California Intensive Care Outcomes (CALICO) project were to assess the feasibility of, potential benefits from, and most efficient approach to ICU performance reporting in California. The first objective of CALICO was to evaluate the performance of MPM<sub>0</sub> II, SAPS II, APACHE II, and APACHE III by applying them to a contemporary database (2002-2004) of California ICU patients, including an audit of the reliability of the model variables and customizing the models to the California dataset to improve their goodness-of-fit. The second objective was to develop and evaluate an ICU mortality risk adjustment model based wholly or partially on OSHPD’s Patient Discharge Database (PDD), administrative data currently reported by hospitals. The third objective was to use these models to determine whether there is significant variation among project hospitals in risk-adjusted mortality for ICU patients, and hence potential for improvement in quality of care. The final objective was to compare the available models in terms of their predictive performance versus the burden of data collection—considering both the number of variables used and the sources from which those data are likely to be obtained—to identify the most efficient model or combination of models to report ICU performance.

## Methods

To achieve the project goals, demographic, clinical, and limited therapeutic data were collected on ICU patients from 33 California hospitals that volunteered to join CALICO. Each hospital was instructed to collect data on consecutive, eligible patients and continue until their target sample size, based on ICU case volume, was reached. Eligible patients were adults (18 or older) who were admitted for at least 4 hours into an adult ICU and who were not burn, trauma, or coronary bypass patients. Patients admitted to rule out myocardial infarction who were not found to have a critical illness were excluded.

Data quality was monitored throughout the project through initial and subsequent training of data collectors, automated data quality checks internal to the data collection software, and electronic screens applied to the data following data submissions. In addition, a 400 patient audit was conducted to allow calculation of inter-rater reliability statistics (percent agreement and kappas).

Mortality predictions were calculated for each patient using the four extant models with the coefficients as published by their developers and after re-estimating the models (using the same variables but recalculating the coefficients) on a 60% development sub-sample of the CALICO data. In addition, a simplified APACHE III model was developed newly for this project. This model used the APACHE III variables but, rather than using APACHE III's 94 specific reasons for admission, involved reclassifying each patient's reason for admission into one of the nine categories, eight by body system and one for overdose/poisoning (hereafter we refer to this model as the APACHE III System model).

Two models were developed that used variables available from the PDD. The first used as predictors only variables in the PDD: age, gender, primary reason for hospital admission, and other conditions present on hospital admission. The second model (PDD+ clinical) used these PDD data plus clinical variables that would be easy to collect via chart abstraction. Each of these clinical variables (heart rate, blood pressure, Glasgow coma score (GCS), need for mechanical ventilation, presence of an intracranial mass, and type of ICU admission) came from the MPM<sub>0</sub> II model. This model includes only data collected at the time of the ICU admission. This approach creates less chart abstraction burden than collecting the variables in the other models, which must be assessed over the first full day of ICU admission. To improve the calibration of the PDD+ clinical model, heart rate, blood pressure and GCS were treated as continuous variables instead of being dichotomized as they were in MPM<sub>0</sub> II.

The performance of each hospital was evaluated using standardized mortality ratios (SMRs). The expected mortality was calculated for each hospital using the re-estimated coefficients from the ICU risk-adjustment models. To get an SMR for each hospital, the observed mortality was divided by the model-specific expected mortality.

The ability of each of the models to identify outliers was evaluated in several ways. The first was to determine whether the 95% confidence interval of the SMR included 1.0.

The second approach involved a hospital fixed effects model. This method compares each hospital effect versus the un-weighted average of all the hospitals. Logistic regression was used to estimate the effect of each hospital on the overall model. Finally, for each hospital a “contrast” test between that hospital’s effect and the average effect of all the hospitals was performed.

## Results

Data were received from volunteer hospitals on 10,398 eligible patients. Of these, there were 9,441 for whom data was sufficiently complete and accurate that a risk calculation could be done across the four extant models. The analyses were performed only on these patients.

**Data quality.** An audit was performed to determine if the data collected by the CALICO hospital participants were reported with sufficient reliability to justify inclusion in the risk adjustment models. In general, agreement between the CALICO auditors and the hospital data collectors was high (>90%) and the corresponding kappa substantial (0.70 or greater) for the audited physiological variables. The most problematic variable was the Reason for Admission variable used in APACHE III, which had a kappa of only 0.51. However, the derived Reason for Admission variable that grouped Reasons for Admission by body system affected was more reliable (agreement 80%, kappa 0.73).

**Model Comparisons: Data Burden.** The PDD model would create no additional data burden for the hospitals. Of the other models, the PDD + clinical and the MPM<sub>0</sub> II require by far the least in terms of data collection costs, both because they have many fewer variables than the APACHE models and because, unlike SAPS II and the APACHE models, they only require data collection at the time of admission, rather than over the first day in the ICU. SAPS II is less burdensome than the APACHE models, because it has many fewer variables and does not require the most difficult task, selecting a reason for admission.

**Model Comparisons: Predictive Accuracy.** The original versions of the four extant models had adequate discrimination (0.811 – 0.880) but inadequate calibration ( $p < 0.001$  on Hosmer-Lemeshow C and H tests). Statistically significant “P-values” for the Hosmer-Lemeshow tests indicate that the model is poorly calibrated. All of the models over-predicted death across the ten strata of mortality risk. The PDD model had inadequate discrimination (0.774) and inadequate calibration. The PDD + clinical model had better discrimination (0.851), but the Hosmer-Lemeshow H test was still significant at the  $P < 0.0001$  level and the C Test was borderline significant ( $P = .058$ ). Visual inspection of the calibration curve showed marked over-prediction of mortality in the two highest mortality strata.

In terms of predictive accuracy, the results for the four extant models, re-estimated, and the simplified APACHE III model, suggest that the MPM<sub>0</sub> II has substantially worse discrimination than the other models and the APACHE III has the best discrimination. The calibration curves and Hosmer-Lemeshow statistics suggest that the SAPS II and the APACHE III System show better fit than any of the other models, and the APACHE III

model appears to have the worst fit considering both the HL tests and visual inspection of the calibration curves.

**Hospital comparisons.** In contrast to the fairly close agreement on each hospital's risk-adjusted mortality rate across the risk-adjustment models, there are large and statistically significant differences among the hospitals in their mortality performance using the SMR calculation. For hospitals that submitted over 100 patients, SMRs varied from approximately 0.5 to approximately 2.0, regardless of the risk-adjustment model used. This corresponds to risk-adjusted mortality rates from about 7% to about 31%.

In terms of the models' identification of hospital outliers, for 32 of the 33 hospitals, the 95% confidence intervals for the point estimates of the SMR overlapped for all models. Four of 33 hospitals were identified as outliers by all models (two high mortality outliers, two low mortality outliers) using the SMR method. The results for fixed effects models and the contrast test were similar to the SMR results in terms of the number of outliers identified and the significance of hospital-specific effects.

## Conclusions

In summary, there is sufficient evidence to justify moving forward with measuring and reporting ICU performance. In terms of risk adjustment model selection, the PDD model is the most immediately feasible, but has severe limitations in terms of discrimination and calibration and probably should not be adopted. MPM<sub>0</sub> II has significantly worse discrimination than the models other than PDD, but is much less burdensome, while APACHE III has slightly better discrimination and is preferred by the Joint Commission on Accreditation of Healthcare Organizations (JCAHO), but may have calibration and data reliability issues. SAPS II, APACHE II, and APACHE III System are between MPM<sub>0</sub> II and APACHE III in terms of discrimination but have better calibration and data reliability. Since SAPS II does better than APACHE II on all criteria, it seems reasonable to drop this model from consideration. The PDD plus clinical model is similar to SAPS II in its discrimination but has worse calibration in a way that would be expected to influence which hospitals are labeled outliers.

In considering congruence with the national use of ICU risk models, JCAHO has announced that it will be developing ICU core measures. JCAHO plans to include a mortality model and is currently most seriously considering the APACHE III model. JCAHO has also recently noted data reliability problems with the APACHE III reason for admission variable.

Thus, the choice is among:

- PDD plus clinical: least burdensome model to get good discrimination at the cost of calibration,
- MPM<sub>0</sub> II: similar burden with worse discrimination but slightly less problematic calibration,
- SAPS II: better predictive accuracy than MPM<sub>0</sub> II with less burden than APACHE or APACHE III System, but not in JCAHO's plans,

- APACHE III: burdensome, with data reliability issues, but aligned with JCAHO, or
- APACHE III System: between SAPS II and APACHE III in burden, good discrimination, calibration, and data reliability and entirely calculable from the variables JCAHO is currently beta testing.

The CALICO investigators believe this choice should be made after further public discussion. This discussion should include consideration of additional testing of these models and the reliability of the reason for admission versus the system variable in a broader sample of hospitals.

# Literature Summary

## Importance of ICU

Intensive care units became common in hospitals in the United States during the polio epidemics of the 1950s. With the development of mechanical ventilators to assist breathing, a separate, specialized unit within the hospital was needed. Further developments in medical technology necessitated the expansion of intensive care to facilitate the handling of other types of severe organ dysfunction.

The modern intensive care unit (ICU) is the highest mortality unit in any hospital. There are approximately 4 million ICU admissions per year in the United States with average mortality rates reported ranging from 8-19%, or about 500,000 deaths annually.<sup>1,2,3,4</sup> This mortality rate is higher than for any condition or procedure, with the exception of myocardial infarction, for which California hospital performance reports have previously been developed. There is wide variability in the performance of ICUs with studies showing 2- to 3-fold variations in mortality rates, after adjusting for patient risk factors.<sup>5,6,7</sup> The ICU is also one of the sites in which medical errors are most likely to occur because of the complexity of care.<sup>8,9,10,11</sup> Since the patient population is severely ill and undergoes multiple complex interventions at the same time, these patients are extremely vulnerable to experiencing adverse outcomes.<sup>12,13,14</sup> In addition to its impact on mortality, the ICU is an expensive component of the national healthcare budget, accounting for approximately 10% of the total hospital budget.<sup>15</sup>

Based on the clinical significance of the ICU, the Joint Commission on Accreditation of Hospital Organizations (JCAHO), and the Leapfrog Group have decided that ICU care is a priority. The National Quality Forum and the Leapfrog Group are considering requiring or recommending ICU performance reporting. JCAHO has already announced its intention to make ICU risk-adjusted mortality a core measure of performance. Given the levels of mortality and apparent variations in performance, much may be gained from public reporting of ICU outcomes. Hospitals and clinicians would have benchmarks to use in setting quality improvement goals. Consumers could incorporate the information into decisions about their choice of hospital and perhaps choice of health plan or primary physician. However, these potential benefits can only be realized if ICU performance can be assessed accurately.

## Extant ICU Mortality Models

Clinicians and researchers have long recognized how important ICU performance is to overall hospital mortality. A significant amount of work has already been done to develop tools to assess ICU performance. This work has focused primarily on the development of general predictive models to compare observed versus expected mortality

rates across a wide range of patients. Disease-specific models and organ dysfunction/failure models have also been developed. The disease-specific models have the advantage of identifying variables for a particular disease that will affect outcome, theoretically improving the predictive power of a model compared to models developed for all ICU patients. However, studies have shown mixed results, in terms of improving discrimination, for disease-specific models versus general models.<sup>16</sup> Since disease-specific models only predict outcomes for a specific disorder, they cannot be used for all ICU patients. As a consequence, unlike the general models, they cannot be used to examine mortality of an entire ICU population or compare overall performance among ICUs.

Organ dysfunction/failure models were developed for the quantification of multiple organ dysfunction syndrome and have shown a good correlation between the presence and duration of organ failure and outcome.<sup>17</sup> These models are often relatively simple and are used to describe individual organ dysfunction/failure in a continuous form, from mild dysfunction to severe failure. They can be used over a period of time to monitor the progression of individual (or aggregated) organ dysfunction.<sup>16</sup> These models are intended to be used to describe morbidity and have not been shown to be accurate for mortality prediction. Consequently, general ICU mortality prediction models are currently the most effective in evaluating the performance of ICUs.

Four general ICU mortality risk adjustment models are widely used; the Acute Physiology and Chronic Health Evaluation, versions II and III (APACHE II and III), the Simplified Acute Physiology Score II (SAPS II) and the Mortality Probability Model II at admission or “zero hours” (MPM<sub>0</sub> II).<sup>4,18,19,20</sup> The models have been used in more than 2,000 publications in the medical literature. These second and third generation models represent an effort to improve the performance of the original models, which are no longer in wide use. The performance of these four models is summarized in Table 2.1. Chapter 8 details the specific steps in model development.

These models were developed in the mid-1980s to early 1990s. While the models do utilize some of the same variables, they differ in the number of variables collected, the type of variables, and specification of the variables. Table 2.2 lists the variables collected in each model. The difference in the number of variables collected and the difficulty in abstracting the variables has a major impact on the time required to abstract the variables and ultimately the cost of generating mortality predictions. The APACHE II model includes 9 physiologic variables, 9 chronic health variables, age, source of admission, and 50 reason for admission categories. The revised APACHE III model contains 13 physiologic variables, 8 chronic health variables, 7 categories for age, source of admission, and 94 reason for admission categories.<sup>4</sup> SAPS II uses 12 physiologic variables, age, a chronic health variable, and type of admission (surgical vs. medical).<sup>19</sup> MPM<sub>0</sub> II uses 10 physiologic variables, 3 chronic health variables, age and type of admission.<sup>20</sup> Neither the MPM<sub>0</sub> II model nor the SAPS II requires the data collector to determine a reason for admission to the ICU. Since the development of the models, their developers have revised and expanded their work, although the risk adjustment models are still calculated in essentially the same manner. Data collection to revise the SAPS model is currently ongoing in Europe (see [www.saps3.org](http://www.saps3.org)); the developers are assessing

the possibility of adding a single reason for admission as well as the presence of infection at admission. However, most of their revisions are aimed at developing a multi-dimensional model of ICU operations and investigating other outcomes and cost-effectiveness.

## **Use of ICU Mortality Models for Performance Evaluation**

Among the general ICU models, no model is obviously superior to the others for the purpose of ICU performance evaluation. The APACHE models are the most widely used, in part because their complexity gives them greater clinical plausibility. To compare the models, one must assess the calibration and discrimination of each. Discrimination is assessed with the area under the receiver operating characteristic curve (AUC). Calibration is evaluated by Hosmer-Lemeshow statistics. These statistics test whether one can reject the null hypothesis that the model fits well across deciles of risk, so that on average, people with high predicted values have comparable mortality rates, and contrariwise for those with low predicted values. The relative importance of model calibration versus model discrimination depends on the intended use of the model. Both assessments are needed to identify a well-fit model. Hosmer and Lemeshow have argued that if a model does not calibrate well, it is meaningless to examine discrimination.<sup>21</sup> Calibration is especially important if the model is anticipated to be used to compare predicted and actual death rates, and thus compare performance across hospitals, especially if the risk profile varies among hospitals.

All the general models seem to have performed well on the populations on which they were developed, although we were unable to find reports of the calibration of the APACHE models on their developmental or validation datasets. For all models, AUCs were 0.74 or better. The MPM<sub>0</sub> II and SAPS II studies, using Hosmer-Lemeshow statistics, demonstrated good calibration (MPM<sub>0</sub> II model P= 0.623 and P= 0.327 in the developmental and validation samples respectively, SAPS II, P= 0.883 and P=0.104). However, when applied to populations other than the ones on which they were developed and validated, all four models discriminate adequately but calibrate poorly.<sup>22,23,24,25,26,27,28,29,30,31,32,33,34</sup> When comparing APACHE III to other models, especially APACHE II, the additional variables added to APACHE III that distinguish it from APACHE II lead to increased discrimination. However, prior investigators have found that this discriminatory power comes at a cost of poorer calibration in most cases.<sup>32</sup> Table 2.3 is a summary of studies assessing and comparing the calibration and discrimination of the models.

A number of hypotheses have been put forward to account for these findings, including differences in the definition and collection of the data, real differences in the patient populations (case-mix), lead time bias, lack of important predictive variables or interactions between the variables in the models, pre-ICU or post-ICU management, or a lack of validity of the dependent variable.<sup>16,35,36</sup> Some studies have shown that the models do not calibrate well because they underestimate the mortality of low-risk patients and overestimate the mortality of high-risk patients.<sup>35</sup> Other researchers have shown the problematic nature of utilizing the worst physiologic variables over the first 24 hours of an ICU admission, knowing that physiologic data can be strongly influenced by medical

and nursing intervention.<sup>37</sup> As a result, patients treated inappropriately in the first 24 hours after admission may receive higher mortality scores, even when their risk of mortality at admission was lower. Lead-time bias is an issue, as a large proportion of the patients admitted to an ICU come from the emergency department. Consequently, differences in the treatment of individuals in the emergency department will affect the degree of a patient's physiologic derangements at ICU admission and thus their mortality score.<sup>38</sup> Therefore, hospitals that do a superior job stabilizing their patients in the emergency department may appear to have a case-mix with lower predicted mortality. Finally, critical care practice, technology and knowledge have changed significantly since the development of the newest models more than 10 years ago and the coefficients used in the models need to be reassessed to reflect modern practices and outcomes.

Approaches to resolving problems through model innovation, such as restricting mortality assessment to patients above a specific risk threshold, eliminating transfer patients, or assessing clinically defined subgroups from the assessed population, have not, to our knowledge, been attempted for ICU mortality models. Although some of the original models have been shown to have reasonable discrimination at the patient level and adequate calibration among deciles of patients, there is legitimate concern about their usefulness when comparing ICUs to each other without explicitly considering ICU thresholds for admission. Some hospitals may have estimated risks above or below their true risks because they use different admission criteria for their ICUs patients.

Highly predictive risk models have already been developed for pneumonia and myocardial infarction in Californian populations, as well as congestive heart failure (CHF) and pneumonia in national populations.<sup>39,40,41,42,43,44,45</sup> In reviewing the literature on condition-specific mortality risk models for conditions that have a high prevalence in the ICU and/or that result in high mortality, several key risk factors for these conditions were not included in the general ICU mortality models (e.g., ejection fraction <40, diabetic complications). A model that includes some of these factors may perform better than any of the existing models.

**Table 2.1**  
**Performance of the models in the original datasets**

Model	MPM <sub>0</sub> II	SAPS II	APACHE II	APACHE III
Year of Publication	1993	1993	1985	1992
Hospital Mortality	20.8%	Varied from 13.8% in Switzerland to 32.4% in the UK	Varied by diagnosis	Varied from 0.9% for drug overdose to 65.9% for cardiogenic shock
Discrimination (AUC)*	0.837 for developmental 0.824 for validation	0.88 for developmental 0.86 for validation	0.863	0.90
Calibration (H-L Statistic)†	P = 0.623 developmental P = 0.327 validation	P = 0.883 developmental P = 0.104 validation	Not given	Not given
Data Reliability (% agreement)	96-99% for dichotomous variables, 63% for chronic renal insufficiency, 80-85% for other variables	81% for potassium; >87% for other variables	96% on physiologic variables; less on preadmission data	90% on APS‡; 85.7 -99.5% for other variables

\*= Area under the receiver operating characteristic curve

†= Hosmer-Lemeshow Statistic; C test for MPM, H test for SAPS

‡= Acute Physiology Score

**Table 2.2**  
**Variables collected by each model**

VARIABLES	MPM <sub>0</sub> II	SAPS II	APACHE II	APACHE III
<b>Chronic Health Status</b>				
AIDS		✓	✓	✓
Cirrhosis	✓		✓	✓
Lymphoma			✓	✓
Hematologic malignancy		✓	✓	
Leukemia			✓	✓
Hepatic failure			✓	✓
Metastatic cancer	✓	✓	✓	✓
Immunosuppression			✓	✓
Chronic renal insufficiency	✓		✓	✓
<b>Physiology</b>				
Temperature		✓	✓	✓
Heart rate	✓	✓	✓	✓
Respiratory rate			✓	✓
Blood pressure	✓	✓	✓	✓
White blood cell count		✓	✓	✓
Albumin				✓
Bilirubin		✓		✓
Electrolytes		✓	✓	✓
Blood urea nitrogen		✓		✓
Creatinine			✓	✓
Urine output		✓		✓
Blood gas		✓	✓	✓
Glasgow coma score	✓	✓	✓	✓
<b>Acute Diagnoses</b>				
Acute renal failure	✓			
Arrhythmias	✓			
Cerebrovascular accident	✓			
GI bleeding	✓			
Leukemia	✓			
Infection				
Intracranial mass effect	✓			
Select one from a list of ...			50 diagnoses	94 diagnoses
<b>Other</b>				
Age	✓	✓	✓	✓
Patient origin			✓	✓
CPR prior to ICU admission	✓			
Mechanical ventilation	✓			
Vasoactive drug therapy	✓			

**Table 2.3**  
**Summary of studies assessing performance of the models**

Study Authors	Pub. Year	Location	Hospital Mortality	# of Patients	Model	AUCs*	H-L Statistic†	Test	P-value
Livingston et. al.	2000	Scotland	29.40%	10,334	SAPS II	0.78	142.0	C	< 0.05
				10,393	MPM <sub>0</sub> II	0.74	451.8	C	< 0.05
				9,848	APACHE II	0.76	67.4	C	< 0.05
				9,848	UK APACHE II ‡	0.76	236.8	C	< 0.05
				10,326	APACHE III	0.80	365.7	C	< 0.05
Zimmerman et. al.	1998	USA	12.35%	36,668	APACHE III	0.89	35.8	C	< 0.0001
							48.7	H	< 0.0001
Moreno et. al.	1998	Europe	20.00%	10,027	SAPS II	0.82	208.4	C	< 0.0001
							218.2	H	< 0.0001
					MPM <sub>0</sub> II	0.79	368.2	C	< 0.0001
							437.1	H	< 0.0001
Pappachan et. al.	1999	England	25.90%	12,793	APACHE III	0.89	332.9	C	< 0.01
							312.5	H	< 0.01
Beck et. al.	1997	UK	26.11%	1,144	APACHE II	0.80	98.6	C	< 0.05
					APACHE III	0.85	129.8	C	< 0.05
Markgraf et. al.	2000	Germany	18.50%	2,661	SAPS II	0.85	20.5	C	< 0.01
				2,795	APACHE II	0.83	11.8	C	> 0.1
				2,661	APACHE III	0.85	48.1	C	< 0.001
Rowan et. al.	1993	Britain & Ireland	27.70%	8,796	APACHE II	0.83	79.8	H	< 0.05
Castella	1995	Europe & N. America	21.80%	4,099	SAPS II	0.85	n/a	C	0.0244
							n/a	H	0.1019
					MPM <sub>0</sub> II	0.81	n/a	C	0.072
							n/a	H	0.0148
					APACHE II	0.86	n/a	C	0.0245
							n/a	H	0.0074
APACHE III	0.86	n/a	C	n/a					
n/a	H	n/a							
Apolone et. al.	1996	Italy	34.00%	1,393	SAPS II	0.8	71.0	H	< 0.001
Moreno et. al.	1997	Portugal	32.00%	982	SAPS II	0.82	28.3	C	0.002
							29.7	H	0.001
					APACHE II	0.79	49.7	C	< 0.0001
							32.7	H	0.0003
Metnitz et. al.	1999	Austria	19.50%	1,733	SAPS II	0.81	91.8	C	< 0.0001
							89.1	H	< 0.0001
Bastos et. al.	1996	Brazil	34.00%	1,734	APACHE II	0.82	n/a	n/a	n/a
Rivera-Fernandez	1998	Spain	21.10%	10,929	APACHE III ‡	0.83	12.27	H	> 0.5

\* = Area under the receiver operating characteristic curve      † = Hosmer-Lemeshow statistic ; C Test or H Test  
‡ - these models were modified by having the coefficients for the variables re-estimated based on the study population

## References:

- <sup>1</sup> Wu AW, Pronovost P, Morlock L. ICU incident reporting systems. *J Crit Care.* Jun 2002;17(2):86-94.
- <sup>2</sup> Angus DC, Linde-Zwirble WT, Sirio CA, et al. The effect of managed care on ICU length of stay: implications for Medicare. *JAMA.* Oct 2 1996;276(13):1075-1082.
- <sup>3</sup> Young MP, Birkmeyer JD. Potential reduction in mortality rates using an intensivist model to manage intensive care units. *Eff Clin Pract.* Nov-Dec 2000;3(6):284-289.
- <sup>4</sup> Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest.* Dec.1991;100(6):1619-1636.
- <sup>5</sup> Gunning K, Rowan K. ABC of intensive care: outcome data and scoring systems. *BMJ.* Jul 24 1999;319(7204):241-244.
- <sup>6</sup> Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. *Ann Intern Med.* Mar 1986;104(3):410-418.
- <sup>7</sup> Shortell SM, Zimmerman JE, Gillies RR, et al. Continuously improving patient care: practical lessons and an assessment tool from the National ICU Study. *QRB Qual Rev Bull.* May 1992;18(5):150-155.
- <sup>8</sup> Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients. Results of the Harvard Medical Practice Study I. *N Engl J Med.* Feb 7 1991;324(6):370-376.
- <sup>9</sup> Cullen DJ, Sweitzer BJ, Bates DW, Burdick E, Edmondson A, Leape LL. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. *Crit Care Med.* Aug 1997;25(8):1289-1297.
- <sup>10</sup> Thomas EJ, Studdert DM, Burstin HR, et al. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care.* Mar 2000;38(3):261-271.
- <sup>11</sup> Andrews LB, Stocking C, Krizek T, et al. An alternative strategy for studying adverse events in medical care. *Lancet.* Feb 1 1997;349(9048):309-313.
- <sup>12</sup> Giraud T, Dhainaut JF, Vaxelaire JF, et al. Iatrogenic complications in adult intensive care units: a prospective two-center study. *Crit Care Med.* Jan 1993;21(1):40-51.
- <sup>13</sup> Pronovost P, Wu AW, Dorman T, Morlock L. Building safety into ICU care. *J Crit Care.* Jun 2002;17(2):78-85.
- <sup>14</sup> Joint Commission on Accreditation of Healthcare Organizations. What every hospital should know about sentinel events. Oakbrook Terrace, IL: *JCAHO*; 2000.
- <sup>15</sup> Wagner DP, Wineland TD, Knaus WA. The hidden costs of treating severely ill patients: charges and resource consumption in an intensive care unit. *Health Care Financ Rev.* Fall 1983;5(1):81-86.
- <sup>16</sup> Moreno R. The "new" scores: what problems have been fixed, and remain? *Current Opinion in Critical Care.* 2000;6:158-165.
- <sup>17</sup> Knaus WA, Draper EA, Wagner DP, Zimmerman JE. Prognosis in acute organ-system failure. *Ann Surg.* Dec 1985;202(6):685-693.
- <sup>18</sup> Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med.* Oct 1985;13(10):818-829.
- <sup>19</sup> Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA.* Dec 22-29 1993;270(24):2957-2963.
- <sup>20</sup> Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA.* Nov 24 1993;270(20):2478-2486.
- <sup>21</sup> Mourouga P GC, Rowan KM. Does it fit? Assessment of scoring systems. *Current Opinion in Critical Care.* 2000;6:176-180.

- 
- <sup>22</sup> Apolone G, Bertolini G, D'Amico R, et al. The performance of SAPS II in a cohort of patients admitted to 99 Italian ICUs: results from GiViTI. Gruppo Italiano per la Valutazione degli interventi in Terapia Intensiva. *Intensive Care Med.* Dec 1996;22(12):1368-1378.
- <sup>23</sup> Moreno R, Morais P. Outcome prediction in intensive care: results of a prospective, multicentre, Portuguese study. *Intensive Care Med.* Feb 1997;23(2):177-186.
- <sup>24</sup> Moreno R, Reis Miranda D, Fidler V, Van Schilfgaarde R. Evaluation of two outcome predictors on an independent database. *Crit Care Med.* 1998;26:50-61.
- <sup>25</sup> Metnitz PG, Valentin A, Vesely H, et al. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. Simplified Acute Physiology Score. *Intensive Care Med.* Feb 1999;25(2):192-197.
- <sup>26</sup> Rowan KM, Kerr JH, Major E, McPherson K, Short A, Vessey MP. Intensive Care Society's APACHE II study in Britain and Ireland--II: Outcome comparisons of intensive care units after adjustment for case mix by the American APACHE II method. *BMJ.* Oct 16 1993;307(6910):977-981.
- <sup>27</sup> Bastos PG, Sun X, Wagner DP, Knaus WA, Zimmerman JE. Application of the APACHE III prognostic system in Brazilian intensive care units: a prospective multicenter study. *Intensive Care Med.* Jun 1996;22(6):564-570.
- <sup>28</sup> Rivera-Fernandez R, Vazquez-Mata G, Bravo M, et al. The Apache III prognostic system: customized mortality predictions for Spanish ICU patients. *Intensive Care Med.* Jun 1998;24(6):574-581.
- <sup>29</sup> Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C, Knaus WA. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. *Crit Care Med.* Aug 1998;26(8):1317-1326.
- <sup>30</sup> Castella X, Artigas A, Bion J, Kari A. A comparison of severity of illness scoring systems for intensive care unit patients: results of a multicenter, multinational study. The European/North American Severity Study Group. *Crit Care Med.* Aug 1995;23(8):1327-1335.
- <sup>31</sup> Livingston BM, MacKirdy FN, Howie JC, Jones R, Norrie JD. Assessment of the performance of five intensive care scoring models within a large Scottish database. *Crit Care Med.* Jun 2000;28(6):1820-1827.
- <sup>32</sup> Markgraf R, Deutschinoff G, Pientka L, Scholten T. Comparison of acute physiology and chronic health evaluations II and III and simplified acute physiology score II: a prospective cohort study evaluating these methods to predict outcome in a German interdisciplinary intensive care unit. *Crit Care Med.* Jan 2000;28(1):26-33.
- <sup>33</sup> Beck DH, Taylor BL, Millar B, Smith GB. Prediction of outcome from intensive care: a prospective cohort study comparing Acute Physiology and Chronic Health Evaluation II and III prognostic systems in a United Kingdom intensive care unit. *Crit Care Med.* Jan 1997;25(1):9-15.
- <sup>34</sup> Pappachan JV, Millar B, Bennett ED, Smith GB. Comparison of outcome from intensive care admission after adjustment for case mix by the APACHE III prognostic system. *Chest.* Mar 1999;115(3):802-810.
- <sup>35</sup> Beck DH, Smith GB, Taylor BL. The impact of low-risk intensive care unit admissions on mortality probabilities by SAPS II, APACHE II and APACHE III. *Anaesthesia.* Jan 2002;57(1):21-26.
- <sup>36</sup> Glance LG, Osler TM, Papadakos P. Effect of mortality rate on the performance of the Acute Physiology and Chronic Health Evaluation II: a simulation study. *Crit Care Med.* Oct 2000;28(10):3424-3428.
- <sup>37</sup> Boyd O, Grounds RM. Physiological scoring systems and audit. *Lancet.* Jun 19 1993;341(8860):1573-1574.
- <sup>38</sup> Tunnel RD, Millar BW, Smith GB. The effect of lead time bias on severity of illness scoring, mortality prediction and standardized mortality ration in intensive care – a pilot study. *Anaesthesia.* 1998;53:1045-1053.

- 
- <sup>39</sup> Haas J, Romano P, Dean ML, Hung Y, Bacchetti P. Report for the California Hospital Outcomes Project, Community-Acquired Pneumonia, 1996. Sacramento, California: Office of Statewide Planning and Development, November 2000.
- <sup>40</sup> Haas J, Dean ML, Hung Y, Rennie DJ. Differences in mortality among patients with community-acquired pneumonia in California by ethnicity and hospital characteristics. *Am J Med.* Jun 1 2003;114(8):660-664.
- <sup>41</sup> Luft HS, Romano PS. Report on Heart Attack, 1991-1993, Volume 3: Detailed Statistical Results. Sacramento, CA. California Office of Statewide Health Planning and Development, December 1997.
- <sup>42</sup> Rosenthal GE, Harper DL, Shah A, Covinsky KE. A regional evaluation of variation in low-severity hospital admissions. *J Gen Intern Med.* Jul 1997;12(7):416-422.
- <sup>43</sup> Alexander M, Grumbach K, Selby J, Brown AF, Washington E. Hospitalization for congestive heart failure. Explaining racial differences. *JAMA.* Oct 4 1995;274(13):1037-1042.
- <sup>44</sup> Fine MJ, Orloff JJ, Arisumi D, et al. Prognosis of patients hospitalized with community-acquired pneumonia. *Am J Med.* May 1990;88(5N):1N-8N.
- <sup>45</sup> LaCroix AZ, Lipson S, Miles TP, White L. Prospective study of pneumonia hospitalizations and mortality of U.S. older people: the role of chronic conditions, health behaviors, and nutritional status. *Public Health Rep.* Jul-Aug 1989;104(4):350-360.

# Hospital and Patient Sampling

The participating hospitals were volunteers responding to a series of letters sent to all acute care hospitals in the state. An effort was made to ensure the sample included a diverse group of hospitals comprising teaching and non-teaching, urban and rural, and a variety of governance and ownership arrangements.

Hospitals collected data on consecutive patients who were discharged from the hospital after having an eligible stay in the ICU. Hospitals with up to 1,200 ICU admissions per year were asked to provide data on 200 patients, hospitals with between 1,200 and 2,400 admissions per year provided data on 400 patients and hospitals with more than 2,400 ICU admissions per year provided data on 600 patients. Patient exclusions are described in detail in the next chapter.

Overall, 33 hospitals participated in the study. Twenty-four hospitals participated in the 2002-2003 data collection and 22 hospitals in the 2003-2004 data collection. Eleven hospitals participated in both years. Three hospitals had more than 2,400 ICU admissions a year, 11 hospitals had 1,200-2,400 admissions, and 19 had 1,200 or fewer admissions. Six hospitals provided less than 100 patient records that could be used across all models; these hospitals were not included in hospital-to-hospital comparative analyses.

## **Sample Size**

Statistical concerns led to the adoption of a minimum targeted sample size of at least 200 patients per hospital. Below this level, confidence intervals on the observed over expected deaths at the hospital level may get wide. A graded sample size requirement was used in order to explore the effect of severity and case-mix differences at larger hospitals and to avoid unduly burdening smaller hospitals. If all the hospitals contributed only 200 patients, more than 80 hospitals would be needed to achieve a total sample size similar to that used by APACHE III. Furthermore, larger hospitals are likely to have more heterogeneity in patient populations and may be more likely to have concentrations of unusual patients (e.g., organ transplant patients, oncology patients). Addressing these concerns required more than 200 patients from larger hospitals.

# Patient Inclusion and Exclusion Criteria

Each hospital collected information on consecutive patients admitted to their intensive care unit. The intent of the study was to evaluate the performance of existing ICU risk-adjusted mortality prediction models. As a result, the inclusion and exclusion criteria reflect the parameters already established by these pre-existing models. Patients were excluded from the analysis of a model if they did not have all the required data elements to calculate a mortality score for that model. Analysis was carried out both by comparing the models using a group of patients who met inclusion criteria for all four models and by evaluating the models separately using all patients who met inclusion criteria for each particular model, regardless of their inclusion/exclusion status in the other models.

## **Inclusion criteria for all models:**

### **1. Age 18 or older**

The study included adults only. The APACHE models were developed on a patient population  $\geq 16$  years old.<sup>1,2</sup> SAPS II<sup>3</sup> and MPM II<sup>4</sup> were developed on a population  $\geq 18$  years old. The clinical spectrum of diseases for children is significantly different from adult illnesses and would require recalibration of the existing models.

### **2. 1:1 or 2:1 patient: RN staffing**

California law restricts patient:nurse ratios in ICU to 1:1 or 2:1 staffing. Secondary to bed availability issues, patients sometimes “board” in the ICU, where they are physically in the ICU but do not require ICU care. For these patients, the patient:nurse ratio is typically greater than 2:1. The study’s purpose is to evaluate how these models predict ICU performance, so only patients requiring ICU care were included. Consequently, we excluded patients admitted to an ICU room with a ratio of more than two patients per RN.

### **3. Admitted to an adult ICU**

Patients admitted to pediatric ICUs may have significantly different risk of mortality for a given condition compared to adult ICUs due to the vastly different spectrum of disease and clinical expertise.

### **4. Stay in the ICU for at least four hours**

The outcomes of individuals admitted to the ICU for less than four hours often reflect the care prior to the ICU admission. Such short stays are usually ended either by death (often reflecting irresolvable problems prior to admission) or transfer to another unit (often reflecting a change in the patient’s clinical status).

## **Exclusion criteria for all models:**

### **1. Burn patients**

Burn patients were excluded from the original SAPS II, MPM II, and APACHE model development populations. Physiologic and clinical variables to predict mortality in burn patients are considerably different than those used to predict mortality in a general ICU population. Often these patients are treated in separate, specialized units. Furthermore, specific prognostic systems have been previously developed for this subset of patients.

### **2. Trauma patients**

Trauma patients were excluded from the CALICO data collection, even though they were included in the original SAPS II, MPM II, and APACHE models. Currently, in most parts of the United States, trauma patients who are critically ill go to designated regional trauma centers. Thus, those centers would have trauma patients while other hospitals in the region would not. Since the goal of public performance reports is to allow consumers and others to compare hospitals on their treatment of similar groups of patients, it seemed inappropriate to include trauma patients. Furthermore, specific prognostic systems have been previously developed for trauma patients and would be more useful for assessing the performance of regional trauma centers (if this were desired) than general ICU models.

### **3. Coronary artery bypass grafting surgery (CABG)**

CABG patients represent a specialized group whose physiologic derangements do not predict the same risk of mortality as other patients in the ICU. California already has a public reporting system for CABG, and the condition-specific risk adjustment model OSHPD uses in these reports is likely to have better predictive power than any general ICU model. Therefore, these patients were excluded from CALICO.

### **4. Patients admitted to rule out myocardial infarction that are found within 24 hours of ICU admission to *not* have a myocardial infarction or another critical illness**

Individuals who “rule out” for myocardial infarction (MI) essentially are admitted to the ICU for monitoring of chest pain or a similar symptom. When this symptom is not due to myocardial ischemia (or another accepted reason for ICU admission, such as rupture of a thoracic aortic aneurysm), their risk of death is close to zero. Thus, variation in hospital policies about what percentage of patients are admitted to rule out for MI could have a large influence on calculated performance (hospitals that admitted many such patients would have lower than predicted mortality). Since such policies are known to vary and could significantly affect performance, we excluded rule out MI patients from the model if they did not have an MI or other critical illness.

## 5. Readmissions

Readmissions to the ICU during the same stay were collected in the 2002-2003 dataset but not used for modeling since interventions during the first admission may impact the patient's risk of mortality in the second admission. Patients who were readmitted to the ICU were not collected in the 2003-2004 dataset.

## MPM II and SAPS II Exclusion Criteria:

### 1. Cardiac Surgery

Both MPM and SAPS excluded cardiac surgery patients. In APACHE III, cardiac surgery patients are included but have a separate prediction model that has not been released to the public.

## References:

---

<sup>1</sup> Knaus WA, Draper EA, Wagner DP, Zimmerman J. APACHE II: A severity of disease classification system. *Crit Care Med.* 1985;13:818-829.

<sup>2</sup> 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.

<sup>3</sup> Le Gall J, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA.* Dec 22-29 1993;270(24):2957-2963.

<sup>4</sup> Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA.* 1993; 270:2478-2486.

# Hospital Selection and Patient Population

## Hospital Selection

All hospitals with an eligible ICU in California were sent a recruitment packet including support letters from OSHPD, a letter from the project Principal Investigator, and materials describing the project. In addition, conference calls and presentations were made at the hospital level and at various ICU-related meetings. Follow-up materials further explaining the CALICO project were sent to hospitals that expressed interest. The Project Director called each potential ICU participant. If there was interest in the project, the Principal Investigator participated in a conference call with the appropriate hospital staff, including the ICU physician in charge, ICU nurse managers and quality improvement staff from the hospital. Particular attention was paid to the recruitment of hospital systems, as decision making can be more complex at these institutions. A summary of the characteristics of the 33 hospitals that participated are found in Table 5.1.

## Patient Population

A total of 11,612 charts were abstracted. We excluded 25 patients for being under 18 years old or having a missing age; 279 patients because they were burn, trauma, or CABG patients; 570 patients because they were readmissions to the ICU; 42 patients due to missing data on vital status at hospital discharge; 119 patients who had duplicate records; and 94 patients whose ICU admission was less than 4 hours. The patient's status at discharge was also checked against OSHPD's patient discharge database. If there were discrepancies between vital status at discharge the patient was excluded. This resulted in an additional 85 exclusions. The exclusions are summarized in Table 5.2. After the exclusions, there was a database of 10,398 patients, from 33 different hospitals. In order to compare the performance of the models, patients were excluded if they were missing data required to calculate a mortality prediction for any of the models or if there was specific exclusion for one of the models. Consequently, an additional 957 patients were excluded, leaving 9,441 patients to use for comparison of the models. The additional exclusions are summarized in Table 5.3.

The mean hospital mortality in the population was 15.2%. The mean age of patients was 62.1 years (standard deviation 17.4 years) with a median of 64 years and mode of 74 years. The youngest patient was 18 and the oldest 104. Approximately 53% of the patients were male. Seventy-nine percent of the population was admitted for a medical reason, while 17% were admitted for elective surgery and 4% for emergency surgery. A summary of the characteristics of the population by vital status at hospital discharge is presented in Table 5.4.

**Table 5.1**  
**Characteristics of hospitals included in the study**

<b>Hospital Characteristics</b>	<b># of hospitals (N=33)</b>	<b>%</b>
Rural	4	12.1%
Teaching	5	15.2%
Separate CCU	17	51.5%
<b>Licensed beds</b>		
0-99	2	6.3%
100-149	4	12.5%
150-199	5	15.6%
200-299	9	28.1%
300-499	7	21.9%
500+	5	15.6%
<b>Total critical care beds</b>		
0-20	16	48.5%
21-40	10	30.3%
41-60	4	12.1%
60+	3	9.1%
<b>Location (HSA)</b>		
Northern California	2	6.3%
Golden Empire	1	3.1%
North Bay	2	6.3%
East Bay	5	15.6%
North San Joaquin	2	6.3%
Santa Clara	1	3.1%
Mid-coast	2	6.3%
Central	5	15.6%
Santa Barbara/Ventura	1	3.1%
Los Angeles County	5	15.6%
Inland Counties	3	9.4%
Orange County	1	3.1%
San Diego/Imperial	2	6.3%

**Table 5.2**  
**Patients excluded from the CALICO database**

<b>Total patient charts abstracted</b>	<b>11,612</b>
<b>Reason for exclusion</b>	
Age less than 18, missing age	25
CABG, burn, and trauma patients	279
Duplicate records	119
Readmissions to the ICU	570
ICU admission less than 4 hours	94
Missing outcome variable (alive or dead at hospital discharge)	42
Outcome variable inconsistent with OSHPD PDD	85
<b>Total exclusions</b>	<b>1,214</b>
<b>Total patients in CALICO database</b>	<b>10,398</b>

**Table 5.3**  
**Patients excluded from the comparative analysis of the models**

<b>Total patients in the CALICO database</b>	<b>10,398</b>
<b>Reason for exclusion*</b>	
ICU stay less than 8 hours (APACHE II)	129
Missing APACHE II diagnosis	191
Missing APACHE III diagnosis	341
Cardiac surgery (MPM <sub>0</sub> II and SAPS)	129
Incomplete GCS score (APACHE III)	263
<b>Total exclusions</b>	<b>1,053 (957<sup>†</sup>)</b>
<b>Total patients used for comparative analysis of the models</b>	<b>9,441</b>

\* = Exclusion criteria listed are not mutually exclusive  
<sup>†</sup> = 957 total patients were excluded from the final analysis, some patients had multiple exclusions

**Table 5.4**  
**Summary of patient population characteristics**

<b>Patient Characteristics</b>	<b># of patients (N=9,441)</b>	
Age ≥65	4,594	48.7%
Male	5,033	53.5%
Deaths	1,433	15.2%
<b>ICU Admitting Diagnoses (selected)</b>		
Acute myocardial infarction	837	8.9%
Rhythm disturbance	561	6.0%
Pneumonia, bacterial	428	4.5%
Gastrointestinal bleed	452	4.8%
Congestive heart failure	403	4.3%
Sepsis	369	3.9%
Chronic obstructive pulmonary dis.	335	3.6%
Overdose/poisoning	315	3.3%
Intracranial hemorrhage	242	2.6%
Diabetic ketoacidosis	226	2.4%
Unstable angina	212	2.3%
<b>Location admitted to ICU from:</b>		
Emergency room	4,711	49.9%
Post acute care unit (recovery room)	1,973	20.9%
Inpatient floor	2,111	22.4%
Transfer (another hospital)	355	3.8%
Other	291	3.1%
<b>Patient type</b>		
Medical	7,449	78.9%
Elective surgery	1,589	16.8%
Emergency surgery	403	4.3%

# Data Collection

## Training

For each hospital involved in the CALICO project, at least one data collector (most hospitals sent more than one) was required to attend a day-long training session. The training session involved an overview of the project, an extensive didactic portion describing each of the data elements collected, and information intended to promote data quality. In addition, data collectors were trained on the use of the data entry software developed specifically for the CALICO project. Incorporated into the software are automated checks on the quality of data, including alerts for unexpected or impossible values for physiologic variables. Each data collector was required to perform data abstraction on sample medical records and submit their results to CALICO to demonstrate their proficiency with the data collection process before beginning actual chart abstraction.

## Data Collection Process

At the individual hospitals, data were abstracted from patient medical records and were either directly entered into the CALICO data-entry software or entered into a paper-version of the data collection instrument and later transferred to the software. Periodically, the hospitals uploaded the patient information to a secure, password-protected FTP site, where it was downloaded by the CALICO project. The dataset was then read into the SAS statistical software program.

The data elements collected included all variables required to calculate mortality predictions for the SAPS II, MPM<sub>0</sub> II, APACHE II, and APACHE III models. Additional variables were collected to be evaluated for use in the generation of a new model. Some of the variables added included specific location of myocardial infarctions, a detailed description of the type of pneumonia, and an additional 12 chronic health conditions. See Table 6.1 for a listing of additional variables.

The reason for admission to the intensive care unit was abstracted by data collectors who determined a primary and, if present, a secondary diagnosis from a listing of over 1,000 diagnoses adapted from a coding method developed by the Intensive Care National Audit and Research Center (ICNARC). ICNARC is a center established to undertake comparative audit and evaluative research of intensive care in the United Kingdom. The ICNARC Coding Method was derived empirically from textual data describing the reason for admission for 10,806 patients from the Intensive Care Society's UK APACHE II study.<sup>1</sup> Due to its five-tiered hierarchy, the ICNARC Coding Method allows for stepwise analysis to investigate the potential value that each level of diagnostic information adds to a prognostic model. The five tiers include a description for the type of diagnosis (surgical vs. medical), body system involved, specific site involved, process,

and condition. For surgical patients, CALICO has augmented the original ICNARC Coding Method by adding a sixth tier to describe the surgical procedure performed. ICNARC codes were mapped to appropriate APACHE II and APACHE III diagnostic codes.

**Table 6.1**  
**Listing of additional variables in the CALICO data collection**

<b>Physiologic Variables</b>
INR
Platelet count
Body mass index
Hemoglobin
<b>Chronic Health Conditions</b>
Hypertension
Previous myocardial infarction
Peripheral vascular disease
Congestive heart failure
Last known ejection fraction <40%
HIV+, but no AIDS related illnesses
Chronic obstructive pulmonary disease
Diabetes
Chronic complications of diabetes (nephropathy, neuropathy, etc.)
Transient ischemic attack
Cerebrovascular accident
Previous carotid artery surgery
<b>Reason for Admission</b>
Over 1000 ICNARC codes for reason for admission
Nosocomial versus community-acquired pneumonia
Location of myocardial infarction
Timing of myocardial infarction ( $\leq 24$ or $> 24$ hrs prior to admission)

## References

<sup>1</sup> De Keizer NF, Bonsel GJ, Goldfad, Rowan KM. The added value that increasing levels of diagnostic information provide in prognostic models to estimate hospital mortality for adult intensive care patients. *Intensive Care Med.* 2000;26:577-584.

# Data Quality

An audit was performed to determine if the data collected by CALICO Project hospital participants was reported with sufficient reliability to justify inclusion in the risk adjustment models. The audit included key variables in all models.

For all variables reviewed in the audit, the CALICO Project staff auditors' data was compared to the data abstracted by the participating hospitals, with the exception of the discharge status of the patient from the hospital (alive or dead). For this outcome variable, because of its impact on all models, if the hospital and the PDD did not agree on discharge status, and the discrepancy could not be resolved, then that patient was not included in the audit or in the main study. This PDD screen was not available for 6.9% of the patients; either because the PDD was not yet compiled for patients hospitalized in 2004 or they could not be matched. There were only 22 patients with PDD and hospital discrepancies on discharge status in the 9,759 patients screened (0.2%) and no discrepancies in discharge status in the audited data.

## **Design of the Audit**

### **Selection of hospitals for audit**

Participation in both the project and in the audit was voluntary. All hospitals that submitted at least 100 cases by the time of the audit were asked to participate. Twenty-eight out of 29 eligible hospitals were audited. One hospital was audited initially for training purposes in the first year of the data collection, but was not re-audited due to scheduling difficulties.

### **Record selection within hospitals**

A total of 15 patient charts were requested from each hospital. This is the lowest number which, when sampled from the 28 hospitals expected to provide audit records, would generate the sample of 400 records requested by OSHPD. Of these 15, six were sampled from patients who died, nine from those who lived. This sampling plan results in approximately two and one-half fold over-sampling of deaths (mean death rate of CALICO ICU patients was 15.4%), thus increasing the sample of patients with rare high-risk conditions or physiological abnormalities that are more likely to be present in patients who died. Patients were stratified on discharge status alive or dead, and then randomly sampled within each participating hospital.

### **Variables for audit**

Table 7.1 lists the variables selected for audit and the models associated with the variables. In addition to the MPM variables, the specification of the physiologic

variables was from the APACHE III model, except for GCS, PaO<sub>2</sub>, pH and FiO<sub>2</sub>. The FiO<sub>2</sub> is used in the models as a part of a ratio, so it was grouped by increments of 0.1, (i.e., 0.21-0.3, 0.31-0.4, etc.). The other three variables were weighted using the APACHE II specifications. The GCS was collected as specified by the MPM. The physiological variables included in SAPS and APACHE II are also in either MPM or APACHE III.

### **Audit procedures and data quality**

Auditors participated in a training program that included an overview of the research project, audit principles, review of the data collection tool and accompanying manuals, and software review. Each auditor then had to successfully complete four training case studies. The auditors' case studies were reviewed by the physician trainers and the auditors were given additional education on the data collection process if needed. The majority of the audits were done on-site, but 10 hospitals sent copies of the entire patient record to our project office to be audited. In the Southern California area, our lead physician auditor worked with our RN auditor as a team on the hospital audits. The Northern California audit staff of registered nurses first reviewed charts at the project office while working in teams, with staff physician oversight, before doing any audits on-site at Northern California hospitals. The physician staff that trained hospital data collectors and then project auditors were available to answer any questions that came up during both the on-site and "chart copy" audits.

### **Inter-rater reliability**

Five charts were reviewed by all five auditors. Our physician/trainer then reviewed these records. Agreement statistics were then calculated between the physician "gold standard" data and the auditors' data. The overall agreement for all variables across the five auditors was 94%, ranging from 94% to 95%.

### **Audit Results**

The audited data was compared to the data originally submitted to the CALICO project by the participating hospitals. Table 7.1 lists the variables audited, the percent agreement, and Cohen's kappa where appropriate. We present Cohen's kappa for assistance in interpreting the agreement statistics. When the prevalence of a risk factor is very low or very high, high agreement can result just by chance; the kappa statistic corrects for this issue.

In general, agreement was high (> 90%) and the corresponding kappa substantial (0.70 or greater) for the APACHE III variables with a few notable exceptions as shown in Table 7.1. All APACHE III physiologic variables except the Glasgow Coma Score at admission had agreement above 91% and a kappa statistic greater than 0.70, indicating substantial agreement. The same was true for the type of admission and the location before hospitalization and before admission to the ICU. The Reason for Admission, the Glasgow Coma Score and the Past Medical History variables are discussed below.

**Table 7.1**  
**Inter-rater reliability of variables collected**

	<b>% Agreement</b>	<b>Kappa Statistic*</b>	<b>% of patients with condition†</b>
<b>MPM<sub>0</sub> II variables</b>			
Coma or deep stupor	91.8%	0.63	12.0%
Heart rate $\geq$ 150 beats/min	98.0%	0.63	2.5%
Systolic blood pressure $\leq$ 90 mmHg	91.3%	0.74	22.0%
Chronic renal insufficiency	99.0%	0.93	7.5%
Cirrhosis	97.5%	0.71	4.0%
Metastatic neoplasm	95.8%	0.30	2.8%
Acute renal failure	97.8%	0.56	3.0%
Cardiac dysrhythmia	94.0%	0.55	6.5%
Cerebrovascular accident	98.0%	0.81	5.5%
GI bleeding	97.5%	0.65	4.3%
Intracranial mass effect	94.5%	0.42	3.8%
CPR prior to admission	94.5%	0.71	10.8%
Mechanical ventilation	90.5%	0.79	33.0%
Emergency surgery	92.5%	0.73	18.3%
<b>Medical history variables-APACHE III</b>			
Chronic cardiovascular disease	94.3%	0.24	1.8%
Cirrhosis	97.0%	0.63	3.5%
Portal hypertension	98.5%	0.62	1.3%
Hepatic encephalopathy	98.0%	0.49	1.5%
Chronic renal insufficiency	93.5%	0.38	4.8%
Dialysis	99.0%	0.93	7.5%
Radiation treatment	96.5%	0.29	1.3%
Chemotherapy	98.5%	0.66	2.3%
Metastatic neoplasm	96.3%	0.33	2.5%
CML or CLL	99.5%	0.50	0.3%
Lymphoma	99.5%	0.66	0.8%
AIDS	99.8%	0.80	0.5%
Chronic severe respiratory disease	93.3%	0.28	2.8%
Steroid treatment	96.8%	0.36	2.8%
<b>APACHE III variables</b>			
Reason for admission	52.3%	0.51	
Location prior to hospital admission	97.0%	0.75	
Unit prior to ICU admission (e.g., ER, PACU)	92.3%	0.87	
Elective surgery, emergency surgery, or medical	94.0%	0.80	

\* Unweighted Kappa Statistic

† Percentage of patients with the selected condition of the 400 patients audited

**Table 7.1**  
**Inter-rater reliability of variables collected (continued)**

	% Agreement	Kappa Statistic*
<b>APACHE III physiologic variables</b>		
Temperature	97.6%	0.72
Respiratory rate	95.1%	0.77
Heart rate	96.6%	0.87
Mean arterial pressure	92.8%	0.79
Glasgow Coma Score at admission	86.0%	0.55
Urine output in 1 <sup>st</sup> 24 hours in ICU	91.5%	0.76
Creatinine	98.8%	0.96
BUN	97.1%	0.92
Sodium	99.2%	0.94
Albumin	97.5%	0.82
Bilirubin	98.1%	0.71
Glucose	96.9%	0.77
Hematocrit	97.1%	0.85
pH	97.9%	0.87
PaO <sub>2</sub>	95.3%	0.78
FiO <sub>2</sub>	96.3%	0.75

\* Weighted Kappa statistic

The Reason for Admission (RFA) variable had the lowest inter-rater reliability, with agreement of 52.3% and a corresponding kappa of 0.51. Because the CALICO project used a series of codes designating the system, site, process, and condition—which allowed cross-walking to both APACHE II and APACHE III codes but which might have been more complicated than either system alone—the agreement on RFA could be understated. However, project staff found the reason for admission the single most difficult variable for data collectors to abstract even with physician supervision available. Two of the common problems were the presence of two or more eligible reasons for admission of similar importance and difficulty discerning the underlying reason for ICU admission. Even considering only the 94 categories currently available for the APACHE RFA, it is likely that an OSHPD-directed, statewide agreement on how these categories should be coded and extensive training would be necessary before they would be accurately used. Some of the factors that have made a complicated system such as the ICD-9 coding relatively reliable are mandatory use, professional coders, a method of resolving problematic issues and changes in practice through the “Coding Clinics”, the influence of the payment mechanisms, and periodic audits. A similar effort may be necessary if the APACHE III RFA is to be used more widely.

To help determine if there was an effective, more reliable and simpler method of obtaining needed information about the reason for admission, as previously suggested in the literature, project staff substituted eight body systems comprising clinically related

reasons for admission and one additional category. The percent agreement of the body system variable was 80% with a corresponding kappa of 0.73. This variable was effective and reliable enough in the CALICO study to be used for model building.

Two different methods of specifying the Glasgow Coma Score (GCS) resulted in reasonable reliability. First, to approximate the MPM coma or deep stupor variable, a GCS of less than 6 was used to indicate this condition. This resulted in an agreement of 91.8% and a kappa statistic of 0.63. We independently confirmed that the effect sizes estimated for each separate GCS score showed as step-shaped increases below the value of 6, and that is consistent with the MPM model. The second method was as described using the APACHE II.

The MPM<sub>0</sub> II variables abstracted were all above 90.5% agreement with kappas ranging from 0.30 to 0.93. The abstraction of metastatic neoplasm and intracranial mass effect were the most potentially problematic but again there were few patients with these conditions (2.8% and 3.8% respectively).

The agreement of the APACHE III Past Medical History variables ranged from 93.3% to 99.8%, with kappas ranging from 0.24 to 0.93. In most instances, the number of patients with the Past Medical History variable is very small so the results in this section are difficult to interpret.

In summary, most variables could be collected with a high degree of inter-rater reliability. Among those variables for which inter-rater reliability was lower, the primary reasons seemed to be either genuine medical uncertainty (especially determining a single reason for admission) or problems with documentation that required some inference on the part of the data collector. For example, in many cases, the Glasgow Coma Score is not explicitly listed, but must be assessed from descriptions of the patient's verbal, motor and eye responses. This is significant because those variables for which documentation is the primary issue are likely to improve if public reporting is adopted, while the reason for admission may not. For instance, once providers understand that getting credit for chronic cardiovascular disease as a risk factor requires documentation of symptoms at rest or with minimal exertion, they are more likely to be explicit about the relevant symptoms in their notes.

# Model Descriptions and Development

## Risk-Adjusted ICU Mortality Models

In order to compare the extant ICU risk-adjusted mortality models, the data elements required to calculate SAPS II, MPM<sub>0</sub> II, APACHE II, and APACHE III scores and mortality probabilities were collected. All of these models are based on multiple logistic regression equations. Programs were written for SAS to calculate SAPS II, MPM<sub>0</sub> II, APACHE II, and APACHE III scores and mortality probabilities. Descriptions of the equations as well as the coefficients calculated for each variable were obtained from the published literature. In addition, for the APACHE III model, information was obtained from personal communications with the producers of APACHE III. At the time of initiation of the CALICO project, the coefficients and diagnostic categories used in the APACHE III equations were proprietary and required specific authorization by the producers of APACHE III for use in the CALICO project, but these have now been released to the public. When questions arose concerning the APACHE III equation, the APACHE/Cerner Corporation was contacted directly. A summary of the initial development of the ICU-risk adjusted mortality models and the variables contained in each model are found in the following section.

### MPM<sub>0</sub> II

The Mortality Probability Model II on admission (MPM<sub>0</sub> II) was developed as an updated and revised version of the Mortality Probability Model. The goal of the developers was to construct a model that would accurately predict the mortality experience of a patient sample using the fewest variables required to discriminate and calibrate well.<sup>1</sup> Only variables that had clear definitions, could be easily obtained, and could be reliably collected were included in the final model. The model did not require the data collectors to obtain a primary reason for admission. All variables were collected at the time of ICU admission.

Bivariate analyses were carried out with each of the prospective variables to test for associations with hospital mortality. Chi-square tests were used to assess for associations with hospital mortality for categorical variables, while the Student's *t* test and Wilcoxon's Rank Sum tests were used for continuous variables. Variables were eligible to be included in the multiple logistic regression model if they were significantly associated with hospital mortality at  $P < 0.1$  and were present in at least 2% of the sample population. Variables were then placed in a multivariate model and eliminated if they were not significant at  $P \leq 0.05$ . Calibration of the multivariate model was assessed using the Hosmer-Lemeshow goodness-of-fit test and discrimination assessed by the area under the receiver operating characteristic curve. In an effort to reduce the number of variables in the model, variables whose exclusion improved calibration while not

substantially impacting discrimination were considered for omission from the model. Interactions between the variables were also assessed. For an interaction term to be included in the final model, it needed to have a  $P \leq 0.05$ , be present in at least 1% of the sample population, and be clinically plausible. No interaction terms met these criteria. The final model included 15 variables as detailed in Table 8.1.

**Table 8.1**  
**MPM<sub>0</sub> II variables and coefficients**

Variable	Coefficient ( $\beta$ )	SE
Constant	-5.46836	-
Physiology		
Coma or deep stupor	1.48592	(0.079)
Heart rate $\geq$ 150 beats/min	0.45603	(0.145)
Systolic blood pressure $\leq$ 90 mmHg	1.06127	(0.079)
Chronic Diagnoses		
Chronic renal insufficiency	0.91906	(0.105)
Cirrhosis	1.13681	(0.126)
Metastatic neoplasm	1.19979	(0.098)
Acute Diagnoses		
Acute renal failure	1.48210	(0.089)
Cardiac dysrhythmia	0.28095	(0.068)
Cerebrovascular accident	0.21338	(0.089)
GI bleeding	0.39653	(0.094)
Intracranial mass effect	0.86533	(0.088)
Other		
Age (per 10 years)	0.03057	(0.002)
CPR prior to admission	0.56995	(0.112)
Mechanical ventilation	0.79105	(0.056)
Non-elective surgery	1.19098	(0.074)

*\* Adapted from Lemeshow et. al. "Mortality Probability Models (MPM II) "*

## SAPS II

The Simplified Acute Physiology Score (SAPS II) was developed from a large multi-center European/North American study that enrolled 14,745 patients from more than 12 countries and 137 ICUs.<sup>2</sup> The coordinators of the SAPS study initially chose 37 variables based on clinical reasons to predict mortality. For physiologic variables, the worst value in the first 24 hours after ICU admission was used. Bivariate analyses were used to identify the variables that were associated with hospital mortality. Significant variables were entered into a logistic regression in a step-wise fashion. If the variable did not improve the goodness-of-fit of the model, it was excluded from the study. The final group of 17 variables was used to generate a SAPS score. Continuous physiologic variables were grouped into ranges and given weights. The variables were plotted against vital status at hospital discharge. The LOWESS (locally weighted least squares) smoothing function was used to propose ranges for each variable.<sup>2</sup> Dummy variables were assigned to each range and placed in a multiple logistic regression. The coefficients of the dummy variables were used to assign points to each range, generally multiplying the coefficient by 10 and rounding to the nearest integer. Table 8.2 details the SAPS variables and point assignments. For each patient, the points for all of the variables were summed to produce a SAPS II score. The resultant SAPS II scores were placed in a logistic regression with inpatient mortality. Since the distribution of SAPS II scores was

highly skewed in the developmental dataset, a shrinking power transformation,  $\ln(\text{SAPS II score} + 1)$ , was incorporated into the mortality prediction. The final multiple logistic equation and conversion to determine hospital mortality are noted below.

$$\text{logit} = -7.7631 + 0.0737(\text{SAPS II score}) + 0.9971 [\ln(\text{SAPS II score} + 1)]$$

$$\text{Probability of Inpatient Mortality} = e^{\text{logit}} / 1 + e^{\text{logit}}$$

**Table 8.2**  
**SAPS II variables and point assignments**

Variable	Ranges and Points					
Age	<40 0	40-59 7	60-69 12	70-74 15	75-79 16	≥ 80 18
HR	< 40 11	40-69 2	70-119 0	120-159 4	≥ 160 7	
Systolic BP	<70 13	70-99 5	100-199 0	≥ 200 2		
Temp. (°C)	<39° 0	≥39° 3				
PaO <sub>2</sub> (m Hg)/ FiO <sub>2</sub>	< 100 11	100-199 9	≥200 6			
Urine output (L/day)	<0.5 11	0.5-999 4	≥1 0			
BUN (mg/dL)	<28 0	28-83 6	≥ 84 10			
WBC (10 <sup>3</sup> /mm <sup>3</sup> )	< 1.0 12	1.0-19.9 0	≥ 20 3			
Serum K <sup>+</sup> (mmol/L)	<3.0 3	3.0-4.9 0	≥5 3			
Serum Na <sup>+</sup> (mmol/L)	< 125 5	125-144 0	≥ 145 1			
Serum HCO <sub>3</sub> (mEq/L)	<15 6	15-19 3	≥ 20 0			
Bilirubin (mg/dL)	<4.0 0	4.0-5.9 4	≥ 6 9			
Glasgow coma score	<6 26	6 to 8 13	9 to 10 7	11 to 13 5	14-15 0	
Chronic disease	AIDS 17		Metastatic cancer 9		Hematologic malignancy 10	
Type of admission	Medical 6		Scheduled surgical 0		Unscheduled surgical 8	

## The APACHE II Model

The APACHE II model is based on an earlier prototype system, APACHE (Acute Physiology and Chronic Health Evaluation). The fundamental basis of the APACHE model is that the severity of acute illness can be quantified by the degree of abnormality in multiple physiologic variables.<sup>3</sup> In the original APACHE model, a variation of the nominal group process was used to choose and weight physiologic variables.<sup>4</sup> Thirty-four variables were chosen and each given a weight of 0-4 depending on the degree of physiologic derangement. In the APACHE II model, the number of physiologic variables was reduced to 12.<sup>5</sup> Infrequently collected variables such as lactic acid and serum osmolarity were excluded, as well as variables that were similar markers of disease, e.g.,

BUN and creatinine. For each deleted measurement, a multivariate comparison of the original APACHE system with each proposed change was evaluated to assess the impact on the statistical precision of the model. Ultimately, the fewest physiologic variables that would reflect physiologic derangement for all organ systems, while maintaining precision, were retained.<sup>5</sup> Coma has been shown to have significant impact on mortality and since Glasgow Coma Score was the only measure of neurologic function, it received greater weight than the other variables.<sup>6</sup> The loss of renal function is also known to be a strong indicator of poor prognosis, so serum creatinine was given double weight in patients with acute renal failure.<sup>7</sup> If multiple values for a given variable exist in the first 24 hours after ICU admission, the value with the worst derangement is used. (See Table 8.3 for listings of variables and weights.)

The APACHE II score also includes markers of diminished physiologic reserve. Both age and severe chronic disease reduce the probability of survival during an acute illness. From the original APACHE study, it was noted that patients who were non-operative or who had emergency surgery were at greater risk of death secondary to their prior organ system insufficiency than elective surgical patients. The hypothesis was that patients with the most severe chronic health conditions may not be considered candidates for elective surgery.<sup>5</sup> As a result, emergency surgery patients and non-operative patients are given a higher weight for severe chronic organ dysfunction than elective surgical patients. Table 8.4 lists the definitions of chronic severe organ dysfunction as well as the weights assigned to severe chronic organ dysfunction and age.

The weights assigned to physiology variable derangements were combined to produce the Acute Physiology Score (APS). The APS was then added to the weights for age and chronic health to derive the APACHE II score. Using multiple logistic regression, coefficients were derived for 53 diagnostic categories (Table 8.5), the APS, and post-emergency surgery patients.

R = the risk of hospital death

$$\ln (R/1-R) = -3.517 + (APACHE II Score * 0.146) + (0.603, \text{ only if post emergency surgery}) + (\text{Diagnostic category coefficient})$$

**Table 8.3**  
**The APACHE II severity of disease classification system**

Physiologic Variable	High Abnormal Range					Low Abnormal Range			
	+4	+3	+2	+1	0	+1	+2	+3	+4
Temperature – rectal (°C)	≥ 41°	39°-40.9°		38.5°-38.9°	36°-38.4°	34°-35.9°	32°-33.9°	30°-31.9°	≤ 29.9°
Mean Arterial Pressure – mm Hg	≥ 160	130-159	110-129		70-109		50-69		≤ 49
Heart Rate	≥ 180	140-179	110-139		70-109		55-69	40-54	≤ 39
Respiratory Rate – ventilated or non-ventilated	≥ 50	35-49		25-34	12-24	10-11	6-9		≤ 5
Oxygenation A. if FiO <sub>2</sub> ≥ 0.5 , record A-aDO <sub>2</sub> B. if FiO <sub>2</sub> < 0.5 , record only PaO <sub>2</sub>	≥ 500	350-499	200-349		<200 > 70	61-70		55-60	<55
Arterial pH	≥ 7.7	7.6-7.69		7.5-7.59	7.33-7.49		7.25-7.32	7.15-7.24	<7.15
Serum Sodium – mMol/L	≥ 180	160-179	155-159	150-154	130-149		120-129	111-119	≤ 110
Serum Potassium – mMol/L	≥ 7	6-6.9		5.5-5.9	3.5-5.4	3-3.4	2.5-2.9		<2.5
Serum creatinine – mg/100 cc (Double pts for Acute Renal Failure)	≥ 3.5	2-3.4	1.5-1.9		0.6-1.4		<0.6		
Hematocrit (g%)	≥ 60		50-59.9	46-49.9	30-45.9		20-29.9		<20
White Blood Count total/mm <sup>3</sup> (x 10 <sup>3</sup> )	≥ 40		20-39.9	15-19.9	3-14.9		1-2.9		<1
Glasgow Coma Score – Score= 15 minus actual GCS									
Serum HCO <sub>3</sub> (venous –mMol/l) Use if No ABGs	≥ 52	41-51.9		32-40.9	22-31.9		18-21.9	15-17.9	<15

*\*Adapted from Knaus et. al. "Apache II: A severity of disease classification system"*

**Table 8.4**  
**APACHE II weights for age and chronic health**

Age (yrs)	Points
≤ 44	0
45-54	2
55-64	3
65-74	5
≥ 75	6

If a patient has a history of severe organ system insufficiency or is immuno-compromised	
Non-operative or emergency postoperative patients	5
Elective postoperative patients	2

**Table 8.5**  
**APACHE II diagnoses and coefficients**

NONOPERATIVE PATIENTS		POSTOPERATIVE PATIENTS	
<i>Respiratory failure or insufficiency from:</i>		Multiple trauma	-1.684
Asthma/allergy	-2.108	Admission due to chronic CV disease	-1.376
COPD	-0.367	Peripheral vascular surgery	-1.315
Pulmonary edema (non-cardiogenic)	-0.251	Heart valve surgery	-1.261
Postrespiratory arrest	-0.168	Craniotomy for neoplasm	-1.245
Aspiration/poisoning/toxic	-0.142	Renal surgery for neoplasm	-1.204
Pulmonary embolus	-0.128	Renal transplant	-1.042
Infection	0	Head trauma	-0.955
Neoplasm	0.891	Thoracic surgery for neoplasm	-0.802
		Craniotomy for ICH/SDH/SAH	-0.788
<i>Cardiovascular failure or insufficiency from:</i>		Laminectomy and other spinal cord surgery	-0.699
Hypertension	-1.798	Hemorrhagic shock	-0.682
Rhythm disturbance	-1.368	GI bleeding	-0.617
Congestive heart failure	-0.424	GI surgery for neoplasm	-0.248
Hemorrhagic shock/hypovolemia	0.493	Respiratory insufficiency after surgery	-0.140
Coronary artery disease	-0.191	GI perforation/obstruction	0.060
Sepsis	0.113	Sepsis	0.113
Postcardiac arrest	0.393	Postcardiac arrest	0.393
Cardiogenic shock	-0.259	Postrespiratory arrest	-0.168
Dissecting thoracic/abdominal aneurysm	0.731		
		<i>If not in one of the specific groups above, then use major organ system affected:</i>	
<i>Trauma:</i>		Metabolic/renal/hematologic	-0.196
Multiple trauma	-1.228	Respiratory	-0.610
Head trauma	-0.517	Neurologic	-1.150
		Cardiovascular	-0.797
<i>Neurologic:</i>		Gastrointestinal	-0.613
Seizure disorder	-0.584		
ICH/SDH/SAH	0.723		
<i>Other:</i>			
Drug overdose	-3.353		
Diabetic ketoacidosis	-1.507		
GI bleeding	0.334		
<i>If not in one of the specific groups above, then use major organ system:</i>			
Metabolic/renal/hematologic	-0.885		
Respiratory	-0.890		
Neurologic	-0.759		
Cardiovascular	0.470		
Gastrointestinal	0.501		

\* Adapted from Knaus et. al. "Apache II: A severity of disease classification system"

## **The APACHE III Model**

The APACHE III model attempted to improve on the APACHE II model by re-evaluating the selection and weighting of physiologic variables, examining how differences in patient admission criteria and timing of admission to ICUs related to outcome variations among hospitals, and examining issues regarding the selection of patients and the timing of scoring.<sup>8</sup> APACHE III also expanded the size and representativeness of the developmental database used to generate the model.

Unlike APACHE II, the weights given to physiologic variables were not empirically assigned. To assign weights to the physiologic variables, variables were first placed into clinically appropriate ranges. Multiple logistic regression was then used to determine weights for these groups. The derived weights were adjusted to comply with clinical and physiologic principles and validated on a separate portion of the database.

In evaluating the optimal time to obtain the physiologic variables, individuals were given APACHE III scores by using the initial value for the physiologic variables during the first hour of admission, the worst value over the remaining 23 hours of the first day of ICU admission, and the worst value over the initial 24 hours of ICU admission. The absolute differences between the scores were not statistically significant using these different timeframes. The developers decided to use the worst value over the initial 24 hours in the APACHE III model since this time frame had the lowest proportion of missing values.

Interaction between the physiologic variables and diagnostic categories was assessed by using disease-specific weighting of the physiologic variables. The new weights did not substantially improve the explanatory power of the model.

The APACHE III developers found that the more time spent in another inpatient location before ICU admission, the higher the risk of hospital mortality. This was presumably because despite medical treatment these patients still had a deterioration in their clinical status. As a result, length of stay in other inpatient care areas before ICU admission was added to the model.

A list of 212 disease categories was developed to determine the reason for ICU admission. The stability of the weights was assessed with regard to the weights derived and used in APACHE II and the clinical experience of the developers of the model. The criteria used for assessing stability included homogeneity, cell size, and the impact of the disease on short term outcome. The initial set of categories was reduced to 78 categories, but has since been expanded to 94 diagnoses.

The final APACHE III mortality prediction is determined by an equation including weights for APS, age, chronic health conditions, pre-ICU length of stay, location admitted from, reason for ICU admission, and whether the patient had emergency surgery. A complete listing of the APACHE III reason for admission codes, hospital mortality prediction equation and calculation of the APS can be found in the public information section of the APACHE III / Cerner Corporation Web page at [www.apache-web.com](http://www.apache-web.com).

## References

- 
- <sup>1</sup> Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients. *JAMA*. 1993; 270:2478-2486.
- <sup>2</sup> Le Gall J, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA*. Dec 22-29 1993; 270(24):2957-2963.
- <sup>3</sup> Knaus WA, Wagner DP, Draper EA. Relationship between acute physiologic derangement and risk of death. *J Chron Dis*. 1985; 38: 295-300.
- <sup>4</sup> Knaus WA, Zimmerman JE, Wagner DP. APACHE – acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med*. 1981; 9: 591-597.
- <sup>5</sup> Knaus WA, Draper EA, Wagner DP, Zimmerman J. APACHE II: A severity of disease classification system. *Crit Care Med*. 1985; 13:818-829.
- <sup>6</sup> Teres D, Brown RB, Lemeshow S. Predicting mortality of intensive care patients: The importance of coma. *Crit Care Med*. 1982; 10:86-95.
- <sup>7</sup> Sweet SJ, Glenney CU, Fitzgibbons JP. Synergistic effect of acute renal failure and respiratory failure in the surgical intensive care unit. *Am J Surg*. 1981; 141: 492-496.
- <sup>8</sup> 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; 100:1619-1636.15.

# Customization of the Models

In other studies in which the ICU risk-adjustment models have been applied to populations distinct from the ones on which they were developed, each model has maintained adequate discrimination but showed poor calibration. To improve the calibration of these models, we re-estimated the coefficients in the models using the California ICU population. The methods used were similar to prior studies that customized the models to new populations.<sup>1,2,3,4,5,6,7</sup> The CALICO dataset was divided into a developmental dataset (60% of the sample) and validation dataset (40% of the sample).

For the MPM<sub>0</sub> II model, the coefficients for each of the 15 variables were re-estimated using logistic regression. For the SAPS II model, there are only two coefficients in the risk equation, the coefficient preceding the SAPS score and the coefficient preceding the natural logarithm of the SAPS score. Both of these coefficients were re-estimated. The internal SAPS score, including weights assigned to various physiologic variables, was not altered. Previous efforts to customize the models have kept the internal weighting of the physiology score the same and have re-estimated only the coefficients in the regression equation.

For the APACHE II model, the coefficients of the variables for the Acute Physiology Score (APS) and emergency surgery were re-estimated. Similar to the SAPS II model, the internal APS was not changed. In the APACHE II model, there is also a separate coefficient for each diagnostic category (reason for admission to the ICU). In the CALICO dataset, there were insufficient numbers of patients in some of the categories to re-estimate the coefficients. As a result, categories that did not contain approximately 1% (approximately 100 patients) of the total sample of CALICO patients were combined with other categories before re-estimating the coefficients for the diagnostic categories. Categories were combined if they were clinically similar and had similar crude mortality rates. In addition, when the number of patients in categories was very small, the original APACHE coefficients were used as a guide to combine categories. Appendix Table A.11 displays the combined APACHE II diagnostic categories.

The APACHE III model's coefficients were re-estimated for age, past medical history, APS, location prior to admission, and type of admission. As in the APACHE II model, there were insufficient numbers of patients in some of the categories to re-estimate the diagnostic coefficients. Categories that did not contain approximately 0.7% (65 patients) of the total sample were combined with other categories using the same method as in the APACHE II model. Appendix Table A.12 displays the combined APACHE III diagnostic categories.

## References

- 
- 1 Zhu B-P, Lemeshow S, Hosmer DW, et al. Factors affecting the performance of the models in the Mortality Probability Model II system and strategies of customization. *Crit Care Med.* Jan 1996;24(1):57-63.
  - 2 Murphy-Filkins RL, Teres D, Lemeshow S, et al. Effect of changing patient mix on the performance of an intensive care unit severity-of-illness model. *Crit Care Med.* Dec 1996;24(12):1968-1973.
  - 3 Glance LG, Osler TM, Dick AW. Identifying quality outliers in a large, multiple-institution database by using customized versions of the Simplified Acute Physiology Score II and the Mortality Probability Model II. *Crit Care Med.* 2002;30(9):1995-2002.
  - 4 Markgraf R, Deuschinoff G, Pientka L, et al. Performance of the score systems Acute Physiology and Chronic Health Evaluation II and III at an interdisciplinary intensive care unit, after customization. *Crit Care Med.* Jan 2001;5(1):31-36.
  - 5 Rivera-Fernandez R, Vazquez-Mata G, Bravo M, et al. The Apache III prognostic system: customized mortality predictions for Spanish ICU patients. *Intensive Care Med.* 1998;24:574-581.
  - 6 Metnitz PGH, Valentin A, Vesely H, et al. Prognostic performance and customization of the SAPS II: results of a multicenter Austrian study. *Intensive Care Med.* 1999;25:192-197.
  - 7 Sirio CA, Shepardson LB, Rotondi AJ, et al. Community-wide assessment of intensive care outcomes using a physiologically based prognostic measure. *Chest.* 1999;115(3):793-801.

# Patient Discharge Database (PDD) Risk-adjustment Model

While the extant models have been widely used, they all have the limitation of needing clinical data from chart abstraction. In some instances, such as with the APACHE II and APACHE III models, the amount of data collection is quite significant. These models require the abstractor to search through the lab values and vital signs during the first twenty four hours of ICU admission to find the values reflecting the greatest physiologic derangement and to determine the reason for ICU admission. To address this limitation, we developed and tested three novel models for ICU-risk adjustment using administrative data from the OSHPD patient discharge.

The analysis database was created by matching the 2002 and 2003 OSHPD patient discharge database (PDD) with the CALICO database. Of the 9,935 patients in the CALICO database with hospital discharges in the 2002-2003 calendar years, 9,759 (98%) were successfully matched with a PDD record. Of these, 85 were excluded because of discrepancies between the PDD and CALICO data as to (1) the hospital mortality outcome, or (2) difference in age of more than one year. An additional 122 were excluded because of missing clinical variables, 20 because principle diagnosis was not present at admission, and 23 because the principle diagnosis was unclassifiable. The resulting analysis sample of 9,509 had 15.4% mortality.

The first model developed from the patient discharge database (PDD model) used as predictors only variables routinely available from administrative data. Included in the model were age, gender, primary reason for hospital admission, and other conditions present on hospital admission. The primary reason for hospital admission was determined by the following method. First, the primary diagnosis ICD-9 code was placed in one of 258 categories using the SAS format library for the Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS). These were reviewed for internal consistency and modified slightly. Within each body system separately, all categories with fewer than 55 patients were combined into one category per body system. The value 55 was selected because in a 60% random subset and at the average mortality rate of 15% found in our data the predicted number of deaths would be 5, a minimum number considered adequate for estimating a death rate. (The method did not require that the actual number of deaths exceed 5, because that could introduce bias.)

The secondary diagnostic conditions present on admission were used to create a set of condition indicator variables using the components of the Deyo et al. adaptation of the Charlson chronic condition score. More than one Charlson category variable could be applicable. To avoid double-counting the risk due to a condition, a table of exclusions was constructed to indicate which secondary conditions would be ignored in the presence of a given primary condition affecting the same body system.

A priori, it was hypothesized that a model with solely administrative data would not have the predictive accuracy of a clinical model. As a result a second model was developed that contained the variables included in the PDD model plus clinical variables that would be easy to collect. Unlike the other models, the clinical variables in the MPM<sub>0</sub> II model are collected at the time of the ICU admission. Since this method creates the least chart abstraction burden, the variables in the MPM<sub>0</sub> II were considered for addition to the PDD model. The MPM<sub>0</sub> II model's variables for acute and chronic conditions were not included because acute and chronic conditions were already captured by the PDD reasons for admission and Charlson score. The MPM<sub>0</sub> II variables for the highest heart rate, lowest systolic blood pressure, Glasgow coma score (GCS), mechanical ventilation, intracranial mass, the type of admission (medical, emergency surgery, or elective surgery), and CPR prior to admission were added to the PDD model.

To improve the calibration of the model, the variables for heart rate, blood pressure and GCS were treated as continuous variables instead of being dichotomized as they were in the MPM<sub>0</sub> II model. To determine an appropriate model for the continuous clinical variables, first a model was estimated without the clinical variables. Then for each clinical variable a collection of indicator variables were added, each representing a short interval of the clinical measure in question. The shapes of the resulting effect curves were then examined to determine reasonable modeling methods. Heart rate was modeled as a U-shaped broken line, with the segment between 60 and 80 bps having weight of zero, the segments below 60 and above 80 being attached lines whose slopes were fitted to the data. Systolic blood pressure was modeled as a U-shaped broken line, with the segment between 110 and 160 mm Hg having a weight of zero, the segments below 110 and above 160 being attached lines whose slopes were fitted to the data. GCS was modeled by grouping scores 3 to 5, and scores 6 to 14, with the perfect score 15 used as a reference score.

# APACHE III System Model

The reason for ICU admission used in the APACHE models is one of the most difficult variables to collect reliably. Determining the single reason for ICU admission can be complicated as ICU patients often have complex medical problems that affect more than one system and are interrelated. Our data collectors consistently reported having difficulty discerning the ICU reason for admission. Our data audit illustrated that the inter-rater reliability of this variable was much lower than other variables tested (see Table 7.1) with a kappa of 0.51 and a 52.3% agreement.

To address this problem, we determined whether a simplified APACHE III model, using the organ system affected as the clinical grouper rather than a specific reason for admission, could be abstracted more reliably without compromising the predictive accuracy of the model. Previously, de Keiezer et. al.<sup>1</sup> have shown that simplifying the APACHE II model, by extending the admission type and substituting the 53 UK APACHE II diagnostic categories with nine body systems, did not alter the discriminatory power or calibration of the model.

To simplify the APACHE III reason for admission, we first separated the reason for admission into eight system categories and one category for overdose/poisoning. With this reclassification, the inter-rater reliability improved with an observed agreement of 80% and a kappa of 0.73.

However, the risk of mortality in medical patients is very different than in surgical patients. As a result, we further split the system categories into medical and surgical. For the musculoskeletal category, the majority of the patients were surgical and for the hematology/oncology category, the majority of the patients were medical. Splitting these two groups by medical/surgical status would have resulted in categories with less than 1% of the total CALICO database, so they were kept combined. Table 11.1 shows the crude mortality by system and medical/surgical status. With the addition of medical/surgical status, the observed agreement and kappa were relatively unchanged compared to the model using system alone (Table 11.2).

**Table 11.1**  
**Crude mortality by system and medical/surgical status**

<b>Category</b>	<b>N</b>	<b>Crude mortality</b>
Respiratory, surgical	270	4.4%
Respiratory, medical	1805	23.3%
Cardiology, surgical	558	5.0%
Cardiology, medical	3223	17.0%
GI, surgical	529	15.3%
GI, medical	655	13.3%
Neurology, surgical	444	4.7%
Neurology, medical	834	22.5%
GU, surgical	115	1.7%
GU, medical	133	21.8%
Metabolic, surgical	148	0.0%
Metabolic, medical	414	5.8%
Hematology/Oncology	176	26.7%
Musculoskeletal	129	9.3%
Overdose/Poisoning	367	3.3%

**Table 11.2**  
**Inter-rater reliability of the APACHE III models**

<b>Model*</b>	<b>Observed Agreement</b>	<b>Expected Agreement</b>	<b>Kappa</b>
<b>APACHE III</b>	52.3%	3.3%	0.51
<b>APACHE III System</b>	77.3%	20.8%	0.71

\* = The models represent the extant ICU-risk adjustment models re-estimated using the CALICO database

## References:

<sup>1</sup> de Keizer NF, Bonsel GJ, Goldfad C, Rowan KM. The added value that increasing levels of diagnostic information provide in prognostic models to estimate hospital mortality for adult intensive care patients. *Intensive Care Med.* 2000 May; 26(5):577-84.

# Results I: Performance of the Original Models

## Discrimination

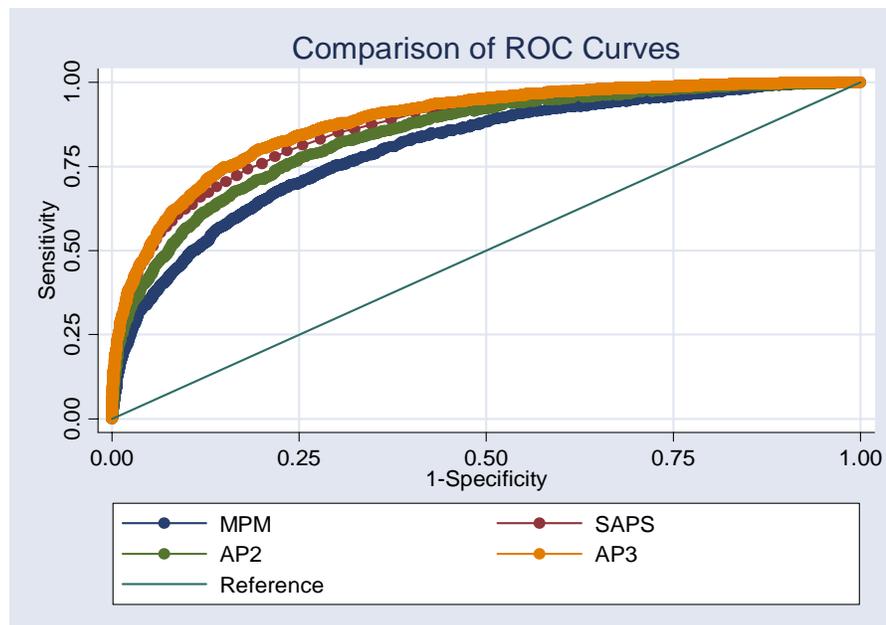
Discrimination was assessed by using the area under the receiver operating characteristic curve (AUC). The minimum AUC that was considered reasonable discrimination was 0.80.<sup>1</sup> The AUC was determined using the 9,441 patients in the CALICO dataset that could be used across all models. MPM<sub>0</sub> II, SAPS II, APACHE II, and APACHE III all showed reasonable discrimination with AUCs of 0.803, 0.865, 0.843, and 0.882 respectively. In pair-wise comparisons of the AUCs of the models using the DeLong method, the models were all statistically significantly different from each other with P<0.001.<sup>2</sup> The AUC for the APACHE III model was the highest, followed by SAPS II, APACHE II, and lastly the MPM<sub>0</sub> II model.

**Table 12.1**  
**Summary of discrimination – Original models**

Model	AUC*	95% CI
MPM <sub>0</sub> II	0.803	(0.790-0.815)
SAPS II	0.865	(0.855-0.875)
APACHE II	0.843	(0.832-0.854)
APACHE III	0.882	(0.872-0.891)

\* = Area under the receiver operating characteristic curve

**Figure 12.1**  
**Receiver operator curves for the four models**



## Calibration

Calibration was assessed using Hosmer-Lemeshow (H-L) goodness-of-fit tests and calibration curves. Both Hosmer-Lemeshow C tests and H tests were performed. Analyses using the C test divide patients into deciles (i.e., equal number of patients) in ascending order of death. The range of predicted risk of mortality within each decile is determined by the patients in that decile. The H test forms 10 groups based on fixed, equal deciles of risk (i.e., 0.0-0.09%, 0.1%-0.19%, etc.) with variable numbers of patients in each group. The difference between the observed and expected mortality for each strata is summarized by the Pearson chi-square statistic. The statistics are summed over the ten deciles and are compared to a chi-square distribution. The degrees of freedom equal N-2, where N= the number of groups, when used on an estimation dataset. However, when used on an application dataset, one in which the coefficients used are not recalculated using the dataset being analyzed, typically the degrees of freedom are the same as the number of groups (10 degrees of freedom).<sup>1</sup>

All of the models tested had calibration limitations when using their original coefficients. A complete summary of the Hosmer-Lemeshow goodness-of-fit tests can be found in Appendix Tables A.1 and A.2. The four models all had significant P-values for their Hosmer-Lemeshow statistics, indicating a significant difference between the observed and predicted mortality and therefore poor calibration. However, it should be noted that large sample size by itself is more likely to generate significant P-values. Although exact comparisons cannot be made among the Hosmer-Lemeshow statistics, MPM<sub>0</sub> II and APACHE III had the lower values for the Hosmer-Lemeshow statistics compared to APACHE II and SAPS II.

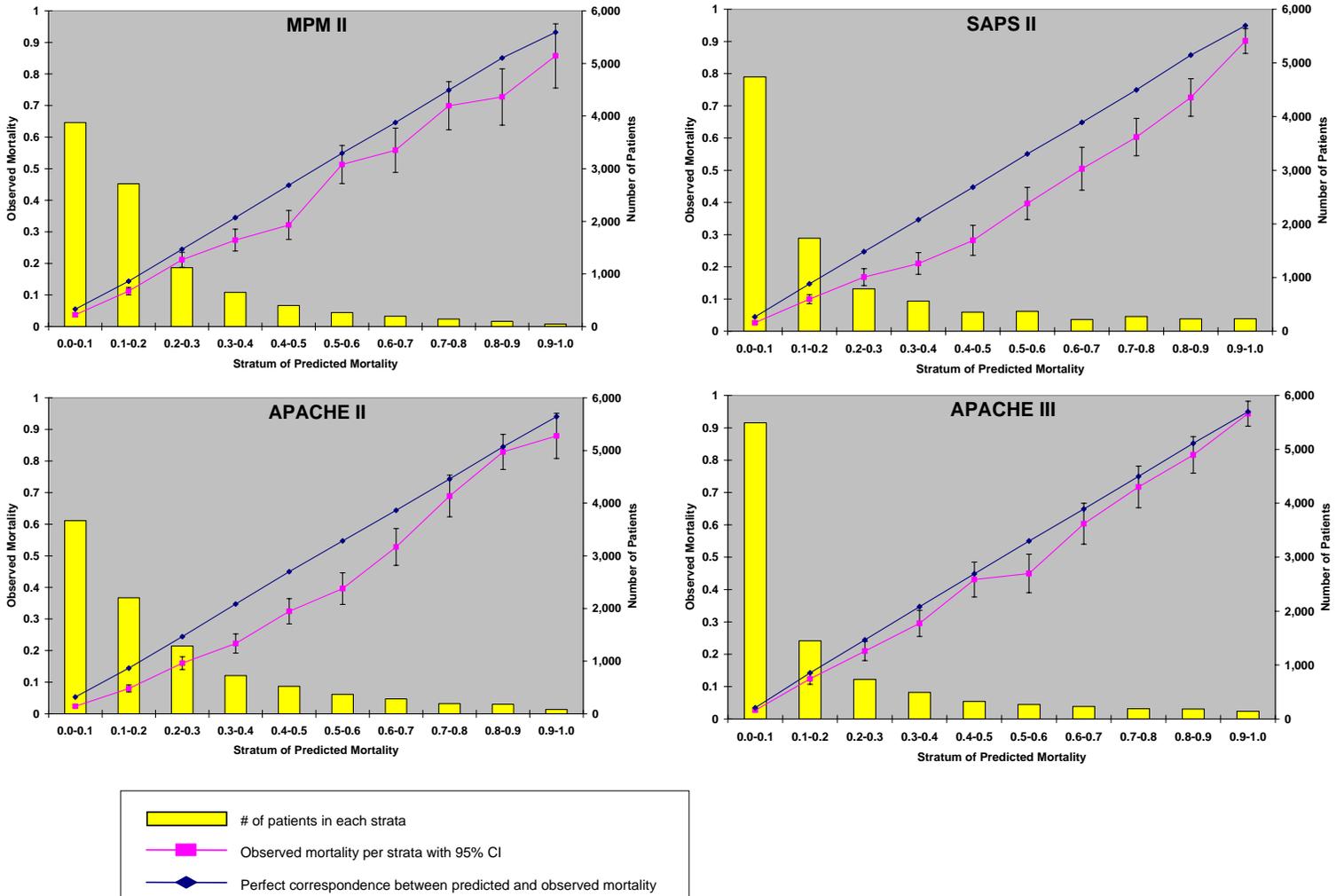
The calibration curves for the respective models are shown in Figure 12.2. In the calibration curves, patients are divided into strata based upon their predicted risk of mortality (0-10%, 11-20%, etc). The actual mortality rate of patients in each stratum (the number of deaths divided by the number of patients in each stratum) is plotted and compared to the line where observed mortality = expected mortality. Models showing good calibration should approximate this line. All the models over-predicted death across the ten strata of mortality risk. Comparing the models in Figure 12.2, MPM<sub>0</sub> II and APACHE III more closely approximate the observed = expected line (O=E line) than the other models. All the models are close to the O=E line in the two lowest deciles of risk. However, there is significant variation among the other deciles of mortality risk illustrating poor uniformity of fit across deciles of predicted mortality.

**Table 12.2**  
**Summary of calibration – Original models**

Model	H-L* Statistic	
	C Test	H Test
<b>MPM<sub>0</sub> II</b>	110.00 (P<0.001)	119.06 (P<0.001)
<b>SAPS II</b>	310.16 (P<0.001)	311.20 (P<0.001)
<b>APACHE II</b>	324.06 (P<0.001)	328.81 (P<0.001)
<b>APACHE III</b>	38.17 (P<0.001)	40.51 (P<0.001)

\* = Hosmer-Lemeshow statistic; df 10

**Figure 12.2**  
**Calibration curves - Original models**



**Summary**

Overall, MPM<sub>0</sub> II had the worst discrimination of the four models, and APACHE III had the best. The models with the best calibration were the MPM<sub>0</sub> II and APACHE III models. Although these models showed better calibration than the SAPS II and APACHE II model, calibration was still poor. Before models can be used to compare the performance of ICUs, the models should be recalibrated to the current dataset.

## References

---

<sup>1</sup> Mourouga P, Goldfrad C, Rowan KM. Does it fit? Is it good? Assessment of scoring systems. *Current Opinion in Critical Care*. 2000;6:176–180.

<sup>2</sup> H DeLong ER, DeLong DM, Clarke-Pearson DL.H Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988 Sep;44(3):837-45.

# Results II: Performance of the Customized Models

The performance of the re-estimated models was assessed by evaluating their discrimination and calibration in a validation subset of the CALICO database. The re-estimated models resulted in significant improvement in the Hosmer-Lemeshow statistics (calibration). The original models all showed poor fit ( $P < 0.05$ ) for both the C and H tests. In the re-estimated models, the calibration was improved. The H-L statistics for the APACHE III model still reached statistical significance ( $P < 0.05$ ); however, this may be due in part to the large sample size of the validation dataset ( $> 3,000$ ). A known limitation of Hosmer-Lemeshow statistics is that P-values become smaller as sample size increases.

**Table 13.1**  
**Performance of the re-estimated models in the validation sample**

Model	AUC <sup>†</sup> (95% CI)	H-L <sup>‡</sup> Statistic	
		C Test	H Test
<b>MPM II</b>			
Original	0.809 (0.789 – 0.828)	52.9 ( $P < 0.001$ )	61.5 ( $P < 0.001$ )
Re-estimated Model	0.811 (0.791 – 0.830)	11.3 ( $P = 0.33$ )	13.6 ( $P = 0.19$ )
<b>SAPS II</b>			
Original	0.870 (0.854 – 0.887)	139.6 ( $P < 0.001$ )	143.5 ( $P < 0.001$ )
Re-estimated Model	0.870 (0.854 – 0.887)	15.2 ( $P = 0.12$ )	6.9 ( $P = 0.73$ )
<b>APACHE II</b>			
Original	0.841 (0.823 – 0.859)	155.0 ( $P < 0.001$ )	157.6 ( $P < 0.001$ )
Re-estimated Model	0.864 (0.848 – 0.879)	15.2 ( $P = 0.12$ )	16.0 ( $P = 0.10$ )
<b>APACHE III</b>			
Original	0.881 (0.866 – 0.895)	32.2 ( $P < 0.001$ )	37.3 ( $P < 0.001$ )
Re-estimated Model	0.880 (0.865 – 0.894)	20.4 ( $P = 0.026$ )	27.1 ( $P = 0.002$ )

<sup>†</sup>= Area under the receiver operator curve

<sup>‡</sup>= Hosmer-Lemeshow statistic; *df* 10 for developer model; *df* 8 for re-estimated models

The calibration of the re-estimated models was also assessed with calibration curves (Figure 13.1). The validation dataset of the first 60/40 split of the CALICO database was used to generate the calibration curves. The curves demonstrate that with re-estimation, all of the models except APACHE III clearly approximate the observed deaths more closely than the original models. In the original models, the 95% confidence interval for observed mortality across the deciles of risk was often below the line for perfect correspondence, representing over-prediction of death. In general, omnibus test approaches are not as good at finding departures as more focused tests. Several data points in a row that are all well above or well below the predicted line may suggest more inaccuracy in the model than if they were spread out seemingly at random. This may imply more inaccuracy in the MPM II and APACHE III models.

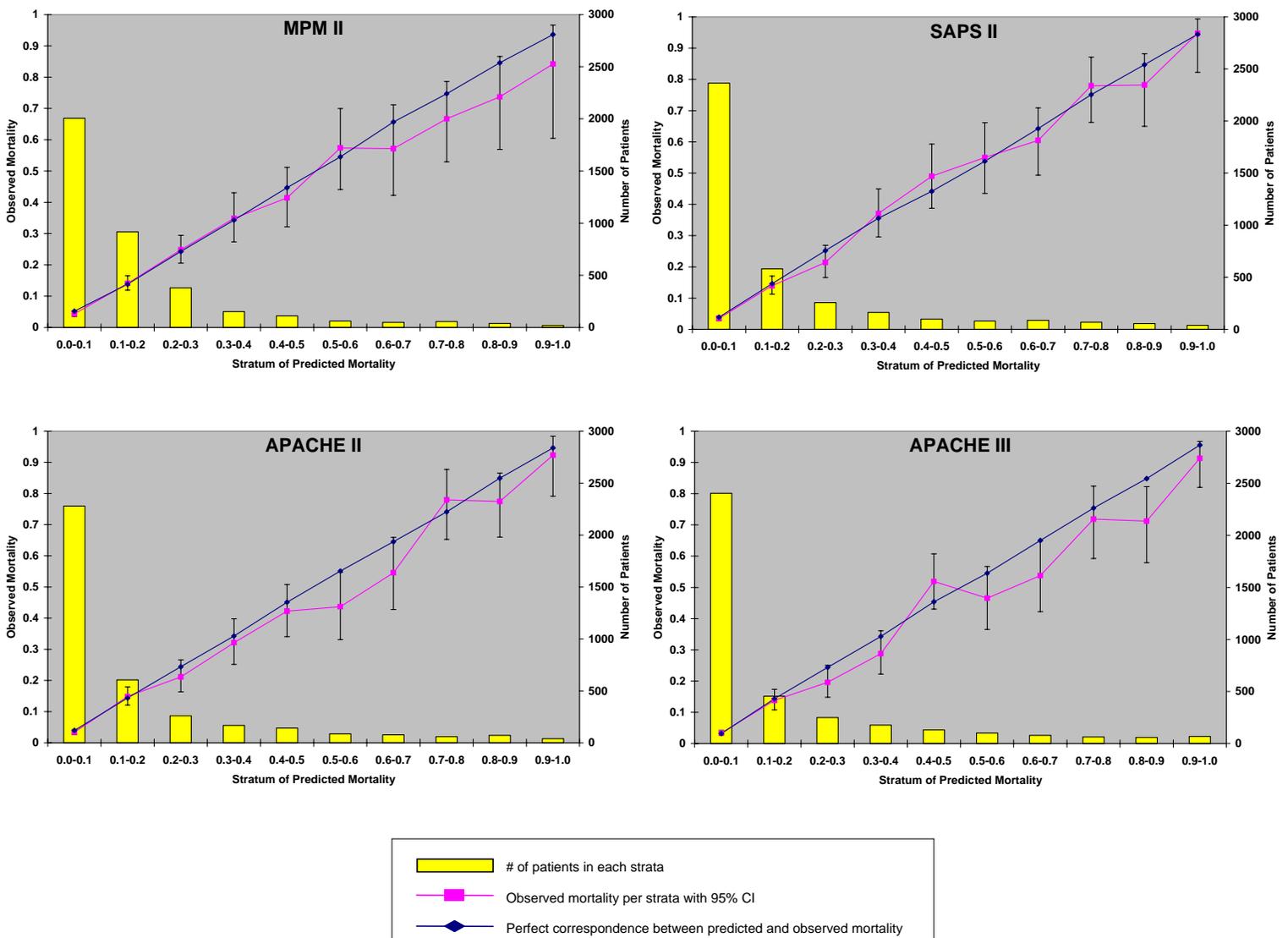
## Coefficients in the Customized Models

The re-estimated coefficients are displayed in the appendix for both the estimation sample as well as using the entire CALICO database.

## Summary

With re-estimation, all the models demonstrated reasonably good calibration, although MPM II and APACHE III are slightly less well calibrated than SAPS II and APACHE II. APACHE III appeared to have a higher discrimination than the other the models in pairwise comparisons using the DeLong method ( $P < 0.001$ ). While calibration can be improved with re-estimation of coefficients, discrimination cannot be improved without changing the model. In selecting the optimal model, the predictive performance of the models should be compared to their data collection burden.

**Figure 13.1**  
**Calibration curves – Re-estimated models**



# Results III: Performance of the PDD Models

The performance of the PDD and PDD + clinical models were compared to the four re-estimated extant risk-adjustment models. The PDD model demonstrated lower discrimination than all the other models with an AUC of 0.774 (95% CI 0.755-0.793). The PDD plus clinical variables model had an AUC that was higher than the MPM<sub>0</sub> II model but slightly lower than the other models. This discrimination was still adequate with an AUC of 0.851 (95% CI 0.835 -0.867). Table 14.1 details the discrimination of all six models.

**Table 14.1**  
**Discrimination of the PDD and re-estimated models**

<b>Model*</b>	<b>AUC†</b>	<b>95% CI</b>
<b>MPM<sub>0</sub> II</b>	0.811	(0.791-0.830)
<b>SAPS II</b>	0.870	(0.854-0.887)
<b>APACHE II</b>	0.864	(0.848-0.879)
<b>APACHE III</b>	0.880	(0.865-0.894)
<b>PDD</b>	0.774	(0.755-0.793)
<b>PDD + Clinical</b>	0.851	(0.835-0.867)

\* = The models represent the extant ICU-risk adjustment models re-estimated using the CALICO database  
 † = Area under the receiver operating characteristic curve

Calibration of the PDD and PDD + clinical model was inferior to the extant re-estimated risk-adjustment ICU models. The Hosmer-Lemeshow goodness-of-fit statistics for the PDD model indicated poorer fit. For the PDD plus clinical model, the H test had  $p < 0.001$  indicating poor calibration, while the C test for the PDD plus clinical model had a P value of 0.058, suggesting borderline miscalibration. For the Hosmer-Lemeshow tests, significant P-values indicate poor performance. Table 14.2 shows a summary of the calibration of all six models. It is important to note that the Hosmer-Lemeshow statistics of the different models cannot be compared directly. One can only calculate the probability of observing the difference in deaths predicted by the model and observed deaths, if the model fit the data. In addition, a known limitation of this statistical method is that with large sample sizes, even very small percentage differences between the predicted and observed death rates can become large enough in the difference in absolute number of predicted versus observed deaths to make the Hosmer-Lemeshow statistics significant.

Given these limitations with quantitative assessments of calibration by Hosmer-Lemeshow statistics, visual assessment of calibration using calibration curves are often used to evaluate the fit of models when developed and validated in large populations

(Figure 14.1). The PDD model's calibration curve shows that the observed mortality (pink line) is above the predicted mortality (blue line) in the low risk patients but the model over-predicts mortality in the higher-risk patients. The PDD plus clinical model is better than the PDD model; however, in the higher strata of predicted mortality, this model also over-predicts deaths. In comparison to the calibration curves of the re-estimated extant models, the fit of the PDD model is clearly inferior. The PDD model plus clinical variables is inferior to the SAPS model, but appears to have similar fit compared to MPM and the APACHE models, as these models also have problems with fit in the highest strata of predicted mortality.

Since the PDD models are based on an individual's reason for admission (RFA) to the hospital, and not the RFA to the ICU as in the four extant models, we hypothesized that a portion of the calibration problems might be due to patients who developed a different medical condition after hospitalization and were subsequently admitted to the ICU for treatment of that medical condition. To capture the risk associated with these other medical conditions, while trying to keep the PDD model simple and easy to collect, we designed another PDD model that included the seven APACHE III diagnostic categories with the highest odds ratios for mortality. The categories were included as dichotomous variables. Cross-maps were made between these seven conditions and the hospital reason for admission so that if a person was admitted to the hospital with one of these conditions, the condition would not be counted twice. Despite these additions, the new model showed no significant improvement in discrimination and calibration. With the added conditions, the AUC was .853 (compared to .851), the Hosmer-Lemeshow H statistic was 16.6 (compared to 17.8), and the H-L C statistic was 35.8 (compared to 40.6).

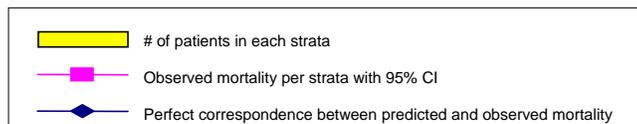
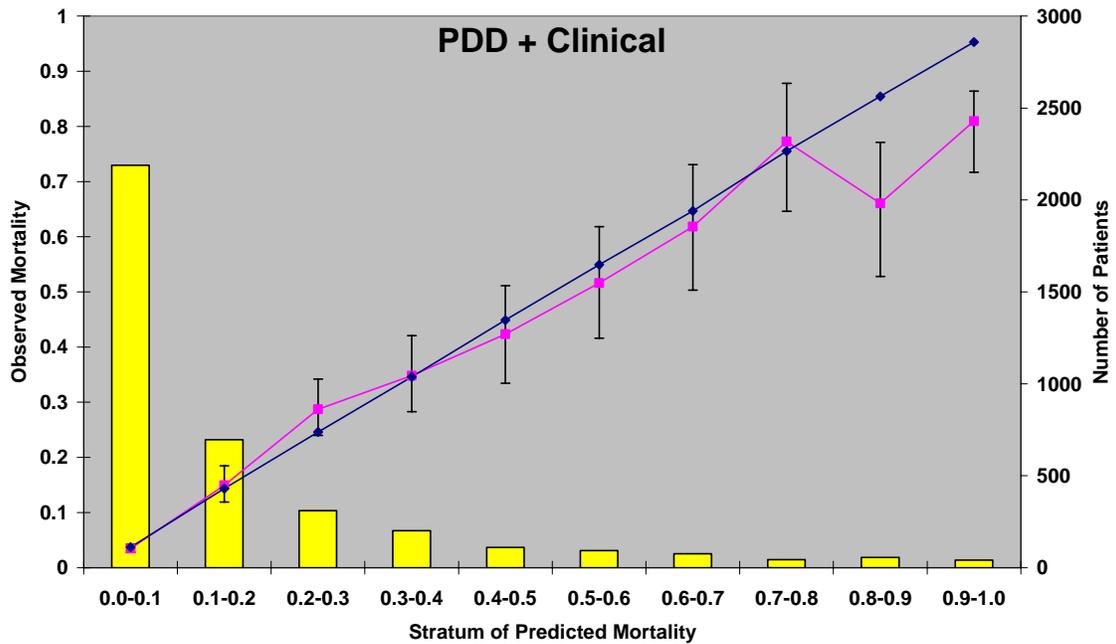
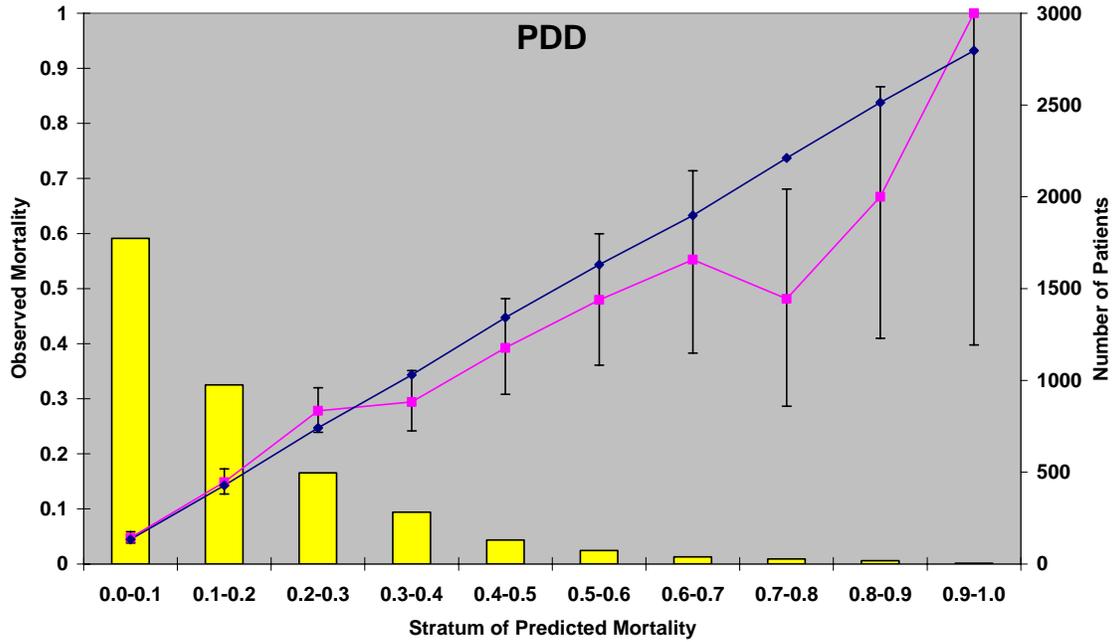
**Table 14.2**  
**Calibration of the PDD models**

Model*	H-L <sup>†</sup> Statistic	
	C Test	H Test
MPM <sub>0</sub> II	11.3 (P=0.33)	13.6 (P=0.19)
SAPS II	15.2 (P=0.12)	6.9 (P=0.73)
APACHE II	15.2 (P=0.12)	16.0 (P=0.10)
APACHE III	20.4 (P=0.026)	27.1 (P=0.025)
PDD	24.43 (P=0.007)	23.32 (P=0.010)
PDD + Clinical	17.85 (P=0.058)	40.58 (P<0.001)

\* = The models represent the extant ICU-risk adjustment models re-estimated using the CALICO database

† = Hosmer-Lemeshow statistic; df 10

**Figure 14.1**  
**Calibration curves of the PDD models**



# Results IV: Performance of the APACHE III System Model

We examined the performance of the model that used only the system associated with the primary reason for admission relative to the model using all the APACHE III reason for admission categories (recalibrated to California ICU data). The discrimination was very similar between these two models (Table 15.1). The calibration of the system-based model was better than that of the APACHE III model using all APACHE III diagnostic categories, as shown in both the Hosmer-Lemeshow goodness-of-fit statistics (Table 15.2) and calibration curves (Figure 15.1). Our results indicate that by simplifying the APACHE III reason for admission categories into systems, the resultant model will have higher inter-rater reliability with only minor compromise in discrimination and with improvement in calibration. We hypothesize that the improvement in calibration may result from the reduction of misclassification using system instead of more specific reason for admission.

**Table 15.1**  
**Discrimination of the APACHE III System model**

Model*	AUC†	95% CI
<b>MPM<sub>0</sub> II</b>	0.811	(0.791-0.830)
<b>SAPS II</b>	0.870	(0.854-0.887)
<b>APACHE II</b>	0.864	(0.848-0.879)
<b>APACHE III</b>	0.880	(0.865-0.894)
<b>APACHE III System</b>	0.873	(0.857-0.889)

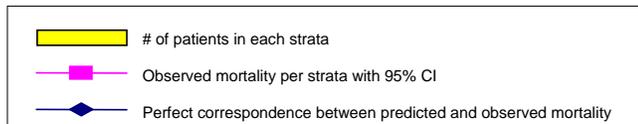
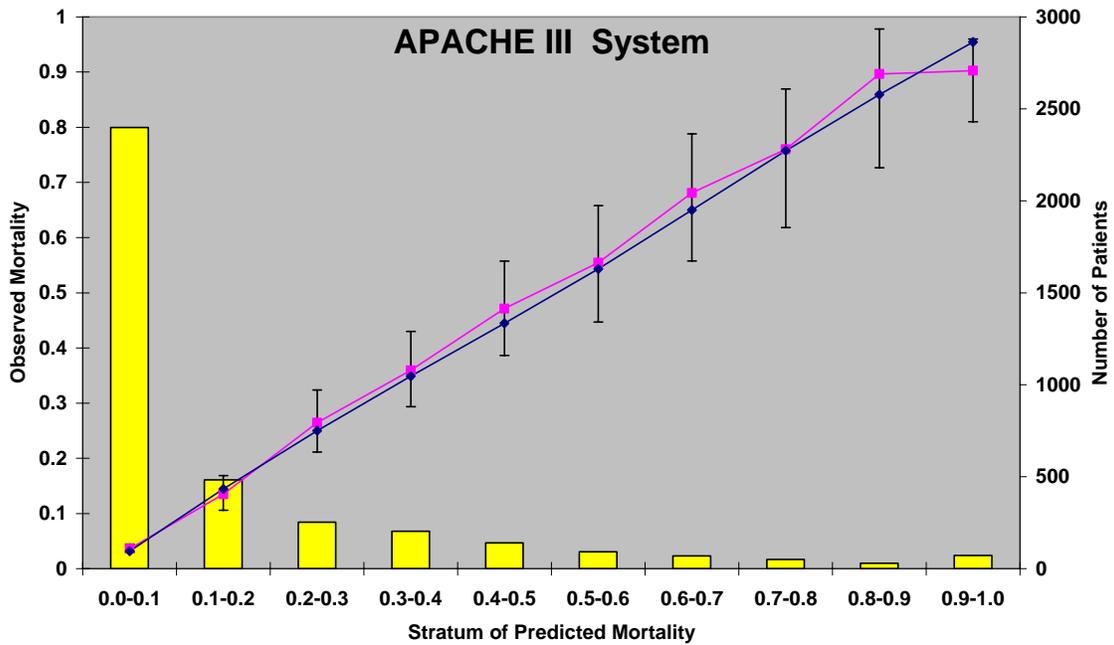
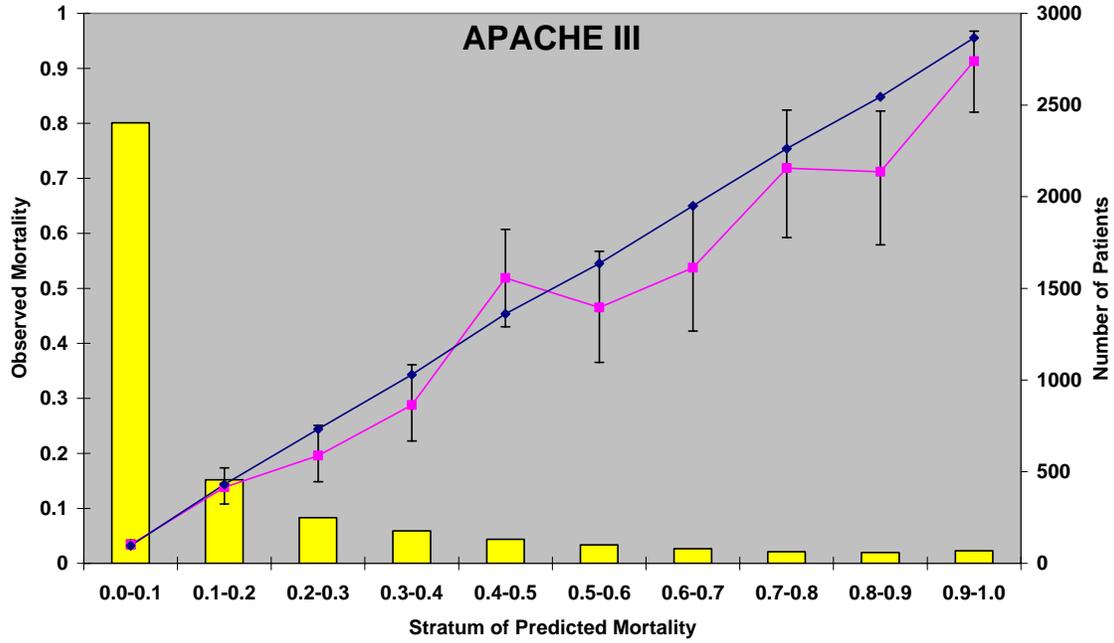
\* = The models represent the extant ICU-risk adjustment models re-estimated using the CALICO database  
† = Area under the receiver operating characteristic curve

**Table 15.2**  
**Calibration of the APACHE III System model**

Model*	H-L† Statistic	
	C Test	H Test
<b>MPM<sub>0</sub> II</b>	11.3 (P=0.33)	13.6 (P=0.19)
<b>SAPS II</b>	15.2 (P=0.12)	6.9 (P=0.73)
<b>APACHE II</b>	15.2 (P=0.12)	16.0 (P=0.10)
<b>APACHE III</b>	20.4 (P=0.03)	27.1 (P<0.01)
<b>APACHE III System</b>	16.4 (P=0.09)	8.0 (P=0.63)

\* = The models represent the extant ICU-risk adjustment models re-estimated using the CALICO database  
† = Hosmer-Lemeshow statistic; df 10

**Figure 15.1**  
**Calibration curves of the APACHE III Models**



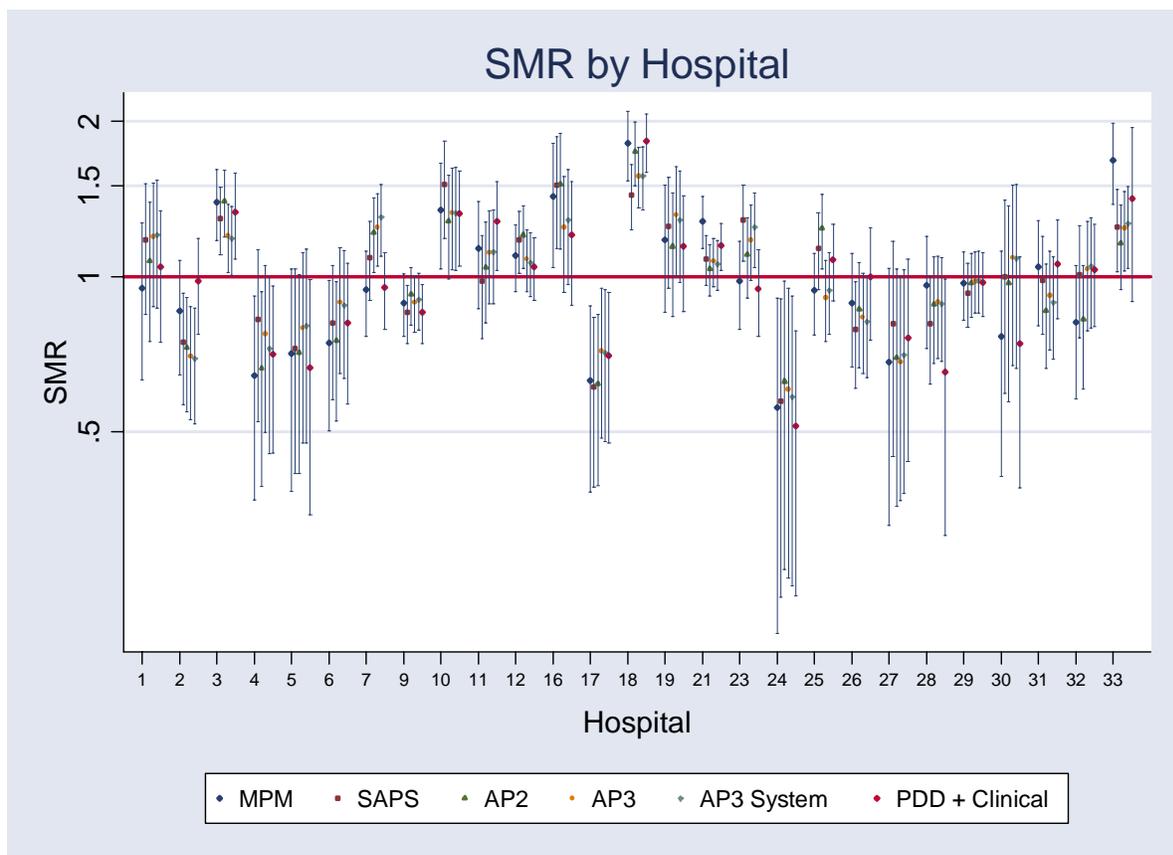
# Results V: Comparative Performance of the Hospitals

The performance of each hospital was evaluated using standardized mortality ratios (SMR). The expected mortality was calculated for each hospital using the re-estimated coefficients from the ICU risk-adjustment models. To get an SMR for each hospital, the observed mortality was divided by the model-specific expected mortality.

Each hospital's SMR with 95% confidence interval for six models is displayed in Figure 16.1. For 32 of the 33 hospitals, the 95% confidence intervals for the point estimates of the SMR overlapped between models. For Hospital 33, the MPM<sub>0</sub> II and APACHE II SMR confidence intervals did not overlap. Regardless of the model used, each hospital had a similar SMR.

In contrast to the fairly similar performance of the risk-adjustment models, there are large and statistically significant differences among the hospitals in terms of SMR. For hospitals that submitted over 100 patients, SMRs varied from approximately 0.5 to approximately 2.0, regardless of the risk-adjustment model used. This corresponds to risk-adjusted mortality rates from about 7% to about 31%.

**Figure 16.1**  
**SMR and 95% confidence interval of the hospitals (Hospitals n >100 patients)**



The ability of each of the models to identify outliers was evaluated using a hospital fixed effect model using the entire sample excluding hospitals with data submitted on less than 100 patients. A logistic model was fitted with patient characteristics (different for each model) plus hospital as a class variable. Then each hospital effect was contrasted versus the “average effect” of all the hospital effects. The “average effect” was a simple average, not weighted by the number of ICU patients sampled from each hospital. A  $P < 0.05$  for the contrast test was used to identify a hospital as an outlier. Appendix Table A.15 shows hospital outliers using the SMR. High outliers (higher than expected mortality) were defined as hospitals whose lower 95% confidence bound for their SMR was greater than 1. Conversely, low outliers (lower than expected mortality) were defined as hospitals whose upper 95% confidence bound for their SMR was lower than 1.

Note that  $P < 0.05$  does not completely rule out the possibility that a hospital appears to be better or worse just by chance, and we do not currently have a means to confirm that a hospital’s quality rating as judged by its mortality performance truly reflects its degree of adherence to best practices. Also note that larger hospitals are more likely to reach  $P < 0.05$  than smaller ones with the same actual quality, simply because of higher statistical power.

Table 16.1 shows which hospitals were identified as outliers by each model using the contrast test method with fixed hospital effects. Unfortunately, there is no gold standard at this time to assess the sensitivity and specificity of each of the models in predicting outliers. No model predicted an outlier hospital in which another model predicted the same hospital as an outlier on the opposite end of the spectrum (i.e., in no instance was a hospital labeled a high outlier by one model and a low outlier by another model). There were two hospitals that were identified as outliers regardless of the risk-adjustment model utilized.

The outliers appeared to have clinically important differences in risk-adjusted mortality rates from the average of hospitals. The absolute difference in risk compared to average mortality for all hospitals was -2.3% to -8.6% for low outliers and +3.2% to +16.0% for high outliers.

We also ranked all the hospitals that submitted valid data on at least 100 patients by their SMR as calculated using each model (see Table 16.2). The hospital with the lowest SMR with a given model received a rank of 1 and the hospital with the highest SMR a rank of 27. In addition, in Table 16.2, the hospitals were grouped into quartiles by hospital performance.

Regardless of the model used, the hospitals were ranked similarly. When hospitals changed quartiles between models, generally they moved only a single quartile. The exceptions were Hospitals 7, 11, 25, and 30, which moved two quartiles between models. No hospitals moved from the top quartile to the bottom quartile or vice versa. Spearman rank correlation coefficients between the models are displayed in Table 16.3. The rank orders produced by the models were highly correlated regardless of the models compared. In a general sense, as the discriminatory power of the model at the individual level (the AUC) increased, the correlation with crude mortality performance as assessed for a hospital’s entire population decreased, suggesting that there is a need for risk adjustment in the ICU population. The high correlations between the MPM<sub>0</sub> II and the PDD + clinical models are to be expected (since the latter includes elements of

the former). The correlation between APACHE III and APACHE III System is so high as to suggest that there is almost no difference in hospital rankings when switching to the System reason for admission, while inter-rater reliability is increased.

**Table 16.1**  
**Statistically significant high and low hospital outliers\***  
**(Contrast test using fixed hospital effects; hospitals with N>100)**

Hospital	N	MPM <sub>0</sub> II	SAPS II	APACHE II	APACHE III	APACHE III System	PDD + Clinical
24	169	L	L	L	L	L	L
17	345	L	L	L	L	L	L
4	183	L		L		L	L
27	152						
5	147						L
6	351	L		L			
30	206						
32	265						
2	437		L	L	L	L	
9	697		L		L		L
26	370		L		L	L	
7	661			H		H	
25	445			H			
1	130						
28	328		L				L
29	738						
23	435		H			H	
31	434						
12	523		H	H			
11	338						H
19	126				H		
21	905	H					H
10	183	H	H	H	H	H	H
3	142	H	H	H	H	H	H
16	117	H	H	H			
33	165	H	H		H	H	
18	143	H	H	H	H	H	H

\* Hospitals arranged from lowest to highest SMR by MPM<sub>0</sub> II model  
 Note: H= high mortality outlier hospital; L= low mortality outlier hospital

**Table 16.2**  
**Rank order of hospitals' SMRs (sorted by crude mortality rank)**

Hospital	N	Crude Mortality	MPM <sub>0</sub> II	SAPS II	APACHE II	APACHE III	APACHE III System	PDD + Clinical
24	169	1	1	1	2	1	1	1
17	345	2	2	2	1	4	4	4
30	206	3	7	14	13	17	17	6
27	152	4	4	6	4	2	3	7
6	351	5	6	8	7	9	8	8
4	183	6	3	9	3	5	5	5
2	437	7	9	4	6	3	2	13
31	434	8	18	13	9	12	10	18
5	147	9	5	3	5	6	6	3
7	661	10	13	17	22	24	25	11
26	370	11	11	5	10	7	7	14
28	328	12	15	7	11	10	9	2
11	338	13	20	12	16	18	18	23
32	265	14	8	15	8	14	14	15
29	738	15	16	11	14	13	13	12
23	435	16	17	23	18	19	21	10
1	130	17	14	20	17	20	20	16
25	445	18	12	18	23	11	12	19
21	905	19	22	16	15	15	15	21
9	697	20	10	10	12	8	11	9
10	183	21	23	27	24	26	26	24
12	523	22	19	19	21	16	16	17
19	126	23	21	22	19	25	23	20
16	117	24	25	26	26	23	24	22
33	165	25	26	21	20	22	22	26
18	143	26	27	25	27	27	27	27
3	142	27	24	24	25	21	19	25

Legend:

TOP QUARTILE (Lowest SMR)	2 <sup>nd</sup> QUARTILE (2 <sup>nd</sup> Lowest SMR)
3 <sup>rd</sup> QUARTILE (2 <sup>nd</sup> Highest SMR)	BOTTOM QUARTILE (Highest SMR)

**Table 16.3**  
**Spearman rank correlation coefficients**

	Crude	MPM <sub>0</sub> II	SAPS II	APACHE II	APACHE III	APACHE III system
MPM <sub>0</sub> II	0.8675	-	-	-	-	-
SAPS II	0.8217	0.8193	-	-	-	-
APACHE II	0.8492	0.8541	0.9084	-	-	-
APACHE III	0.7418	0.8284	0.9170	0.8852	-	-
APACHE III System	0.7466	0.8053	0.9206	0.8956	0.9902	-
PDD + Clinical	0.8046	0.8724	0.7772	0.7796	0.7369	0.7012

\* P ≤ 0.001 for all values

# Summary and Conclusions

Thirty-three hospitals with annual ICU admissions ranging from 1,200 to more than 2,400 patients a year participated in the CALICO Project. A database of 10,398 patients was assembled and 9,441 patients were eligible for inclusion in the analyses across all four extant ICU models, MPM<sub>0</sub> II, SAPS II, APACHE II and APACHE III. As detailed in Chapters 12, 13, and 15, all four models required recalibration and varied in their subsequent discrimination and calibration. In addition to the four models, models based wholly or primarily on the Patient Discharge Abstract and a simplification of the APACHE III model were developed. The following are the conclusions reached based on the analyses of these models.

## **Evidence of Variation in ICU Outcomes**

CALICO has yielded several findings for those interested in comparing hospitals' ICU outcomes. The most important of these is that, regardless of the approach used to risk adjustment, there is substantial variation among hospitals in their mortality rates after risk adjustment. The various models we investigated included the measures of acute physiology, chronic health and comorbidities, reason for admission, and type of admission that experts in the field have maintained are the most important aspects of case mix; and all models studied show at least 2-fold differences in mortality performance across hospitals. Thus, although it is still possible that this variation reflects some as yet unmeasured case mix differences, these differences would have to be due to factors the field has not identified previously. It is unlikely that such factors would entirely explain performance differences of this magnitude. The only major category of risk adjusters not included in the models are socioeconomic factors, and these have never been shown to cause outcome differences of this magnitude for any acute care condition. Therefore, it is reasonable to expect that publishing risk-adjusted ICU outcomes for hospitals will offer consumers, providers, and other stakeholders clinically meaningful and useful information about hospital performance.

## **ICU Mortality Risk Adjustment Methods**

If public reporting of ICU outcomes in the state of California is warranted, then, there remains the issue of choosing a preferred model for ICU performance assessment. The original versions of the four extant models had calibration problems that could make their use for comparing ICU performance very problematic. The preferred ICU risk adjustment model, therefore, is likely to come from among the recalibrated versions of MPM<sub>0</sub> II, SAPS II, APACHE II, APACHE III, and the PDD models and the APACHE III System variant developed in CALICO.

The primary criteria on which California policymakers should compare these five models are: predictive accuracy (i.e., calibration and discrimination), data reliability (percent

agreement and kappas for the variables used in each model), data collection burden, and alignment with (or opportunity to influence) national decisions about how to measure ICU mortality. Because the data in CALICO come from a group of volunteer hospitals that comprise only a small proportion of all California hospitals and that were not randomly selected, information about these criteria from CALICO must be considered estimates. However, our findings on the four extant models are consistent with prior literature.

In terms of predictive accuracy, the results for these models suggest that MPM<sub>0</sub> II and the PDD model have substantially worse discrimination than the other models and APACHE III has the best discrimination (although it is only slightly better than SAPS II, APACHE II, APACHE III System, and the PDD + clinical model). The calibration curves suggest that MPM<sub>0</sub> II, APACHE II, and APACHE III and especially the PDD models are less well calibrated than SAPS II and APACHE III System. In terms of how these calibration issues might reflect which hospitals are labeled outliers, it is especially worrisome that the calibration issues for the PDD models and APACHE III are concentrated in the highest risk patients, as these might not be allocated randomly among hospitals. It should be noted that, although it was beyond the scope of this project (since there is a very large number of possible combinations), it is possible that an alternative approach to the PDD plus clinical model, developed by expanding and/or modifying the prognostic risk factors, could calibrate better.

The calibration issues with APACHE III may reflect the difficulty of collecting some of its variables reliably. The data reliability issues for APACHE III may be a manifestation of the original developers' desire for the model to capture—through an approach that incorporates both a large number of variables and some variables for which there are many possible response options—the complexity and heterogeneity of the ICU patient population. While admirable from a clinical perspective, the result of this approach is that different data collectors may disagree about the correct assessment of some elements of APACHE III that have large weights in the model's predictions, especially reason for admission. For the other models, the data can be collected significantly more reliably. In addition, the APACHE III System model addresses the most problematic variable of the APACHE III by replacing the specific ICU reason for admission with a simpler variable (designating the primary system affected) that can be collected with reasonable inter-rater agreement. Further study of the APACHE III Reason for Admission variable may be indicated.

There also are large differences among the models in the burden of data collection. The PDD models are the least burdensome, of course. Of the other models, MPM<sub>0</sub> II requires by far the least in terms of data collection costs, both because it has many fewer variables than the APACHE models and because, unlike SAPS II and the APACHE models, it only requires data collection at the time of admission, rather than over the first day in the ICU. SAPS II is less burdensome than the APACHE models, because it has many fewer variables and does not require the most difficult task, selecting a reason for admission.

At this point, nationally, the JCAHO has favored APACHE III (and its successor APACHE IV, which differs from APACHE III, in terms of the variables used, only in that it includes a larger number of choices for reason for admission). However, data reliability issues have recently surfaced with use of this model and JCAHO may be willing to consider alternatives as they move forward in their efforts to develop an ICU mortality model.

## Summary

In summary, there is sufficient evidence to justify moving forward with measuring and reporting ICU performance. In terms of risk adjustment model selection, the PDD model is the most immediately feasible, but has severe limitations in terms of discrimination and calibration and probably should not be adopted. MPM<sub>0</sub> II has significantly worse discrimination than the models other than PDD, but is much less burdensome, while APACHE III has slightly better discrimination and is preferred by JCAHO, but may have calibration and data reliability issues. SAPS II, APACHE II, and APACHE III System are between MPM<sub>0</sub> II and APACHE III in terms of discrimination but have better calibration and data reliability. Since SAPS II does better than APACHE II on all criteria, it seems reasonable to drop this model from consideration. The PDD plus clinical model is similar to SAPS II in its discrimination but has worse calibration in a way that would be expected to influence which hospitals are labeled outliers.

Thus, the choice is among:

- PDD + clinical: least burdensome model to get good discrimination at the cost of calibration,
- MPM<sub>0</sub> II: similar burden with worse discrimination but slightly less problematic calibration,
- SAPS II: better predictive accuracy than MPM<sub>0</sub> II with less burden than APACHE III or APACHE III System, but not in JCAHO's plans,
- APACHE III: burdensome, with data reliability issues, but aligned with JCAHO, or
- APACHE III System: between SAPS II and APACHE III in burden, good discrimination, calibration, and data reliability, and entirely calculable from the variables JCAHO is currently beta testing.

The CALICO investigators believe this choice should be made after further public discussion. This discussion should include consideration of additional testing of these models and the reliability of the reason for admission versus the system variable in a broader sample of hospitals.

# Appendix

**Table A.1**  
**Hosmer-Lemeshow Goodness-of-Fit (C test) - Original models**

## MPM II

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic
0.007 - 0.034	466	2	10.3	464	455.7	6.88
0.034 - 0.052	461	17	19.7	444	441.3	0.39
0.052 - 0.073	454	27	28.3	427	425.7	0.07
0.073 - 0.100	478	24	41.5	454	436.5	8.07
0.100 - 0.127	450	35	51.4	415	398.6	5.89
0.127 - 0.154	469	53	65.8	416	403.2	2.90
0.154 - 0.202	465	66	82.1	399	382.9	3.83
0.203 - 0.287	461	93	111.2	368	349.8	3.94
0.288 - 0.445	463	138	162.5	325	300.5	5.70
0.445 - 0.968	463	255	287.9	208	175.1	9.94
Total	4630	710	860.8	3920	3769.2	47.61

**C= 47.61      df 10, P < 0.0001**

## SAPS II

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic
0.000 - 0.017	478	4	4.9	474	473.1	0.18
0.020 - 0.033	470	9	12.3	461	457.7	0.89
0.037 - 0.047	428	12	17.6	416	410.4	1.84
0.052 - 0.072	519	16	32.4	503	486.6	8.84
0.079 - 0.106	467	21	43.2	446	423.8	12.53
0.117 - 0.153	444	32	59.0	412	385.0	14.28
0.167 - 0.230	435	53	84.7	382	350.3	14.70
0.247 - 0.370	438	78	130.5	360	307.5	30.12
0.392 - 0.661	497	183	253.7	314	243.3	40.23
0.681 - 0.997	454	302	379.4	152	74.6	96.21
Total	4630	710	1017.7	3920	3612.3	219.83

**C= 219.83      df 10, P < 0.0001**

**Table A.1 (continued)**  
**Hosmer-Lemeshow Goodness-of-Fit (C test) – Original models**

**APACHE II**

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic
0.001 - 0.033	468	2	8.9	466	459.1	5.46
0.033 - 0.056	457	9	20.4	448	436.6	6.63
0.056 - 0.081	463	16	31.0	447	432.0	7.74
0.081 - 0.109	465	13	43.9	452	421.1	24.05
0.110 - 0.146	467	32	59.5	435	407.5	14.54
0.146 - 0.201	456	47	78.5	409	377.5	15.24
0.201 - 0.263	462	62	106.2	400	355.8	23.87
0.265 - 0.387	468	85	148.5	383	319.5	39.75
0.389 - 0.577	463	165	220.7	298	242.3	26.84
0.577 - 0.996	461	279	342.1	182	118.9	45.09
Total	4630	710	1059.5	3920	3570.5	209.20

**C= 209.20    *df* 10, P < 0.0001**

**APACHE III**

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic
0.000 - 0.009	463	2	2.4	461	460.6	0.07
0.009 - 0.017	463	5	5.8	458	457.2	0.11
0.017 - 0.030	463	8	10.7	455	452.3	0.69
0.030 - 0.048	463	11	17.7	452	445.3	2.61
0.048 - 0.074	463	28	27.5	435	435.5	0.01
0.074 - 0.117	463	34	43.7	429	419.3	2.36
0.117 - 0.186	463	45	68.8	418	394.2	9.69
0.186 - 0.315	463	86	111.1	377	351.9	7.48
0.315 - 0.571	463	166	198.2	297	264.8	9.13
0.571 - 0.997	463	325	353.8	138	109.2	9.96
Total	4630	710	839.7	3920	3790.3	42.11

**C= 42.11    *df* 10, P < 0.0001**

**Table A.2**  
**Hosmer-Lemeshow Goodness-of-Fit (H test) – Original models**

**MPM II**

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic	
0.007 - 0.100	1859	70	99.9	1789	1759.1	9.46	
0.100 - 0.200	1348	149	192.0	1199	1156.0	11.24	
0.200 - 0.300	555	115	135.5	440	419.5	4.10	
0.301 - 0.400	323	97	110.9	226	212.1	2.65	
0.400 - 0.499	185	58	83.0	127	102.0	13.64	
0.500 - 0.598	131	70	71.7	61	59.3	0.09	
0.600 - 0.694	103	58	66.4	45	36.6	3.02	
0.701 - 0.800	65	48	48.6	17	16.4	0.03	
0.803 - 0.896	48	34	40.7	14	7.3	7.29	
0.904 - 0.968	13	11	12.1	2	0.9	1.59	
Total	4630	710	860.8	3920	3769.2	53.10	
<b>H= 53.10</b>						<b>df 10, P &lt; 0.0001</b>	

**SAPS II**

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic	
0.000 - 0.097	2259	57	99.3	2202	2159.7	18.88	
0.106 - 0.196	835	66	122.2	769	712.8	30.26	
0.213 - 0.285	375	60	92.7	315	282.3	15.31	
0.306 - 0.392	266	57	92.3	209	173.7	20.65	
0.415 - 0.484	185	51	82.6	134	102.4	21.81	
0.507 - 0.598	167	69	92.3	98	74.7	13.16	
0.619 - 0.681	108	59	69.8	49	38.2	4.72	
0.700 - 0.799	163	84	121.8	79	41.2	46.39	
0.813 - 0.897	145	95	124.0	50	21.0	46.98	
0.905 - 0.997	127	112	120.7	15	6.3	12.51	
Total	4630	710	1017.7	3920	3612.3	230.67	
<b>H= 230.67</b>						<b>df 10, P &lt; 0.0001</b>	

**Table A.2 (continued)**  
**Hosmer-Lemeshow Goodness-of-Fit (H test) – Original models**

**APACHE II**

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic
0.001 - 0.099	1720	36	90.0	1684	1630.0	34.21
0.100 - 0.199	1052	82	151.2	970	900.8	37.03
0.201 - 0.300	650	95	159.0	555	491.0	34.08
0.300 - 0.399	323	62	111.8	261	211.2	33.96
0.402 - 0.498	267	88	120.3	179	146.7	15.79
0.500 - 0.598	192	80	105.6	112	86.4	13.80
0.602 - 0.699	161	73	103.9	88	57.1	25.95
0.700 - 0.798	101	66	75.3	35	25.7	4.49
0.801 - 0.898	120	91	101.0	29	19.0	6.29
0.902 - 0.996	44	37	41.3	7	2.7	7.10
Total	4630	710	1059.5	3920	3570.5	212.70

**H= 212.70    *df* 10, P < 0.0001**

**APACHE III**

Predicted Mortality Within Decile (%)	# of Admissions	Observed Deaths	Expected Deaths	Observed Survivors	Expected Survivors	H-L Statistic
0.000 - 0.100	2608	68	89.3	2540	2518.7	5.28
0.100 - 0.200	701	78	100.3	623	600.7	5.80
0.200 - 0.299	360	64	87.3	296	272.7	8.18
0.300 - 0.400	234	65	81.5	169	152.5	5.12
0.400 - 0.499	162	60	72.7	102	89.3	4.05
0.500 - 0.598	147	73	81.1	74	65.9	1.79
0.601 - 0.699	125	72	80.8	53	44.2	2.72
0.700 - 0.800	110	78	82.9	32	27.1	1.16
0.801 - 0.900	100	75	85.0	25	15.0	7.87
0.901 - 0.997	83	77	78.8	6	4.2	0.77
Total	4630	710	839.7	3920	3790.3	42.74

**H= 42.74    *df* 10, P < 0.0001**

**Table A.3**  
**MPM<sub>0</sub> II variables and re-estimated coefficients**  
 (Estimation samples)

Variable	Original	Re-estimated Model			
		Split #1	Split #2	Split #3	
Intercept	<b>-5.468</b>	-6.012	-6.081	-6.017	*
Coma or deep stupor	<b>1.486</b>	1.576	1.458	1.543	*
Heart Rate $\geq$ 150 beats/min	<b>0.456</b>	0.980	1.211	0.990	*
Systolic blood pressure $\leq$ 90 mmHg	<b>1.061</b>	0.893	0.919	0.772	*
Chronic renal insufficiency	<b>0.919</b>	0.892	0.816	0.816	*
Cirrhosis	<b>1.137</b>	0.963	1.175	1.054	*
Metastatic neoplasm	<b>1.200</b>	1.119	1.387	0.996	*
Acute renal failure	<b>1.482</b>	1.071	0.745	1.016	*
Cardiac dysrhythmia	<b>0.281</b>	-0.732	-0.626	-0.467	1,2
Cerebrovascular accident	<b>0.213</b>	0.601	0.597	0.491	1,2
GI bleeding	<b>0.397</b>	-0.218	-0.115	-0.516	NS
Intracranial mass effect	<b>0.865</b>	0.688	0.534	0.653	*
CPR prior to admission	<b>0.570</b>	1.279	1.396	1.481	*
Mechanical ventilation	<b>0.791</b>	0.674	0.591	0.666	*
Non-elective Surgery	<b>1.191</b>	1.100	1.257	1.210	*
Age	<b>0.031</b>	0.037	0.038	0.037	*

\* = P < 0.05 in all three splits  
 1 = P < 0.05 in Split #1  
 2 = P < 0.05 in Split #2

**Table A.4**  
**SAPS II variables and re-estimated coefficients**  
 (Estimation samples)

Variable	Original	Re-estimated Model			
		Split #1	Split #2	Split #3	
Intercept	<b>-7.763</b>	-7.558	-7.050	-8.987	*
SAPS II Score	<b>0.074</b>	0.067	0.075	0.062	*
LOG SAPS II Score	<b>0.997</b>	0.842	0.632	1.280	NS

\* = P < 0.05 in all three splits

**Table A.5**  
**APACHE II variables and re-estimated coefficients**  
**(Estimation samples)**

Variables	Original	Re-estimated Model (1) <sup>†</sup>				Re-estimated Model (2) <sup>‡</sup>			
		Split #1	Split #2	Split #3		Split #1	Split #2	Split #3	
Intercept	<b>-3.517</b>	-4.847	-4.905	-4.878	*	-4.724	-4.784	-4.713	*
APACHE II score	<b>0.146</b>	0.172	0.175	0.174	*	0.171	0.173	0.172	*
Emergency surgery	<b>0.603</b>	0.500	0.514	0.481	NS	0.848	0.869	0.807	*
Diagnostic categories <sup>§</sup>	<b>n/a</b>	0.475	0.492	0.514	*	-	-	-	
Coronary artery disease	<b>-0.191</b>	-	-	-		0.012	-0.042	-0.095	NS
Metabolic/renal/hematologic non-operative	<b>-0.885</b>	-	-	-		-0.595	-0.523	-0.632	*
Rhythm disturbance	<b>-1.368</b>	-	-	-		-0.675	-0.504	-0.570	1,3
Congestive heart failure	<b>-0.424</b>	-	-	-		-0.549	-0.484	-0.645	1,3
GI bleeding	<b>0.334</b>	-	-	-		-0.514	-0.216	-0.347	NS
Neurologic non-operative	<b>-0.759</b>	-	-	-		0.153	0.212	0.195	NS
ICH/SDH/SAH	<b>0.723</b>	-	-	-		0.834	0.667	0.662	*
COPD	<b>-0.367</b>	-	-	-		-0.379	-0.205	-0.260	NS
Sepsis	<b>0.113</b>	-	-	-		-0.060	0.009	-0.155	NS
Cardiovascular post-operative	<b>-0.797</b>	-	-	-		-0.870	-0.755	-0.916	3
Pulmonary edema (non-cardiogenic)	<b>-0.251</b>	-	-	-		-0.748	-0.462	-0.456	1
Drug overdose	<b>-3.353</b>	-	-	-		-1.594	-1.933	-1.869	*
Cardiovascular non-operative	<b>0.47</b>	-	-	-		-0.237	-0.273	-0.351	NS
Metabolic/renal/hematologic post-operative	<b>-0.196</b>	-	-	-		-2.346	-2.436	-2.546	*
Gastrointestinal non-operative	<b>0.501</b>	-	-	-		-0.008	0.234	0.141	NS
Craniotomy for neoplasm	<b>-1.245</b>	-	-	-		-1.073	-1.295	-1.523	3
Respiratory non-operative	<b>-0.89</b>	-	-	-		0.430	0.330	0.308	NS
Thoracic surgery for neoplasm	<b>-0.802</b>	-	-	-		-0.761	-0.577	-0.713	NS
GI perforation/obstruction	<b>0.06</b>	-	-	-		-0.603	-0.511	-0.336	NS
Hemorrhagic shock/hypovolemia	<b>0.493</b>	-	-	-		-1.130	-0.440	-0.520	NS
Gastrointestinal post-operative	<b>-0.613</b>	-	-	-		-0.608	-0.505	-0.445	NS
Peripheral vascular surgery	<b>-1.315</b>	-	-	-		-0.224	-0.809	-1.019	NS
GI surgery for neoplasm	<b>-0.248</b>	-	-	-		-0.076	-0.517	-0.720	NS
Other diagnoses <sup>  </sup>	<b>n/a</b>	-	-	-		1.101	1.137	1.234	*

\* = P < 0.05 in all three splits

1 = P < 0.05 in Split #1

3 = P < 0.05 in Split #3

<sup>†</sup> = Diagnostic category coefficients were NOT re-estimated

<sup>‡</sup> = Diagnostic category coefficients were re-estimated

<sup>§</sup> = Relative contribution of the diagnostic categories as a whole in predicting mortality compared to the other variables in the risk equation

<sup>||</sup> = Diagnostic categories with insufficient patients to generate a coefficient were combined into a single category

**Table A.6**  
**APACHE III variables and re-estimated coefficients**  
**(Estimation samples)**

Variable		Original	Re-estimated Model (1) <sup>†</sup>			Re-estimated Model (2) <sup>‡</sup>				
			Split #1	Split #2	Split #3	Split #1	Split #2	Split #3		
Intercept:		<b>-6.413</b>	-5.874	-5.994	-5.100	*	-5.844	-5.891	-5.080	*
Age:	45-54	<b>0.342</b>	0.181	-0.185	-0.089	NS	0.187	-0.203	-0.105	NS
	55-59	<b>0.321</b>	0.632	0.329	0.422	<sup>1</sup>	0.685	0.344	0.451	<sup>1</sup>
	60-64	<b>0.613</b>	0.646	0.613	0.266	<sup>1,2</sup>	0.672	0.630	0.292	<sup>1,2</sup>
	65-69	<b>0.757</b>	0.750	0.450	0.623	<sup>1,3</sup>	0.783	0.459	0.624	<sup>1,3</sup>
	70-74	<b>1.006</b>	1.215	0.911	1.123	*	1.248	0.912	1.135	*
	75-84	<b>1.127</b>	1.221	1.081	1.097	*	1.276	1.093	1.116	*
	≥85	<b>1.495</b>	1.543	1.235	1.166	*	1.576	1.243	1.180	*
Past Medical History:	AIDS	<b>1.024</b>	-1.081	-1.654	-1.340	NS	-1.154	-1.670	-1.320	NS
	Hepatic failure	<b>1.13</b>	0.189	0.274	0.959	<sup>3</sup>	0.134	0.205	0.943	NS
	Lymphoma	<b>1.005</b>	1.386	1.331	1.589	*	1.359	1.359	1.579	*
	Metastatic cancer	<b>0.886</b>	0.561	0.872	0.292	<sup>2</sup>	0.583	0.869	0.323	<sup>1,2</sup>
	Leukemia/multiple myeloma	<b>0.756</b>	0.274	0.824	-0.026	NS	0.259	0.761	-0.055	NS
	Immunosuppression	<b>0.321</b>	0.714	0.249	-0.455	NS	0.741	0.258	-0.451	NS
	Cirrhosis	<b>0.860</b>	0.735	0.698	0.124	<sup>2</sup>	0.674	0.655	0.131	NS
Acute Physiology Score:	APS	<b>0.088</b>	0.046	0.065	0.022	<sup>1,2</sup>	0.048	0.065	0.023	<sup>2</sup>
	APS3ALL1	<b>-0.309</b>	0.265	0.077	0.579	<sup>3</sup>	0.255	0.075	0.568	NS
	APS3ALL2	<b>0.513</b>	-0.545	-0.248	-1.133	<sup>3</sup>	-0.533	-0.240	-1.112	<sup>3</sup>
	APS3ALL3	<b>-0.353</b>	0.492	0.377	1.007	<sup>3</sup>	0.500	0.359	0.984	<sup>3</sup>
	APS3ALL4	<b>0.451</b>	-0.363	-0.229	-0.762	NS	-0.402	-0.163	-0.724	NS
Location:	Admitted to ICU from floor	<b>0.048</b>	0.272	0.133	0.211	NS	0.306	0.154	0.237	NS
	Transfer	<b>0.206</b>	0.621	0.805	0.841	<sup>2,3</sup>	0.592	0.824	0.869	<sup>2,3</sup>
	Admit to ICU from OR	<b>-0.238</b>	0.004	-0.309	-0.471	NS	-0.162	-0.406	-0.561	NS
Other:	Emergency surgery	<b>0.079</b>	0.337	0.232	1.021	<sup>3</sup>	0.359	0.216	1.021	<sup>3</sup>
	Pre-ICU LOS (days)	<b>0.141</b>	0.112	0.165	0.159	*	0.111	0.162	0.160	*
	Excess hospital LOS	<b>0.069</b>	0.101	0.093	0.066	<sup>1</sup>	0.098	0.091	0.062	NS
Diagnosis:	Diagnostic categories <sup>§</sup>	n/a	0.869	0.736	0.874	*	-	-	-	NS
	Acute myocardial infarction	<b>0.678</b>	-	-	-		0.438	0.082	0.466	NS
	Respiratory medical other	<b>0.220</b>	-	-	-		-0.173	-0.171	-0.298	NS
	Pneumonia, bacterial	<b>0.357</b>	-	-	-		0.126	0.333	0.251	NS
	Rhythm disturbance	<b>-0.236</b>	-	-	-		-0.465	-0.308	-0.075	NS
	Cardiovascular medical other	<b>-0.292</b>	-	-	-		-0.627	-0.392	-0.555	NS
	COPD	<b>0.438</b>	-	-	-		-0.013	0.181	0.195	NS
	Sepsis	<b>0.354</b>	-	-	-		-0.065	-0.175	0.037	NS
	Drug intoxication / overdose	<b>-1.528</b>	-	-	-		-0.954	-1.850	-1.544	<sup>2</sup>
	Intracranial hemorrhage	<b>1.521</b>	-	-	-		1.412	1.190	1.450	*
	Diabetic ketoacidosis	<b>-1.920</b>	-	-	-		-2.482	-0.787	-1.504	<sup>1,3</sup>
	Other diagnoses <sup>  </sup>	n/a	-	-	-		0.728	0.754	0.823	*

\* = P < 0.05 in all three splits

<sup>1</sup> = P < 0.05 in Split #1

<sup>2</sup> = P < 0.05 in Split #2

<sup>3</sup> = P < 0.05 in Split #3

<sup>†</sup> = Diagnostic category coefficients were NOT re-estimated

<sup>‡</sup> = Diagnostic category coefficients were re-estimated

<sup>§</sup> = Relative contribution of the diagnostic categories as a whole in predicting mortality compared to the other variables in the risk equation

<sup>||</sup> = Diagnostic categories with insufficient patients to generate a coefficient were combined into a single category

**Table A.7**  
**MPM<sub>0</sub> II re-estimated coefficients and odds ratios**  
**(100% sample)**

Variable	Original		Re-estimated Model			
	Coefficient	OR	Coefficient	OR	95% CI	
Intercept	<b>-5.468</b>	n/a	<b>-5.632</b> *	n/a		
Coma or deep stupor	<b>1.486</b>	4.42	<b>1.376</b> *	3.96	3.17	- 4.95
Heart Rate ≥ 150 beats/min	<b>0.456</b>	1.58	<b>0.909</b> *	2.48	1.82	- 3.38
Systolic blood pressure ≤ 90 mmHg	<b>1.061</b>	2.89	<b>1.055</b> *	2.87	2.49	- 3.32
Chronic renal insufficiency	<b>0.919</b>	2.51	<b>0.967</b> *	2.63	2.05	- 3.39
Cirrhosis	<b>1.137</b>	3.12	<b>1.378</b> *	3.97	2.96	- 5.32
Metastatic neoplasm	<b>1.200</b>	3.32	<b>1.024</b> *	2.78	2.14	- 3.62
Acute renal failure	<b>1.482</b>	4.40	<b>0.768</b> *	2.16	1.54	- 3.02
Cardiac dysrhythmia	<b>0.281</b>	1.32	<b>-0.576</b> *	0.56	0.43	- 0.73
Cerebrovascular accident	<b>0.213</b>	1.24	<b>0.568</b> *	1.76	1.35	- 2.31
GI bleeding	<b>0.397</b>	1.49	<b>-0.374</b> *	0.69	0.51	- 0.93
Intracranial mass effect	<b>0.865</b>	2.38	<b>0.673</b> *	1.96	1.47	- 2.61
CPR prior to admission	<b>0.570</b>	1.77	<b>1.289</b> *	3.63	2.82	- 4.68
Mechanical ventilation	<b>0.791</b>	2.21	<b>0.792</b> *	2.21	1.91	- 2.55
Non-elective surgery	<b>1.191</b>	3.29	<b>1.024</b> *	2.78	2.24	- 3.45
Age	<b>0.031</b>	1.03	<b>0.031</b> *	1.03	1.03	- 1.04

\* = Coefficient was significant (P < 0.05) in logistic regression

**Table A.8**  
**SAPS II re-estimated coefficients and odds ratios**  
**(100% sample)**

Variable	Original		Re-estimated Model			
	Coefficient	OR	Coefficient	OR	95% CI	
Intercept	<b>-7.763</b>	n/a	<b>-8.205</b> *	n/a		
SAPS II Score	<b>0.074</b>	1.08	<b>0.068</b> *	1.07	1.05	- 1.09
LOG SAPS II Score	<b>0.997</b>	2.71	<b>1.030</b> *	2.80	1.27	- 6.20

\* = Coefficient was significant (P < 0.05) in the logistic regression

**Table A.9**  
**APACHE II re-estimated coefficients and odds ratios**  
**(100% sample)**

Variables	Original		Re-estimated Model			
	Coefficient	OR	Coefficient	OR	95% CI	
Intercept	<b>-3.517</b>	n/a	<b>-4.687</b>			
APACHE II Score	<b>0.146</b>	1.16	<b>0.181</b>	*	1.20	1.19 - 1.21
Emergency surgery	<b>0.603</b>	1.83	<b>0.332</b>	NS	1.39	0.95 - 2.04
Sepsis	<b>0.113</b>	1.12	<b>-0.228</b>	NS	0.80	0.58 - 1.10
Peripheral vascular surgery	<b>-1.315</b>	0.27	<b>-1.350</b>	*	0.26	0.09 - 0.74
Craniotomy for neoplasm	<b>-1.245</b>	0.29	<b>-1.560</b>	*	0.21	0.06 - 0.69
Thoracic surgery for neoplasm	<b>-0.802</b>	0.45	<b>-1.130</b>	*	0.32	0.13 - 0.83
GI surgery for neoplasm	<b>-0.248</b>	0.78	<b>-0.009</b>	NS	0.99	0.52 - 1.87
Resp. insufficiency after surgery	<b>-0.14</b>	0.87	<b>0.079</b>	NS	1.08	0.56 - 2.11
GI perforation/obstruction	<b>0.06</b>	1.06	<b>-0.100</b>	NS	0.91	0.52 - 1.59
Respiratory surgical, other	<b>-0.61</b>	0.54	<b>-1.592</b>	*	0.20	0.06 - 0.68
Asthma/allergy	<b>-2.108</b>	0.12	<b>-0.240</b>	NS	0.79	0.54 - 1.15
Pulmonary edema (non-cardiogenic)	<b>-0.251</b>	0.78	<b>-0.223</b>	NS	0.80	0.51 - 1.25
Aspiration/poisoning/toxic	<b>-0.142</b>	0.87	<b>-0.778</b>	*	0.46	0.27 - 0.77
Hypertension	<b>-1.798</b>	0.17	<b>-2.388</b>	*	0.09	0.02 - 0.39
Rhythm disturbance	<b>-1.368</b>	0.25	<b>-0.617</b>	*	0.54	0.37 - 0.78
Congestive heart failure	<b>0.47</b>	1.60	<b>-0.611</b>	*	0.54	0.38 - 0.78
Hemorrhagic shock/hypovolemia	<b>0.493</b>	1.64	<b>-0.628</b>	*	0.53	0.32 - 0.89
Coronary artery disease	<b>-0.191</b>	0.83	<b>-0.436</b>	*	0.65	0.47 - 0.90
Seizure disorder	<b>-0.584</b>	0.56	<b>-2.017</b>	*	0.13	0.05 - 0.38
Drug overdose	<b>-3.353</b>	0.03	<b>-1.538</b>	*	0.22	0.10 - 0.48
GI bleeding	<b>0.334</b>	1.40	<b>-0.637</b>	*	0.53	0.35 - 0.80
Neurologic medical, other	<b>-0.759</b>	0.47	<b>-0.056</b>	NS	0.95	0.66 - 1.35
Gastrointestinal medical, other	<b>0.501</b>	1.65	<b>0.156</b>	NS	1.17	0.72 - 1.90
Low-risk respiratory medical, other			<b>-0.651</b>	NS	0.52	0.26 - 1.04
Miscellaneous medical, other	<b>-0.885</b>	0.41	<b>-0.908</b>	*	0.40	0.29 - 0.57
Cardiac surgery, other	<b>-0.797</b>	0.45	<b>-1.056</b>	*	0.35	0.19 - 0.63
Miscellaneous surgery, other	<b>-0.196</b>	0.82	<b>-2.631</b>	*	0.07	0.02 - 0.25
Neurologic surgery, other	<b>-1.15</b>	0.32	<b>-1.051</b>	*	0.35	0.14 - 0.89
GI surgery, other	<b>-0.613</b>	0.54	<b>-0.096</b>	NS	0.91	0.51 - 1.63
High-risk respiratory medical, other			<b>0.388</b>	NS	1.47	1.00 - 2.18
Cardiogenic shock/cardiac arrest			<b>0.705</b>	*	2.02	1.22 - 3.35
Cardiac medical, other	<b>-0.797</b>	0.45	<b>-0.556</b>	*	0.57	0.34 - 0.96
ICH/SDH/SAH	<b>0.723</b>	2.06	<b>0.409</b>	*	1.51	1.07 - 2.11

\* = Coefficient was significant (P< 0.05) in logistic regression

**Table A.10**  
**APACHE III re-estimated coefficients and odds ratios**  
**(100% sample)**

Variables		Original		Re-estimated Model				
		Coefficient	OR	Coefficient	OR	95% CI		
	Intercept	<b>-6.4134</b>	n/a	<b>-5.6202</b>	*	n/a		
Age:	45-54	<b>0.3428</b>	1.41	<b>0.0261</b>	NS	1.026	0.753 - 1.4	
	55-59	<b>0.3214</b>	1.38	<b>0.4878</b>	*	1.629	1.167 - 2.273	
	60-64	<b>0.6134</b>	1.85	<b>0.6274</b>	*	1.873	1.338 - 2.62	
	65-69	<b>0.7568</b>	2.13	<b>0.6351</b>	*	1.887	1.363 - 2.612	
	70-74	<b>1.0062</b>	2.74	<b>0.9085</b>	*	2.481	1.82 - 3.38	
	75-84	<b>1.1268</b>	3.09	<b>1.0711</b>	*	2.919	2.22 - 3.837	
	≥85	<b>1.4950</b>	4.46	<b>1.1394</b>	*	3.125	2.28 - 4.283	
Past Medical	AIDS	<b>1.0241</b>	2.78	<b>0.6953</b>	NS	2.004	0.854 - 4.704	
History:	Hepatic failure	<b>1.1334</b>	3.11	<b>1.2448</b>	*	3.472	1.951 - 6.181	
	Lymphoma	<b>1.0048</b>	2.73	<b>1.0195</b>	*	2.772	1.42 - 5.41	
	Metastatic cancer	<b>0.8859</b>	2.43	<b>0.8523</b>	*	2.345	1.627 - 3.379	
	Leukemia/multiple myeloma	<b>0.7557</b>	2.13	<b>0.5377</b>	NS	1.712	0.923 - 3.174	
	Immunosuppression	<b>0.3214</b>	1.38	<b>0.5212</b>	*	1.684	1.196 - 2.371	
	Cirrhosis	<b>0.8605</b>	2.36	<b>0.9764</b>	NS	2.655	1.729 - 4.076	
	Acute	APS	<b>0.0880</b>	1.09	<b>0.0504</b>	*	1.052	1.023 - 1.082
Physiology	APS3ALL1	<b>-0.3095</b>	0.73	<b>0.1477</b>	NS	1.159	0.846 - 1.587	
	Score:	APS3ALL2	<b>0.5125</b>	1.67	<b>-0.3177</b>	NS	0.728	0.412 - 1.285
	APS3ALL3	<b>-0.3533</b>	0.70	<b>0.3288</b>	NS	1.389	0.845 - 2.285	
	APS3ALL4	<b>0.4508</b>	1.57	<b>-0.1878</b>	NS	0.829	0.375 - 1.829	
Location:	Admitted to ICU from floor	<b>0.0480</b>	1.05	<b>0.2756</b>	*	1.317	1.099 - 1.58	
	Transfer	<b>0.2056</b>	1.23	<b>0.4349</b>	NS	1.545	0.962 - 2.481	
	Admit to ICU from OR	<b>-0.2375</b>	0.79	<b>-1.0007</b>	*	0.368	0.138 - 0.98	
Other:	Emergency surgery	<b>0.0787</b>	1.08	<b>0.3429</b>	NS	1.409	0.906 - 2.191	
	Pre-ICU LOS (days)	<b>0.1415</b>	1.15	<b>0.1249</b>	*	1.133	1.065 - 1.205	
	Excess hospital LOS	<b>0.0690</b>	1.07	<b>0.1137</b>	*	1.12	1.056 - 1.189	
Diagnosis:	Acute myocardial infarction	<b>1.0455</b>	2.84	<b>0.2155</b>	NS	1.24	0.816 - 1.886	
	Asthma	<b>-0.6017</b>	0.55	<b>-1.6479</b>	NS	0.192	0.025 - 1.474	
	Pneumonia, aspiration	<b>-0.3654</b>	0.69	<b>-0.4652</b>	NS	0.628	0.349 - 1.132	
	Pneumonia, bacterial	<b>0.2923</b>	1.34	<b>0.4665</b>	*	1.594	1.059 - 2.401	
	Shock, cardiogenic	<b>1.2797</b>	3.60	<b>1.0245</b>	*	2.786	1.715 - 4.526	
	Coma/change in LOC	<b>-0.6575</b>	0.52	<b>-0.8172</b>	*	0.442	0.211 - 0.925	
	Emphysema/bronchitis	<b>0.1474</b>	1.16	<b>0.3595</b>	NS	1.433	0.916 - 2.24	
	Cardiovascular medical, other	<b>-0.2502</b>	0.78	<b>-0.327</b>	NS	0.721	0.461 - 1.127	
	Diabetic ketoacidosis	<b>-1.8568</b>	0.16	<b>-2.2645</b>	*	0.104	0.032 - 0.335	
	Musculoskel. medical & surgical, other	<b>-0.2299</b>	0.79	<b>-0.1313</b>	NS	0.877	0.277 - 2.777	
	GI bleed, upper	<b>-0.1537</b>	0.86	<b>-0.3003</b>	NS	0.741	0.404 - 1.357	
	GI bleed, lower	<b>-0.5844</b>	0.56	<b>0.2994</b>	NS	1.349	0.574 - 3.168	
	GI bleed, esophageal varices	<b>0.1766</b>	1.19	<b>-0.4509</b>	NS	0.637	0.288 - 1.409	
	GI medical, other	<b>-0.1357</b>	0.87	<b>0.3428</b>	NS	1.409	0.818 - 2.428	
	Hem/onc medical, other	<b>0.3988</b>	1.49	<b>0.2057</b>	NS	1.228	0.719 - 2.1	
	High risk respiratory medical, other		1.00	<b>0.7367</b>	*	2.089	1.168 - 3.738	
	High risk GI surgical, other		1.00	<b>1.1935</b>	NS	3.299	0.958 - 11.353	
	Hypertension, uncontrolled	<b>-0.3131</b>	0.73	<b>-0.8879</b>	NS	0.412	0.16 - 1.061	
	Hemorrhage/hematoma, intracranial	<b>1.4402</b>	4.22	<b>1.3812</b>	*	3.979	2.475 - 6.398	
	Metabolic disorder	<b>-0.3188</b>	0.73	<b>-0.1868</b>	NS	0.83	0.391 - 1.76	
	Neurologic medical, other	<b>-0.2673</b>	0.77	<b>0.5983</b>	NS	1.819	0.943 - 3.509	
	Drug withdrawal or overdose	<b>-1.2749</b>	0.28	<b>-0.9133</b>	*	0.401	0.176 - 0.914	
	ARDS	<b>0.8648</b>	2.37	<b>0.5249</b>	NS	1.69	0.953 - 2.997	
	Genitourinary medical, other	<b>-0.3645</b>	0.69	<b>-0.3444</b>	NS	0.709	0.387 - 1.296	
	Respiratory- medical, other	<b>0.0474</b>	1.05	<b>0.2903</b>	NS	1.337	0.872 - 2.05	
	Rhythm disturbance	<b>-0.3131</b>	0.73	<b>-0.2171</b>	NS	0.805	0.516 - 1.254	
	Graft, aorto-iliac and fem-pop bypass	<b>-0.5905</b>	0.55	<b>0.6701</b>	NS	1.954	0.434 - 8.798	
	Subarachnoid hemorrhage/AVM	<b>1.5737</b>	4.82	<b>1.7443</b>	*	5.722	2.857 - 11.458	

**Table A.10 (continued)**  
**APACHE III re-estimated coefficients and odds ratios**  
**(100% sample)**

Variables	Original		Re-estimated Model					
	Coefficient	OR	Coefficient	OR	95% CI			
Epidural and subdural hematomas	<b>-0.0180</b>	0.98	<b>0.4237</b>	NS	1.528	0.602	-	3.877
Cardiovascular surgery, other	<b>-0.2202</b>	0.80	<b>0.3116</b>	NS	1.366	0.345	-	5.403
Endarterectomy, carotid	<b>-0.6211</b>	0.54	<b>-0.7756</b>	NS	0.46	0.05	-	4.235
Neoplasm-cranial, surgery for	<b>-0.0157</b>	0.98	<b>0.3069</b>	NS	1.359	0.288	-	6.422
Seizures	<b>-1.1798</b>	0.31	<b>-1.6148</b>	*	0.199	0.056	-	0.706
Aortic aneurysm	<b>-0.3857</b>	0.68	<b>0.3486</b>	NS	1.417	0.372	-	5.4
Sepsis	<b>0.3471</b>	1.41	<b>0.2623</b>	NS	1.3	0.85	-	1.988
GI cancer	<b>-0.4887</b>	0.61	<b>1.5034</b>	*	4.497	1.355	-	14.92
GI obstruction, surgery for	<b>-0.2789</b>	0.76	<b>1.0406</b>	NS	2.831	0.809	-	9.902
GI surgery, other	<b>-1.1071</b>	0.33	<b>1.1037</b>	NS	3.015	0.994	-	9.15
Spinal cord surgery, other	<b>-0.1578</b>	0.85	<b>0.8252</b>	NS	2.282	0.474	-	11
Neurologic surgery, other	<b>0.1988</b>	1.22	<b>1.5833</b>	*	4.871	1.199	-	19.785
GU surgery, other	<b>-1.4836</b>	0.23	<b>-1.5874</b>	NS	0.204	0.022	-	1.916
Respiratory surgery, other	<b>0.0963</b>	1.10	<b>0.7054</b>	NS	2.025	0.503	-	8.152
Resp. cancer, surgery for	<b>-0.3233</b>	0.72	<b>0.7541</b>	NS	2.126	0.582	-	7.764
Intracranial hemorrhage, surgery for		1.00	<b>1.2372</b>	NS	3.446	0.95	-	12.503
Stroke	<b>0.4174</b>	1.52	<b>0.9925</b>	*	2.698	1.382	-	5.268
Unstable angina	<b>-0.1629</b>	0.85	<b>-1.1662</b>	NS	0.312	0.093	-	1.043

**Table A.11**  
**Combined APACHE II diagnostic categories**

APACHE II Diagnostic Category	Code	N	Crude Mortality Rate	Combined APACHE II Diagnostic Category
Asthma/allergy	R25	74	0.01	lowRESPmed
Pulmonary embolus	R13	69	0.17	lowRESPmed
Diabetic ketoacidosis	M29	25	0.04	MISCmed
Metabolic/renal/hematologic non-operative	M/K/H	719	0.14	MISCmed
Admission due to chronic CV disease after surgery	C35	13	0.08	CARDsurg
Cardiovascular post-operative	C	283	0.05	CARDsurg
Heart valve surgery	C11	2	0.50	CARDsurg
Hemorrhagic shock	C22	18	0.06	CARDsurg
Renal surgery for neoplasm	K02	21	0.00	MISCsurg
Renal transplant	K33	4	0.00	MISCsurg
Metabolic/renal/hematologic post-operative	M/K/H	257	0.01	MISCsurg
Neurologic post-operative	N	124	0.02	NEUROsurg
Head Trauma	N03	1	0.00	NEUROsurg
Laminectomy and other spinal cord surgery	N14	62	0.03	NEUROsurg
GI bleeding	G23	25	0.16	GISurg
Gastrointestinal post-operative	G	131	0.12	GISurg
Post-respiratory arrest	R24	14	0.36	highRESPMED
Respiratory non-operative	R	215	0.28	highRESPMED
Neoplasm	R02	34	0.44	highRESPMED
Cardiogenic shock	C19	99	0.53	cardioshock
Post-cardiac arrest	C24	31	0.81	cardioshock
Cardiovascular non-operative	C	248	0.09	CARDmed
Dissecting thoracic/abdominal aneurysm	C12	36	0.06	CARDmed
ICH/SDH/SAH	N06	393	0.27	ICH
Craniotomy for ICH/SDH/SAH	N06	82	0.18	ICH

**Table A.12**  
**Combined APACHE III diagnostic categories**

APACHE III Diagnostic Code	Description	APACHE III Coefficient	N	Mortality Rate	APACHE III Combined Code
CARDIOG	Shock, cardiogenic	0.78	98	0.52	CARDIOG
CARDARR	Cardiac arrest	0.55	62	0.73	CARDIOG
CARDIOMY	Cardiomyopathy	1.36	27	0.15	CARDIOG
MEDAORT	Aneurysm, dissecting aortic	1.26	30	0.07	CARDIOG
HYPERT	Hypertension, uncontrolled	-0.13	117	0.03	HYPERT
PERIART	Aneurysm/pseudoaneurysm, other, arterial thrombosis	0.13	47	0.06	HYPERT
CVOTH	Cardiovascular medical, other	-0.29	506	0.12	CVOTH
SEPTICUT	Sepsis, renal/UTI	-0.48	54	0.17	CVOTH
SCARDOTH	Vascular surgery, other	-0.25	51	0.04	SCARDOTH
SAORTDIS	Aneurysm, abdominal aortic; with dissection	0.96	13	0.23	SCARDOTH
SFEMAORT	Graft, aorto-femoral bypass	-0.63	17	0.06	SCARDOTH
SPERISC	Embolectomy, thrombectomy, or dilatation	-0.06	18	0.00	SCARDOTH
SRUPAOR	Aneurysm, thoracic aortic; with rupture	0.24	1	0.00	SCARDOTH
PANCRE	Pancreatitis	-0.13	4	0.00	SRENOTH
SRENTNAN	Kidney transplant	-1.4	4	0.00	SRENOTH
SOBHYST	Hysterectomy	-0.46	20	0.00	SRENOTH
SRENCA	Surgery for urinary tract cancer	-0.63	34	0.03	SRENOTH
SRENOTH	Genitourinary surgery, other	-1.4	12	0.00	SRENOTH
GIOTHER	GI medical, other	0.2	51	0.20	GIOTHER
HEPATF	Hepatic failure, acute	0.29	22	0.36	GIOTHER
GIINFLA	Inflammatory bowel disease, diverticulitis, cholangitis, peritonitis, or GI abscess	-0.01	40	0.25	GIOTHER
GINEOP	Cancer, other GI	1.39	16	0.31	GIOTHER
GIOBSTR	GI obstruction	0.25	19	0.16	GIOTHER
GIPERF	GI perforation/rupture	-0.21	7	0.29	GIOTHER
GIVASC	GI vascular insufficiency	0.19	6	0.17	GIOTHER
SGIOTH	GI surgery, other	-0.23	206	0.06	SGIOTH
SABIACES	Fistula/abscess, surgery for (not inflammatory bowel disease)	-0.25	18	0.22	SGIOTH
SGIBLEE	Bleeding - GI, surgery for	-0.29	26	0.15	SGIOTH
SPANCRE	Pancreatitis, surgery for	-0.28	7	0.14	SGIOTH
SPERITON	Peritonitis, surgery for	-0.28	7	0.43	SGIOTH
SGICHOL	Cholecystectomy/cholangitis, surgery for	-0.75	40	0.08	SGIOTH
SLIVERTR	Liver transplant	-1.2	8	0.00	SGIOTH
SGIINFL	Inflammatory bowel disease or diverticular disease, surgery for	-0.28	24	0.13	SGIOTH
SGIPERF	GI perforation/rupture, surgery for	0.48	64	0.20	highriskSGI
SGIVASC	GI vascular ischemia, surgery for (resection)	0.67	15	0.27	highriskSGI
COAGTHRO	Thrombocytopenia, pancytopenia, neutropenia, or coagulopathy	1.09	11	0.45	HEMONCOTH
HEMAMISC	Hematologic medical, other	0.4	183	0.23	HEMONCOTH
NEUROTH	Neurologic medical, other	0.13	59	0.12	NEUROTH
NEONEUR	Neoplasm, neurologic	0.17	37	0.08	NEUROTH
NEURINF	Meningitis	0.69	43	0.14	NEUROTH
NEURMUS	Neuromuscular medical, other	-0.22	15	0.13	NEUROTH
SICH	Hemorrhage/hematoma-intracranial, surgery for	1.06	22	0.27	SSTROKE
SSAH	Arteriovenous malformation, surgery for	0.31	36	0.11	SSTROKE
SSAH	Subarachnoid hemorrhage/intracranial aneurysm, surgery for	0.83	37	0.08	SSTROKE
AIROB	Obstruction-airway (i.e., acute epiglottitis, post-extubation edema, foreign body, etc.)	0.91	64	0.22	HIRISKRESP
BACVPNEU	Pneumonia, viral	0.52	12	0.17	HIRISKRESP
PARAPNEU	Pneumonia, fungal or parasitic	1.24	25	0.44	HIRISKRESP
RESPCA	Cancer, lung or airway	1.11	32	0.41	HIRISKRESP
RESPOTH	Respiratory medical, other	0.21	383	0.20	RESPOTH
PULEMB	Embolus, pulmonary	0.21	69	0.17	RESPOTH
RESPARR	Arrest, respiratory (without cardiac arrest)	0.3	9	0.33	RESPOTH
SRESOTH	Respiratory surgery, other	0.21	64	0.03	SRESPTH
SRESPINF	Respiratory infection/abscess, surgery for	-0.23	28	0.11	SRESPOTH
SRESPCA	Respiratory cancer, surgery for	-0.11	146	0.05	SRESPCA
SRESPLAR	Cancer-laryngeal, tracheal, oral, or sinus, surgery for	-0.22	30	0.00	SRESPCA

**Table A.13**  
**PDD model estimated coefficients and odds ratios**  
(Estimation sample)

Variable		Coefficient	P value	OR	95% CI		
	Intercept	-5.56	<.0001	n/a	-	-	-
	Age (years)	0.03	<.0001	1.03	1.02	-	1.04
	Male	-0.08	0.39	0.92	0.77	-	1.10
Primary reason for admission	Acute myocardial infarction (3, other)	-0.42	0.15	0.66	0.38	-	1.16
	Adult respiratory failure	-0.10	0.69	0.90	0.54	-	1.50
	Pneumonia	0.13	0.61	1.14	0.68	-	1.91
	Septicemia	0.51	0.05	1.67	1.00	-	2.81
	Other oncology/hematology	0.57	0.06	1.77	0.97	-	3.21
	Acute myocardial infarction (2, inferior-post)	-0.34	0.31	0.71	0.37	-	1.37
	Other cerebrovascular disease	-0.31	0.36	0.73	0.38	-	1.42
	Other genitourinary	-0.09	0.77	0.91	0.49	-	1.71
	Other gastrointestinal	-0.15	0.66	0.86	0.45	-	1.66
	Coronary artery disease	-2.20	0.00	0.11	0.03	-	0.49
	Diabetes mellitus with complications	-0.64	0.16	0.53	0.22	-	1.30
	GI hemorrhage	-0.62	0.10	0.54	0.26	-	1.12
	Acute myocardial infarction (1, anterior)	-0.47	0.19	0.63	0.31	-	1.25
	Complication, device	0.09	0.80	1.09	0.56	-	2.11
	Other Neurology	-0.10	0.80	0.90	0.40	-	2.03
	Acute cerebrovascular dz. (2, intracerebral)	1.08	0.00	2.95	1.53	-	5.70
	Dysrhythmia (1, supraventricular)	-1.84	0.00	0.16	0.05	-	0.47
	Other infection	0.50	0.13	1.65	0.87	-	3.12
	Precerebral occlusion w/o infarct	-0.87	0.25	0.42	0.10	-	1.86
	Aspiration pneumonia	0.11	0.72	1.12	0.60	-	2.08
	Poisoning due to medication	-2.80	0.02	0.06	0.01	-	0.70
	Acute cerebrovascular dz. (3, precerebral)	1.30	0.00	3.69	1.90	-	7.15
	COPD	0.02	0.96	1.02	0.50	-	2.06
	Other nutritional disorder	-13.66	0.98	<0.001	-	-	-
	Acute cerebrovascular dz. (1, SAH/SDH)	0.77	0.04	2.17	1.02	-	4.62
	Aneurysm	0.25	0.57	1.28	0.54	-	3.04
	Bronchogenic/lung cancer	0.33	0.46	1.40	0.58	-	3.35
	Other respiratory	-0.16	0.71	0.85	0.36	-	1.99
	Complication of procedure	-0.06	0.91	0.95	0.37	-	2.43
	Secondary malignancy	0.17	0.69	1.19	0.51	-	2.75
	Other benign neoplasm	-13.86	0.98	<0.001	-	-	-
	Heart valve disorder	-0.20	0.77	0.82	0.21	-	3.18
	Hypertension with complication	-0.58	0.20	0.56	0.23	-	1.37
	Peripheral atherosclerosis	0.90	0.02	2.46	1.16	-	5.22
Alcoholic liver disease	0.84	0.08	2.32	0.90	-	5.98	
Other psychiatric	-14.18	0.98	<0.001	-	-	-	
Diverticulosis	-0.02	0.97	0.98	0.38	-	2.54	
Poisoning psychiatric	-0.90	0.24	0.41	0.09	-	1.81	
Brain / nervous system cancer	-0.85	0.42	0.43	0.05	-	3.39	
Acute renal failure	0.41	0.32	1.51	0.68	-	3.35	
Biliary disorder	-0.10	0.86	0.90	0.29	-	2.82	
Other liver diagnosis	1.72	<.0001	5.56	2.59	-	11.94	
Intracranial injury	0.35	0.49	1.42	0.53	-	3.85	
Epilepsy / seizures	-1.07	0.15	0.34	0.08	-	1.45	

**Table A.13 (continued)**  
**PDD model estimated coefficients and odds ratios**  
**(Estimation sample)**

	Variable	Coefficient	P value	OR	95% CI		
	Fluid / electrolyte disorder	-0.11	0.84	0.90	0.32	-	2.52
	Intestinal obstruction	0.63	0.21	1.87	0.71	-	4.95
	Pancreatic disorder	-0.14	0.81	0.87	0.27	-	2.80
	Colon cancer	0.27	0.59	1.31	0.49	-	3.54
	Conduction disorder	-1.36	0.20	0.26	0.03	-	2.06
	Pulm heart dx	0.92	0.05	2.51	0.99	-	6.39
	Pleurisy	-1.00	0.15	0.37	0.09	-	1.46
	Chest pain	-16.88	0.98	<0.001	-	-	-
	Carditis	-1.89	0.09	0.15	0.02	-	1.34
	Back problem	-1.61	0.16	0.20	0.02	-	1.92
	Other obstetrical / gynecological	-13.73	0.99	<0.001	-	-	-
Charlson chronic conditions	Angina 1 (chonic)	-0.42	0.12	0.66	0.39	-	1.11
	Angina 2 (unstable)	-1.05	0.21	0.35	0.07	-	1.82
	Arrhythmia 1 (minor)	0.29	0.01	1.33	1.08	-	1.64
	Arrhythmia 2 (severe)	1.01	<.0001	2.74	1.89	-	3.97
	Vascular heart disease 1 (valve disorder)	-0.13	0.45	0.88	0.63	-	1.23
	Vascular heart disease 2 (valve replacement)	0.18	0.60	1.19	0.62	-	2.31
	Myocardial infarction (acute)	0.62	0.01	1.87	1.20	-	2.90
	Myocardial infarction (old)	-0.27	0.22	0.77	0.50	-	1.17
	CHF	0.40	0.00	1.49	1.20	-	1.85
	Peripheral vascular disease	0.18	0.29	1.20	0.86	-	1.69
	Cerebrovascular disease	0.20	0.46	1.22	0.72	-	2.08
	COPD 1 (without pulmonary hypertension)	0.01	0.92	1.01	0.82	-	1.25
	COPD 2 (with pulmonary hypertension)	0.09	0.74	1.09	0.66	-	1.79
	Neurologic other	0.95	<.0001	2.59	1.78	-	3.76
	Dementia	-0.31	0.22	0.73	0.44	-	1.21
	Paralysis	-0.16	0.67	0.85	0.42	-	1.75
	Endocrine other	-0.12	0.42	0.89	0.66	-	1.19
	Diabetes	0.17	0.13	1.19	0.95	-	1.48
	Diabetic ketoacidosis	-1.36	0.02	0.26	0.08	-	0.82
	Diabetes with sequelae	0.31	0.07	1.37	0.98	-	1.91
	Chronic renal failure 1 (not on dialysis)	0.51	0.04	1.66	1.03	-	2.66
	Chronic renal failure 2 (on dialysis)	0.29	0.28	1.33	0.79	-	2.23
	Various cirrhoses	1.08	<.0001	2.95	1.80	-	4.84
	Moderate-severe liver disease	0.72	0.03	2.06	1.08	-	3.90
	Ulcers	0.04	0.89	1.04	0.59	-	1.83
	Various cancers	0.15	0.38	1.16	0.83	-	1.64
	Metastatic solid tumors	0.56	0.01	1.74	1.16	-	2.63
Hypertension 1 (no complications)	-0.35	0.00	0.70	0.58	-	0.86	
Hypertension 2 (with complications)	0.16	0.63	1.18	0.60	-	2.30	
Physiologic variables	HR <60 (at admission)	0.02	0.17	1.02	0.99	-	1.06
	HR >80 (at admission)	0.01	<.0001	1.02	1.01	-	1.02
	Systolic blood pressure <110 (at admission)	0.03	<.0001	1.03	1.03	-	1.04
	Systolic blood pressure >160 (at admission)	0.02	0.01	1.02	1.00	-	1.03
	Glasgow Coma Score 3 to 5	1.32	<.0001	3.75	2.29	-	6.14
Glasgow Coma Score 6 to 14	0.20	0.10	1.22	0.96	-	1.54	

**Table A.13 (continued)**  
**PDD model estimated coefficients and odds ratios**  
**(Estimation sample)**

	<b>Variable</b>	<b>Coefficient</b>	<b>P value</b>	<b>OR</b>	<b>95% CI</b>		
Miscellaneous	Admitted from Emergency Room	0.27	0.23	1.31	0.84	-	2.05
	Medical or Emergency Surgery	0.68	<.0001	1.96	1.43	-	2.71
	CPR within 24 hours prior to ICU admission	1.16	<.0001	3.20	2.24	-	4.58
	Intracranial mass effect at ICU admission	0.72	0.00	2.05	1.32	-	3.18
	Mechanical ventilation at ICU admission	0.76	<.0001	2.13	1.74	-	2.60

**Table A.14**  
**PDD model estimated coefficients and odds ratios**  
**(100% sample)**

	Variable	Coefficient	P value	OR	95% CI		
	Intercept	-5.97	<.0001	n/a			
	Age (years)	0.03	<.0001	1.03	1.03	-	1.04
	Male	-0.05	0.46	0.95	0.83	-	1.09
Primary reason for admission	Acute myocardial infarction (3, other)	-0.07	0.75	0.93	0.60	-	1.44
	Adult respiratory failure	-0.07	0.72	0.93	0.62	-	1.39
	Pneumonia	0.34	0.09	1.41	0.94	-	2.10
	Septicemia	0.72	0.00	2.05	1.37	-	3.07
	Other oncology/hematology	0.83	0.00	2.29	1.44	-	3.66
	Acute myocardial infarction (2, inferior-post)	-0.36	0.18	0.70	0.41	-	1.17
	Other cerebrovascular disease	-0.05	0.85	0.95	0.58	-	1.58
	Other genitourinary	0.10	0.69	1.11	0.67	-	1.82
	Other gastrointestinal	0.27	0.28	1.31	0.80	-	2.15
	Coronary artery disease	-1.87	0.00	0.15	0.05	-	0.45
	Diabetes mellitus with complications	-0.59	0.12	0.55	0.26	-	1.16
	GI hemorrhage	-0.35	0.21	0.71	0.41	-	1.23
	Acute myocardial infarction (1, anterior)	-0.21	0.46	0.81	0.47	-	1.40
	Complication, device	0.12	0.68	1.12	0.65	-	1.94
	Other neurology	0.12	0.71	1.12	0.60	-	2.10
	Acute cerebrovascular dz (2, intracerebral)	1.11	<.0001	3.04	1.83	-	5.06
	Dysrhythmia (1, supraventricular)	-1.85	<.0001	0.16	0.07	-	0.38
	Other infection	0.47	0.08	1.60	0.95	-	2.70
	Precerebral occlusion w/o infarct	-0.82	0.18	0.44	0.13	-	1.47
	Aspiration pneumonia	0.34	0.16	1.41	0.87	-	2.27
	Poisoning due to medication	-1.80	0.01	0.17	0.05	-	0.62
	Acute cerebrovascular dz (3, precerebral)	1.22	<.0001	3.38	2.02	-	5.66
	COPD	-0.10	0.74	0.91	0.51	-	1.63
	Other nutritional disorder	-1.45	0.16	0.23	0.03	-	1.77
	Acute cerebrovascular dz (1, SAH/SDH)	0.74	0.01	2.10	1.16	-	3.82
	Aneurysm	0.18	0.65	1.19	0.56	-	2.54
	Bronchogenic/lung cancer	0.71	0.04	2.03	1.05	-	3.92
	Other respiratory	0.09	0.79	1.09	0.58	-	2.08
	Complication of procedure	0.01	0.97	1.01	0.51	-	2.04
	Secondary malignancy	0.63	0.05	1.88	1.00	-	3.53
	Other benign neoplasm	-13.57	0.97	<0.001	-	-	-
	Heart valve disorder	0.30	0.55	1.36	0.50	-	3.68
	Hypertension with complication	-0.37	0.33	0.69	0.33	-	1.45
	Peripheral atherosclerosis	1.26	<.0001	3.51	1.92	-	6.44
	Alcoholic liver disease	1.13	0.00	3.08	1.50	-	6.35
	Other psychiatric	-13.94	0.97	<0.001	-	-	-
	Diverticulosis	0.24	0.49	1.27	0.64	-	2.52
	Poisoning psychiatric	-1.20	0.11	0.30	0.07	-	1.29
	Brain / nervous system cancer	-0.17	0.79	0.84	0.24	-	2.93
	Acute renal failure	0.69	0.03	1.99	1.08	-	3.68
	Biliary disorder	0.24	0.54	1.27	0.59	-	2.72
	Other liver diagnosis	1.72	<.0001	5.57	3.07	-	10.11
	Intracranial injury	0.64	0.08	1.89	0.92	-	3.92
	Epilepsy / seizures	-1.33	0.05	0.26	0.07	-	0.98
	Fluid / electrolyte disorder	0.10	0.80	1.11	0.50	-	2.43
	Intestinal obstruction	0.85	0.03	2.33	1.10	-	4.95

**Table A.14 (continued)**  
**PDD model estimated coefficients and odds ratios**  
**(100% sample)**

	Variable	Coefficient	P value	OR	95% CI	
	Pancreatic disorder	0.18	0.69	1.19	0.50	- 2.84
	Colon cancer	0.31	0.47	1.36	0.59	- 3.12
	Conduction disorder	-0.97	0.14	0.38	0.11	- 1.37
	Pulm heart dx	0.76	0.05	2.14	1.00	- 4.59
	Pleurisy	-0.33	0.52	0.72	0.26	- 1.96
	Chest pain	-15.16	0.98	<0.001	-	-
	Carditis	-2.23	0.04	0.11	0.01	- 0.89
	Back problem	-1.61	0.13	0.20	0.03	- 1.61
	Other obstetrical / gynecological	-0.51	0.52	0.60	0.12	- 2.89
Charlson chronic conditions	Angina 1 (chronic)	-0.31	0.12	0.74	0.50	- 1.09
	Angina 2 (unstable)	0.05	0.91	1.05	0.44	- 2.49
	Arrhythmia 1 (minor)	0.17	0.04	1.19	1.01	- 1.40
	Arrhythmia 2 (severe)	0.82	<.0001	2.28	1.71	- 3.04
	Vascular heart disease 1 (valve disorder)	-0.06	0.62	0.94	0.73	- 1.21
	Vascular heart disease 2 (valve replacement)	0.15	0.57	1.16	0.70	- 1.93
	Myocardial infarction (acute)	0.46	0.01	1.59	1.12	- 2.25
	Myocardial infarction (old)	-0.42	0.01	0.65	0.47	- 0.92
	CHF	0.34	<.0001	1.40	1.19	- 1.66
	Peripheral vascular disease	0.29	0.03	1.34	1.03	- 1.74
	Cerebrovascular disease	0.40	0.04	1.50	1.03	- 2.19
	COPD 1 (without pulmonary hypertension)	-0.02	0.80	0.98	0.83	- 1.15
	COPD 2 (with pulmonary hypertension)	0.06	0.75	1.07	0.71	- 1.60
	Neurologic other	0.94	<.0001	2.56	1.92	- 3.41
	Dementia	-0.31	0.12	0.74	0.50	- 1.08
	Paralysis	-0.03	0.90	0.97	0.58	- 1.63
	Endocrine other	-0.08	0.51	0.93	0.74	- 1.16
	Diabetes	0.18	0.04	1.19	1.01	- 1.42
	Diabetic ketoacidosis	-0.48	0.21	0.62	0.30	- 1.31
	Diabetes with sequelae	0.34	0.01	1.40	1.09	- 1.80
	Chronic renal failure 1 (not on dialysis)	0.70	0.00	2.02	1.42	- 2.88
	Chronic renal failure 2 (on dialysis)	0.43	0.03	1.54	1.04	- 2.27
	Various cirrhoses	0.92	<.0001	2.52	1.73	- 3.66
	Moderate-severe liver disease	0.63	0.01	1.88	1.16	- 3.05
	Ulcers	-0.09	0.68	0.91	0.59	- 1.42
	Various cancers	0.38	0.00	1.47	1.14	- 1.89
	Metastatic solid tumors	0.65	<.0001	1.92	1.40	- 2.63
Hypertension 1 (no complications)	-0.31	<.0001	0.73	0.63	- 0.85	
Hypertension 2 (with complications)	0.05	0.86	1.05	0.61	- 1.82	
Physiologic variables	HR <60 (at admission)	0.00	0.72	1.01	0.98	- 1.03
	HR >80 (at admission)	0.02	<.0001	1.02	1.01	- 1.02
	Systolic blood pressure <110 (at admission)	0.03	<.0001	1.03	1.03	- 1.03
	Systolic blood pressure >160 (at admission)	0.01	0.20	1.01	1.00	- 1.02
	Glasgow Coma Score 3 to 5	1.20	<.0001	3.31	2.29	- 4.79
	Glasgow Coma Score 6 to 14	0.18	0.05	1.20	1.00	- 1.45
Miscellaneous	Admitted from Emergency Room	0.60	0.00	1.82	1.30	- 2.56
	Medical or Emergency Surgery	0.87	<.0001	2.40	1.87	- 3.08
	CPR within 24 hours prior to ICU admission	1.14	<.0001	3.14	2.38	- 4.13
	Intracranial mass effect at ICU admission	0.78	<.0001	2.18	1.56	- 3.05
	Mechanical ventilation at ICU admission	0.78	<.0001	2.18	1.86	- 2.54

**Table A.15**  
**Hospital Outliers by SMR method\***

Hospital	N	MPM <sub>0</sub> II	SAPS II	APACHE II	APACHE III	APACHE III System	PDD + Clinical
8	35						
24	169	L	L	L	L	L	L
15	57					L	
17	345	L	L	L	L	L	L
4	183	L		L		L	L
27	152				L		
20	64						
5	147						L
6	351	L		L			
30	206						
32	265						
2	437		L	L	L	L	
9	697		L				L
26	370		L				
25	445			H			
7	661			H	H	H	
1	130						
28	328						L
29	738						
23	435		H			H	
22	77						
31	434						
12	523		H	H			
11	338						H
19	126				H		
21	905	H					H
10	183	H	H		H	H	H
3	142	H	H	H	H	H	H
16	117	H	H	H			
33	165	H	H		H	H	
13	51						
18	143	H	H	H	H	H	H
14	22	H		H			H

H = High Outlier ; L = Low Outlier

\*Hospitals arranged from lowest to highest SMR by MPM<sub>0</sub> II model