Society of Thoracic Surgeons 30-Day Predicted Risk of Mortality Score Also Predicts Long-Term Survival

John D. Puskas, Patrick D. Kilgo, Vinod H. Thourani, Omar M. Lattouf, Edward Chen, J David Vega, William Cooper, Robert A. Guyton and Michael Halkos

From the Clinical Research Unit, Division of Cardiothoracic Surgery, Emory University School of Medicine, Atlanta, GA, USA.

Correspondence:

John D. Puskas, MD, MSc, FACS, FACC
Professor of Surgery, Division of Cardiothoracic Surgery, Emory University
Chief of Cardiac Surgery, Emory University Hospital Midtown
6th Floor Medical Office Tower
550 Peachtree Street
Atlanta, GA, USA
Tel: 404-686-2513
Fax: 404-686-4959
Email: john.puskas@emoryhealthcare.org

Abstract:

Purpose: The Society of Thoracic Surgeons Predicted Risk of Mortality (PROM) score is a well-validated predictor of 30-day mortality after cardiac procedures. However, the role of PROM in predicting longer-term survival has not been investigated.

Methods: From 1/1/1996 to 12/31/2009, 24,222 patients who had PROM scores underwent cardiac procedures at a US academic center. Long-term all-cause mortality was determined by referencing the national Social Security Death Master File. Logistic and Cox survival regression analyses evaluated the long-term predictive utility of the PROM. Area under the receiver operator characteristic (AUROC) curve measured the discrimination of PROM at 1, 3, 5 and 10 years. Kaplan-Meier curves were stratified by quartiles of PROM risk to compare long-term survival. All analyses were performed for both the whole sample and for 30-day survivors.

Results: Overall 30-day mortality was 2.78% (674/24222). As expected, PROM predicted 30-day mortality extremely well (AUROC=0.794). Interestingly, PROM also predicted longer-term survival almost as well (Table). Among all patients and 30-day survivors AUROC values for PROM at 1, 3, 5 and 10 years were remarkably similar to the 30-day endpoint for which PROM is calibrated. Moreover, PROM was highly predictive of Kaplan-Meier survival, even when this analysis was restricted to patients surviving beyond 30 days (Figure). Among 30-day survivors, each percent increase in PROM score was associated with a 9.6% increase (95% CI 9.3%-10.0%) in instantaneous hazard of death ($p<0.001$).

Conclusions: The STS Predicted Risk of Mortality algorithm--developed to predict mortality within 30-days of specific cardiac procedures -- accurately predicts mortality both at 30-days

and during 14 years of follow-up with almost equally strong discriminatory power. This may have profound implications for informed consent as well as for longitudinal comparative effectiveness studies.

Introduction

STS Predicted Risk of Mortality: The STS 30-day risk models were developed to provide clinicians and hospitals with a tool to evaluate risk-adjusted outcomes and to guide quality improvement initiatives. The scores themselves are simply predicted probabilities (ranging from 0 to 1) calculated from a multivariable logistic regression model calibrated on STS data within fixed time periods. Periodic updates of the model coefficients are undertaken in an effort to make the predictions commensurate with evolving technology and generally improved outcomes over time. The STS Predicted Risk of Mortality (PROM) score, most recently calibrated by Shroyer [1], is known to discriminate well between 30-day survivors and non-survivors (c-index=0.78) and has a high degree of agreement between predicted and observed mortality (calibration).

The PROM score can be calculated for five different procedures including 1) isolated primary coronary artery bypass grafting (CABG), 2) isolated aortic valve replacement (AVR), 3) isolated mitral valve repair (MV repair) or replacement (MVR), 4) combined CABG and AVR and, 5) CABG and MV Repair or MVR. The weighting of various risk factors is recalibrated with each new version of the STS Adult Cardiac Database on the basis of the most recent data uploaded to the STS National Cardiac Database (STS NCDB) by the more than 900 participating cardiac

surgical programs in the United States. PROM is used for analysis and comparison of clinical outcomes in comparative effectiveness research, in quality assurance initiatives and recently in various pay-for-performance programs. While there have been other predictive algorithms widely applied to predict short-term outcomes after cardiac surgery, including the Parsonnet score [2] and Euroscore [3], none have proved as accurate nor been as rigorously recalibrated as the STS PROM score, which has become the global standard.

Interestingly, while a great deal of effort has been focused on predicting short-term outcomes after cardiac surgery—driven by demands for improvement in operative and perioperative care processes—there has been relatively little effort to develop a statistical algorithm to predict long-term survival after cardiac procedures. The additive and logistic EuroSCORE has been reported to be predictive of long-term survival in a series of 180 patients undergoing mitral valve surgery [4]; the additive, but not the logistic EuroSCORE was predictive of mid-term survival in 233 patients who had aortic valve replacement and CABG [5]. Moreover, the role of PROM in predicting longer-term survival has not been systematically investigated to date. This is at least in part because the STS NCBD has not included data on survival beyond 30-days. Nonetheless, a reliable predictor of long-term survival would be enormously useful and would have powerful implications for patients, physicians, administrators and society at large as decisions are made about individual treatments, alternative therapies and healthcare funding. Further, such a score might be useful in risk adjustment when evaluating the long-term effects of different treatments. The goal of this study was to test and validate the PROM score as a predictor of long-term survival.

Methods

Subjects and Sample: From January 1, 1996 to December 31, 2009, 30636 patients underwent cardiac surgery at Emory University hospitals. Of these, 24222 (79.1%) had one of the five procedures for which PROM models have been developed: 1) isolated primary coronary artery bypass grafting (CABG), 2) isolated aortic valve replacement (AVR), 3) isolated mitral valve repair (MV repair) or replacement (MVR), 4) combined CABG and AVR and, 5) CABG and MV Repair or MVR. Mitral valve repair patients were included only from 2008 forward since PROM was not calibrated on these patients before that date.

Measurements: Prior to analysis, the 30 pre-operative risk factors used to calculate the PROM score were identified and harvested from the Emory University institutional STS Adult Cardiac Database. For descriptive purposes, each variable was summarized in the exact manner in which it is included in the PROM predictive model for each procedure type (see Table 1). Patient age was used both as a continuous and a dichotomous measure (age > 66, age ≤ 66). Race was represented by three dichotomous variables: Black, Hispanic, or Other race. Chronic lung disease, which had been measured by different scales during the study period, was summarized dichotomously in this study. Ejection fraction was dichotomized as either < 50% or greater than ≥ 50%. Status was measured as either elective or non-elective. Previous incidence of sternotomy was measured dichotomously across two variables: first reoperation and multiple reoperations. New York Heart Association (NYHA) classification score was measured dichotomously as Class IV or non-Class IV.

Short-term (30-day, operative) mortality was measured directly and extracted from the

institutional STS database. Long-term all-cause mortality was determined by referencing the Death Master File via the US National Social Security Death Index (SSDI) by a HIPAA-compliant mechanism. This study was approved by the Emory University Institutional Review Board. On March 14th, 2010, the survival for each of the patients in this study was verified by querying SSDI; patients still alive on this date were considered censored in survival analyses. The sensitivity of the SSDI (92.2%) is comparable to that of the National Death Index among American-born persons (87% to 98%) [6].

For analysis purposes, PROM was treated as a percentage (0 to 100) rather than a probability (0 to 1) so that meaningful interpretations of unit increases in risk could be posited. To evaluate the relationship between PROM and long-term survival endpoints, and to validate it for use as a predictor, a variety of analytical approaches were performed using logistic and survival regression methods.

Model Performance: Survival to fixed points in time (30-days, 1, 3, 5, and 10 years) was analyzed separately for patients who were operated on early enough in the study period to observe the endpoint. In this manner, each eligible patient was classified as either dead or alive at the specified time point. Logistic regression models were constructed that related survival as a function of PROM for each time point and procedure combination and for all procedures combined. In all, 30 models are evaluated (5 time points x 5 procedures + all procedures combined), though the 30-day mortality models were only included for comparison purposes.

Each model was evaluated with respect to discrimination and calibration. *Discrimination* is the model's ability to separate survivors and non-survivors. This was assessed

using the Area Under the Receiver Operating Characteristic Curve (AUROC). AUROC ranges from 0.50 to 1.00; higher values portend better discrimination while values closer to 0.5 indicate that the model's discrimination is essentially random, like flipping a coin. One useful interpretation of the AUROC is as follows: If a randomly selected survivor is paired with a randomly selected non-survivor, then the AUROC is the probability that the non-survivor will have a higher model-predicted risk of death than the survivor.

Calibration refers to the degree of agreement between observed and predicted outcomes. Normally, the Hosmer-Lemeshow (H-L) statistic is recommended to evaluate calibration. However, the H-L statistic is known to be underpowered and overly sensitive to large sample sizes [7-8]. Instead, calibration curves, similar to those originally reported for PROM, were visually inspected for each model. Each curve is a scatter plot of observed and predicted probabilities of death averaged by decile of PROM. The connected points should closely track with the line of identity (y=x).

Further, for each model, the odds ratio was noted, which represents the increase in odds of death at a fixed time point for each unit increase in PROM percentage. Ninety-five percent confidence intervals (CI) were also constructed. Also, the point biserial correlation between survival (dichotomous) and PROM (numerical) was reported for each model to observe how the correlation increases with the length of the fixed time point.

Predictive Validation: Validation in a large sample with only one predictor can be taken for granted in most cases since the real danger of a predictive model formulation is over-fitting (which by definition requires more than one predictor). However, to demonstrate that PROM is

internally valid, two general approaches were employed – a bootstrapping validation and a split-sample validation [9].

The bootstrapping approach was employed for each time point/ procedure subset (hereafter, "the original sample") and involved repeated sampling with replacement of the eligible patients used for each model. A total of 1000 bootstrap samples of size *n* (where n is the number of eligible patients for that model) were collected. For each bootstrap sample, a logistic regression model was fit and the model estimates (intercept and slope) were collected and applied to the original sample of size *n* to estimate predicted probabilities of death for each patient. These predicted probabilities were then used as independent variables in the original sample and model performance statistics (AUROC and point biserial correlation) were computed. This process was completed 1000 times, each time collecting the performance statistics. After all 1000 bootstrap samples were analyzed and applied to the original sample, the 500th ordered value of each performance statistic was considered the best estimate of the true value of the statistic and the 25th and 975th ordered values served as 95% confidence bounds. Also, bootstrapped estimates of the model parameters were collected for reporting purposes.

The split-sample validation approach for each model consisted of dividing the original sample into two halves in a random fashion. Unlike bootstrapping, split-sample validation is not a re-sampling algorithm. The first sample (half of the original sample), called the test set, was used to fit a logistic regression model that related dichotomous survival to PROM. The model parameters were then collected and applied to the second half of the data, called the

holdout set, and predicted probabilities of death were calculated for each patient in the holdout sample. Using these predicted probabilities, model performance estimates were calculated including AUROC, point biserial correlation and calibration curves were generated for each model [10]. Because of the nature of the approach, no confidence intervals are calculated for the performance statistics.

Once estimates of the model performance statistics were calculated using the two validation approaches, they were compared with the analogous measures from the original sample.

Survival Analysis: To evaluate PROM as a predictor of long-term survival, PROM was divided into deciles of risk and ten Kaplan Meier curves were constructed by decile. Similar curves were constructed separately for 30-day survivors to assess PROM's predictive validity apart from the early deaths for which PROM was originally intended.

The PROM score was further evaluated in a Cox proportional hazards regression model for each procedure and for all procedures combined (6 models total). Associated hazard ratios (HR) and 95% CI were computed. The proportional hazards assumption was checked by examining the correlation between ranked survival time and the Schoenfeld residuals for uncensored patients [11]. Adjusted survival estimates were generated from the Cox models and median survival estimated for various values of PROM [12. A smoothing algorithm from a quadratic regression equation was used to create curves where estimated median survival was calculated for each value of PROM.

Results

The patient sample included 24,222 patients including 20014 (82.6%) patients who underwent isolated CABG, 1781 (7.4%) patients who had isolated AVR, 945 (3.9%) patients who had isolated mitral procedures, 423 (1.8%) patients who had CABG/MVR and 1059 (4.4%) patients who had CABG/AVR.

Preoperative characteristics that informed PROM are listed in Table 1 by procedure. Patients undergoing concomitant CABG with valves procedures had the highest average PROM scores, tended to be older, and exhibited more pre-operative comorbidities.

Overall 30-day mortality was 2.8% (674/24222). As expected, PROM discriminated 30-day mortality moderately well overall (AUROC=0.794). In the isolated CABG group PROM exhibited comparable discrimination of 30-day mortality (AUROC=0.769) to that of the original STS cohort on which the PROM score was calibrated (AUROC=0.780) [1]. PROM also discriminated well in the other procedure subgroups; isolated AVR (AUROC=0.763), isolated mitral (AUROC=0.816), CABG + AVR (AUROC=0.749) and CABG + mitral (0.717). These results suggest that this patient population is generally representative of the larger, national population with respect to early survival risk. See Table 2.

For longer term survival, the PROM score showed excellent face validity as demonstrated in Kaplan-Meier analysis. After classifying PROM into deciles of risk, K-M curves were generated for each decile. The ten curves were remarkably sequential, with the first decile having the best survival and each subsequent decile having survival lower than the previous (Figure 1). Further, the same analysis was performed separately for 30-day survivors,

the endpoint for which PROM was originally designed. Strata of PROM deciles exhibited similar sequentiality in 30-day survivors, suggesting a predictive robustness of PROM beyond its original intent. See Figure 2. Survival estimates overall and by procedure are shown in Table 3. Patients undergoing concomitant CABG procedures generally had poorer long-term survival than their isolated counterparts.

When considering survival to fixed time points (1, 3, 5, and 10 years), PROM demonstrated an only slightly diminished ability to discriminate between survivors and non-survivors. For all patients, AUROC for survival to thirty days was 0.794; for 1-year, 0.789; for 3-years, 0.767; for 5-years, 0.763; for 10 years, 0.762. This similarity remains intact in the procedure subgroups of CABG and AVR with the lowest AUROC at more distant time points seldom meaningfully lower (and sometimes higher) than AUROC at 30-days; in isolated CABG cases (0.769 to 0.755), isolated AVR (0.763 to 0.790), CABG + AVR (0.749 to 0.728). In cases where mitral procedures are performed either in isolation of in combination with CABG, there is a more pronounced decline in discrimination; isolated mitral (0.816 to 0.741), CABG + mitral (0.717 to 0.626). PROM is weakest as a predictor of long-term survival in CABG + mitral patients and is a relatively poor discriminator of longer-term survival in these patients. See Table 2.

For each unit increase in PROM percentage, the additional increase in odds of mortality at each fixed time point was calculated. For the entire sample, the odds ratios generally increased for longer-term survival endpoints, being at its highest at 10 years. This was also true in the AVR and CABG procedures; isolated CABG (1-year OR=1.12, 10-year OR=1.89), isolated

AVR (1-year OR=1.11, 10-year OR=1.44), and CABG + AVR (1-year OR=1.12, 10-year OR=1.43). Changes in the OR over time were not necessarily sequential, nor as pronounced, in the mitral subgroups; isolated mitral (1-year OR=1.12, 10-year OR=1.15), CABG + mitral (1-year OR=1.08, 10-year OR=1.06). See Table 2.

Results from both internal validation techniques confirmed the discriminative ability of PROM for fixed survival time points. Not surprisingly, in the bootstrap validation algorithm, the resulting estimated AUROC was very similar to the AUROC calculated on the original sample. This was true in all procedures and for the entire sample overall. Additionally, similar to the odds ratios, the Spearman correlation coefficients between predicted survival and actual survival (point biserial) increased as the survival time point increased (with the exception of the CABG + mitral group, where it decreased). See Table 4.

Results from the split-sample validation are consistent with the original sample and the bootstrap results with respect to the AUROC and the Spearman coefficients. See Table 4. Additionally, in the split-sample approach, calibration curves were fit to each model to assess observed and predicted rate agreement. Calibration for each fitted model was moderate to good for the entire cohort and generally moderate for procedure subgroups. Also, calibration tended to improve at later fixed time points ("worst" at 1-year, "best" at 10 years). Calibration curves for the entire sample at each post-operative time point are presented in Figure 3. However, the inclusion of a squared PROM term into the model markedly improved calibration when applied to the holdout sample while not affecting discrimination at all. This suggests a curvilinear association between PROM and the log odds of survival at each endpoint, probably

attributable to the fact that PROM was originally built to address early peri-operative risk, which is known to be higher due to the "trauma" of the procedure. See Figure 4. The estimated model parameters for each combination of procedure (including all procedures) and time point are included in Table 5, along with equations for predicting survival that utilize the squared PROM term.

PROM was further evaluated as a predictor of long-term survival in Cox proportional hazards models where it was determined to be highly associated with survival (HR=1.066, 95% CI 1.062-1.070, $p<0.001$). This relationship persisted even among 30-day survivors (HR=1.065, 95% CI 1.060-1.070, $p<0.001$). Median survival curves across all levels of PROM generally revealed higher median survival for isolated CABG and mitral procedures and markedly lower long-term survival for AVR procedures. Figure 5 allows the surgeon to predict the median survival of any patient undergoing one of the five procedures for which PROM models have been developed.

Discussion

For many legitimate reasons, there has been much attention and study focused on early outcomes after cardiac surgery. Risk-adjusted comparisons of early outcomes between groups of patients undergoing specific cardiac procedures have been a fundamental part of clinical research, comparative effectiveness studies of alternative techniques and therapies, quality improvement initiatives and institutional/local/regional and national benchmarking in cardiac surgery. Moreover, the ability to predict with reasonable accuracy the short-term risk of

mortality after cardiac procedures is essential to the process of informed consent, which is a foundation of the surgeon-patient relationship.

However, as the US population ages, decisions regarding acute surgical intervention depend increasingly on our imperfect ability to predict long-term survival after cardiac surgery. This applies to the individual patient and to society as a whole, as difficult choices are made both individually and on a societal level. To our knowledge, there has been no user-friendly, effective tool to predict long-term survival for specific patients after specific cardiac procedures.

We sought to test the hypothesis that the STS Predicted Risk of Mortality algorithm could not only predict the likelihood of mortality within 30 postoperative days, as it was designed and calibrated to do, but could also predict long-term survival, for which it has never been intended or utilized. Not surprisingly, the STS PROM was a powerful predictor of 30-day survival after all types of procedures for which it is calibrated. However, the extraordinary power of the STS 30-day PROM to predict survival over 14 years of postoperative follow-up was an unexpected finding. Indeed, PROM performed very nearly as well as a predictor of long-term survival as it did predicting 30-day mortality. While intuitively it is clear that comorbid conditions influencing early mortality after cardiac surgery will also impact long-term survival, it is not intuitive that the algorithmic weighting of these risk factors, calibrated for 30-day events, should so precisely predict long-term survival. This essentially indicates that the 30 different preoperative patient risk factors incorporated into the STS PROM mathematical model impact long-term survival in a manner almost identical to the manner in which they impact a patient's likelihood of surviving 30 days after a specific procedure.

It must be acknowledged that the demographic and comorbidity variables of patients undergoing cardiac surgery have evolved over the 14 years during which patients were prospectively enrolled into the Emory University institutional STS database from which our analyses were performed. However, this does not limit the relevance of the STS PROM for each patient, irrespective of the year of surgery for any given patient. This is due to the fact that the STS PROM is recalibrated every 6 months. Short and long-term survival for each patient is compared to that patient's own contemporary PROM score. Thus, changes in demographics and care processes over time are accommodated by the use of the contemporaneous PROM score for each patient. Interestingly, the recalibration of the STS database is based upon 30-day outcome data; nonetheless the iterative recalibrations have yielded a mathematical algorithm that predicts long-term survival over 14 years with remarkable consistency.

Moreover, the model predicts long-term survival for patients having each of the five different procedures for which the STS PROM has been calibrated, especially patients undergoing CABG and AVR procedures. It is not intuitively obvious why a group of different mathematical models calibrated to predict patients' ability to survive specific procedures should perform so uniformly well in also predicting long-term survival.

Limitations:

This is a single-center, retrospecctive study, potentially serious limitations because our patient population and their average risk profile might not represent those of other centers. Although over 24,000 individual patients are included in these analyses, the generalizability of these findings to the entire nation is not conclusively demonstrated. While the Society of Thoracic

Surgeons has developed PROM algorithms for these different cardiac procedures, there are many combinations and permutations of complex cardiac procedures for which there exists no such predictive model. Analytically, these findings were not *externally* validated; internal validation, while important, is almost assured when using a model with just a single predictor variable. Also, the cause of death in this study is unknown and non-cardiac deaths were assumed to be equally distributed among the patient subgroups.

Conclusions: The STS Predicted Risk of Mortality algorithm--developed to predict mortality within 30-days of cardiac surgery-- accurately predicts mortality both at 30-days and during 14 years of follow-up with almost equally strong discriminatory power for most procedure subgroups. Thus, these mathematical models, based upon the same preoperative risk variables routinely collected for hundreds of thousands of US patients annually, can be used both to estimate the likelihood of long-term survival for specific patients and to adjust survival estimates in comparative effectiveness studies after specific cardiac procedures. This may have profound implications for the informed consent process, comparative effectiveness studies and healthcare policy.

References:

1. Shroyer ALW, Plomondon ME, Grover FL and Edwards FH, for The Society of Thoracic Surgeons' National Database Committee. The 1996 Coronary Artery Bypass Risk Model: The Society of Thoracic Surgeons Adult Cardiac National Database. Ann Thorac Surg 1999; 67: 1205– 8.
2. Parsonnet V, Dean D and Berstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. Circulation 1989;79(2):3–12.
3. Nashef SAM, Roques F, Michel P et al. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg* 1999;**16**(1):9–13.
4. Heikkinen J, Biancari F, Satta J, Salmela E, Mosorin M, Juvonen T and Lepojarvi M. Predicting immediate and late outcome after surgery for mitral valve regurgitation with EuroSCORE. J Heart Valve Disease 2007;16:116-121.
5. Kobayashi KJ, Williams JA, Nwakanma LU, Weiss ES, Gott VL, Baumgartner WA and Conte JV. EuroSCORE predicts short- and mid-term mortality in combined aortic valve replacement and coronary artery bypass patients. J Card Surg 2009;24:637-643.
6. Schisterman EF, Whitcomb BW. Use of the social security administration death master file for ascertainment of mortality status. Popul Health Metr 2004;2:2.

7. Allison, Paul D. Logistic Regression using the SAS system: Theory and Application. The SAS Institute 1999, pgs 55-56
8. Collett, David. Modelling Binary Data, 2nd edition. Chapman and Hall/CRC, 2000, pg. 88
9. Harrell, Frank E. Regression Modeling Strategies. Springer 2001, pgs 90-97
10. Ibid Pg 78
11. Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text, 2nd edition. Springer 2005, pgs 151-153
12. ibid, pages 103-107

Table 1: Preoperative PROM Characteristics by Procedures

Variables	ISOLATED CABG N=20014	ISOLATED AVR N=1781	ISOLATED MV N=945	CABG + AVR N=1059	CABG + MV N=423
PROM Percentage (SD)	2.24 (3.32)	4.22 (4.90)	4.84 (6.51)	5.89 (5.49)	9.60 (8.26)
Patient Age (SD)	62.9 (10.9)	63.5 (14.6)	58.2 (14.3)	70.3 (10.1)	66.2 (10.5)
Patient Age > 66 (%)	7900 (39.5)	830 (46.6)	306 (32.4)	714 (67.4)	216 (51.1)
Aortic Stenosis (%)*	249 (1.2)	877 (49.2)	17 (1.80)	588 (55.5)	9 (2.1)
Black Race (%)*	2894 (14.7)	278 (15.8)	221 (23.6)	99 (9.5)	72 (17.2)
Body Surface Area (SD)*	2.03 (0.26)	1.98 (0.28)	1.89 (0.26)	1.99 (0.25)	1.95 (0.28)
Congestive Heart Failure (%)	3617 (18.1)	887 (49.8)	586 (62.0)	505 (47.7)	275 (65.0)
Chronic Lung Disease (%)	1456 (7.3)	68 (3.8)	57 (6.0)	53 (5.0)	34 (8.0)
Cerebrovascular Accident (%)	1707 (8.5)	141 (7.9)	107 (11.3)	106 (10.0)	57 (13.5))
Diabetes (%)	7249 (36.2)	404 (22.7)	152 (16.1)	331 (31.3)	124 (29.3)
Ejection Fraction < 50% (%)*	6237 (35.4)	385 (26.7)	164 (21.0)	290 (32.2)	175 (47.3)
Elective Status (%)	15071 (75.3)	1398 (78.5)	716 (75.8)	824 (77.8)	279 (66.0)
First Reoperation (%)	1140 (5.7)	259 (14.5)	203 (21.5)	140 (13.2)	57 (13.5)
Hispanic Race (%)*	142 (0.7)	22 (1.3)	16 (1.7)	13 (1.3)	4 (1.0)
Dyslipidemia (%)	8454 (42.2)	615 (34.5)	216 (22.9)	496 (46.8)	206 (48.7)
Hypertension (%)	15485 (77.4)	1239 (69.6)	537 (56.8)	830 (78.4)	331 (78.3)
Preop Intra-aortic Balloon Pump (%)	678 (3.4)	4 (0.2)	13 (1.4)	13 (1.2)	20 (4.7)
Immunosuppressive RX (%)	548 (2.7)	88 (4.9)	48 (5.1)	45 (4.3)	18 (4.3)
Left Main >= 50 Percent (%)	4723 (23.6)	25 (1.4)	8 (0.9)	144 (13.6)	60 (1.2)
Male (%)	14355 (71.7)	1075 (60.4)	386 (40.9)	765 (72.2)	251 (59.3)
Mitral insufficiency (%)*	4716 (50.7)	847 (75.0)	799 (95.8)	514 (75.4)	376 (97.4)
Multiple Reoperations (%)	95 (0.5)	35 (2.0)	51 (5.4)	9 (0.8)	5 (1.2)
New York Heart Assoc Class IV (%)*	2619 (28.0)	218 (18.8)	133 (22.2)	142 (20.9)	109 (39.9)
Other Race (%)*	428 (2.2)	17 (1.0)	28 (3.0)	13 (1.3)	11 (2.6)
Myocardial Infarction (%)	10282 (51.4)	245 (13.8)	108 (11.4)	305 (28.8)	218 (51.5)
PTCA < 6hrs (%)*	64 (0.3)	0 (0.0)	0 (0.0)	2 (0.2)	0 (0.0)
Peripheral Vascular Disease (%)	1945 (9.7)	119 (6.7)	36 (3.8)	128 (12.1)	55 (13.0)
Rena Failure with Dialysis (%)	370 (1.9)	76 (4.3)	46 (4.9)	36 (3.4)	20 (4.7)
Cardiogenic Shock (%)	263 (1.3)	12 (0.7)	18 (1.9)	9 (0.9)	18 (4.3)
Smoker (%)	5196 (26.0)	290 (16.3)	167 (17.7)	164 (15.5)	110 (26.0)
Triple Vessel Disease (%)*	12379 (66.8)	105 (7.0)	50 (6.6)	363 (37.7)	195 (50.6)

*Denotes that the variable contains some missing data

Table 2: Survival and Predictive Ability by Procedure for Fixed Post-Operative Points in Time

Fixed Survival Endpoint	Procedure	#Deaths/total (%)*	Odds Ratio For Death (95% CI) for each Unit Increase in PROM Percentage	Area Under the ROC Curve	Spearman Rank Correlation
30 Days	**Isolated CABG**	423 /20014 (2.1%)	1.12 (1.11-1.14)	0.769	0.16
1-Year	**Isolated CABG**	1037/19053 (5.4%)	1.16 (1.15-1.18)	0.776	0.22
3-Year	**Isolated CABG**	1696/16384 (10.4%)	1.21 (1.19-1.23)	0.757	0.27
5-Year	**Isolated CABG**	2244/13777 (16.3%)	1.29 (1.27-1.31)	0.755	0.33
10-Year	**Isolated CABG**	2368/6781 (34.9%)	1.89 (1.80-1.98)	0.760	0.43
30 Days	**Isolated AVR**	76/1781 (4.3%)	1.11 (1.08-1.14)	0.763	0.20
1-Year	**Isolated AVR**	154/1582 (9.7%)	1.15 (1.12-1.18)	0.786	0.29
3-Year	**Isolated AVR**	205/1273 (16.1%)	1.23 (1.18-1.28)	0.777	0.35
5-Year	**Isolated AVR**	231/1013 (22.8%)	1.31 (1.24-1.38)	0.790	0.42
10-Year	**Isolated AVR**	175/465 (37.6%)	1.44 (1.30-1.60)	0.777	0.46
30 Days	**Isolated Mitral**	49/945 (5.2%)	1.12 (1.09-1.15)	0.816	0.24
1-Year	**Isolated Mitral**	90/852 (10.6%)	1.12 (1.09-1.16)	0.770	0.29
3-Year	**Isolated Mitral**	124/682 (18.2%)	1.13 (1.09-1.17)	0.741	0.32
5-Year	**Isolated Mitral**	148/572 (25.9%)	1.15 (1.10-1.20)	0.748	0.38
10-Year	**Isolated Mitral**	124/305 (40.7%)	1.15 (1.07-1.24)	0.774	0.47
30 Days	**CABG + AVR**	83/1059 (7.8%)	1.12 (1.09-1.16)	0.749	0.21
1-Year	**CABG + AVR**	158/983 (16.1%)	1.17 (1.13-1.21)	0.728	0.29
3-Year	**CABG + AVR**	193/788 (24.5%)	1.20 (1.15-1.25)	0.728	0.34
5-Year	**CABG + AVR**	204/595 (34.3%)	1.30 (1.22-1.38)	0.741	0.40
10-Year	**CABG + AVR**	173/282 (61.4%)	1.43 (1.27-1.62)	0.747	0.42
30 Days	**CABG + Mitral**	43/423 (10.2%)	1.08 (1.05-1.12)	0.717	0.25
1-Year	**CABG + Mitral**	77/381 (20.2%)	1.10 (1.07-1.14)	0.735	0.33
3-Year	**CABG + Mitral**	95/281 (33.8%)	1.09 (1.06-1.13)	0.693	0.32
5-Year	**CABG + Mitral**	97/222 (43.7%)	1.08 (1.04-1.12)	0.659	0.27
10-Year	**CABG + Mitral**	70/111 (63.1%)	1.06 (1.00-1.12)	0.626	0.21
30 Days	**All Procedures**	674/24222 (2.8%)	1.13 (1.12-1.14)	0.794	0.18
1-Year	**All Procedures**	1516/22851 (6.6%)	1.16 (1.15-1.17)	0.789	0.25
3-Year	**All Procedures**	2313/19408 (11.9%)	1.20 (1.19-1.21)	0.767	0.30
5-Year	**All Procedures**	2924/16179 (18.1%)	1.26 (1.24-1.28)	0.763	0.35
10-Year	**All Procedures**	2910/7944 (36.6%)	1.54 (1.49-1.59)	0.762	0.44

*For each endpoint, only those patients with surgery dates early enough to observe the survival endpoints were studied.

Table 3: Survival Estimates by Procedure

Procedure	30-Day Survival	1-Year Survival	3-Year Survival	5-Year Survival	10-Year Survival
Isolated CABG	97.9%	94.6%	89.7%	83.7%	64.8%
Isolated AVR	95.7%	90.5%	83.4%	77.0%	60.3%
Isolated Mitral	94.8%	89.5%	82.6%	74.9%	57.9%
CABG + AVR	92.2%	84.1%	76.5%	67.1%	40.8%
CABG + Mitral	89.8%	80.4%	69.7%	61.7%	39.7%
All Procedures	97.2%	93.4%	88.1%	81.9%	63.0%

Table 4: Model Performance Statistics by Validation Approach and Procedure

		Bootstrap Sample Validation		Split Sample Validation	
Fixed Survival Endpoint	**Procedure**	**AUROC in Bootstrap Sample (95% CI)**	**Spearman Rank Correlation (95% CI)**	**AUROC in Split Sample**	**Spearman Rank Correlation**
1-Year	**Isolated CABG**	0.777 (0.765-0.788)	0.218 (0.207-0.227)	0.779	0.218
3-Year	**Isolated CABG**	0.757 (0.749-0.766)	0.272 (0.261-0.281)	0.761	0.275
5-Year	**Isolated CABG**	0.755 (0.746-0.751)	0.326 (0.314-0.335)	0.754	0.324
10-Year	**Isolated CABG**	0.760 (0.748-0.771)	0.430 (0.408-0.446)	0.767	0.443
1-Year	**Isolated AVR**	0.788 (0.756-0.810)	0.296 (0.260-0.328)	0.817	0.308
3-Year	**Isolated AVR**	0.776 (0.742-0.800)	0.352 (0.307-0.387)	0.827	0.405
5-Year	**Isolated AVR**	0.789 (0.766-0.813)	0.420 (0.379-0.454)	0.813	0.441
10-Year	**Isolated AVR**	0.775 (0.744-0.808)	0.463 (0.409-0.519)	0.787	0.486
1-Year	**Isolated Mitral**	0.773 (0.726-0.810)	0.290 (0.225-0.332)	0.796	0.324
3-Year	**Isolated Mitral**	0.740 (0.691-0.780)	0.321 (0.255-0.374)	0.772	0.363
5-Year	**Isolated Mitral**	0.752 (0.708-0.788)	0.380 (0.321-0.431)	0.778	0.428
10-Year	**Isolated Mitral**	0.773 (0.731-0.813)	0.466 (0.395-0.534)	0.765	0.449
1-Year	**CABG + AVR**	0.726 (0.685-0.764)	0.289 (0.235-0.338)	0.735	0.304
3-Year	**CABG + AVR**	0.727 (0.694-0.759)	0.339 (0.291-0.388)	0.724	0.341
5-Year	**CABG + AVR**	0.742 (0.713-0.777)	0.394 (0.348-0.460)	0.739	0.403
10-Year	**CABG + AVR**	0.749 (0.695-0.791)	0.418 (0.324-0.495)	0.467	0.772
1-Year	**CABG + Mitral**	0.734 (0.690-0.784)	0.325 (0.258-0.394)	0.794	0.410
3-Year	**CABG + Mitral**	0.686 (0.643-0.741)	0.302 (0.234-0.399)	0.701	0.333
5-Year	**CABG + Mitral**	0.658 (0.590-0.722)	0.272 (0.154-0.383)	0.666	0.284
10-Year	**CABG + Mitral**	0.625 (0.500-0.724)	0.205 (0.030-0.369)	0.604	0.178
1-Year	**All Procedures**	0.788 (0.781-0.797)	0.248 (0.240-0.258)	0.793	0.249
3-Year	**All Procedures**	0.767 (0.758-0.775)	0.300 (0.290-0.309)	0.781	0.313
5-Year	**All Procedures**	0.763 (0.756-0.769)	0.350 (0.340-0.359)	0.773	0.365
10-Year	**All Procedures**	0.763 (0.752-0.771)	0.439 (0.421-0.452)	0.761	0.436

Table 5: Model Parameter Estimates to Calculate the Probability of Mortality at Specific Fixed Post-Operative Time Points.

Fixed Survival Endpoint	Procedure	Logit = Intercept + B1(PROM)	Logit = Intercept + B1(PROM) + B2(PROM^2)
1-Year	**Isolated CABG**	Logit = -3.33105 + 0.15141(PROM)	Logit=-3.6332 + 0.2859 (PROM) + -0.00486(PROM^2)
3-Year	**Isolated CABG**	Logit = -2.68945 + 0.18870(PROM)	Logit=-3.0096 + 0.3511 (PROM) + -0.00752(PROM^2)
5-Year	**Isolated CABG**	Logit = -2.26289 + 0.25177(PROM)	Logit=-2.5664 + 0.4259 (PROM) + -0.01050(PROM^2)
10-Year	**Isolated CABG**	Logit = -1.75771 + 0.63431(PROM)	Logit=-1.9444 + 0.8108 (PROM) + -0.02550(PROM^2)
1-Year	**Isolated AVR**	Logit = -2.99381 + 0.14254(PROM)	Logit=-3.3769 + 0.2561 (PROM) + -0.00420(PROM^2)
3-Year	**Isolated AVR**	Logit = -2.59587 + 0.20303(PROM)	Logit=-2.8366 + 0.2921 (PROM) + -0.00419(PROM^2)
5-Year	**Isolated AVR**	Logit = -2.34928 + 0.28135(PROM)	Logit=-2.5440 + 0.3642 (PROM) + -0.00557(PROM^2)
10-Year	**Isolated AVR**	Logit = -1.70389 + 0.36840(PROM)	Logit=-1.9449 + 0.5094 (PROM) + -0.01170(PROM^2)
1-Year	**Isolated Mitral**	Logit = -2.87498 + 0.12331(PROM)	Logit=-3.1514 + 0.1893 (PROM) + -0.00215(PROM^2)
3-Year	**Isolated Mitral**	Logit = -2.23878 + 0.12946(PROM)	Logit=-2.4803 + 0.1994 (PROM) + -0.00249(PROM^2)
5-Year	**Isolated Mitral**	Logit = -1.78424 + 0.14555(PROM)	Logit=-1.9871 + 0.2118 (PROM) + -0.00267(PROM^2)
10-Year	**Isolated Mitral**	Logit = -0.96160 + 0.13851(PROM)	Logit=-1.2849 + 0.2723 (PROM) + -0.00430(PROM^2)
1-Year	**CABG + AVR**	Logit = -2.70029 + 0.15334(PROM)	Logit=-2.9136 + 0.2060 (PROM) + -0.00201(PROM^2)
3-Year	**CABG + AVR**	Logit = -2.21615 + 0.17856(PROM)	Logit=-2.5464 + 0.2665 (PROM) + -0.00387(PROM^2)
5-Year	**CABG + AVR**	Logit = -2.08073 + 0.26192(PROM)	Logit=-2.4811 + 0.4071 (PROM) + -0.00878(PROM^2)
10-Year	**CABG + AVR**	Logit = -1.16212 + 0.36136(PROM)	Logit=-1.4800 + 0.4994 (PROM) + -0.01200(PROM^2)
1-Year	**CABG + Mitral**	Logit = -2.49867 + 0.09743(PROM)	Logit=-2.6963 + 0.1331 (PROM) + -0.00095(PROM^2)
3-Year	**CABG + Mitral**	Logit = -1.63047 + 0.08768(PROM)	Logit=-1.8023 + 0.1131 (PROM) + -0.00066(PROM^2)
5-Year	**CABG + Mitral**	Logit = -1.04890 + 0.07397(PROM)	Logit=-1.2532 + 0.1163 (PROM) + -0.00125(PROM^2)
10-Year	**CABG + Mitral**	Logit = -0.00687 + 0.05479(PROM)	Logit=-0.0817 + 0.0680 (PROM) + -0.00029(PROM^2)
1-Year	**All Procedures**	Logit = -3.22989 + 0.15008(PROM)	Logit=-3.5349 + 0.2646 (PROM) + -0.00404(PROM^2)
3-Year	**All Procedures**	Logit = -2.61075 + 0.18196(PROM)	Logit=-2.9016 + 0.3062 (PROM) + -0.00541(PROM^2)
5-Year	**All Procedures**	Logit = -2.19517 + 0.23033(PROM)	Logit=-2.4380 + 0.3551 (PROM) + -0.00677(PROM^2)
10-Year	**All Procedures**	Logit = -1.46317 + 0.43106(PROM)	Logit=-1.6364 + 0.5580 (PROM) + -0.01120(PROM^2)

$$\text{Probability of Mortality} = 1 / (1 + e^{-(\text{logit})}) \text{ where } e=2.71828\ldots.$$

PROM is represented as a percentage between 0 and 100.

Figure 1: Long-Term Kaplan-Meier Survival Estimates by Decile of PROM

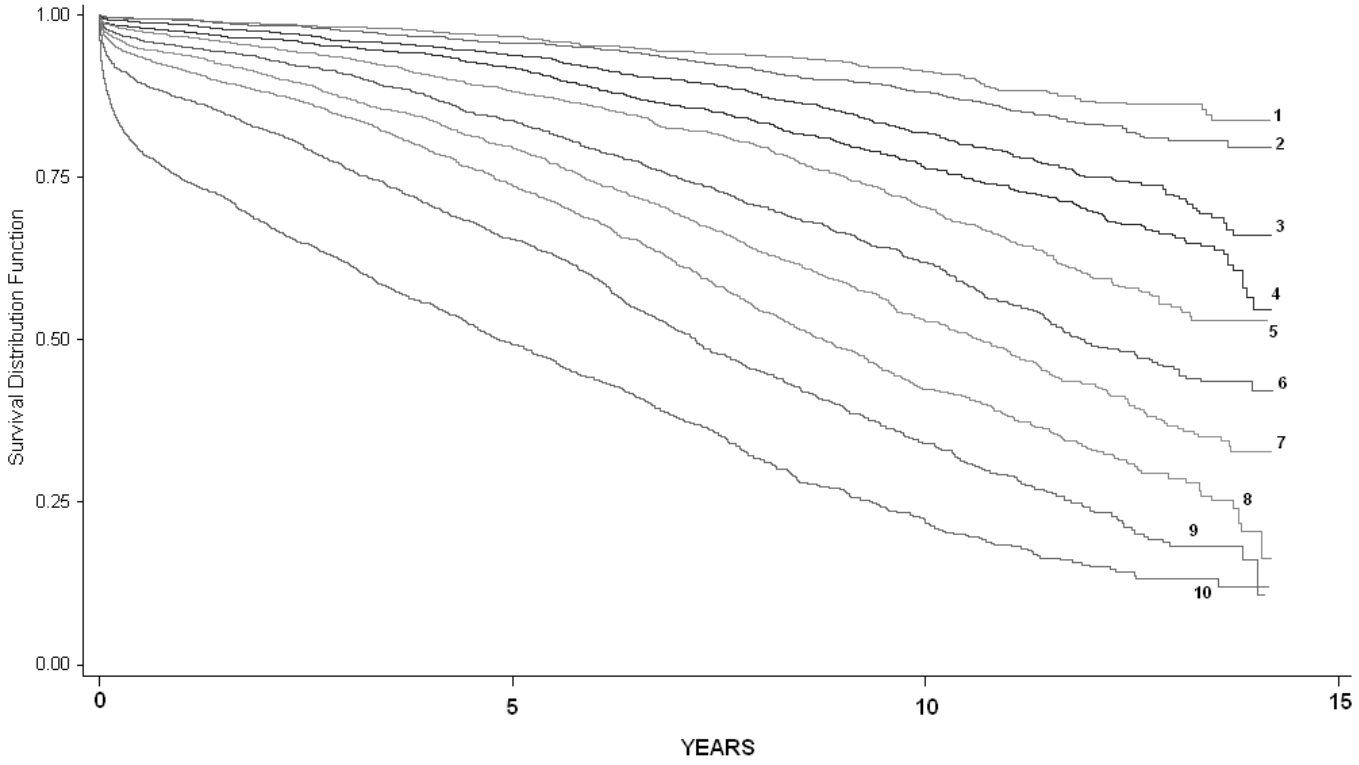

Figure 2: Long-Term Kaplan-Meier Survival Estimates by Decile of PROM Among 30-Day Survivors

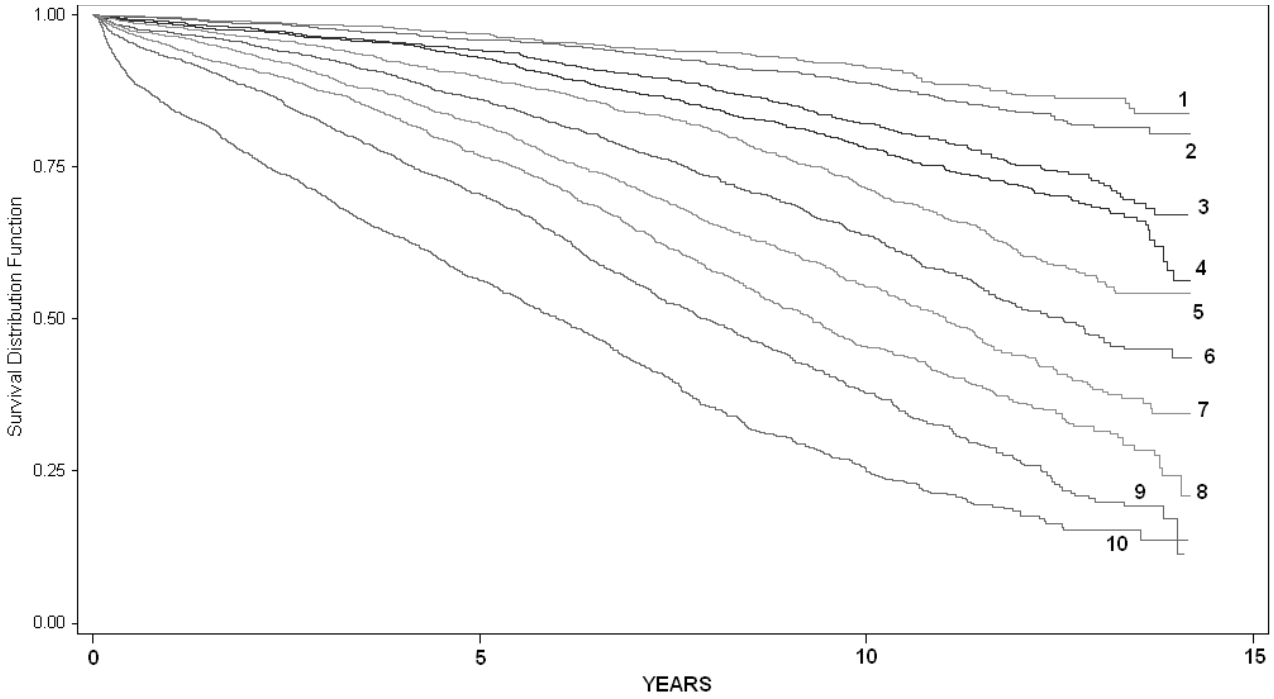

Figure 3: Calibration Curves for All Procedures combined (no squared PROM term)

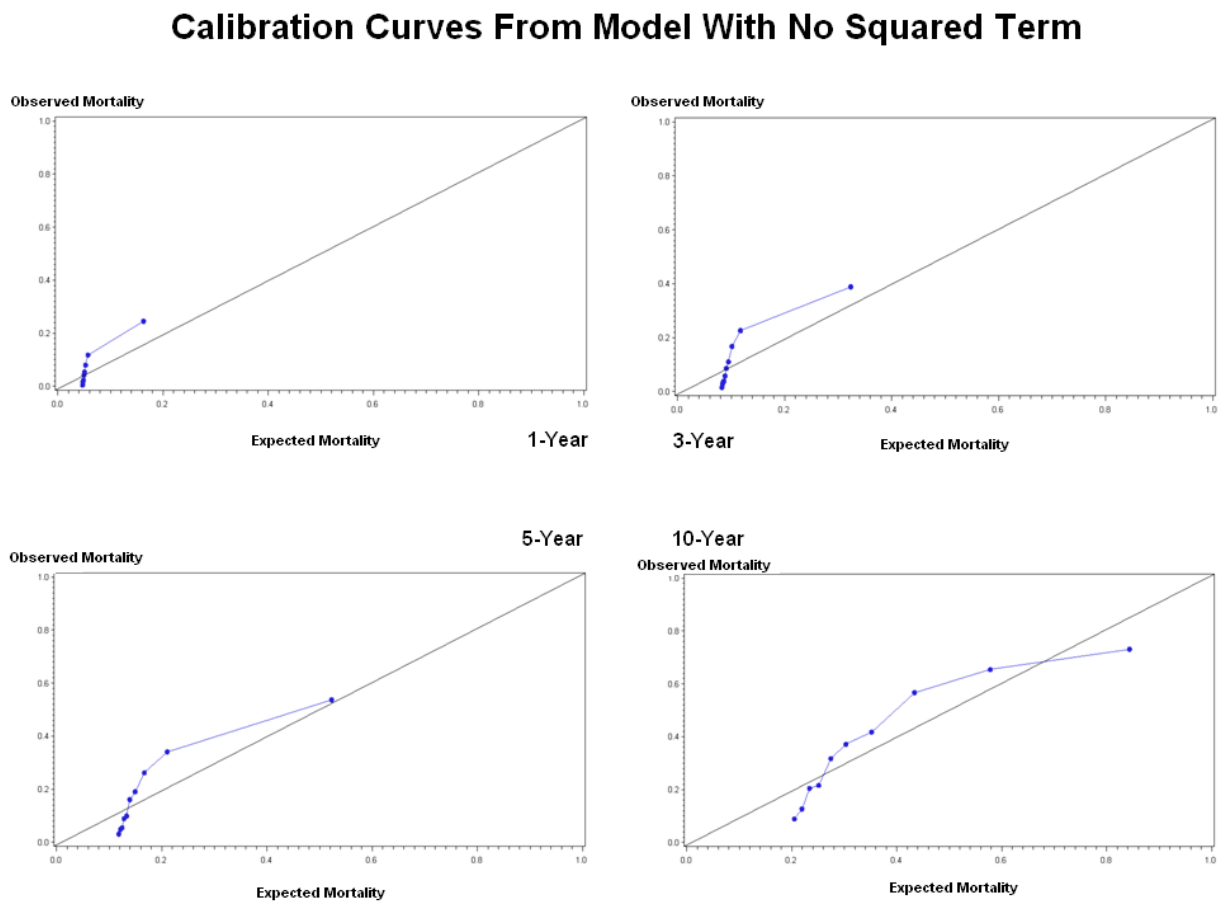

Figure 4: Calibration Curves for All Procedures combined (with squared PROM term). Note that these curves more closely follow the line of identity.

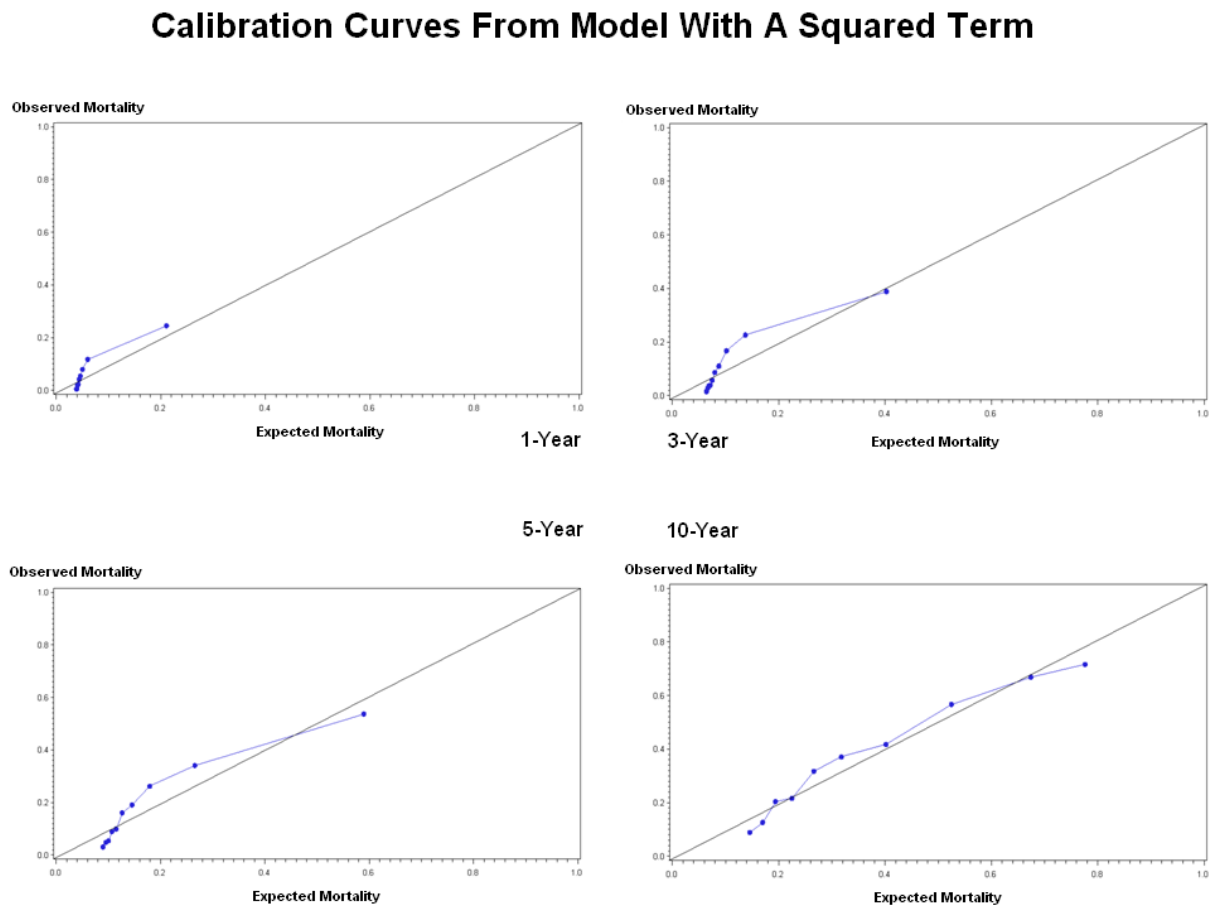

Figure 5: Median Estimated Survival by PROM Percentages