Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biomedical Informatics 38 (2005) 367–375
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

Discrimination and calibration of mortality risk prediction models in interventional cardiology

M.E. Matheny [a,b,c,*], L. Ohno-Machado [a,b], F.S. Resnic [a,d]

[a] Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[b] Health Sciences and Technology Division of Harvard/MIT, Boston and Cambridge, MA, USA
[c] Brigham Internal Medicine Associates, Brigham and Women's Hospital, Boston, MA, USA
[d] Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA, USA

Received 3 January 2005
Available online 26 March 2005

Abstract

Objectives. Using a local percutaneous coronary intervention (PCI) data repository, we sought to compare the performance of a number of local and well-known mortality models with respect to discrimination and calibration.

Background. Accurate risk prediction is important for a number of reasons including physician decision support, quality of care assessment, and patient education. Current evidence on the value of applying PCI risk models to individual cases drawn from a different population is controversial.

Methods. Data were collected from January 01, 2002 to September 30, 2004 on 5216 consecutive percutaneous coronary interventions at Brigham and Women's Hospital (Boston, MA). Logistic regression was used to create a local risk model for in-hospital mortality in these procedures, and a number of statistical methods were used to compare the discrimination and calibration of this new and old local risk models, as well as the Northern New England Cooperative Group, New York State (1992 and 1997), University of Michigan consortium, American College of Cardiology-National Cardiovascular Data Registry, and The Cleveland Clinic Foundation risk prediction models. Areas under the ROC (AUC) curves were used to evaluate discrimination, and the Hosmer–Lemeshow (HL) goodness-of-fit test and calibration curves assessed applicability of the models to individual cases.

Results. Multivariate risk factors included in the newly constructed local model were: age, prior intervention, diabetes, unstable angina, salvage versus elective procedure, cardiogenic shock, acute myocardial infarction (AMI), and left anterior descending artery intervention. The area under the ROC curve (AUC) was 0.929 (SE = 0.017), and the p value for the Hosmer–Lemeshow (HL) goodness-of-fit was 0.473. This indicates good discrimination and calibration. Bootstrap re-sampling indicated AUC stability. Evaluation of the external models showed an AUC range from 0.82 to 0.90 indicating good discrimination across all models, but poor calibration (HL p value $\leqslant 0.0001$).

Conclusions. Validation of AUC values across all models suggests that certain risk factors have remained important over the last decade. However, the lack of calibration suggests that small changes in patient populations and data collection methods quickly reduce the accuracy of patient level estimations over time. Possible solutions to this problem involve either recalibration of models using local data or development of new local models.

Keywords: Percutaneous coronary angioplasty; Coronary angiography; Heart catheterization; Stents; Myocardial infarction; Registries; Risk adjustment; Statistical models; Hospital mortality; Risk assessment; Multivariate analysis; Survival rate; Predictive models; Model discrimination; Model calibration

* Corresponding author. Fax: +1 617 739 3672.
E-mail address: mmatheny@dsg.harvard.edu (M.E. Matheny).

1532-0464/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2005.02.007

1. Background

In the last decade, significant emphasis has been placed on the development of statistical models to help predict risk in various patient populations. In addition to providing the basis for quality scorecards [1,2], these risk profiles can be helpful on the procedural level to both patients and physicians. Numerous studies have shown that subjective prediction of risk tends to be poor at very low and very high probabilities [3,4]. The use of various statistical methods can provide an objective estimation of outcome risk.

There has been conflicting literature on whether or not these models can be used outside of their development population. Initial validation is usually based on patients from a given geography and time frame. These evaluations are only directly applicable in that respect, and concerns have been raised about the applicability of a model when patient demographics change with geography, clinical practice changes with time, and disease prevalence changes with both. Some of this concern stems from prior analyses showing deleterious effects on accuracy by changes in geography and time [5]. A study comparing the demographics of percutaneous coronary intervention (PCI) patients in two registries collected 12 years apart found significant differences in age, lesion severity, thrombolytic use, stent use, and death that highlight how much the characteristics of a population can change with a decade of medical advances [6].

Continuous evaluation of model performance is important to ascertain that classification performance does not degrade with time. Some models are re-developed periodically to adjust for temporal trends [7]. Also important is validation of a model on geographically or temporally distant populations [8]. Constructing a model using a large numbers of patient encounters across a wide variety of geographic areas increases the probability that the model will be suited for different populations, but the only way to determine the model's applicability is to verify the performance empirically in representative sample.

In the field of cardiology, one of the most widely studied areas of risk stratification has been coronary angiography. This article seeks to build on prior work on the applicability of risk models in different geographies and over time. Several prior studies of PCI risk models have shown good external validation with respect to both calibration and discrimination [9–13]. Others have shown a loss in either discrimination, calibration [14], or both [15]. In the present study, we consider the hypothesis that models exhibit differences in discrimination and calibration over space and time.

2. Methods

2.1. Data collection

Brigham and Women's Hospital (BWH), Boston Massachusetts has maintained a detailed database of all cases of PCI since 1997. The dataset is based on the American College of Cardiology National Data Repository dataset [16], with a variety of additional, detailed, data elements. The registry is part of the quality assessment and quality improvement program of Brigham and Women's Hospital, and was approved by the hospital Institutional Review Board. All catheterization laboratory procedures performed are included in the database, and real-time data acquisition is accomplished through a dedicated team of trained nurses, physicians, and technologists. A total of 5216 PCI procedures were recorded between January 01, 2002 and September 30, 2004 on all patients who underwent PCI at BWH. This dataset serves as the source for the evaluation of each model in this study.

2.2. Model evaluation

Evaluation of all models was done with χ^2 and maximum log likelihood methods. Discrimination was assessed with the area under the receiver operating characteristic curve (AUC) [17,18]. Calibration was evaluated with Hosmer–Lemeshow goodness-of-fit χ^2-estimates using deciles [19]. 95% confidence intervals for these parameters were computed with the non-parametric bootstrapping method of STATA [20,21]. These CIs were reported using the percentile method, or bias corrected method if the estimation bias was greater than 25% of the standard error [22].

2.3. External validation of risk models

Six external and one local previously described multivariate post-PCI in-hospital mortality risk models were evaluated using the BWH dataset: the Northern New England Cooperative Group (NNE 1999) [23], the New York State (NY 1992 & NY 1997) [24,25], University of Michigan Consortium (MI 2001) [26], the American College of Cardiology-National Cardiovascular Data Registry (ACC 2002) [27], the Cleveland Clinic Foundation Multi-Center (CC 1997) [28], and the Brigham and Women's Hospital (BWH 2001) [29] models. Table 1 shows the independent risk variables and corresponding odds ratios for all multivariate models. Pair-wise comparison of the area under the ROC curve for each model was performed by Analyse-It 1.71 (Leeds, England, UK) (see Table 2).

2.4. Local model development

To test the hypothesis that time and space degrade the accuracy of a risk model, a new local model was

Table 1
Overview of the odds ratios for the variables in each model evaluated

Published	NY [23] 1992	NY [24] 1997	CC [27] 1997	NNE [22] 1999	MI [25] 2001	BWH [28] 2001	ACC [26] 2002
Shock	67.6	18.31	12.7		11.5	8.33	8.49
Unstable	24.9	4.154					
Salvage							13.38
Emergent				7.71			5.75
Urgent				2.19			1.78
Pre-PCI IABP		2.394		3.91			1.68
Male			0.55				
Female	2.45	1.308			1.82		
Age(log)⟨C⟩			24.87				
Age ⟨C⟩		1.062					
Age 50–59				NS			2.61
Age 60–69				NS			3.75
Age 70–79				3.32	2.24		6.44
Age > 75						2.53	
Age > 79				3.72	2.65		11.25
Diabetes		1.410					1.41
EF	0.93						
LVEF < 50					1.66		
LVEF 50–59				2.53			
LVEF 40–49				3.32			NS
LVEF < 40				5.16			
LVEF 20–40		1.490					
LVEF 30–39							NS
LVEF 20–29							2.04
LVEF < 20		3.681					
LVEF 10–19							3.43
LVEF < 10							3.93
Hx CHF		2.381		3.01			
CHF NYHA3/4						8.14	
AMI < 6 h		5.220					
AMI 6–24 h		3.672					
AMI < 24 h			4.75		2.8		1.31
AMI 1–7 days		2.101					
Peri-Op MI						1.83	
MI Therapy				1.85			
Hx Arrest					3.65		
SCAI LC-II							1.64
SCAI LC-III							1.87
SCAI LC-IV							2.11
Left main PCI						6.59	
Lesion type C				1.94			
Tachycardia						2.77	
Unstable angina						1.69	
L. Main lesion							2.04
Prox LAD Les.							1.97
# Vessels Dis.			1.32		1.54		
Renal failure		3.514					3.04
Cr > 2.0				2.33			
Cr > 1.5					5.5		
CRI						2.71	
COPD							1.33
Thrombus					1.67		
Thrombolytics							1.39
Non-Stent Dev.							1.64
PVD		1.775		2.12	1.57		
Prior Angio.		0.594					
>1 Vessels		1.817					
Prior CABG		1.431					
Stent use						0.51	
Lesion complex			1.63				

Definition details can be found in the respective sources.

Table 2
Summary of the training datasets for the models used in this study

Model	Dates		Location	Sample	AUC	HL(p)	Validation type
NNE [22] 1999	1/1/1994	12/31/1996	NH, ME, MA, VT (7)	15331	0.88	0.09	Bootstrap resampling
NY [23] 1992	1/1/1991	6/30/1991	NY	5827	0.884	NA	Subset significance
NY [24] 1997	1/1/1991	12/31/1994	NY	62670	0.892	0.11	Subset significance
MI [25] 2001	10/1/1999	8/30/2000	Detroit, MI	10796	0.90	0.5	Training/test
ACC [26] 2002	1/1/1998	9/30/2000	National	100253	0.89	0.133	Training/test
BWH [28] 2001	1/1/1997	12/31/1999	Boston, MA	2804	0.86	0.11	Training/test
CC [27] 1997	1/11993	12/31/1994	Cleveland, OH (5)	12985	0.846	NA	Bootstrap resampling

Sample, sample size. AUC, area under the receiver operating characteristic. HL(p), Hosmer–Lemeshow p value.

Table 3
Univariate association of factors with in-hospital mortality and registry demographics

Factor	% Pts	% Deaths	OR	95% CI	p
Age					
<50	11.0	0.2	1.00	Ref.	
50–59	21.6	0.4	2.55	0.30–39.9	0.392
60–69	27.8	0.9	5.20	0.68–39.9	0.112
70–79	27.6	1.5	8.91	1.199–66.3	0.033
>79	11.9	4.8	29.3	3.98–215.5	0.001
Gender					
Male	70.7	1.4	1.00	Ref.	
Female	29.3	1.4	1.02	0.61–1.70	0.952
Diabetes	31.7	1.8	1.58	0.99–2.5	0.058
PVD	9.5	2.4	1.97	1.05-3.69	0.034
COPD	10.6	2.0	1.55	0.81–2.97	0.183
Shock	1.7	37.4	82.0	48.1–139.8	<0.001
Unstable angina	4.9	11.6	15.8	9.7–25.7	<0.001
Urgency					
Elective	49.9	0.3	1.00	Ref.	
Urgent	37.9	0.9	2.98	1.3–6.9	0.010
Emergent	11.8	5.7	19.6	9.1–42.6	<0.001
Salvage	0.4	45.4	270.3	91–803.2	<0.001
LVEF					
>39	91.3	1.1	1.00	Ref.	
20–39	7.6	3.6	3.22	1.77–5.84	<0.001
<20	1.1	5.5	5.04	1.53–16.6	0.008
Tachycardia	2.4	13.5	14.5	8–17–25.9	<0.001
Pre-PCI IABP	0.7	19.4	19.3	8.15–45.7	<0.001
AMI 24 h	10.6	5.2	6.1	3.8–9.8	<0.001
Cr > 2.0 mg/dL	5.3	5.0	4.5	2.5–8.3	<0.001
CHF	10.1	4.0	3.9	2.3–6.5	<0.001
Prior PCI	33.8	0.5	0.28	0.14–0.57	<0.001
Prior CABG	1101	1.1	0.76	0.41–1.42	0.385
Lesion risk					
Low	66.3	0.5	1.00	Ref.	
High	33.7	3.0	5.5	3.24–9.4	<0.001
Intervention					
LAD	42.4	1.9	1.87	1.17–3.01	0.010
Disease location					
Proximal LAD	47.2	2.2	3.34	1.95–5.72	<0.001
RCA	52.3	1.7	1.69	1.03–2.76	0.036
Diseased vessels					
0	9.0	0.4	1.00	Ref.	
1	52.6	0.9	2.07	0.49–8.8	0.323
2	25.5	1.9	4.5	1.06–19.1	0.041
3	12.8	3.0	7.3	1.7–31.2	0.008

% Pts, percent of sample population. % Deaths, percent of deaths within the sub-population. OR, odds ratio. 95% CI, 95% confidence interval. p, p value. PVD, peripheral vascular disease. COPD, chronic obstructive pulmonary disease. LVEF, left ventricular ejection fraction. PCI, percutaneous coronary intervention. IABP, intra-aortic balloon pump. AMI, acute myocardial infarction. Cr, creatinine. CHF, congestive heart failure. CABG, coronary artery bypass grafting. LAD, left anterior descending. RCA, right coronary artery.

developed using the same BWH data that was used to evaluate the discrimination and calibration of existing models. Standard univariate methods were used to generate odds ratios (ORs) with 95% confidence intervals (CIs) and p values to select variables that would be included in the new model [30]. Additionally, all available covariates which have been shown to be univariate risk factors in previous studies were included in the analysis (Table 3). Backward stepwise logistic regression was performed using STATA [19]. Variables were first removed using a residual Wald χ^2 p value of 0.1, and then considered for inclusion based on a p value of 0.05. Since there was no independent test set, the evaluation was based on bootstrap resampling with 1000 samples [31].

3. Results

3.1. Local multivariate prediction rule development

After full backward stepwise variable selection, the variables associated with an increased risk included older age, diabetes, unstable angina, salvage procedure, cardiogenic shock, AMI, and any intervention on the left anterior descending artery as shown in Table 4. The AUC was 0.929 revealing excellent discriminatory ability of the new model, and bootstrap re-sampling the data to obtain a 95% CI of 0.90–0.96 with an SE of 0.017, indicating a good ability to discriminate with respect to the outcome of death. The model had an adequate goodness of fit (HL $\chi^2 = 7.61$ with 8 df, $p = 0.473$).

3.2. External validation

The external model performances on the BWH dataset are shown in Table 5. During the study period, there were 71 observed deaths (1.36%). BWH 2004 very closely approximated this with 70.5 deaths, NY 1992, CC 1997, and BWH 2001 over predicted, and the remainder under predicted. The AUC indicates excellent discrimination across all models, with the worst being the New York State 1992 model and the best being the new local model. A summary view of the AUC for all models is shown in Fig. 1. Of the external models, the best AUC was obtained by the ACC 2002 model.

Pair-wise AUC comparisons were performed as well, shown in Table 6, by using the method described by Hanley and McNeil [32]. Overall, the best discrimination was obtained by the new local model, which attained significance with respect to every model but ACC

Table 4
Multivariate analysis of factors significantly associated with in-hospital mortality in the new BWH model

Factor	OR	95% CI	β	p
Prior PCI	0.30	0.12–0.74	−1.20	0.009
Age (years)				
60–69	4.41	1.31–14.84	1.48	0.016
70–79	8.25	2.58–26.34	2.11	<0.001
80+	21.39	6.76–66.97	3.06	<0.001
Diabetes	1.82	1.02–3.26	0.60	0.042
Unstable	5.46	2.82–10.52	1.70	<0.001
Salvage	19.25	5.06–73.24	2.96	<0.001
Shock	14.86	7.39–29.87	2.70	<0.001
AMI Present	1.72	1.37–2.17	0.54	<0.001
Any LAD PCI	1.72	0.97–3.07	0.54	0.066

OR, odds ratio. 95% CI, 95% confidence interval. β, beta coefficient. p, p value. Constant (intercept) = −7.777; Hosmer and Lemeshow goodness-of-fit $\chi^2 = 7.61$; $p = 0.473$; AUC = 0.929. AMI, acute myocardial infarction. LAD, left anterior descending. PCI, percutaneous coronary intervention.

Table 5
Summary of discrimination and calibration performance for each model

Curve	Deaths	AUC	95% CI	HL χ^2	95% CI	HL(p)	95% CI
NY 1992	96.7	0.82	0.76–0.88	31.1	13.9–50.0	<0.001	<0.001–0.003
NY 1997	61.6	0.88	0.81–0.92	32.2	16.4–45.5	<0.001	<0.001–0.004
CC 1997	78.8	0.88	0.82–0.93	27.8	19.6–38.7	<0.001	<0.001–0.013
NNE 1999	56.2	0.89	0.84–0.94	45.9	31.9–67.4	<0.001	<0.001–<0.001
MI 2001	61.8	0.86	0.81–0.90	30.4	16.7–43.1	<0.001	<0.001–0.011
BWH 2001	136.1	0.89	0.84–0.93	39.7	23.2–73.3	<0.001	<0.001–0.001
ACC 2002	49.9	0.90	0.84–0.95	42.0	24.9–63.3	<0.001	<0.001–0.002
BWH 2004	70.5	0.93	0.89–0.96	7.61	1.5–14.2	0.473	0.073–0.992

Deaths, estimated deaths. AUC, area under the receiver operating characteristic curve. 95% CI, 95% confidence interval. HL χ^2, Hosmer–Lemeshow χ^2. HL(p), Hosmer–Lemeshow prob > χ^2 value.

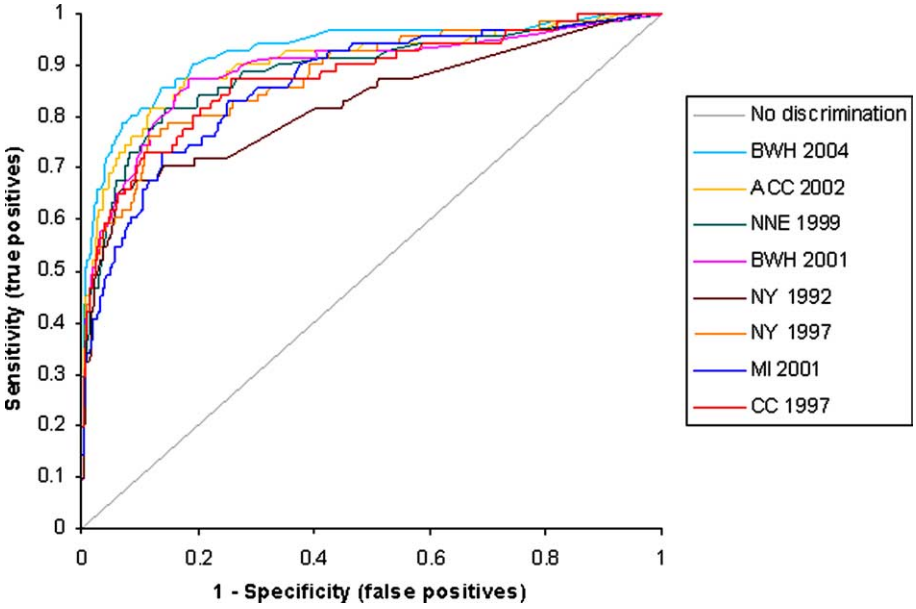

Fig. 1. AUC for all models. The grey line shows no discrimination.

Table 6
Pair-wise discrimination model comparison

	NY 1992		NY 1997		CC 1997		NNE 1999		MI 2001		BWH 2001		ACC 2002	
	Diff	*p*	Diff	*p*	Diff	*p*	Diff	*p*	Diff	*p*	Diff	*p*	Diff	*p*
NY 1992														
NY 1997	0.056	0.007												
CC 1997	0.051	0.101	0.004	0.859										
NNE 1999	0.062	0.013	−0.007	0.712	0.011	0.644								
MI 2001	0.041	0.165	−0.015	0.485	−0.01	0.627	−0.022	0.310						
BWH 2001	0.066	0.019	0.011	0.602	0.015	0.551	0.004	0.849	0.026	0.287				
ACC 2002	0.080	0.002	0.025	0.145	0.03	0.176	0.018	0.254	0.040	0.045	0.014	0.519		
BWH 2004	0.105	0.001	0.049	0.007	0.053	0.011	0.043	0.048	0.064	0.003	0.038	0.050	0.024	0.176

Diff, AUC difference. *p*, *p* value of difference.

2002. The second best performance was by the external model constructed with the largest training set (ACC 2002), followed by the old local model (BWH 2001). Significant differences were noted between NY 1992 and every model but MI 2001, as well as between MI 2001 and ACC 2002. This indicates that the NY 1992 model, and to a lesser extent the MI 2001 model, is the least discriminatory.

The Hosmer–Lemeshow goodness-of-fit test reveals poor calibration ($p < 0.05$) for all the models but the newly developed one. Calibration for all models was further explored by plotting the observed to expected frequency of death for each quintile of every model. Fig. 2B is provided to more clearly show the relationships for the low risk population. As shown in Fig. 2, the NY 1992 model underestimated the risk of death for low scoring patients, and over estimated this risk for high scoring patients. NY 1997 performed fairly well for low risk patients, but overestimated the probability of death for high risk patients. ACC 2002 performed well under low risk conditions, but significantly underestimated the probability of death for high risk conditions. NNE 1999 consistently under predicted deaths, and CC 1997 as well as BWH 2001 consistently overestimated mortality risk. As expected, BWH 2004 performs well, but since this is not an independent test sample, this result should be interpreted with caution.

4. Discussion

Interventional cardiology practice has changed significantly over the last decade. Procedural skill development, pharmacology, and device development have all contributed to the evolution of the field, and patient outcomes have changed over that time period in response to these advances. There as been a substantial reduction in risk of death and major adverse cardiac events (MACE) [33] over the past decade. All of these factors create a moving target for any risk stratification model.

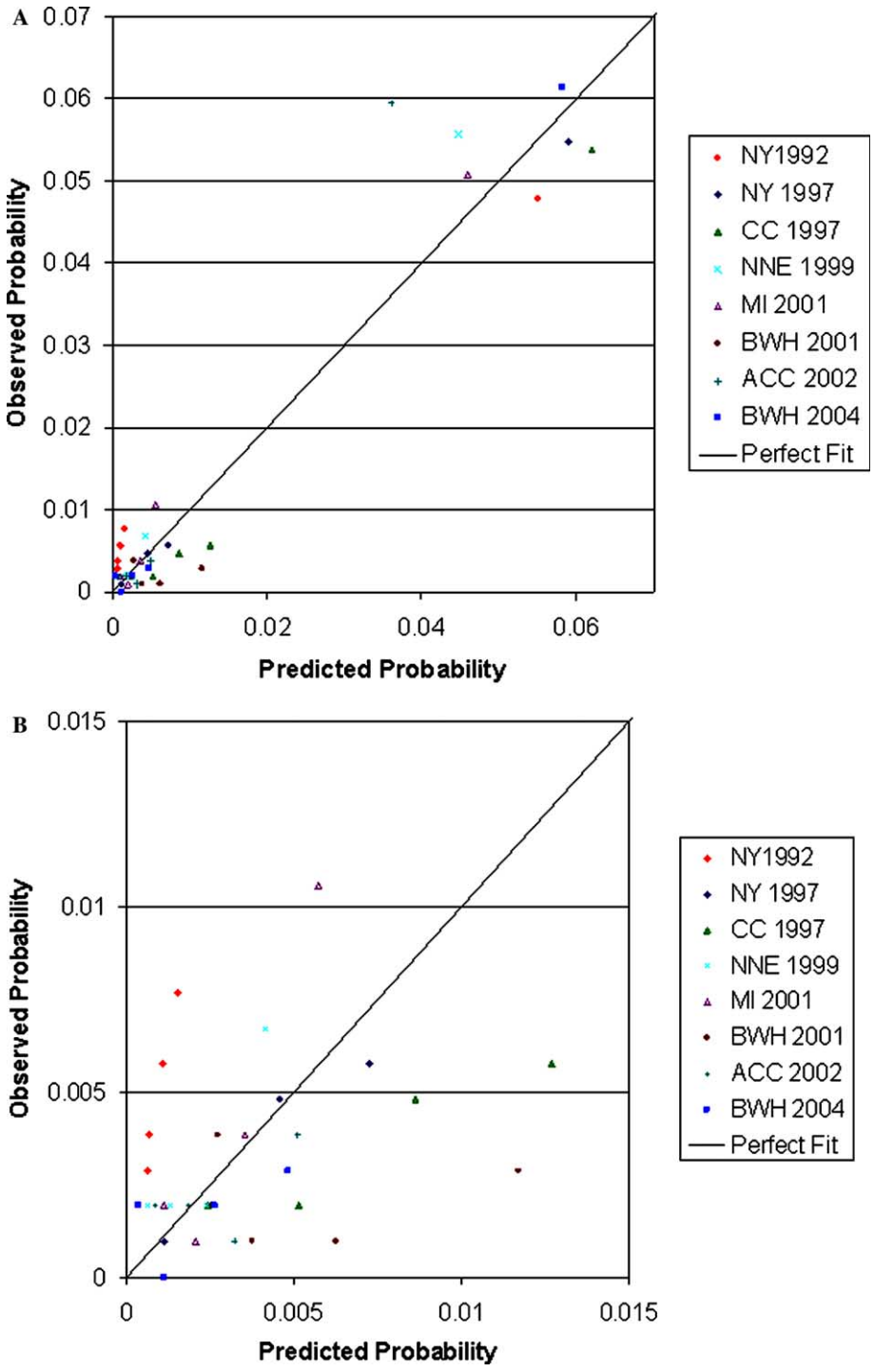

Fig. 2. (A) The observed and expected mortality rates for each quintile of patient risk. Each risk quintile contains approximately 1050 patients. The diagonal line represents a perfect agreement between observed and expected mortality estimates. (B) Expanded view from 0 to 0.15 of the observed to expected probability ratios.

All the external models evaluated on the BWH dataset showed good discrimination. The model with the worst discrimination was NY 1992, which was to be expected due to the age of the study, and small sample size with which the model was developed. The best external model was the one developed on a national database with the most patient records, suggesting that geographic issues may be related to discrimination. Although these results are promising, it is important to note that discrimination is not the only (and possibly not the most important) factor in determining the applicability of a prognostic model from the perspective of

physicians and patients. A model can exhibit perfect discrimination but still be useless for application on individual cases. Good calibration is essential for this type of application. All models, except possibly the one developed locally with recent data, but including a model derived locally, showed poor calibration for our test set, suggesting that time may play an important role in the applicability of a model.

Similar findings have been previously reported. Some techniques have been suggested to recalibrate the model [34]. One of these techniques was employed by Kizer et al. [14] and Peterson et al. [35] with some success, and may offer a strategy to maintain discrimination and improve calibration.

This study supports routine evaluation of any risk model, including aging local models, prior to local implementation. Discrimination was maintained for most risk adjustment models, though those more recently published and those based on the largest original datasets appeared to have the most robust discrimination when applied to a current clinical dataset. The preservation of discrimination supports the use of these models for generic risk stratification, but the poor calibration indicates that they are not useful for application in individual cases: the estimated risk of death for a single patient that is produced by these models is incorrect.

The poor calibration of the prior models suggests that variations in practice and patient demographics as well as clinical features over time have a large effect at the patient level on the risk estimate's accuracy. Further study is required to identify the optimum frequency of model recalibration.

This research was supported in part by Grants 1-T15-LM-07092 and R01-LM08142-01 from the National Library of Medicine of the National Institutes of Health. The authors have no conflicts of interest to disclose.

References

[1] Topol EJ, Block PC, Holmes DR, Klinke WP, Brinker JA. Readiness for the scorecard era in cardiovascular medicine. Am J Cardiol 1995;75:1170–3.
[2] McNeil BJ, Pedersen SH, Catsonis C. Current issues in profiling quality of care. Inquiry 1992;29:298–307.
[3] Poses RM, Smith WR, McGlish DK, Huber EC, Clemo FLW, Schmitt BP, et al. Physician's survival predictions for patients with acute congestive heart failure. Arch Intern Med 1997;157:1001–7.
[4] Perkins HS, Jonsen AR, Epstein WV. Providers as predictors: using outcome predictions in intensive care. Crit Care Med 1986;14:105.
[5] deDombal FT. Computer-aided diagnosis and decision-making in the acute abdomen. J R College Physicians London 1975;9(3):211–8.
[6] Williams DO, Holubkov R, Yeh W, Bourassa MG, Al-Bassam M, Block PC, et al. Percutaneous coronary intervention in the current era compared with 1985–1986: the national heart, lung, and blood institute registries. Circulation 2000;102(24):2910–4.
[7] Shaw RE, Anderson HV, Brindis RG, Krone RJ, Klein LW, McKay CR, et al. Updated risk adjustment mortality model using the complete 1.1 dataset from the American college of cardiology national cardiovascular data registry (ACC-NCDR). J Invasive Cardiol 2003;15(10):578–80.
[8] Altman DG, Royston P. What do we mean by validating a prognostic model. Stat Med 2000;19:453–73.
[9] Holmes DR, Selzer F, Johnston JM, Kelsey SF, Holubkov R, Cohen HA, et al. Modeling and risk prediction in the current era of interventional cardiology. Circulation 2003;107:1871–6.
[10] Holmes Jr DR, Berger PB, Garratt KN, Mathew V, Bell MR, Barsness GW, et al. Application of the New York State PTCA mortality model in patients undergoing stent implantation. Circulation 2000;102:517–22.
[11] Moscucci M, O'Connor GT, Ellis SG, Malenka DJ, Sievers J, Bates ER, et al. Validation of risk adjustment models for in-hospital percutaneous transluminal coronary angioplasty mortality on an independent data set. J Am Coll Cardiol 1999;34:692–7.
[12] Rihal CS, Grill DE, Bell MR, Berger PB, Garratt KN, Holmes Jr DR. Prediction of death after percutaneous coronary interventional procedures. Am Heart J 2000;139:1032–8.
[13] Singh M, Rihal CS, Selzer F, Kip KE, Detre K, Holmes DR. Validation of Mayo Clinic risk adjustment model for in-hospital complications after percutaneous coronary interventions, using the national heart, lung, and blood institute dynamic registry. J Am Coll Cardiol 2003;42(10):1722–8.
[14] Kizer JR, Berlin JA, Laskey WK, Schwartz JS, Sauer WH, Krone RJ, et al. Limitations of current risk-adjustment models in the era of coronary stenting. Am Heart J 2003;145:683–92.
[15] Hannan EL, Wu C. Assessing quality and outcomes for percutaneous coronary intervention: choosing statistical models, outcomes, time periods, and patient populations. Am Heart J 2003;145(4):571–4.
[16] Cannon CP, Battler A, Brindis RG, Cox LJ, Ellis SG, Every NR, et al. ACC key data elements and definitions for measuring the clinical management and outcomes of patients with acute coronary syndromes: reference guide: a report of the American College of Cardiology Task Force on Clinical Data Standards (Acute Coronary Syndromes Writing Committee) 2001. Available from: http://www.acc.org/clinical/data_standards/ACS/acs_index.htm.
[17] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Diagn Radiol 1982;143:29–36.
[18] Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285–93.
[19] Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115(1):92–106.
[20] STATA 8.2 (College Station, TX).
[21] Margolis DJ, Bilker W, Boston R, Localio R, Berlin JA. Statistical characteristics of area under the receiver operating characteristic curve for a simple prognostic model using traditional and bootstrapped approaches. J Clin Epidemiol 2002;55:518–24.
[22] Carpenter J, Bithell J. Bootstrap Confidence intervals: when, which, what? A practical guide for medical statisticians. Stat Med 2000;19:1141–64.
[23] O'Connor GT, Malenka DJ, Quinton H, Robb JF, Kellett MA, Shubrooks S, et al. Multivariate prediction of in-hospital mortality after percutaneous coronary interventions in 1994–1996. J Am Coll Cardiol 1999;34:681–91.
[24] Hannan EL, Arani DT, Johnson LW, Kemp HG, Lukacik G. Percutaneous transluminal coronary angioplasty in New York state. JAMA 1992;268:3092–7.
[25] Hannan EL, Racz M, Ryan TJ, McCallister BD, Johnson LW, Arani DT, et al. Coronary angioplasty volume-outcome relationships for hospitals and cardiologists. JAMA 1997;279:892–8.

[26] Moscucci M, Kline-Rogers E, Share D, O'Donnell M, Maxwell-Eward A, Leengs WL, et al. Circulation 2001;104:263–8.
[27] Shaw RE, Anderson V, Brindis RG, Krone RJ, Klein LW, McKay CR, et al. J Am Coll Cardiol 2002;39:1104–12.
[28] Ellis SG, Weintraub W, Holmes D, Shaw R, Block PC, King SB. Relation of operator volume and experience to procedural outcome of percutaneous coronary revascularization at hospitals with high interventional volumes. Circulation 1997;95:2479.
[29] Resnic FS, Ohno-Machado L, Selwyn A, Simon DI, Popma JJ. Simplified risk score models accurately predict the risk of major in-hospital complications following percutaneous coronary intervention. Am J Cardiol 2001;88:5–9.
[30] Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown; 1988.
[31] Efron B, Tibshirani R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1986;1:54–77.
[32] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–43.
[33] McGrath PD, Malenka DJ, et al. Changing outcomes in percutaneous coronary interventions: a study of 34752 procedures in northern New England, 1990 to 1997. J Am Coll Cardiol 1999;34:674–80.
[34] DeLong ER, Peterson ED, DeLong DM, Muhlbaier LH, Hackett S, Mark DB. Comparing risk-adjustment methods for provider profiling. Stat Med 1997;16:2645–64.
[35] Peterson ED, DeLong ER, Muhlbaier LH, Rosen AB, Buell HE, Kiefe CI, et al. Challenges in comparing risk-adjusted bypass surgery mortality results. J Am Coll Cardiol 2000;36:2174–84.