CathPCI: Bleeding model (risk-adjusted) Specifications and Testing Overview

Measure Purpose	Intra and post percutaneous coronary intervention (PCI) major bleeding events are common and associated with an increased cost, short- and long-term risk of morbidity and mortality [J Am Coll Cardiol Intv 2013;6:897–904]. The 2011 ACC/AHA guidelines provide a Level IC recommendation for the assessment of bleeding prior to PCI [J Am Coll Cardiol. 2011 Dec 6;58(24):e44–122]. This is grounded in the realization that there are several bleeding avoidance strategies (BAS), such as radial approaches and the use of bivalirudin, that can be applied to mitigate the risk of bleeding, particularly in high-risk patients [J Am Coll Cardiol Intv 2013;6:897–904]. Improvements in these combined processes provide an improved outcome of a reduction in the major bleeding adverse event rate and a reduction in costs associated with these events. *Additional evidence is provided under the validity testing sections of this application.*
Numerator	Patients 18 years of age and older with a post-PCI bleeding event as defined below: Post-PCI bleeding defined as any ONE of the following: 1. Bleeding event w/in 72 hours ; *OR* 2. Hemorrhagic stroke; *OR* 3. Tamponade ; *OR* 4. Post-PCI transfusion for patients with a pre-procedure hgb >8 g/dL and pre-procedure hgb not missing; *OR* 5. Absolute hgb decrease from pre-PCI to post-PCI of >= 3 g/dl AND pre-procedure hgb =<16 g/dL AND pre-procedure hgb not missing.
Denominator	Patients 18 years of age and older with a PCI procedure performed during admission
Model Elements	• Age • Gender • Body Mass Index • ST-segment elevation MI • Lytics • Pre-procedure hemoglobin • PCI Status • Renal Failure • Glomerular filtration rate • Cardiac arrest/in 24 hours • Cerebrovascular disease • Peripheral vascular disease • Chronic lung disease • Prior PCI • Diabetes status • Heart Failure NYHA class • Ejection fraction • Number of diseased vessels • PCI of proximal LAD • PCI of left main • Pre-procedure TIMI flow • SCAI lesion classification • Presence of chronic total occlusion • In- stent thrombosis (previously treated within 1 month)
Exclusions	The only exclusions from the bleeding model are patients undergoing CABG surgery and those who present to the hospital severely anemic and do not have an obvious clinical bleed after their procedure. These exclusions are relatively rare and firmly supported by the clinical rationale that a) bleeding and blood transfusions are common after

	cardiopulmonary bypass surgery and not necessarily related to the safety and quality of the PCI procedure; and b) that patients presenting to the hospital with severe anemia and receiving a blood transfusion may have been likely to be treated with a blood transfusion had they not undergone PCI. There were 7589 CABG patients with 2726 (35.92%) bleeding from a total population of 656744 indicating a prevalence of this exclusion of 1.1% (7589/656744) There were 1843 with HGB<8 with 87 (4.72%) bleeding from a total population of 656744 indicating a prevalence of this exclusion of 0.3% (1843/656744) NCDR does not believe that the exclusions have any impact on the validity, accuracy or interpretability of the risk-adjusted bleeding outcome measure.
Demonstrated Opportunity for Improvement	Number of patients: 636938 Number of PCI procedures per hospital volume: 0-10: 49 hospitals 11-200: 49 hospitals 201-400: 49 hospitals 401-600: 263 hospitals 601-1000: 239 hospitals 1001-2000: 128 hospitals 2001+: 19 hospitals Data range date: 2012QTR1-2012QTR4. Mean: 5.66% Stddev: 1.7% Quartile 1: 4.76% Quartile 3: 6.13% Deciles of bleeding adjusted rates: 1: 4.1% 2: 4.6% 3: 4.9% 4:5.2% 5(median): 5.5% 6: 5.7% 7: 6.0% 8: 6.4% 9: 7.1% This means that at the 10th percentile that 90% of hospitals demonstrated bleeding rates higher than 4.1% although 10 percent of hospitals were able to attain such a low rate of bleeding and could be a target for other hospitals to improve. At the 90th percentile that 10% of hospitals demonstrated bleeding rates higher than 7.1% The observed bleeding events are higher than in the previous version of this model since the new definition more accurately reflects the inclusion of bleeding complications (such as tamponade and transfusions in clinically appropriate groups) there were not included in the prior definition. This bleeding rate definition is consistent with the rate reported in clinical trials, such as the Acute Catheterization and Urgent Intervention Triage Strategy (ACUITY) trail, where the rate of bleeding among patients treated with glycoprotein IIb/IIIa inhibitors was 5.3-5.7% (Stone 2006)

	Stone GW, McLaurin BT, Cox DA, et al. Bivalirudin for Patients with Acute Coronary Syndreomes. N Engl Med 2006;355:220-3-2216. The below information below tables addresses the observed vs. predicted rates of bleeding for various populations that include hospital location, sex, race, gender, and insurance status, and race. **Hospital Location:** There were 83060 Rural, 193266 Suburban, and 360642 Urban patients representing these hospital locations. The bleeding rates were: 5.1% 5.3%, and 5.3% respectfully, representing a non-significant difference (p=.085) **Teaching Hospital:** There were 316680 patients at teaching hospitals and 320288 patients at non-teaching hospitals. The teaching hospitals had a slightly higher bleeding rate 5.4% vs 5.1% (p<.0001) **Gender:** There were 204569 Females and 432399 Males, with Females having a significantly higher bleeding rate (7.8% vs 4.1%, p<.0001). However, since gender is a predictor variable in the model this is to be expected. **Insurance Status:** There were 406664 patients with Private insurance, 144717 with Medicare, 24304 with Medicaid, 14785 with other insurance, and 46498 with no insurance. Bleeding rates were significantly different (p<.0001): 4.9%, 6.2%, 5.5%, 5.3%, and 6.0%. Medicare patients had the highest bleeding rate of 6.2% but this is due to age being a strong predictor of bleeding. **Race:** There were 555094 Caucasians, 53501 African Americans, and 28373 from other races. Bleeding rates were: 5.3%, 5.6%, 5.3% respectfully.
Reliability Testing	We have chose to use different datasets to provide support for different aspects of the proposed measure. 1. Audit data: 01/2009-12/ 2009 has been used to support the inter-rater reliability of the application. 2. A separate cohort of the NCDR CathPCI registry was used to validate the model, which included all data collected from 1365 hospitals in 2012 (01/2012-12/2012). These data were also used to provide test-retest reliability of the data elements for the risk model and further validate the model. ***Assessment of test-retest reliability among patients undergoing 2 procedures within 2012:*** The key data elements for the bleeding risk model tested among patients with 2 procedures in 2012 are shown below: Gender demonstrated excellent reproducibility, with only 12 of 40,197 (0.03%) patients having different genders on the 2 procedures. Age as assessed by Date of Birth was identical in 99.91% of the 40,197 patients on both assessments.

Cerebrovascular disease (CVD) revealed that only 1161 patients had evidence of CVD on the initial visit that was not noted on the second visit. This represents 2.89% of the population being clearly misclassified on one of the assessments.

Peripheral Vascular Disease (PVD) revealed that only 1341 (3.3%) patients who had evidence of PVD at the time of their initial PCI no longer had this recorded at the time of their second procedure and were clearly misclassified on one of the assessments.

Chronic Lung Disease (CLD) was recorded in 1370 (3.4%) of the patients at the time of their initial PCI, but not at the time of the second procedure.

Prior PCI should have been recorded on the second procedure for each of the 40,197 patients. 1200 (2.9%) were not classified as having had a prior PCI.

Diabetes was not recorded among 732 (1.8%) of the patients who were noted to have diabetes at the time of their original procedure.

Because dynamic elements are expected to change over time, the following variables could not have their test-retest reliability assessed by this method: Prior cardiac arrest, GFR, NYHA classification, shock within 24 hours of PCI, indication for PCI, urgency of the procedure, use of lytics prior to PCI, pre- and post-procedure hemoglobin, number and location of diseased vessels, lesion severity as assessed by the SCAI definitions, pre-procedural TIMI flow and acute stent thrombosis.

Assessment of item-level reliability through the Audit Program:

To assess inter-rater reliability of the extracted data elements that comprise this measure, 627 patients at 25 hospitals were reviewed by an independent contractor, hired by ACCF.

The National Cardiovascular Data Registry® (NCDR®) Audit Program's overall purpose is to ensure that data submitted to the NCDR registries are complete, valid, and accurately interpreted and collected. The National Audit program also ensures that benchmarks and comparisons provided to all NCDR participants accurately reflect registry data. A summary of the Audit Program is noted below.

Methodology	• Nationwide program (i.e., all submitting participants in the United States) • Review of data submitted the previous year • Review of a subset of data elements that can rotate each year • Remote review of data combined with couple of onsite visit • Onsite visits are targeted based on the Data Outlier Program • Random selection of sites and records • Blinded data abstraction from medical charts • Inter-rater Reliability Assessment conducted to validate the audit findings • Adjudication step for participant to refute audit findings
Scope	• Review of hospital's medical records for related episodes of care • Assessment of complete submission (Comparison of two lists : hospital list of cases with specific billing codes versus NCDR

	submitted records)
Criteria for selecting sites/records	Remote audit : • Sites passing their quarterly DQR for 2 quarters within audited year • Sites submitting at least the number of records/sites being reviewed Onsite audit • Sites identified with an outlier and not contacted with the data outlier program
Scoring	NCDR uses a grading system for identifying the amount of agreement or matching between the data captured during the medical record review and data submitted to the NCDR.

The kappa and percentage agreement scores were computed and noted below:

CE #	Variable Name	Kappa	Levels	Agreement (%))
2060	Sex	0.937	3	98.1
4035	Prior PCI	0.872	2	94.4
4065	Currently on Dialysis	0.863	2	99.5
4085	Diabetes Mellitus	0.909	2	96.2
4090	Diabetes Therapy	0.815	5	91.2
5000	CAD Presentation	0.553	6	69.1
5010	Thrombolytics	0.787	3	94.4
5040	Heart Failure w/in 2 Weeks	0.512	3	92.2
5045	NYHA Class w/in 2 Weeks	0.375	5	91.5
5060	Cardiogenic Shock w/in 24 Hours	0.565	3	98.6
5065	Cardiac Arrest w/in 24 Hours	0.446	3	98.1
5305	PCI	0	2	99.4
7020	PCI Status	0.563	4	95.1
7026	Pre-PCI Left Ventricular Ejection Fraction Not Assessed	0.43	2	77.8
7030	Cardiogenic Shock at Start of PCI	0.401	3	97.9
8021	Hemorrhagic Stroke	1	3	100.0
8040	RBC/Whole Blood Transfusion	0.559	2	97.6
8050	Bleeding Event w/in 72 Hours	0.471	2	97.6
9000	CABG	0.907	2	99.7
9040	Discharge Status	1	2	100.0

Of the 20 elements that are part of the bleeding model, or outcome, that were evaluated in the audit, the agreement between the auditors and that reported in the data collection form exceeded 90% for all and 97% for all but 5. The kappas vary, in part, because of the rarity of some of these events. Furthermore, the lowest kappas, such as heart failure and NYHA, are often not well documented, independent of the NCDR data collection form, in routine PCI and there may be errors made the abstractors, or the data collection system at the hospital.

There is one instance of kappa = 0 (element: PCI). It is probably a situation of Kappa paradox which functions similar to a negative Kappa. This essentially means that there is a large difference between two options and the answers are universally 1-sided. This

	creates a uneven distribution which can cause a high agreement rate with a very law Kappa score. Collectively, we believe that the audit data and repeat procedure data strongly support the reliability of the data elements used in the model. (Reference: Landis J, Koch G, The measurement of observer agreement for categorical data, *Biometrics*, 1977;33:159-174.)
Validity Testing	<u>Systematic assessment of content validity:</u> Content validity of this process was achieved by the specialized expertise of those individuals who developed this measure as well as the structured discussions that the group conducted. For this particular topic those individuals who were involved in identifying the key attributes and variables for this process measure were leaders and experts in the field of interventional cardiology. Serial phone calls were held to both define the eligible population and given process. These clinical leaders are noted below. NCDR Clinical Subworkgroup ensured the measure demonstrated an opportunity for improvement, had strong clinical evidence, and was a reliable and valid measure. These members included Drs. Jeptha Curtis (Chair), Frederick Masoudi, John Rumsfeld, Issam Moussa, and David Malenka. NCDR Scientific Quality and Oversight Committee—a committee that served as the primary resource for crosscutting scientific and quality of care methodological issues. These members included Drs. Frederick Masoudi (Chair) , David Malenka, Thomas Tsai, Matthew Reynolds, David Shahian, John Windle, Fred Resnic, John Moore, Deepak Bhatt, James Tcheng, Jeptha Curtis, Paul Chan, Matthew Roe, and John Rumsfeld. Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees reviewed and approved these measures for submission to NQF. <u>Evidence</u> Beyond the inherent content validity of this process, we have data showing that the bleeding risk score is highly actionable – a critical feature for moving beyond quality assessment to quality improvement. For example, a comparative effectiveness analysis of bivaluridin use by bleeding risk suggested that bivalirudin was preferentially used in low-risk patients (NNT=224) and least often used in patients at high risk for bleeding (NNT=43; JAMA 2010;303(21):2156-2164). At Saint Luke's Mid America Heart Institute, the original bleeding model was executed prior to non-emergent PCI in all patients undergoing the procedure. Not only was the 'risk-treatment' paradox reversed, but the bleeding rate at that institution decreased by 40% (J Am Coll Cardiol 2013;61: 1847–52). Unpublished data from a 9-center study of providing pre-procedural bleeding risks demonstrated a fully-adjusted 45% lower odds of bleeding when the models were used. The relationship between bleeding and mortality is observed across each indication for PCI, from elective through the acute MI. The review of literature indicates that bleeding is as predictive of 1-year mortality as a clinical history of a previous MI and urgent repeat revascularization. Bleeding is also associated with an increased risk of recurrent ischemic events, length of hospital stay, and increasing cost. Based upon these associations, outcomes, and increased resource utilization, the identification of patients at high risk for bleeding is a clinical imperative. Improved identification of high-risk patients will enable physicians to develop alternative approaches to mitigate the risk of bleed [Circ Cardiovasc Interv 2009;2: 222–229]. Current studies demonstrate a "risk-treatment" paradox exists with respect to the use of bleeding avoidance strategies among patients undergoing PCI. The BAS have been found

to be used the least among the patient populations with the highest bleeding risk. In addition, among high-risk patients, such as those with ST-segment elevation myocardial infarction, these BAS are associated with reduced mortality, underscoring the importance of identifying these high risk patients and applying BAS in these patients most likely to benefit [J Am Coll Cardiol Intv 2013;6:897–904].

Inter-Rater Reliability

Inter-rater reliability from independent audit can also be used to establish the 'validity' of the data in the NCDR. Specifically, we examined the agreement between the 2009 audit data for each available element used to predict bleeding. The bleeding outcome could not be assessed using the original audit data because of changes in definitions between the original bleeding model (Mehta et al, Circ Cardiovasc Intervent 2009;2:222-229) and the current model . The current definition of post-PCI bleeding, using the updated data collection form, was created by a panel of experts to most accurately capture clinically important bleeding events. The new definition includes any of the following occurring within 72 hours after PCI or prior to hospital discharge (whichever occurs first): site-reported arterial access site bleeding (either external or a hematoma >10 cm for femoral access, >5 cm for brachial access, or >2 cm for radial access); retroperitoneal, gastrointestinal, or genitourinary bleeding; intracranial hemorrhage; cardiac tamponade; post-procedure hemoglobin decrease of 3 g/dl in patients with a pre-procedure hemoglobin ≤16 g/dl; or post-procedure non-bypass surgery-related blood transfusion for patients with a pre-procedure hemoglobin ≥8 g/dl. This definition includes events such as intracranial hemorrhage, tamponade, hemoglobin decreases that account for potential hemodilution, and transfusions that account for severe anemia, which were not included in the prior definition. The definitions of all data elements are available at http://www.ncdr.com. Key components of the bleeding definition that were audited, however are included in these comparisons.

The change in bleeding amongst the highest risk patients from increased use of bleeding avoidance therapies attests that the model is capable of identifying those patients with the greatest potential to benefit from therapies and clear modifiability of risk to improve safety and outcomes. The ultimately validity of the model is that the use of the model to target therapy improves outcomes strongly supports the appropriateness and capacity of this model to measure and improve quality.

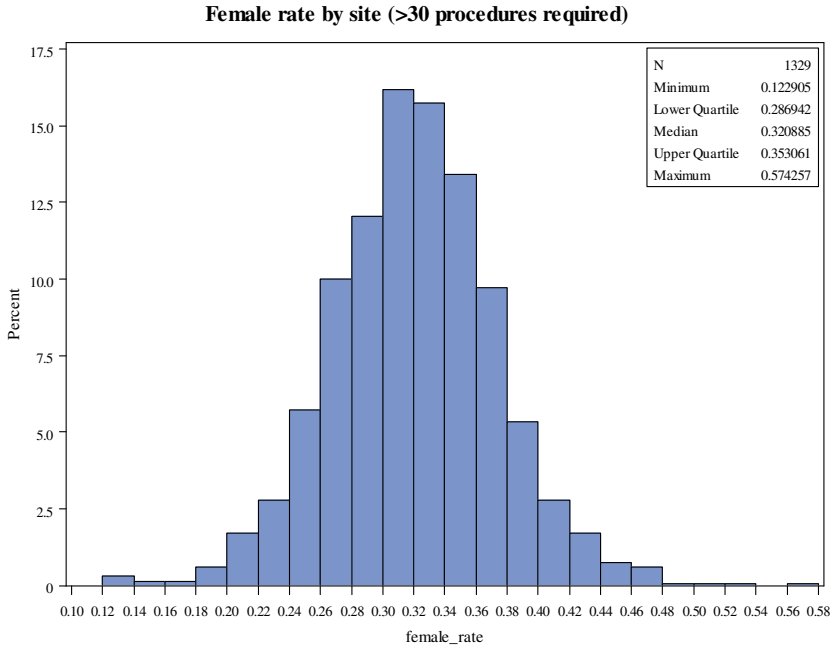
Threats to Validity:

Missing Data Bias: Because of the large amount of data typically contained in registries, it is not feasible to meet the stringent requirements used in clinical trials. However, unlike with administrative claims data, data fields in a registry can be assessed for completeness, consistency, and accuracy to support the central activities of the registry. The NCDR Data Quality Program consists of 3 main components: data completeness, consistency, and accuracy. Completeness focuses on the proportion of missing data within fields, whereas consistency determines the extent to which logically related fields contain values consistent with other fields. Accuracy characterizes the agreement between registry data and the contents of original charts from the hospitals submitting data. The thresholds for all critical elements in a performance measure are set high to ensure data completeness and consistency for the overall calculation of the performance measure. Therefore it is unlikely missing data bias would threaten the validity properties.

Selection Bias: Based on the entity and patient descriptive statistics, there does not appear to be certain subgroups of hospitals or patients who are excluded. Lastly, the exclusion frequencies did not appear to be unusually high.

Risk Adjustment

NCDR is proposing a risk-adjusted peri-procedural bleeding outcome measure to help assess the quality and safety of PCI. It is noteworthy that we included gender and age in the risk-adjustment model, although not race. This is because both gender and age are strongly associated with peri-procedural bleeding and are outside of the locus of control of physicians. Importantly, if the predicted risk is to be used prospectively to improve the use of bleeding avoidance strategies, it is important that clinicians know the true expected risk of bleeding for each and every patient, as recommended by the guidelines. Moreover, there is not a wide variation in the proportion of women (medial 32%, IQRs 29%, 35%) across hospitals as shown in the figure below.



Describe the conceptual/clinical and statistical methods and criteria used to select patient factors used in the statistical risk model or for stratification by risk

There was an extensive process to develop the face and contact validity of the measure. After settling on the outcome definition and candidate variables through serial conference calls with the expert panel, categorical variables were summarized as frequencies and percentages and compared with Pearson chi-squared tests. Continuous variables were summarized as medians (interquartile range) and compared using Wilcoxon rank-sum tests. Ordinal variables were tested using a chi-square test based on the rank of the group mean score.

The study population was then randomly split into a development sample consisting of 80% of PCI procedures and a validation sample consisting of the remaining 20%. Baseline patient characteristics and variables from diagnostic catheterization were considered candidate variables. Candidate variables had less than 0.5% missing data except for estimated glomerular filtration rate (7.8%), pre-procedure hemoglobin (9.5%), and ejection fraction (29.4%). Missing values were imputed to the lower risk group for discrete variables and replaced with gender-specific medians for body mass index (BMI), gender and renal failure/dialysis-specific medians for estimated glomerular filtration rate, median value for hemoglobin, and congestive heart failure (CHF)/cardiogenic shock/prior myocardial infarction (MI)-specific medians for ejection fraction. We used logistic regression with backward selection with a 'stay' criterion of $p<0.05$ to develop a model predicting post-PCI bleeding. Variables that showed non-linear associations with the

outcome were transformed using splines.

We developed a full post-PCI bleeding model using all potential predictive variables. We also developed a risk prediction score by taking the regression coefficients from the pre-procedure model and assigning them an integer that was weighted to the comparative odds ratio associated with the risk factors. While this score is not proposed as a performance measure, we mention it here to show that a tool exists that can be used by hospitals to reduce their bleeding rates and increase the safety of their PCI performance. Covariates selected for the risk score were those with a chi-square >500.

The C-statistic was used to describe the discrimination of the model and replicated in clinically important subgroups of interest, including patients STEMI, females, those aged >70 years, and patients with diabetes. Calibration plots were used to access goodness of fit. A p-value <0.05 was considered statistically significant. All statistical tests were two-sided. All statistical analyses were performed using SAS software (version 9.2, SAS Institute, Cary, NC). Multivariable, hierarchical logistic regression analyses were then performed to retain those with a statistically significant association with bleeding ($p<0.05$ for each).

Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach:

We developed the model in the 80% derivation set and tested its discrimination and calibration (using both the Hosmer-Lemeshow test and the slope of the predicted vs. observed risk). We then replicated this in 2 separate data sets; 20% of the original sample from 2/08-4/11 and in a completely unique set of data from 2012. Given secular trends in bleeding rates, with increasing use of radial approaches and bivalirudin leading to lower bleeding rates, we propose recalibrating the model with a new intercept (no change to the β-weights) each year, as was done for 2012 data.

Statistical Risk Model Discrimination Statistics:

The c-statistic is 0.784, which means that the probability that predicting the outcome is better than chance. This method is used to compare the goodness of fit of logistic regression models. The range is between 0.5 to 1.0. A value of 0.5 indicates that the model is no better than chance at making a prediction of membership in a group and a value of 1.0 indicates that the model perfectly identifies those within a group and those not. Models are typically considered reasonable when the C-statistic is higher than 0.7. (Hosmer & Lemeshow, 2000).

The c-statistics for the original derivation and validation cohorts, as well as clinically important subgroups are provided in the table below:

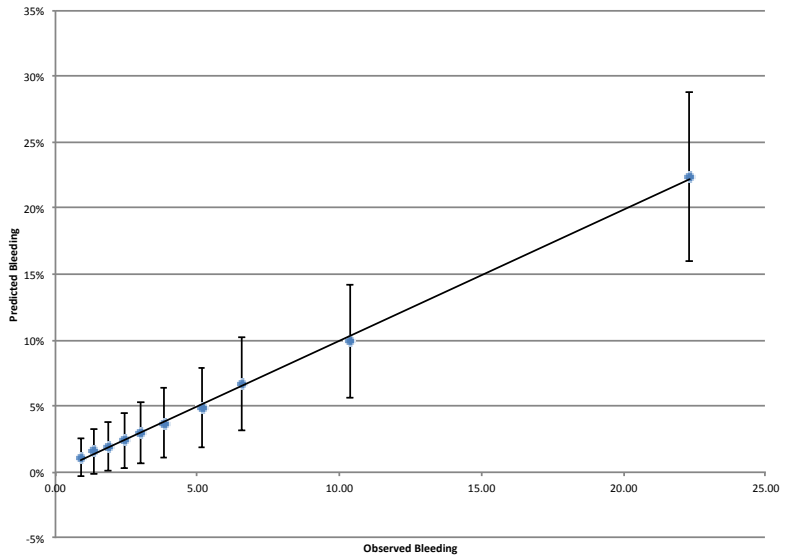
Group	N		Full Model		Risk Score	
	Development Sample	Validation Sample	Development Sample	Validation Sample	Development Sample	Validation Sample
Overall	834,696	209,063	0.78	0.77	0.76	0.75
STEMI	133,649	33,311	0.71	0.71	0.70	0.70
Women	272,357	68,540	0.74	0.74	0.73	0.72
Age >70 years	275,089	69,015	0.76	0.76	0.74	0.74
Diabetes	299,402	75,003	0.78	0.78	0.76	0.76
Excluding in-hospital CABG	824,414	205,510	0.79	0.78	0.76	0.76

In the 2012 data, the c-statistic was 0.78, slightly higher than that observed in the original data.

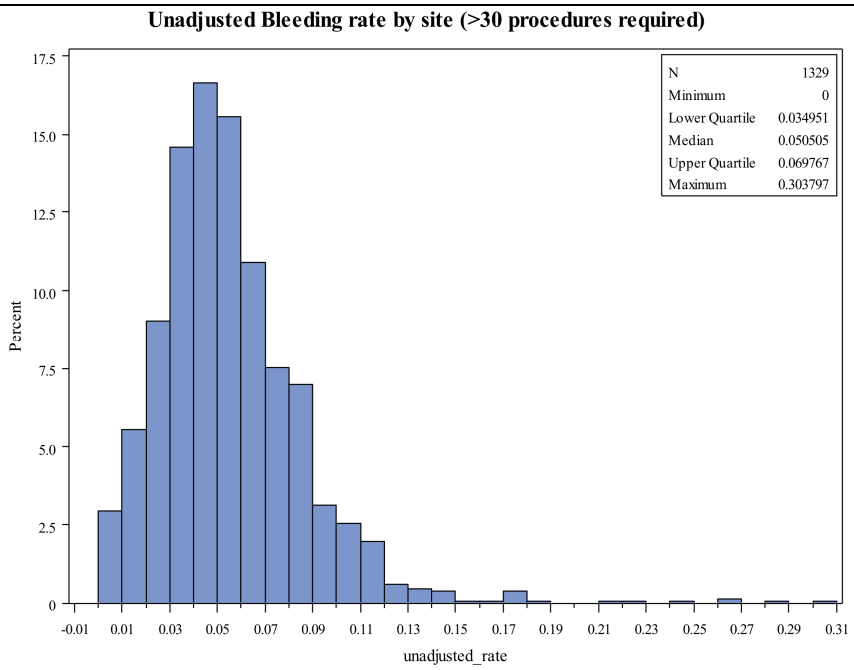
Statistical Risk Model Calibration Statistics:
Before recalibrating the model to the 2012 data the slope of the calibration line was 1.0099 (p=0.069) indicating that the relationship between the independent variables in our model and the bleeding outcome remained consistent, and the intercept of the line was -0.1684 (p<.0001) indicating that the bleeding rate as decreased since the model was developed.

Due to the decreased bleeding rate from model development we recalibrated the model to the 2012 rates and obtained a slope and intercept of 1 and 0 respectively.

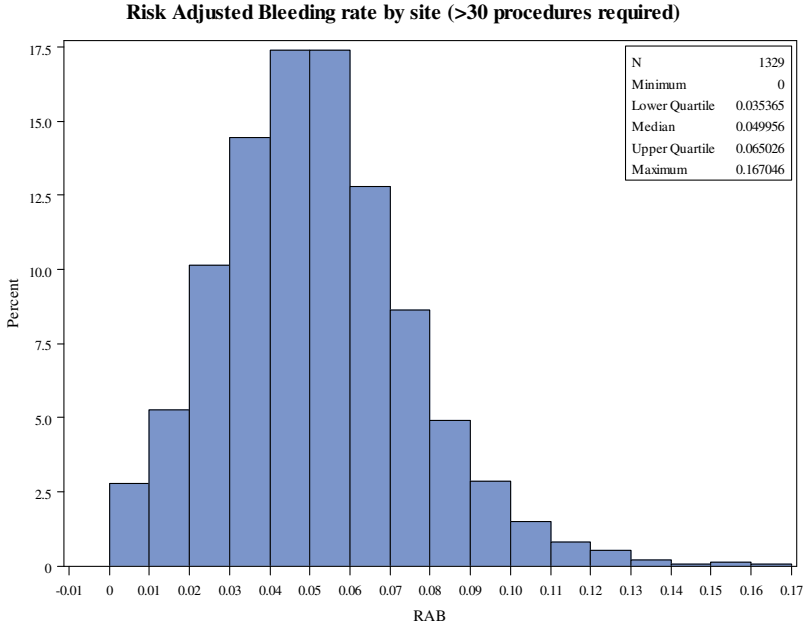
Statistical Risk Model Calibration – Risk decile plots or calibration curves:
The calibration plot in the 20% validation sample is shown below:



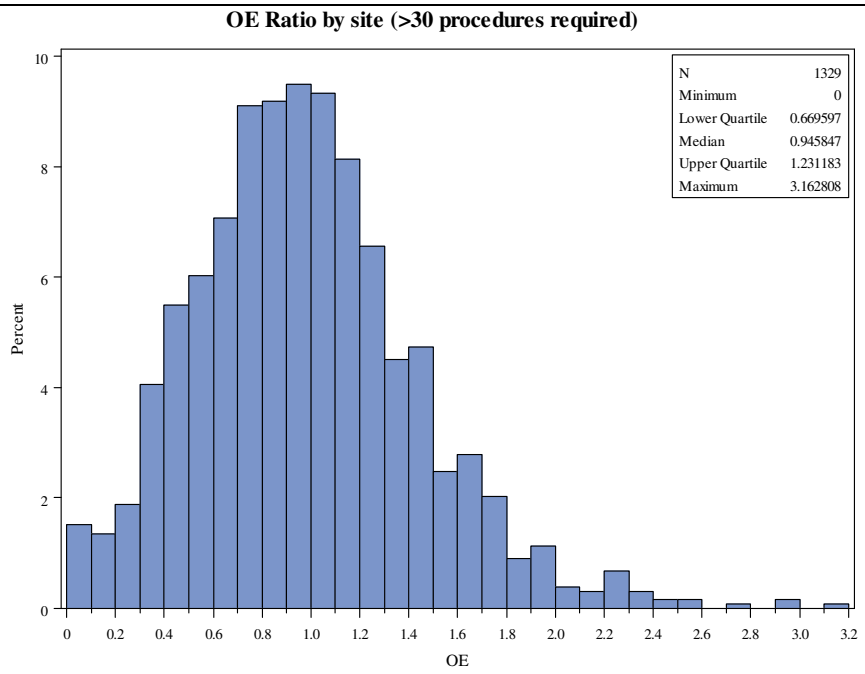
Results of Risk Stratification Analysis:
The risk stratification was able to adequately segregate deciles of risk from <1% to >22% at the patient level. At the hospital level, we observed a broad range of unadjusted risk, which was substantially tightened after adjusting for patient characteristics. The unadjusted distribution of bleeding is shown below:



After adjusting for patient characteristics, we observed a significantly tighter and more normal distribution of bleeding outcomes.



The distribution of sites' observed/expected ratios are shown below:



2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

We believe that the our bleeding model performs very well in adjusting for patient characteristics present prior to the conduct of PCI and is able to discriminate well across a wide variety of important clinical subsets of patients. Moreover, there is substantial hospital variation before and after risk-adjusting for patient characteristics. The distribution of hospitals' O/E ratios show that there are some sites with excellent performance and others with rates of bleeding that are 80% or greater than expected. These would be sites where substantial opportunities to improve patient safety likely exist.

A meaningful difference is one that indicates the potential for improvement in comparison to others. There are no absolute levels of bleeding risk that are significant as compared with others. The average, adjusted bleeding rate was 5% and the upper quartile ranges from 6.5 to 17% bleeding rates. Given an average PCI volume of 410 cases/hospital, this suggests between 6 and 48 extra bleeding events per year among hospitals in the upper quartile as compared with the average hospital. Clinically, this is a large number of events, particularly given that there are readily applied interventions, such as radial approaches and bivalirudin use, to mitigate bleeding. In a recent 9-center study, an OR of 0.55 was attained simply from pre-procedural risk stratification and subsequent changes in bleeding management.

We believe that the use of this model to identify outliers and the ability to pre-procedurally risk stratify patients and tailor therapy to risk holds great promise for improving the quality and safety of PCI.