NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

To navigate the links in the worksheet: Click to go to the link. ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 2459

Corresponding Measures:

De.2. Measure Title: Risk Standardized Bleeding for patients undergoing percutaneous coronary intervention (PCI).

Co.1.1. Measure Steward: American College of Cardiology

De.3. Brief Description of Measure: Risk adjusted rate of intra and post procedure bleeding for all patients age 18 and over undergoing PCI.

1b.1. Developer Rationale: Bleeding is the second most common non-cardiac complication of PCI. It is associated with adverse patient outcomes (e.g. increased mortality, prolonged length of stay and costs) and – most importantly – is modifiable through the use of bleeding avoidance strategies such as radial arterial access. Moreover, studies document under-use of bleeding avoidance strategies in high-risk patients. Thus, as an adverse event that varies widely across providers and is modifiable, the use of risk-adjusted bleeding metrics can provide the foundation for quality improvement initiatives that improve the safety and outcomes of treatment.

References:

Levine, G. N., Bates, E. R., Blankenship, J. C., Bailey, S. R., Bittl, J. A., Cercek, B., Chambers, C. E., Ellis, S. G., Guyton, R. A., Hollenberg, S. M., Khot, U. N., Lange, R. A., Mauri, L., Mehran, R., Moussa, I. D., Mukherjee, D., Nallamothu, B. K., Ting, H. H. (2011) 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelinesand the Society for Cardiovascular Angiography and Interventions. Journal of The American College of Cardiology, 58(24), e44-e122.

Vora, A. N., Peterson, E. D., McCoy, L. A., Garratt, K. N., Kutcher, M. A., Marso, S. P., Roe, M. T., Messenger, J. C., & Rao, S. V. (2016). The impact of bleeding avoidance strategies on hospital-level variation in bleeding rates following percutaneous coronary intervention. Journal of the American College of Cardiology: Cardiovascular Interventions, 9(8), 771-779.

S.4. Numerator Statement: 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. Cardiac Tamponade; OR
4. Post-PCI transfusion for patients with a pre-procedure hemoglobin (Hgb) >8 g/dL and pre-procedure Hgb not missing; OR
5. Absolute Hgb decrease from pre-PCI to post-PCI of >= 4 g/dl AND pre-procedure Hgb =<16 g/dL AND pre-procedure Hgb not missing

S.6. Denominator Statement: Patients 18 years of age and older with a PCI procedure performed during admission

S.8. Denominator Exclusions:

1. Patients who did not have a PCI (episodes of care with a diagnostic catheterization only);
2. Patients who died on the same day of the procedure
3. Patients who underwent CABG during the episode of care

De.1. Measure Type: Outcome

S.17. Data Source: Registry Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Sep 08, 2014 **Most Recent Endorsement Date:** Sep 08, 2014

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

Preliminary Analysis: Maintenance of Endorsement Measure

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation. The emphasis for maintaining endorsement is noted for each criterion.

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

The relationship between the process of evaluating patients for risk of bleeding before percutaneous coronary intervention (PCI) and patient health outcomes is quoted from Levine et al. Evidence derived from seven publications is provided to highlight the direct relationship between periprocedural bleeding and increased mortality. Three additional publications are cited describing the utlity of risk scores associated with bleeding. Seven additional citations with relevant empirical data are provided.

Changes to evidence from last review

☐ **The developer attests that there have been no changes in the evidence since the measure was last evaluated.**

☒ **The developer provided updated evidence for this measure:**

Updates: Updates are in section 1a.2 and include seven new citations providing empirical data.

Question for the Committee:

- *Is there at least one thing that the provider can do to achieve a change in the measure results?*

Guidance from the Evidence Algorithm

1. Does the measures assess performance on a health outcome or PRO (Yes) → 2. Does the SC agree that the relationship between the measured health outcome and at least one healthcare action is demonstrated by empirical data? (Yes) → PASS

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

Updates on the national performance for the risk-standardized bleeding rates are provided for 2015 and 2016. The developer states that the data shows that bleeding events are lower than when the model was first developed. This is because the previous version of this model used a threshold of hemoglobin drop of 3g/dl to reflect a bleeding event, which was raised to 4g/dl in the current iteration of the model to align with the bleeding definitions used in other NCDR registries. A summary of data from the literature indicating an opportunity for improvement was provided. The 2016 data shows that there is substantial variation across hospitals in bleed rate, ranging from a 1.7% rate in the top performing decile to an almost 3-fold greater rate of 5.0% in the worst performing decile. The developer suggests that this is an "important opportunity for improvement due to the observed variability across hospitals."

Disparities

A table of c-indexes of the full model and risk score models in the overall dataset and in pre-specificed subgroups is provided. The absolute rates after patient-level adjustment were clinically marginal, except for gender and age, which are strong risk factors for bleeding.

Questions for the Committee:

- Specific questions on information provided for gap in care.
- Is there a gap in care that warrants a national performance measure?
- If no disparities information is provided, are you aware of evidence that disparities exist in this area of healthcare?

Preliminary rating for opportunity for improvement: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence:

- Moderate
- developers include several references as evidence
- I have concerns about the potential for clinically non-evident drops in Hgb driving the endpoint and therefore diluting the real and important impact of bleeding. The numerator definition for this measure differs somewhat from that used in the trials in which bleeding was linked to mortality. ACUITY study: Major bleeding (not CABG-related) was defined as: intracranial or intraocular; access site bleeding requiring intervention; ≥5-cm diameter hematoma; hemoglobin reduction of ≥4 g/dl without or ≥3 g/dl with an overt source; reoperation for bleeding; or blood product transfusion. REPLACE 2: Major bleeding was defined as intracranial, intraocular, or retroperitoneal hemorrhage, clinically overt blood loss resulting in a decrease in hemoglobin of more than 3 g/dL, any decrease in hemoglobin of more than 4 g/dL, or transfusion of 2 or more units of packed red blood cells or whole blood. TRITON: TIMI major bleeding was defined as intracerebral hemorrhage or clinically overt bleeding (including imaging) associated with a drop in hemoglobin of ≥5 g/dL (corrected for transfusion). Thus, the link between what is measured here and the outcomes of mortality and LOS with which it correlated in these trials becomes more tenuous.
- This is an outcome measure (bleeding) and the evidence is directly related to the outcome
- Risk adjusted bleeding after PCI is important to measure and report

1b. Performance Gap:

- Moderate
- The document demonstrates a clear performance gap
- Performance gap is real and persists
- The 2016 data shows that there is substantial variation across hospitals in bleed rate, ranging from a 1.7% rate in the top performing decile to an almost 3-fold greater rate of 5.0% in the worst performing decile. The opportunity to improve the safety of PCI by reducing bleeding exists for all patients, not just those of a specific race, gender, age or SES status.

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications **and** Testing

2b. Validity: Testing**;** Exclusions**;** Risk-Adjustment**;** Meaningful Differences**;** Comparability**;** Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators: Matt Austin, Mike Stoto, Bijan Borah, Lacy Fabian, Jeff Geppert

Combined reviews

Evaluation of Reliability and Validity (and composite construction, if applicable):

This measure was reviewed by the Scientific Methods Panel. A summary of the measure is provided below.

Reliability

- Reliability testing was conducted at both the data element and measure score levels.
- Data Element testing was conducted for some, but not all, critical data elements
    - Developers conducted a test-retest analysis by reviewing data for CathPCI patients who were readmitted or had a repeat procedure in 2016 (n=42,637). They analyzed 7 data for which values, in general, were not expected to change over the relatively short timeframe (i.e., elements (gender, age, cerebrovascular disease, peripheral vascular disease, chronic lunch disease, prior PCI, and diabetes).
    - Results: Inconsistencies in values for the 7 data elements ranged from 0.06% to 3%.
- Score-level testing was conducted using a signal-to-noise (SNR) analysis (specifically, Adams' beta-binomial method)
    - Results: Developers presented reliability estimates (presumably averages), for all procedures, by hospital volume tertiles, and for hospitals with greater than average volume.
        - Values ranged from .706 to .819.
        - Panel members would have liked to see information about the variation in reliability estimates as well.

Validity

- Empirical validity testing was conducted at the measure score level.
    - NOTE: Developers also described face validity assessments through various means. It is possible that at least one of the assessments described when the measure was initially endorsed conforms to NQF requirements for face validity. However, those results were not presented and therefore were not considered when rating validity (moreover, face validity assessments are less important when results of empirical testing are available).
- Testing of the measure score
    - Developers conducted a construct validation analysis by examining the association of this measure (by quintile) with other outcome including mortality, complications of heart failure and stroke, length of stay, and rates of same-day discharge.
        - Developers hypothesized that that hospitals with higher bleeding rates would have higher rates on these other adverse outcomes measures.
        - Results: Developers found statistically significant associations between quintiles of bleeding rates and the outcomes of interest (higher rates of bleeding were associated with poorer outcomes). These results support the developers' hypothesis.
- This measure is risk-adjusted using hierarchical logistic regression with 32 risk factors.
    - Developers provided a conceptual rationale regarding why they did not include social risk factors in the risk-adjustment approach (i.e., the measure assesses in-hospital bleeding rate).
    - Some panel members questioned the lack inclusion of social risk factors in the risk-adjustment approach.
    - Model discrimination: C-statistic=0.79 for re-calibrated model using data from 2016 for 1,619 hospitals). (NOTE: c-statistic= 0.78 for initial model developed using data from 2/2008-4/2011 for 1,142 hospitals)
    - Model calibration: Developers assessed risk-model calibration by plotting observed versus predicted values. They report a slope=1 and intercept=0.

Questions for the Committee regarding reliability:

- Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?
- The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?

Questions for the Committee regarding validity:

- Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?
- The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?

Preliminary rating for reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Preliminary rating for validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Combined Methods Panel Scientific Acceptability Evaluation

Measure Number: 2459

Measure Title: In-hospital Risk Adjusted Rate of Bleeding Events for patients undergoing PCI

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☐ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☐ **Management Data**

☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**

☐ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**

☐ **Population: Community, County or City** ☐ **Population: Regional and State**

☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☐ **New** ☒ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒ **Yes** ☐ **No**

   Submission document: "MIF_xxxx" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**

   PANEL MEMBER 1: None

PANEL MEMBER 3: None.

PANEL MEMBER 5: No Concerns.

RELIABILITY: TESTING

Submission document: "MIF_xxxx" document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☒ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure**
   ☒ **Yes** ☐ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☐ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   PANEL MEMBER 1: Reliability testing onperformance was based on signal-to-noise analysis. Data element reliability was established by test-retest reliability by reviewing CathPCI patients who were readmitted or had a repeat procedure in 2016. This approach enabled the method developer examine 2 independent abstractions of data for the same patient.
   PANEL MEMBER 2: Assessed the reliability of the measure score using a signal-to-noise analysis. Assessed the reliability of the data elements using test-retest.
   PANEL MEMBER 3: The developer used the beta -binomial method to estimate signal-to-noise.
   PANEL MEMBER 4: Score-level reliability was appropriately tested using using the beta-binomial model. In addition, data were re-abstracted to estimate test-retest reliability for data elements.
   PANEL MEMBER 5: Signal to noise (score) and test-retest (elements)
7. **Assess the results of reliability testing**

   Submission document: Testing attachment, section 2a2.3

   PANEL MEMBER 1: The testing results presented in 2a2 support the reliability of the data elements and measurement scores used in the model.

   PANEL MEMBER 2: No concerns.

   PANEL MEMBER 3: The developer reports an average signal to noise ratio of 0.791 with increasing levels of reliability for facilities performing more procedures.

   PANEL MEMBER 4: The signal to noise ratio analysis demonstrate variability that is attributable to real differences in hospital quality as opposed to measurement error. The finding of no clear misclassification >3.0% for any data element provides strong support for the test-retest reliability of the bleeding risk factors assessed.

   PANEL MEMBER 5: Signal to noise demonstrated performance variation between hospitals. Test-retest showed appropriate classification.
8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.
   Submission document: Testing attachment, section 2a2.2

   ☒**Yes**

   ☐**No**

   ☐**Not applicable** (score-level testing was not performed)
9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?
   Submission document: Testing attachment, section 2a2.2

☒**Yes**

☐**No**

☐**Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☒**High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒**Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☐**Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐**Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    PANEL MEMBER 1: The rationale for my assessment of reliability as "High" is based on my comment in #7.

    PANEL MEMBER 2: Both forms of testing produced significant results.

    PANEL MEMBER 3: The developer reports average reliability. However, reliability is about noise (estimation error) relative to purpose. Therefore, reliability should be reported at the measured entity level in addition to the measure level (that is, the mean and the distribution of the reliability metric).

    PANEL MEMBER 4: Based on appropriate empirical testing of both score-level and data-element level reliability, with strong results.

    PANEL MEMBER 5: No concerns.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

    PANEL MEMBER 1: None.

    PANEL MEMBER 2: None

    PANEL MEMBER 5: No concerns.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Testing attachment, section 2b4.

    PANEL MEMBER 1: None

    PANEL MEMBER 2: None.

    PANEL MEMBER 3: None.

    PANEL MEMBER 5: As never events there should not be any bleeding events; however, variation across hospitals was present, with extrapolation to the impact of these bleeding rates to other clinical improvement opportunities (e.g., length of stay).

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    Submission document: Testing attachment, section 2b5.

    PANEL MEMBER 1: None

    PANEL MEMBER 2: Not applicable.

PANEL MEMBER 3: None.

PANEL MEMBER 5: No concerns.

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

PANEL MEMBER 1: None

PANEL MEMBER 2: None.

PANEL MEMBER 3: None.

PANEL MEMBER 5: No concerns.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☒ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☐ Yes ☒ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☒ Yes ☒ No

PANEL MEMBER 5: It says that social risk factors were not included because they weren't available in the registry; however, clinical variables indicative of severity of illness were used. The component that doesn't seem to be explained is that the inclusion of social risk factors may highlight meaningful gaps for certain populations. Certain populations may have social risk factors indicative of greater clinical severity; therefore, those populations should be a focus of quality improvement. It is unclear to me how such risk factors (e.g., insurance type/race) were unavailable for this measure but do appear to be broadly available in the registry. Further explanation is warranted, as in the risk adjustment section social-factors were assessed.

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☒ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☐ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☒ No

16e**. Assess the risk-adjustment approach**

PANEL MEMBER 1: In order to develop the risk model, the study population was randomly split into a development sample consisting of 80% of PCI procedures and a validation sample consisting of the remaining 20% of admissions. A full post-PCI bleeding model was developed using all potential predictive variables. A logistic regression model with backward selection and a retention criterion of $p<0.05$ was performed to develop the full risk model used in hospital comparisons. The C-statistic was used to describe the discrimination of the model.

The testing result indicates that the model performs very well, accounting for patient characteristics present prior to the conduct of PCI and discriminating within important clinical subsets of patients. Moreover, there is substantial hospital variation before and after risk-adjustment. The distribution of institutional observed/expected ratios identifies some sites with excellent performance and others

with rates of bleeding that are 80% or greater than expected. The latter would be sites where substantial opportunities to improve patient safety likely exist.

PANEL MEMBER 3: Well developed and justified

PANEL MEMBER 4: As documented, the statistical risk adjustment model is appropriate. The results also indicate that the risk models are predictive and are well-calibrated.

I'm not sure that I agree with the decision not to adjust for race and other sociodemographic factors.

PANEL MEMBER 5: The risk-adjustment focused on variables that contributed to different predictability for a bleed event vs. no bleed event. Found performance variation at the facility level even after adjusting patient characteristics away; however, it isn't clear to me why one would expect higher bleed rates—as never events—for women vs. men for example? It seems strong that variation is still present, but it isn't clear to me why this type of variation is being adjusted away, it seems like it is possibly highlighting important disparities in care that shouldn't be adjusted.

VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☒ **Data element** ☐ **Both**
18. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**

19. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    PANEL MEMBER 1: The method of establishing face validity of the measure is detailed in 2b1.2. In essence, various committees and sub-committees under NCDR comprising subject matter experts provided input in the development of this measure. Furthermore, face validity of the measure was also strengthened through the incorporation of comments received during the open comment period from both the registry stakeholders as well as external stakeholders.

    Empirical validity was of the measure was established by examining the association of bleeding rates, by quintiles, with other clinically important outcomes, including mortality, complications of heart failure and stroke, length of stay and rates of same-day discharge. The underlying hypothesis of this association was that patients experiencing a bleeding complication would also be at higher risk for longer post-procedure lengths of stay (because additional observation and treatment, such as transfusions and surgical repairs, would be needed to address the bleeding complication.

    PANEL MEMBER 2: Assessed the predictive validity of the measure with mortality and length of stay.

    PANEL MEMBER 3: The developer examined the facility level correlation between the bleeding outcome measures and other potentially related outcome measures.

    PANEL MEMBER 4: Validity was assessed through a formal process for face validity and an empirical assessment of predictive validity.

    PANEL MEMBER 5: Face validity. Predictive validity

20. **Assess the results(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.3

    PANEL MEMBER 1: The testing results demonstrated both the face validity and the predictive validity of the association of risk-adjusted bleeding with mortality, post-PCI complications (stroke and heart failure exacerbations) and length of stay strongly underscores the importance of this adverse event and supported the hypothesized associations in conducting these analyses (see 2b1.3 and 2b1.4 for details).

PANEL MEMBER 2: Findings supported their hypotheses.

PANEL MEMBER 3: Although the quality construct may or may not be related among the outcome measures that quality construct is not described

PANEL MEMBER 4: The formal process clearly demonstrated face validity. The correlation of risk-adjusted bleeding with mortality, post-PCI complications (stroke and heart failure exacerbations) and length of stay strongly demonstrates predictive validity.

PANEL MEMBER 5: Use of subject matter experts and open public comments appeared appropriate for face validity.

Demonstrated utility as a predictive model to identify actionable quality improvement opportunities at the hospital level.

21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

Submission document: Testing attachment, section 2b1.

☒**Yes**

☐**No**

☐**Not applicable** (score-level testing was not performed)

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?** *NOTE that data element validation from the literature is acceptable.*

Submission document*: Testing attachment, section 2b1.*

☒**Yes**

☐**No**

☒**Not applicable** (data element testing was not performed)

23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

☒**High** (NOTE: Can be HIGH only if score-level testing has been conducted)

☐**Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒**Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐**Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

PANEL MEMBER 1: The rationale for "High" rating is based on my comment in #22.

PANEL MEMBER 2: Found strong relationships with measures that were hypothesized to have a predictive relationships

PANEL MEMBER 3: A demonstration of an implicit quality construct is the lowest level of empirical validity testing. To demonstrate a moderate level, the developer must show an empirical association between the implicit quality construct and the material outcome.

PANEL MEMBER 4: The formal process clearly demonstrated face validity. The correlation of risk-adjusted bleeding with mortality, post-PCI complications (stroke and heart failure exacerbations) and length of stay strongly demonstrates predictive validity.

PANEL MEMBER 5: Face validity and statistical indicators of validity were meaningful.

ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

PANEL MEMBER 1: None noted.

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Specifications:

- No concerns
- i couldn't find the information
- All good for this.
- I don't have concerns about the ability to consistently implement the measure

2a2. Reliability testing:

- No
- is chronic lung disease a real thing? Hard to decide given so little information about the variables in the model
- No
- No concerns. Reliability of the data elements was tested on 42,6137 patients who had 2 PCIs
- Agree – all moderate for relabilty and validity

2b1. Validity testing:

- No
- same answer as above
- I am concerned about the potential for Hgb drops without clinically-evident bleeds may not have the strong relationship to important outcome measures (mortality, LOS). Hgb frequently declines simply due to peri-PCI fluid administration. Further, the measure can be easily gamed by not routinely checking post-PCI Hgb values.
- The developers examined the association of bleeding rates, by quintiles, with other clinically important outcomes, including mortality, complications of heart failure and stroke, length of stay and rates of same-day discharge and found associations with all outcomes.
- Agree – all moderate for relabilty and validity

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data):

- i can't tell, but assume no
- Yes. Concerned with gaming by avoidance of routinely checking post-PCI Hgb values.
- There are significant differences in bleeding rates among the quartiles. I did not identify significant threats to validity. Missing data are not a threat.

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment):

- Standard hierarchical models used here
- no real data presented
- no threats
- Exclusions do not appear to be a threat, and risk adjustment is appropriate. The model does not adjust for SES, and I consider this to be an appropriate decision

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

The developer states that all data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS). According to the developer, there were no difficulties noted with regard to data collection, availability of data, missing data, the frequency of data collection, patient confidentiality, time and cost of data collection, or other feasibility/implementation issues. The developer provides a detailed outline of the NCDR data collection process.

The developer includes information on time of data collection: one full time employee can enter roughly 1200 patient records per year. For calendar year 2017the annual pricing for hospitals, NCDR Analytic and Reporting Services, and licensing of measure specifications ranges from $2900-$50,000.

Measures that are aggregated by ACCF and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

Questions for the Committee:

- Are the required data elements routinely generated and used during care delivery?
- Are the required data elements available in electronic form, e.g., EHR or other electronic sources?
- Is the data collection strategy ready to be put into operational use?

Preliminary rating for feasibility: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility:

- No concerns
- no concerns
- no concerns
- There were no reported feasibility problems for NCDR reporting hospitals
- Moerate to high for feasibility

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☒ **Yes** ☐ **No**

Current use in an accountability program? ☒ **Yes** ☐ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☐ **Yes** ☐ **No**

Accountability program details

The developer lists Blue Distinction Centers for Cardiac Care as the program and Blue Cross Blue Shield Association as the sponsor. The program is a national program with 414 hospitals as accountable entities. The developer states that this measure is used for NCDR public reporting, as well as in the Quality Insight Hospital Program with Anthem.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

Feedback on the measure by those being measured or others

The developer states that feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls. The developer reports that users have not reported any difficulties with reporting this measure and no other feedback was received from other users.

Additional Feedback: The developer states no other feedback was received.

Questions for the Committee:

- How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?
- How has the measure been vetted in real-world settings by those being measured or others?

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

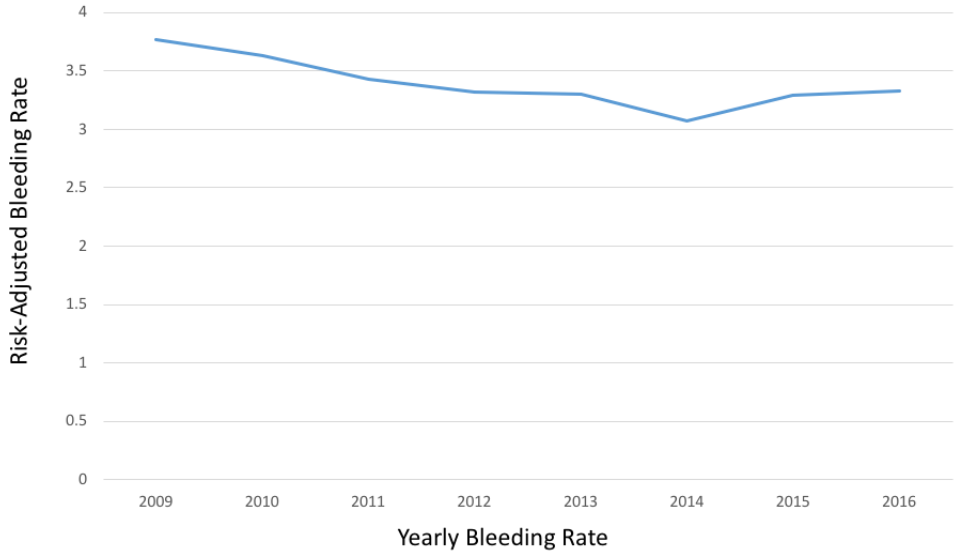
4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

Improvement results

From the developer: Over time, there has been improvement in the rates of peri-procedural bleeding, from 3.8% in 2009 to 3.3% in 2016 ($p<0.0001$). Then annual reduction in the odds of bleeding is 0.963 (95%CI=0.91, 0.965; $p<0.001$), suggesting a 4% reduction in the odds of bleeding per year, on average across all hospitals. The year over year results are shown in the figure below.



4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

Unexpected findings (positive or negative) during implementation

The developer states the most vulnerable aspect of this measure pertains to physician transparency and willingness to report and record adverse events.

Potential harms

The developer does not list any potential harms.

Additional Feedback: The developer does not provide additional feedback.

***Questions for the Committee**:*

- How can the performance results be used to further the goal of high-quality, efficient healthcare?
- Do the benefits of the measure outweigh any potential unintended consequences?

Preliminary rating for Usability and use: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a. Use:

- No concerns
- there appears to be a successful feedback mechanism and some evidence of improvement
- Not clear how many institutions use this for quality improvement given validity concerns.
- NCDR has a public web site and the data are used to identify "Blue Distinction Centers", a national program. Performance results are distributed to all CathPCI registry participants as part of quarterly benchmark reports
- High for usability and use

4b. Usability:

- No concerns. No harms imagined
- unless the model is not properly risk adjusted, no concerns
- none
- Hospitals are provided with feedback and benchmarking. No harms are identified
- High for usability and use

Criterion 5: Related and Competing Measures

Related or competing measures
The developer lists no related or competing measures.
Harmonization
The developer states there are no measure specifications harmonized.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing:

- hard to tell
- no
- There are no related or competing measures
- None

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: 1/25/2019
No comments or support/non-support choices have been submitted as of this date.

Brief Measure Information

NQF #: 2459

Corresponding Measures:

De.2. Measure Title: Risk Standardized Bleeding for patients undergoing percutaneous coronary intervention (PCI).

Co.1.1. Measure Steward: American College of Cardiology

De.3. Brief Description of Measure: Risk adjusted rate of intra and post procedure bleeding for all patients age 18 and over undergoing PCI.

1b.1. Developer Rationale: Bleeding is the second most common non-cardiac complication of PCI. It is associated with adverse patient outcomes (e.g. increased mortality, prolonged length of stay and costs) and – most importantly – is modifiable through the use of bleeding avoidance strategies such as radial arterial access. Moreover, studies document under-use of bleeding avoidance strategies in high-risk patients. Thus, as an adverse event that varies widely across providers and is modifiable, the use of risk-adjusted bleeding metrics can provide the foundation for quality improvement initiatives that improve the safety and outcomes of treatment.

References:

Levine, G. N., Bates, E. R., Blankenship, J. C., Bailey, S. R., Bittl, J. A., Cercek, B., Chambers, C. E., Ellis, S. G., Guyton, R. A., Hollenberg, S. M., Khot, U. N., Lange, R. A., Mauri, L., Mehran, R., Moussa, I. D., Mukherjee, D., Nallamothu, B. K., Ting, H. H. (2011) 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelinesand the Society for Cardiovascular Angiography and Interventions. Journal of The American College of Cardiology, 58(24), e44-e122.

Vora, A. N., Peterson, E. D., McCoy, L. A., Garratt, K. N., Kutcher, M. A., Marso, S. P., Roe, M. T., Messenger, J. C., & Rao, S. V. (2016). The impact of bleeding avoidance strategies on hospital-level variation in bleeding rates following percutaneous coronary intervention. Journal of the American College of Cardiology: Cardiovascular Interventions, 9(8), 771-779.

S.4. Numerator Statement: 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. Cardiac Tamponade; OR
4. Post-PCI transfusion for patients with a pre-procedure hemoglobin (Hgb) >8 g/dL and pre-procedure Hgb not missing; OR
5. Absolute Hgb decrease from pre-PCI to post-PCI of >= 4 g/dl AND pre-procedure Hgb =<16 g/dL AND pre-procedure Hgb not missing

S.6. Denominator Statement: Patients 18 years of age and older with a PCI procedure performed during admission

S.8. Denominator Exclusions:

1. Patients who did not have a PCI (episodes of care with a diagnostic catheterization only);
2. Patients who died on the same day of the procedure

3. Patients who underwent CABG during the episode of care

De.1. Measure Type: Outcome

S.17. Data Source: Registry Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Sep 08, 2014 **Most Recent Endorsement Date:** Sep 08, 2014

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? N/A

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

2459_nqf_evidence_attachment_11.7.18_final.docx

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** 2459

Measure Title: In-hospital Risk Standardized Bleeding Rate for Patients Undergoing PCI

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here:

Date of Submission: 11/1/2018

Instructions

- *Complete 1a.1 and 1a.2 for all measures. If instrument-based measure, complete 1a.3.*
- *Complete EITHER 1a.2, 1a.3 or 1a.4 as applicable for the type of measure and evidence.*
- *For composite performance measures:*
  - *A separate evidence form is required for each component measure unless several components were studied together.*
  - *If a component measure is submitted as an individual performance measure, attach the evidence form to the individual measure submission.*
- All information needed to demonstrate meeting the evidence subcriterion (1a) must be in this form. An appendix of *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the evidence for this measure meets NQF's evaluation criteria.

1a. Evidence to Support the Measure Focus

The measure focus is evidence-based, demonstrated as follows:

- Outcome: 3 Empirical data demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service. If not available, wide variation in performance can be used as

evidence, assuming the data are from a robust number of providers and results are not subject to systematic bias.

- Intermediate clinical outcome: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured intermediate clinical outcome leads to a desired health outcome.
- Process: 5 a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured process leads to a desired health outcome.
- Structure: a systematic assessment and grading of the quantity, quality, and consistency of the body of evidence 4 that the measured structure leads to a desired health outcome.
- Efficiency: 6 evidence not required for the resource use component.
- For measures derived from patient reports, evidence should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.
- Process measures incorporating Appropriate Use Criteria: See NQF's guidance for evidence for measures, in general; guidance for measures specifically based on clinical practice guidelines apply as well.

Notes

3. Generally, rare event outcomes do not provide adequate information for improvement or discrimination; however, serious reportable events that are compared to zero are appropriate outcomes for public reporting and quality improvement.

4. The preferred systems for grading the evidence are the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and/or modified GRADE.

5. Clinical care processes typically include multiple steps: assess → identify problem/potential problem → choose/plan intervention (with patient input) → provide intervention → evaluate impact on health status. If the measure focus is one step in such a multistep process, the step with the strongest evidence for the link to the desired outcome should be selected as the focus of measurement. Note: A measure focused only on collecting PROM data is not a PRO-PM.

6. Measures of efficiency combine the concepts of resource use and quality (see NQF's Measurement Framework: Evaluating Efficiency Across Episodes of Care; AQA Principles of Efficiency Measures).

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Outcome: Reduction in bleeding events for patients undergoing PCI

☐Patient-reported outcome (PRO):

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*):

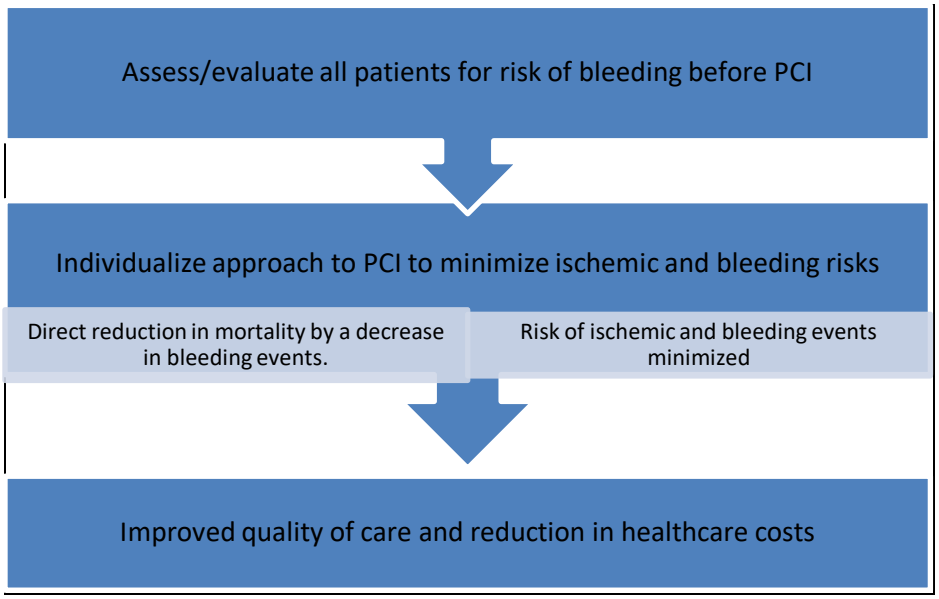
☐ Process:

☐ Appropriate use measure:

☐ Structure:

☐ Composite:

1a.2 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.



CLASS I: All patients should be evaluated for risk of bleeding before PCI. (Level of Evidence: C)

"Intra and post - procedural bleeding is recognized as a major risk factor for subsequent mortality . Bleeding may lead to mortality directly (because of the bleeding event) or through ischemic complications that occur when antiplatelet or anticoagulant agents are withdrawn in response to the bleeding. Bleeding may also be a marker of comorbidities associated with worse prognosis (e.g., occult cancer). The risk of bleeding is associated with a number of patient factors (e.g., advanced age, low body mass index, CKD, baseline anemia), as well as the degree of platelet and thrombin inhibition, vascular access site, and sheath size. The overall approach to PCI should be individualized to minimize both ischemic and bleeding risks."

> *Levine GN, Bates ER, Blankenship JC, et al. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. J Am Coll Cardiol. 2011;58(24):e44-e122. doi:10.1016/j.jacc.2011.08.007.*

1a.3 Value and Meaningfulness: IF this measure is derived from patient report, provide evidence that the target population values the measured ***outcome, process, or structure*** and finds it meaningful. (Describe how and from whom their input was obtained.)

Among patients enrolled in both randomized trials and various clinical registries, there is an approximate 3- to 10-fold increase in in-hospital and 30-day mortality for bleeding versus no bleeding. One retrospective analysis of incidence, predictors, and prognostic impact of periprocedural bleeding and transfusions involving 10,974 PCI patients from 2003 indicated patients who had major bleeding had higher in-hospital and 1-year mortality compared with patients with minor or no bleeding. Bleeding was identified as an independent predictor of in-hospital death [1].

An evaluation of the trends and factors associated with femoral bleeding after PCI was performed from 1994 – 2005 at the Mayo Clinic. A population of 17,901 patients was studied to determine factors were associated with bleeding. A multivariate analysis determined that sheath size, intensity and duration of anticoagulation with heparin, and procedure time were each independent predictors of complications. Major femoral bleeding and blood transfusion were associated with decreased long-term survival [2].

Chhatriwalla et al. recently determined the post PCI bleeding events were associated with increased risk of in-hospital mortality, with an estimated 12.1% of deaths related to bleeding complications. They retrospectively analyzed patient data from 3,386,688 PCI procedures in our CathPCI Registry performed in the US between 2004 and 2011 [3].

Data from the Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel--Thrombolysis in Myocardial Infarction 38 (TRITON-TIMI 38) was analyzed to identify baseline

characteristics that independently predict bleeding and to determine how bleeding events impact the subsequent mortality. The authors determined that major predictors of serious bleeding were a combination of patient and procedural characteristics and antiplatelet therapies. They added that serious bleeding was strongly associated with mortality within the first month after the PCI [4].

Data from the National Heart, Lung, and Blood Institute Dynamic Registry from 2007, suggest that access site complications, especially hematomas requiring transfusions, are a significant predictor of adverse procedural success and patient outcome. This prospective, multi-center, cohort study of consecutive patients undergoing PCI during 3 NHLBI Dynamic Registry recruitment waves (1997-2002) identified that in-hospital mortality and 1-year death rate was 9 and 4.5 times higher respectively in patients experiencing hematomas requiring transfusions compared to those PCI patients without the complication [5].

The REPLACE-2 Trial involving 6,001 patients undergoing PCI, noted 3.2% experienced a major hemorrhage. They determined that a number of baseline and periprocedural factors independently predicted major hemorrhage, including treatment with heparin plus GPI, and in patients undergoing elective or urgent PCI, major hemorrhage was an independent predictor of 1-year mortality [6].

The ACUITY Trial involved 13,819 PCI patients with moderate- and high-risk ACS who were randomized to to heparin (unfractionated or enoxaparin) plus glycoprotein IIb/IIIa inhibition (GPI), bivalirudin plus GPI, or bivalirudin monotherapy (plus provisional GPI). Logistic regression was used to determine predictors of 30-day major bleeding and mortality. Major bleeding was determined to be a powerful independent predictor of 30-day mortality in patients with ACS managed invasively. Several factors independently predict major bleeding, including treatment with heparin plus GPI compared with bivalirudin monotherapy. Knowledge of these findings might be useful to reduce bleeding risk and improve outcomes in ACS [7].

Periprocedural bleeding is recognized to be associated with subsequent mortality and the avoidance of bleeding complications is a critical consideration in performing PCI.

Utility of risk scores associated with bleeding

A total of 17,421 patients with acute coronary syndomes were studied in the ACUITY (Acute Catheterization and Urgent Intervention Triage strategY) and the HORIZONS-AMI (Harmonizing Outcomes with RevasculariZatiON and Stents in Acute Myocardial Infarction) trials. In developing a risk score, the data from these two trials were combined to develop a practical risk score to predict the risk and implications of major bleeding. An integer risk score for major bleeding within 30 days was developed from a multivariable logistic regression model. This practical ACUITY/HORIZONS-AMI scoring system with 6 readily available baseline clinical and laboratory variables plus the anticoagulation regimen used, can be used as a rapid and reliable tool to predict the rate of non–CABG-related major bleeding in patients with ACS and its impact on subsequent mortality within 1 year. The tool and the knowledge derived from its use aids in the accurate prognostication of patients with ACS, facilitating appropriate personalized decision-making for the patient at high risk of bleeding and mortality [8].

Baseline clinical and procedural variables from two contemporary, multicenter, randomized PCI trials were used for risk score development (the REPLACE-2 trial, n = 6002) and validation (the REPLACE-1 trial, n = 1056) to predict the incidence of major peri-procedural bleeding after contemporary PCI using the femoral approach. Variables were identified as independent correlates of major bleeding: (age >55 years, female gender, estimated glomerular filtration rate <60 mL/min/1.73 m(2), pre-existing anemia, administration of low-molecular-weight heparin within 48 h pre-PCI, use of glycoprotein IIb/IIIa inhibitors, and intra-aortic balloon pump use). In the development set, the risk of major bleeding varied from 1.0% in patients without risk factors to 5.4% in high-risk patients. The discriminatory power of this risk model was confirmed in the validation data set, cstat = 0.62). [9]

One additional risk model was developed using a development sample of 71, 277 PCI patients and validated with a sample of 17,857. This CRUSADE (Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA guidelines) model identifies 8 independent

baseline predictors of in-hospital major bleeding among community-treated NSTEMI patients enrolled in this quality Improvement Initiative. The bleeding score quantifies risk for in-hospital major bleeding across all post admission treatments, enhancing baseline risk assessment for the care of patients with NSTEMI diagnoses [10].

[1] Incidence, predictors, and prognostic implications of bleeding and blood transfusion following percutaneous coronary interventions. Kinnaird TD, Stabile E, Mintz GS, Lee CW, Canos DA, Gevorkian N, Pinnow EE, Kent KM, Pichard AD, Satler LF, Weissman NJ, Lindsay J, Fuchs S. Am J Cardiol. 2003;92(8):930.

[2] Major femoral bleeding complications after percutaneous coronary intervention: incidence, predictors, and impact on long-term survival among 17,901 patients treated at the Mayo Clinic from 1994 to 2005. Doyle BJ, Ting HH, Bell MR, Lennon RJ, Mathew V, Singh M, Holmes DR, Rihal CS. JACC Cardiovasc Interv. 2008;1(2):202.

[3] Association between bleeding events and in-hospital mortality after percutaneous coronary intervention. Chhatriwalla AK, Amin AP, Kennedy KF, House JA, Cohen DJ, Rao SV, Messenger JC, Marso SP, National Cardiovascular Data Registry. JAMA. 2013 Mar;309(10):1022-9.

[4] Predictors of bleeding and time dependence of association of bleeding with mortality: insights from the Trial to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition With Prasugrel--Thrombolysis in Myocardial Infarction 38 (TRITON-TIMI 38). Hochholzer W, Wiviott SD, Antman EM, Contant CF, Guo J, Giugliano RP, Dalby AJ, Montalescot G, Braunwald E. Circulation. 2011;123(23):2681.

[5] Access site hematoma requiring blood transfusion predicts mortality in patients undergoing percutaneous coronary intervention: data from the National Heart, Lung, and Blood Institute Dynamic Registry. Yatskar L, Selzer F, Feit F, Cohen HA, Jacobs AK, Williams DO, Slater J. Catheter Cardiovasc Interv. 2007;69(7):961.

[6] Predictors and impact of major hemorrhage on mortality following percutaneous coronary intervention from the REPLACE-2 Trial. Feit F., Voeltz M.D., Attubato M.J., et al;. Am J Cardiol. 2007;100:1364-1369.

[7] Impact of major bleeding on 30-day mortality and clinical outcomes in patients with acute coronary syndromes: an analysis from the ACUITY Trial. J Am Coll Ca he purpose of this study was to determine the predictors of major bleeding and the impact of major bleeding on outcomes, including mortality, in acute coronary syndromes (ACS). Manoukian S.V., Feit F., Mehran R., et al. J Am Coll Cardiol. 2007; 49:1362-1368.

[8] A risk score to predict bleeding in patients with acute coronary syndromes. Mehran R., Pocock S.J., Nikolsky E., et al; J Am Coll Cardiol. 2010;55:2556-2566.

[9] Development and validation of a prognostic risk score for major bleeding in patients undergoing percutaneous coronary intervention via the femoral approach. Nikolsky E., Mehran R., Dangas G., et al; Eur Heart J. 2007;28:1936-1945.

[10] Subherwal S., Bach R.G., Chen A.Y., et al; Baseline risk of major bleeding in non-ST-segment-elevation myocardial infarction: the CRUSADE (Can Rapid risk stratification of Unstable angina patients Suppress ADverse outcomes with Early implementation of the ACC/AHA Guidelines) Bleeding Score. Circulation. 2009;119:1873-1882.

<u>Note</u>: For health outcome performance measures, no further information is required; however, you may provide evidence for any of the structures, processes, interventions, or service identified above.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES - Provide empirical data demonstrating the relationship between the outcome (or PRO) to at least one healthcare structure, process, intervention, or service.

Published trials and observational studies have found that specific processes of care, including the use of radial arterial access, mechanical closure devices when femoral access is used, and bivalirudin for anticoagulation, are associated with lower risks of bleeding.

1. Vora, A. N., Peterson, E. D., McCoy, L. A., Garratt, K. N., Kutcher, M. A., Marso, S. P., Roe, M. T., Messenger, J. C., & Rao, S. V. (2016). The impact of bleeding avoidance strategies on hospital-level variation in bleeding rates following percutaneous coronary intervention. Journal of the American College of Cardiology: Cardiovascular Interventions, 9(8), 771-779.
2. Jolly SS, Amlani S, Hamon M, et al. Radial versus femoral access for coronary angiography or intervention and the impact on major bleeding and ischemic events: a systematic review and meta-analysis of randomized trials. Am Heart J2009;157:132-40.

CrossRefPubMedWeb of ScienceGoogle Scholar

3. Kastrati A, Neumann FJ, Mehilli J, et al. Bivalirudin versus unfractionated heparin during percutaneous coronary intervention. N Engl J Med2008;359:688-96.

CrossRefPubMedWeb of ScienceGoogle Scholar

4. Marso SP, Amin AP, House JA, et al. Association between use of bleeding avoidance strategies and risk of periprocedural bleeding among patients undergoing percutaneous coronary intervention. JAMA2010;303:2156-64.

CrossRefPubMedWeb of ScienceGoogle Scholar

5. Sanborn TA, Ebrahimi R, Manoukian SV, et al. Impact of femoral vascular closure devices and antithrombotic therapy on access site bleeding in acute coronary syndromes: the Acute Catheterization and Urgent Intervention Triage Strategy (ACUITY) trial. Circ Cardiovasc Interv2010;3:57-62.

Abstract/FREE Full TextGoogle Scholar

6. Stone GW, McLaurin BT, Cox DA, et al. Bivalirudin for patients with acute coronary syndromes. N Engl J Med2006;355:2203-16.

CrossRefPubMedWeb of ScienceGoogle Scholar

7. Stone GW, Witzenbichler B, Guagliumi G, et al. Bivalirudin during primary PCI in acute myocardial infarction. N Engl J Med2008;358:2218-30.

CrossRefPubMedWeb of ScienceGoogle Scholar

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES, INCLUDING THOSE THAT ARE INSTRUMENT-BASED) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	
Provide all other grades and definitions from the evidence grading system	
Grade assigned to the **recommendation** with definition of the grade	
Provide all other grades and definitions from the recommendation grading system	
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	
Estimates of benefit and consistency across studies	
What harms were identified?	
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

1a.4.2 What process was used to identify the evidence?

1a.4.3. Provide the citation(s) for the evidence.

1a.1 <u>For Maintenance of Endorsement:</u> Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

Yes

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Bleeding is the second most common non-cardiac complication of PCI. It is associated with adverse patient outcomes (e.g. increased mortality, prolonged length of stay and costs) and – most importantly – is modifiable through the use of bleeding avoidance strategies such as radial arterial access. Moreover, studies document under-use of bleeding avoidance strategies in high-risk patients. Thus, as an adverse event that varies widely across providers and is modifiable, the use of risk-adjusted bleeding metrics can provide the foundation for quality improvement initiatives that improve the safety and outcomes of treatment.

References:

Levine, G. N., Bates, E. R., Blankenship, J. C., Bailey, S. R., Bittl, J. A., Cercek, B., Chambers, C. E., Ellis, S. G., Guyton, R. A., Hollenberg, S. M., Khot, U. N., Lange, R. A., Mauri, L., Mehran, R., Moussa, I. D., Mukherjee, D., Nallamothu, B. K., Ting, H. H. (2011) 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelinesand the Society for Cardiovascular Angiography and Interventions. Journal of The American College of Cardiology, 58(24), e44-e122.

Vora, A. N., Peterson, E. D., McCoy, L. A., Garratt, K. N., Kutcher, M. A., Marso, S. P., Roe, M. T., Messenger, J. C., & Rao, S. V. (2016). The impact of bleeding avoidance strategies on hospital-level variation in bleeding rates following percutaneous coronary intervention. Journal of the American College of Cardiology: Cardiovascular Interventions, 9(8), 771-779.

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

As requested, we have provided updates on national performance for the risk-standardized bleeding rates. These include the entire calendar years of 2015 and 2016. All requested data are provided below:

2015 Data:

Data range date: 2015QTR1-2015QTR4:

Number of patients: 695,232

Number of PCI procedures per hospital volume:

0-10: 5 hospitals

11-200: 428 hospitals

201-400: 472 hospitals

401-600: 273 hospitals

601-1000: 237 hospitals

1001-2000: 125 hospitals

2001+: 13 hospitals

Deciles of Risk-Standardized Bleeding Rates:

Stddev: .0141 (1.4%)

Quartile 1: .0225 (i.e. a less than 2.25% Bleeding rate in the best performing 25% of hospitals)

Mean: .0320 (the average bleeding rate is 3.2%)

Quartile 3: .0390 (i.e. a greater than 3.9% (more than 50% greater than the best performing 25% of hospitals) Bleeding rate in the worst performing 25% of hospitals)

Deciles of bleeding adjusted rates:

1: 1.8%

2: 2.1%

3: 2.4%

4:2.6%

5(median): 2.9%

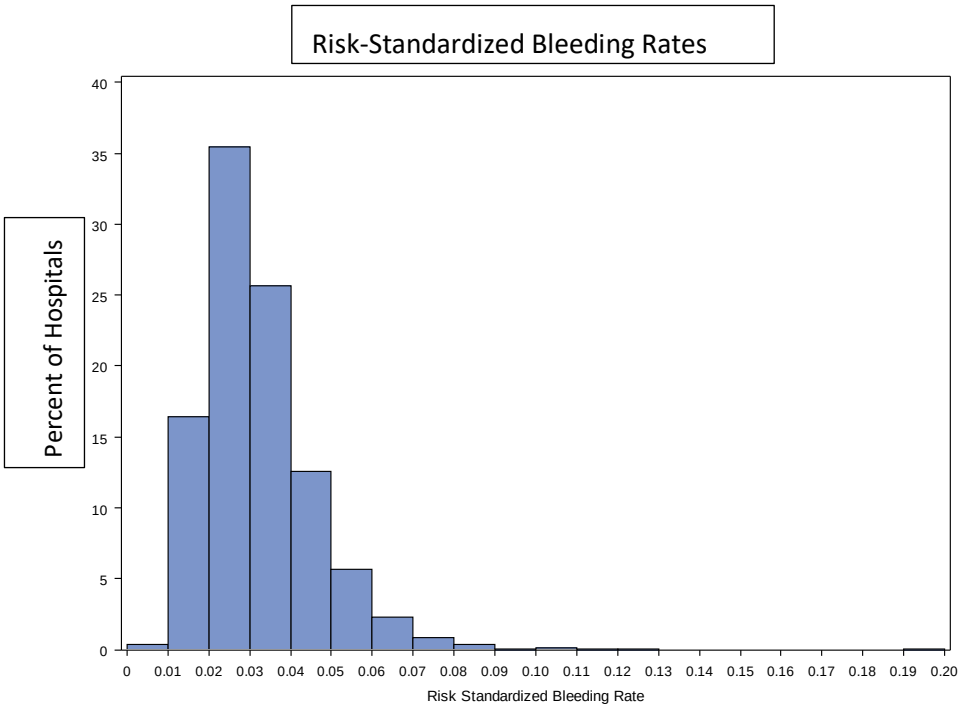
6: 3.3%

7: 3.7%

8: 4.1%

9: 4.9%

The distribution of Risk-Standardized Bleeding Rates across hospitals for 2015 is shown below:



2016 Data:

Data range date: 2016QTR1-2016QTR4:

Number of patients: 717,510

Number of PCI procedures per hospital volume:

0-10: 11 hospitals

11-200: 466 hospitals

201-400: 460 hospitals

401-600: 279 hospitals

601-1000: 253 hospitals

1001-2000: 139 hospitals

2001+: 11 hospitals

Deciles:

Stddev: .0137 (1.3%)

Quartile 1: .0222 (i.e. a less than 2.22% Bleeding rate in the best performing 25% of hospitals)

Mean: .032 (the average bleeding rate is 3.25)

Quartile 3: .0387 (i.e. a greater than 3.9% is the risk-standardized bleeding rate for the worst performing 25% of hospitals)

Deciles of bleeding adjusted rates:

1: 1.7%

2: 2.1%

3: 2.3%

4:2.6%

5(median): 2.9%

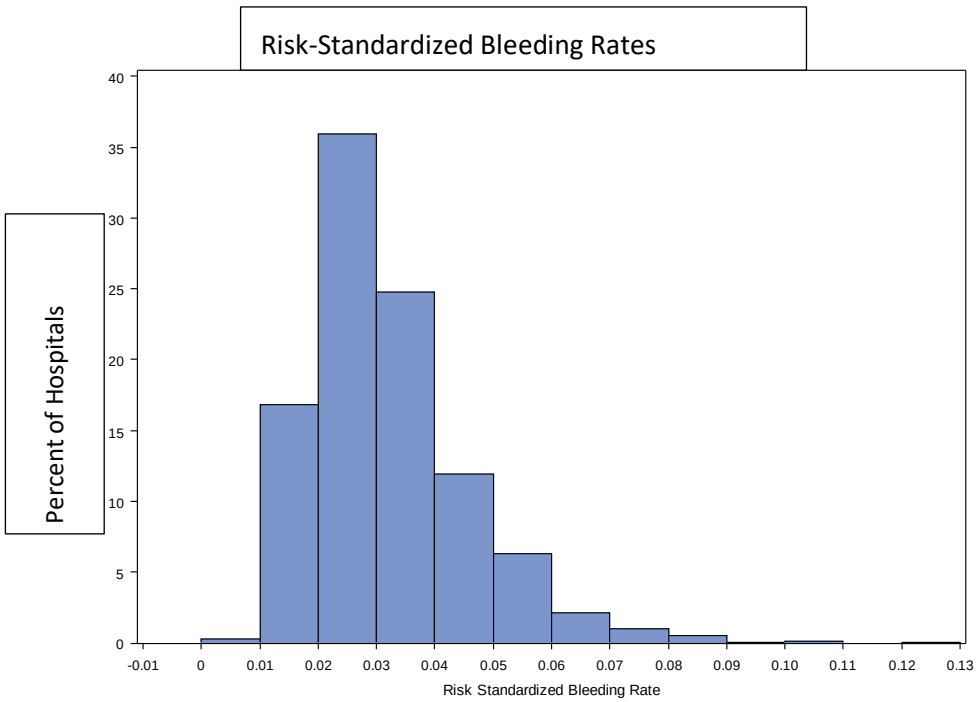
6: 3.2%

7: 3.6%

8: 4.1%

9: 5.0%

The distribution of risk-standardized bleeding across hospitals for 2016 is shown below:



This means that top 10% demonstrated better (lower) than a 1.7% bleeding rate.

This means that top 20% demonstrated better (lower) than a 2.1% bleeding rate.

The observed bleeding events are lower than when the model was first developed, in part, because the previous version of this model used a threshold of hemoglobin drop of 3g/dl to reflect a bleeding event, which was raised to 4g/dl in the current iteration of the model to align with the bleeding definitions used in other NCDR registries. There have also been improvements in performance over time as the use of radial arterial access has increased.

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

As noted above, there is substantial variation across hospitals, ranging from a 1.7% rate in the top performing decile to an almost 3-fold greater rate of 5.0% in the worst performing decile. This suggest an important opportunity for improvement based and decreased variability across hospitals.

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

While we observed some statistically significant differences by gender, race and insurance status. The absolute rates after patient-level adjustment were clinically marginal, except for gender and age which are strong risk factors for bleeding. See testing supplement for details. The model performs equally well in key clinical and socio-demographic populations, as shown in the published table below:

Table 6. c-Indexes of the Full Model and Risk Score Models in the Overall Dataset and in Pre-Specified Subgroups

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 yrs	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

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

The opportunity to improve the safety of PCI by reducing bleeding exists for all patients, not just those of a specific race, gender, age or SES status.

2. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

Cardiovascular, Cardiovascular : Coronary Artery Disease (PCI)

De.6. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

Populations at Risk

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

ACC does not have a measure specific webpage. However more information about the clinical registry that the measure is included in can be found at: https://cvquality.acc.org/NCDR-Home/registries/hospital-registries/cathpci-registry.

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** CathPCI_v4_CodersDictionary_4.4-635230481331385161-635854401108586219.pdf

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. For maintenance of endorsement: Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

Yes

S.3.2. For maintenance of endorsement, please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

The PCI risk-adjusted bleeding model was updated in 2017 to a parsimonious hierarchical model. This means the risk model utilizes less patient variables (than the previous model) to determine individual patient risk of bleeding and the model will take into account the risk relationships within and amongst hospitals (not utilized by the previous model and a hierarchical model feature). Additionally, the hemoglobin parameter used to determine if a Post-PCI bleeding event has occurred, has been revised to assess an absolute hemoglobin (hgb) decrease from pre-PCI to post-PCI of = 4g/dL (previously 3g/dL). The Risk Adjusted Bleeding model provides accurate estimates of post-PCI bleeding risk and is helpful in providing risk-adjusted feedback on bleeding complications, informing clinical decision-making, and directing the use of bleeding avoidance strategies to improve the safety of PCI procedures.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

IF an OUTCOME MEASURE, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

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. Cardiac Tamponade; OR
4. Post-PCI transfusion for patients with a pre-procedure hemoglobin (Hgb) >8 g/dL and pre-procedure Hgb not missing; OR
5. Absolute Hgb decrease from pre-PCI to post-PCI of >= 4 g/dl AND pre-procedure Hgb =<16 g/dL AND pre-procedure Hgb not missing

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

IF an OUTCOME MEASURE, describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The numerator is defined as any patient =18 years of age, with post-PCI bleeding which includes meeting any one of the criteria listed below (as shown below).

1. Bleeding event w/in 72 hours (8050); OR
2. Hemorrhagic stroke (8021); OR
3. Tamponade (8025); OR
4. Post-PCI transfusion (8040) for patients with a pre-procedure hgb >8 g/dL and pre-procedure hgb not missing; OR
5. Absolute hgb decrease (7320 and 7345) from pre-PCI to post-PCI of >= 4 g/dl (excluded if any of the following: pre-procedure (7320) hgb>16g/dl or IABP (5330) = yes or MVSupport (5340) = yes)

Note:

- All data element numbers listed above are included in the attach data dictionary which includes more detailed definitions for the above elements.
- The measure includes risk adjustment to account for differences in case mix across hospitals, thus the ratio determined by the numerator and denominator are modified based upon the adjustment.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

Patients 18 years of age and older with a PCI procedure performed during admission

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The following patients are included in the denominator:

1. Patients 18 years of age or older
2. Patients undergoing PCI during the episode of care
3. Initial PCI procedures for patients who underwent multiple PCI procedures during the episode of care (subsequent PCIs during a single Episode of Care are excluded).

4. Patient with procedures with non-missing values for outcome variables of bleeding event w/in 72 hours (8050) AND transfusion (8040).

Note that all data element numbers listed above are included in the attached data dictionary which includes more detailed definitions for the above elements.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

1. Patients who did not have a PCI (episodes of care with a diagnostic catheterization only);
2. Patients who died on the same day of the procedure
3. Patients who underwent CABG during the episode of care

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

The following patients are excluded from the denominator:

1. Patients who died on the same day of the procedure [Discharge date (9035)=procedure date

(5300) AND discharge status=deceased (9040)]

2. Patients with CABG (9000)=yes

Note that all data element numbers listed above are included in the attached data dictionary which includes more detailed definitions for the above elements.

At the facility level, all data submissions must pass the data quality and completeness reports to be included. Note: For some characteristics, missing values are imputed. In the NCDR data quality program, all key variables in the risk model have a high "inclusion" criteria, meaning that when a hospital submits data, they need to have a high level of completeness (>95%) for those variables. If they are not able to meet the criteria in our data quality program, they do not receive risk-adjusted outcomes for any of the records they submitted for that quarter. Because the high-threshold for inclusion is present, the impact of imputation on hospital-specific rates is minimal, but enables a more complete assessment of hospital performance.

Note that all data element numbers listed above are included in the attach data dictionary which includes more detailed definitions for the above elements.

At the facility level, all data submissions must pass the data quality and completeness reports to be included. Note: If one or two variables are missing, the value is imputed for certain characteristics . In our data quality program, all key variables in the risk model have a high "inclusion" criteria. This means that, when a hospital submits data to us , they need to have a high level of completeness (around 95-99%) for those variables. If they are not able to meet the criteria in our data quality program, they do not receive risk adjusted mortality for the records they submitted for that quarter.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

N/A

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

S.12. Type of score:

Rate/proportion

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

1. Remove hospitals who fail data quality and completeness reports as outlined in the NCDR Data Quality Program (further discussed in the Testing Supplement)
2. Remove hospitals who have do not have at least one patient with a pre-PCI or post-PCI hemoglobin value.
3. Remove patient's subsequent PCIs during the same admission (if the patient had more than one PCI procedure during that episode of care).
4. Remove patients who did not have a PCI (Patient admissions with a diagnostic cath only during that episode of care)
5. Remove patients who died on the same day of the procedure
6. Remove patients who had CABG during the episode of care
7. Remove patients with pre-procedure hemoglobin <8 g/dL patients (severely anemic) who did not also have a documented bleeding event other than transfusion were not counted in the numerator if they received a transfusion.
8. Calculate measure used weight system based on predictive variables as outlined in the accompanying testing documents and supplemental materials.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Registry Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

National Cardiovascular Data Registry CathPCI Registry

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

Available at measure-specific web page URL identified in S.1

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Inpatient/Hospital

If other:

S.22. <u>COMPOSITE Performance Measure</u> - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

2459_testing_form_v7.1_8.1.18_FINAL_-_edits_for_methods_panel_8.16.18_FINAL.docx

2.1 <u>For maintenance of endorsement</u>

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.2 <u>For maintenance of endorsement</u>

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

Yes

2.3 <u>For maintenance of endorsement</u>

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Yes - Updated information is included

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** 2459
Measure Title: In-hospital Risk Adjusted Rate of Bleeding Events for patients undergoing PC
Date of Submission: 08/01/2018

Type of Measure:

☒ **Outcome (*including PRO-PM*)**	☐ **Composite – *STOP – use composite testing form***
☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. *If there is more than one set of data specifications or more than one level of analysis, contact NQF staff* about how to present all the testing information in one form.
- For all measures, sections 1, 2a2, 2b1, 2b2, and 2b4 must be completed.
- For outcome and resource use measures, section 2b3 also must be completed.
- If specified for multiple data sources/sets of specificaitons (e.g., claims and EHRs), section 2b5 also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b1-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 25 pages (*incuding questions/instructions;* minimum font size 11 pt; do not change margins). *Contact NQF staff if more pages are needed.*
- Contact NQF staff regarding questions. Check for resources at Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for version 7.1 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a2. Reliability testing 10 demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing 11 demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument-based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure; 12

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). 13

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions.

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v. $5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (***must be consistent with data sources entered in S.17***)	**Measure Tested with Data From:**
☐ abstracted from paper record	☐ abstracted from paper record
☐ claims	☐ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs
☐ other:	☐ other:

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g.,*

Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

We used the National Cardiovascular Data Registry for CathPCI Registry. This is a national quality improvement registry with more than 1200 participating US hospitals. Participation is largely voluntary though some states and healthcare systems mandate participation. Rigorous quality standards are applied to the data and both quarterly and ad hoc performance reports are generated for participating centers to track and improve their performance.

1.3. What are the dates of the data used in testing?

We have chosen to use different datasets to provide support for different aspects of the proposed measure.

1. Creation of the Bleeding model was performed on all national NCDR data from 02/2008–04/2011 and has been used to provide a description and initial performance characteristics of the model.

2. A validation cohort from the NCDR CathPCI was identified (all cases performed between 01/2016-12/2016). These data were also used to assess test-retest reliability of the risk model covariates and validate the association between the predictor variables and bleeding, including model discrimination and calibration.

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	**Measure Tested at Level of:**
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other:	☐ other:

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Creation of the Bleeding & Validation model:

The model was originally developed using data from 1,142 hospitals. See additional information under section 1.6.

Test-Retest

The 2016 validation sample includes cases from 1,619 hospitals.

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Bleeding & Validation Model:

For the initial derivation and validation of the bleeding risk model, 1,043,759 patients undergoing PCI between 2/2008-4/2011 at 1,142 hospitals were included; 80% were randomly assigned to the derivation cohort with the remaining 20% serving as the validation cohort. Of these, 60,194 PCI procedures had post-procedure bleeding, yielding a post-PCI bleeding event rate of 5.8%. A summary of these patients' clinical characteristics and the hospital characteristics are provided under Table 1:

Table 1. Derivation and Validation Characteristics

	Overall (n=1,043,759)	Development (n=834,696)	Validation (n=209,063)
Demographics			
Median age, y (25^{th}, 75^{th} percentiles)	65.0 (56.0, 74.0)	64.0 (56.0, 74.0)	65.0 (56.0, 74.0)
Female sex	32.7	32.6	32.8
Median BMI, kg/m^2 (25^{th}, 75^{th} percentiles)	29.1 (25.7, 33.3)	29.1 (25.7, 33.3)	29.1 (25.7, 33.3)
Medical conditions			
Diabetes mellitus	35.9	35.9	35.9
Hypertension	81.8	81.8	81.9
Peripheral vascular disease	12.4	12.4	12.4
Chronic kidney disease	3.6	3.6	3.6
Prior PCI	40.3	40.3	40.3
Prior CABG	18.8	18.9	18.7
Median pre-procedure Hgb, g/dl (25^{th}, 75^{th} percentiles)	13.7 (12.4, 14.9)	13.7 (12.4, 14.9)	13.7 (12.4, 14.9)
Procedural characteristics			
Procedure status			
Elective	45.2	45.2	45.1
Urgent	37.5	37.5	37.7
Emergent	17.0	17.0	16.9
Salvage	0.3	0.3	0.3
STEMI	16.0	16.0	15.9
Lytics prior to PCI for STEMI	8.1	8.0	8.2
Shock	2.5	2.5	2.4
Cardiac arrest within 24 hrs of PCI	1.7	1.7	1.7
Hospital characteristics			
Number of beds, median (25^{th}, 75^{th} percentiles)	410.0 (283.0-571.0)	410.0 (283.0-571.0)	409.0 (282.0-569.0)
University hospital (%)	11.3	11.3	11.3
Number of annual PCI cases, median (25^{th}, 75^{th} percentiles)	726.0 (445.1-1177.9)	726.6 (445.1-1183.1)	726.6 (448.0-1177.9)

All p-values >0.05

BMI = body mass index; CABG = coronary artery bypass grafting; Hgb = hemoglobin; PCI = percutaneous coronary intervention; STEMI = ST-segment elevation myocardial infarction

Test-Retest

To test the predictive validity, calibration and test-retest reliability, we used data from 717,510 patients undergoing PCI between 1/2016-12/2016, of whom 23,874 (3.3%) had a bleeding event. A summary of these patients' clinical characteristics are provided in Table 2 below:

Table 2. Predicted Probability of Bleeding (2016)

Bleeding

	Total	Observed Bleed	
	n = 717510	Bleed n = 23874	No Bleed n = 693636
Predicted Bleeding Risk using the proposed risk-adjustment model (multiply by 100 to get % bleeding risk)	0.03707 ± 0.05064	0.10379 ± 0.11061	0.03477 ± 0.04553
Bleeding Variables			
STEMI	121466 (16.9%)	9699 (40.6%)	111767 (16.1%)
Age	65.6 ± 11.9	68.4 ± 12.5	65.5 ± 11.9
BMI	30.2 ± 6.5	28.9 ± 6.8	30.2 ± 6.5
CVD	96012 (13.4%)	4300 (18.0%)	91712 (13.2%)
PVD	88106 (12.3%)	4008 (16.8%)	84098 (12.1%)
CLD	111555 (15.5%)	4956 (20.8%)	106599 (15.4%)
Prior PCI	295177 (41.1%)	7482 (31.3%)	287695 (41.5%)
Insulin Diabetes	116320 (16.2%)	4540 (19.0%)	111780 (16.1%)
Dialysis	21673 (3.0%)	1591 (6.7%)	20082 (2.9%)
GFR 45-60 (Mild)	102839 (14.3%)	4592 (19.2%)	98247 (14.2%)
GFR 30-45 (Moderate)	58134 (8.1%)	4779 (20.0%)	53355 (7.7%)
Lytics Prior to PCI for STEMI	5041 (0.7%)	366 (1.5%)	4675 (0.7%)
Cardiogenic Shock at Start of PCI	16695 (2.3%)	4227 (17.7%)	12468 (1.8%)
Missing	332	8	324
Cardiogenic Shock w/in 24 Hours	14743 (2.1%)	4080 (17.1%)	10663 (1.5%)
Missing	85	2	83
PCI Status			
Elective	251808 (35.1%)	3572 (15.0%)	248236 (35.8%)
Urgent	331775 (46.3%)	9161 (38.4%)	322614 (46.5%)
Emergent	131388 (18.3%)	10331 (43.3%)	121057 (17.5%)
Salvage	2359 (0.3%)	806 (3.4%)	1553 (0.2%)
Missing	180	4	176
In-stent Thrombosis	2420 (0.3%)	245 (1.0%)	2175 (0.3%)
Lesion SCAI Class II/III	352960 (49.2%)	11240 (47.1%)	341720 (49.3%)
Lesion SCAI Class IV	107008 (14.9%)	7304 (30.6%)	99704 (14.4%)
Prox LAD	129513 (18.1%)	5251 (22.0%)	124262 (17.9%)
Left Main	20218 (2.8%)	1780 (7.5%)	18438 (2.7%)
HF NYHA Class IV	24810 (3.5%)	3171 (13.3%)	21639 (3.1%)
HF NYHA Class I/II/III	75463 (10.5%)	3456 (14.5%)	72007 (10.4%)
Cardiac arrest w/in 24 hrs	15060 (2.1%)	3190 (13.4%)	11870 (1.7%)
Lesion: Preprocedure TIMI Flow = NO	141133 (19.7%)	9191 (38.5%)	131942 (19.0%)
Multivessel Disease	297871 (41.5%)	12302 (51.5%)	285569 (41.2%)
PreHGB	13.5 ± 2.0	12.7 ± 2.6	13.5 ± 2.0
Female	223562 (31.2%)	10961 (45.9%)	212601 (30.7%)
History			
IABP	12563 (1.8%)	3207 (13.4%)	9356 (1.3%)

	Total	Observed Bleed	
	n = 717510	Bleed n = 23874	No Bleed n = 693636
Missing	194	4	190
Current/Recent Smoker (w/in 1 year)	182938 (25.5%)	6401 (26.8%)	176537 (25.5%)
Missing	309	16	293
Hypertension	598736 (83.5%)	19532 (81.8%)	579204 (83.5%)
Missing	136	8	128
Dyslipidemia	558972 (78.0%)	16761 (70.3%)	542211 (78.2%)
Missing	449	28	421
Family History of Premature CAD	133606 (18.6%)	3101 (13.0%)	130505 (18.8%)
Missing	330	21	309
Prior MI	218401 (30.4%)	6888 (28.9%)	211513 (30.5%)
Missing	171	9	162
Prior Heart Failure	111292 (15.5%)	5641 (23.6%)	105651 (15.2%)
Missing	206	12	194
Prior Valve Surgery/Procedure	13268 (1.8%)	620 (2.6%)	12648 (1.8%)
Missing	279	14	265
Prior CABG	124219 (17.3%)	3451 (14.5%)	120768 (17.4%)
Missing	79	3	76
Cath Lab Visit			
PCI Indication			
Immediate PCI for STEMI	110008 (15.3%)	8588 (36.0%)	101420 (14.6%)
PCI for STEMI (Unstable, >12 hrs from Sx onset)	7334 (1.0%)	820 (3.4%)	6514 (0.9%)
PCI for STEMI (Stable, >12 hrs from Sx onset)	1556 (0.2%)	103 (0.4%)	1453 (0.2%)
PCI for STEMI (Stable after successful full-dose Thrombolysis)	1556 (0.2%)	43 (0.2%)	1513 (0.2%)
Rescue PCI for STEMI (after failed full-dose lytics)	3060 (0.4%)	261 (1.1%)	2799 (0.4%)
PCI for high risk Non-STEMI or unstable angina	426385 (59.4%)	11019 (46.2%)	415366 (59.9%)
Staged PCI	34899 (4.9%)	744 (3.1%)	34155 (4.9%)
Other	132555 (18.5%)	2295 (9.6%)	130260 (18.8%)
Missing	157	1	156
CAD Presentation			
No symptom, no angina	26571 (3.7%)	755 (3.2%)	25816 (3.7%)
Symptom unlikely to be ischemic	13214 (1.8%)	424 (1.8%)	12790 (1.8%)
Stable angina	91567 (12.8%)	1291 (5.4%)	90276 (13.0%)
Unstable angina	283204 (39.5%)	4932 (20.7%)	278272 (40.1%)
Non-STEMI	181336 (25.3%)	6767 (28.4%)	174569 (25.2%)
ST-Elevation MI (STEMI) or equivalent	121466 (16.9%)	9699 (40.6%)	111767 (16.1%)
Missing	152	6	146
Anginal Classification w/in 2 Weeks			
No symptoms	51063 (7.1%)	2651 (11.1%)	48412 (7.0%)
CCS I	13240 (1.8%)	274 (1.2%)	12966 (1.9%)
CCS II	74955 (10.5%)	1214 (5.1%)	73741 (10.7%)

	Total	Observed Bleed	
	n = 717510	Bleed n = 23874	No Bleed n = 693636
CCS III	275908 (38.5%)	5622 (23.6%)	270286 (39.0%)
CCS IV	300897 (42.0%)	14049 (59.0%)	286848 (41.4%)
Missing	1447	64	1383
Anti-Anginal Medication w/in 2 Weeks	532589 (74.2%)	15753 (66.0%)	516836 (74.5%)
Missing	213	11	202
Heart Failure w/in 2 Weeks	100273 (14.0%)	6627 (27.8%)	93646 (13.5%)
Missing	290	7	283
Cardiomyopathy or Left Ventricular Systolic Dysfunction	91302 (12.7%)	4704 (19.7%)	86598 (12.5%)
Missing	111	2	109
Pre-operative Evaluation Before Non-Cardiac Surgery	13398 (1.9%)	386 (1.6%)	13012 (1.9%)
Missing	107	3	104
Pre-PCI Left Ventricular Ejection Fraction	52.0 ± 12.8	45.5 ± 15.7	52.2 ± 12.6
Missing	216496	9862	206634
Procedure Information			
Contrast Volume	179.9 ± 81.4	198.7 ± 99.2	179.2 ± 80.6
Missing	3382	149	3233
Fluoroscopy Time	15.8 ± 12.4	21.4 ± 18.8	15.6 ± 12.0
Missing	15939	644	15295
Outcomes			
Discharge Status			
Alive	707874 (98.7%)	20522 (86.0%)	687352 (99.1%)
Deceased	9636 (1.3%)	3352 (14.0%)	6284 (0.9%)
Primary Cause of Death			
Cardiac	6948 (72.2%)	2323 (69.5%)	4625 (73.7%)
Neurologic	642 (6.7%)	241 (7.2%)	401 (6.4%)
Renal	79 (0.8%)	23 (0.7%)	56 (0.9%)
Vascular	93 (1.0%)	61 (1.8%)	32 (0.5%)
Infection	274 (2.8%)	122 (3.6%)	152 (2.4%)
Valvular	206 (2.1%)	73 (2.2%)	133 (2.1%)
Pulmonary	535 (5.6%)	149 (4.5%)	386 (6.2%)
Unknown	366 (3.8%)	122 (3.6%)	244 (3.9%)
Other	477 (5.0%)	230 (6.9%)	247 (3.9%)
Missing	16	8	8
Myocardial Infarction (Biomarker Positive)	10766 (1.5%)	1016 (4.3%)	9750 (1.4%)
Missing	26	6	20
Cardiogenic Shock	9697 (1.4%)	3414 (14.3%)	6283 (0.9%)
Missing	7	1	6
Heart Failure	10082 (1.4%)	2304 (9.7%)	7778 (1.1%)
Missing	8	2	6
CVA/Stroke	2164 (0.3%)	759 (3.2%)	1405 (0.2%)

	Total	Observed Bleed	
	n = 717510	Bleed n = 23874	No Bleed n = 693636
Missing	11	4	7
Other Vascular Complications Requiring Treatment	2285 (0.3%)	1143 (4.8%)	1142 (0.2%)
Missing	14	1	13
RBC/Whole Blood Transfusion	12641 (1.8%)	11297 (47.3%)	1344 (0.2%)

Continuous variables compared using Student's T-test.

Categorical variables compared using chi-square or Fisher's exact test.

All p-values were <0.001

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

As noted above, the ACCF used different data sources for different analyses. The original model was developed and validated using data from 1,043,759 patients undergoing PCI between 2/2008-4/2011 at 1,142 hospitals.

A reassessment of model performance was performed using all PCI patients enrolled (n=715,510) in the NCDR CathPCI registry in calendar year 2016. Separately, we identified 42,637patients who underwent 2 PCIs within the 2016 calendar year in whom we were able to assess the test-retest reliability of the data elements used to predict patients' bleeding risks.

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Social risk factors were not used in this risk model for the following reasons. First, as a clincial registry used for quality assessment and improvement, detailed socioeconomic variables are not available. Second, while proxy variables could be considered, these were not felt to be relevant to an inpatient bleeding model, in contrast to a longer-term outcome model where difficulties with access to care, affording medications or cardiac rehabilitation would be more important. Moreover, while it may be true that worse social risk factors might be associated with more severe illness at the time of presentation, we had direct access to detailed clinical variables describing the severity of illness and feel that incorporating such factors (e.g. clinical indication for PCI, Hb, etc.) is a much more accurate means of stratifying risk. Accordingly, we feel that in this model of in-hospital risk-adjusted bleeding rate, given the rich clinical data available through the NCDR CathPCI registry, that social risk factors, which are not readily available, would not likely improve this particular risk model.

2a2. RELIABILITY TESTING

Note*: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☒ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

Performance Measure Score (Signal-to-Noise):

ACCF performed the signal-to-noise analysis on the same cohort of individuals as noted under Section 1.3. (testing method 2). For the signal-to-noise analysis, we followed the methodology as outlined in a Rand Corporation technical report by John L Adams. The document is available at the following URL (https://www.rand.org/content/dam/rand/pubs/technical_reports/2009/RAND_TR653.pdf). This approach uses a beta-binomial model that assumes the physician's score is a binomial random variable conditional on the physician's true value that comes from a beta distribution. The beta distribution is a very flexible distribution on the interval from 0 to 1 and can have any mean within the interval and can be skewed left or right or even U-shaped. It is the most common distribution for probabilities on the 0-1 interval. Signal to Noise analysis for the hospitals participating in 2016 are provided in Table 3. The author used a beta-binomial model, specifically the Betabin SAS macro to output the required parameters in the reliability formula provided.

Data Element (Test-Retest Reliability):

ACCF evaluated the test-retest reliability by reviewing CathPCI patients who were readmitted or had a repeat procedure in 2016. This approach enabled us to examine 2 independent abstractions of data for the same patient. For certain characteristics that would not change (e.g. gender), we would expect near perfect reproducibility. For other characteristics (e.g. diabetes) we would expect that any patient diagnosed with diabetes on the first visit should also have diabetes recorded on the second visit. It is, however, clinically plausible that someone could be diagnosed with diabetes between their first and second visit, so the emergence of diabetes on the second visit is not necessarily an 'error' and no interpretation is made for these scenarios.

Signal to Noise Analysis:

Signal to Noise analysis for the hospitals participating in 2016 are provided in Table 4.

Table 3. Signal to Noise Analysis

Level	Signal-to-Noise
All Procedures	.743
>Q1 (>185 Procedures)	.706
>Q2 (>360 Procedures)	.760
>Q3 (>628 Procedures)	.819
>Average (>470 Procedures)	.791

Assessment of test-retest reliability among patients undergoing 2 procedures within 2016:

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 18 of 42,637 (0.06%) patients having different genders on the 2 procedures.

Age as assessed by Date of Birth was identical in 99.90% of the 42,637 patients on both assessments.

Cerebrovascular disease (CVD) revealed that only 1213 patients had evidence of CVD on the initial visit that was not noted on the second visit. This represents a 2.84% misclassification rate for one of the assessments.

Peripheral Vascular Disease (PVD) revealed that only 1282 (3.0%) 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. This represents a 3.0% misclassification rate for one of the assessments.

Chronic Lung Disease (CLD) was recorded in 1,294 (3.0%) 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 42,637 patients. 1259 (2.95%) were not classified as having had a prior PCI.

Diabetes was not recorded among 745(1.75%) 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 test-retest reliability of the following could not be 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 fibrinolysis 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.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Signal to Noise Analysis:

The signal to noise ratio analysis measures the confidence levels in differentiating performance between hospitals. These numbers demonstrate variability that is attributable to real differences in hospital quality as opposed to measurement error.

Assessment of test-retest reliability among patients undergoing 2 procedures within 2016:

Finding no clear misclassification by test-retest reliability for any assessable risk factor being >3.0% provides strong support for the test-retest reliability of the bleeding risk factors assessed.

Collectively, we believe that the test-retest reliability data and signal to noise analysis strongly support the reliability of the data elements and measurement scores used in the model.

_____

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☐ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of performance measure score as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

We performed 2 different strategies for assessing the validity of this measure. First, we underwent a rigorous process for establishing the face validity of the measure. Because it is a clinically meaningful outcome, we sought to make sure that a broad range of experts and clinicians concurred that this was a clinically important outcome measure. Second, we hypothesized that it would be associated with other clinically important outcomes and sought to establish the predictive validity of the measure. These are described in more detail below:

Systematic Assessment of Face Validity of the Performance Measure:

Bleeding remains one of the most common non-cardiac complications of PCI. It is a serious adverse consequence and, most importantly, is modifiable. The 2011 ACC/AHA guidelines provide for a Level IC

recommendation for the assessment of bleeding prior to PCI. This is grounded in the realization that there are several strategies, such as radial approaches and the use of bivalirudin, that can be applied to mitigate the risk of bleeding, particularly in high-risk patients. The first bleeding risk model was published in 2009 (*Circ Cardiovasc Intervent*. 2009;2:222-229) and was the update was published in 2013 (*JACC Cardiovasc Intervention, 2013;6:897-904).*

Content validity of this outcome – and the specific definition used in defining a bleeding event – was achieved by the specialized expertise of those individuals who developed this model 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 risk model were leaders and experts in the field of interventional cardiology. Multiple conference calls were held to both define a bleeding event and to examine and vet the risk model. These individuals within specific committees and workgroups are noted below:

NCDR Science and Quality Oversight Committee— an ACC leadership oversight committee that serves as the primary resource for crosscutting scientific and quality of care methodological issues – ensured the data dictionaries and metrics are consistent across registries. They also reviewed and approved the methodology and results of the bleeding outcome and model.

These members include

John C. Messenger, MD, FACC (Chair); David M. Shahian, MD, FACC; Thomas T Tsai, MD, MSC; Charles A. Henrikson, MD, MPH; Jeff Jacobs, MD, FACC; John R. Windle, MD, FACC; Amit Amin, MD; John W. M. Moore, MD, FACC; Deepak L. Bhatt, MD, MPH, FACC; Jeffrey Westcott, MD, FACC; Gregory M. Marcus, MD FACC; David J. Slotwiner, MD, FACC; Jeptha P. Curtis, MD, FACC; John Spertus, MD, FACC; Matthew T. Roe, MD, FACC; and Frederick A. Masoudi, MD, MSPH, FACC

NCDR Clinical SubWorkgroup was a designated workgroup that oversaw the initial NQF application. Prior to submission, the group ensured there was variation in care, disparities data, and that the measure is a true reflection of quality care at a particular site and can also be used to improve quality.

Dr. Jeptha Curtis (chair), Dr. Frederick Masoudi, Dr. John Rumsfeld, Dr. David Malenka, and Dr. Issam Moussa.

NCDR Registry Steering Committee provided strategic direction for the Registry and ensures the measures submitted to NQF met key criterion such as reliability, feasibility, and that there is compelling evidence base behind the development and implementation of this measure.

Dr. Issam D. Moussa (chair), Dr. Kirk N. Garratt, Dr. Lloyd W. Klein, Dr. Kendrick A. Shunk, Dr. Samir R. Kapadia, Dr. Robert N. Piana, Dr. Roxana Mehran, Dr. Frederic S. Resnic, Dr. Aaron D. Kugelmass,

Dr. Sunil V. Rao, Dr. W. Douglas Weaver, and Dr. John C. Messenger.

The NCDR Metrics and Reporting Methodology (MRM) Subcommittee of the Science and Quality Oversight Committee, reviews for re-endorsement and a data analytic center is involved in evaluating data, providing corresponding analysis/interpretation of data. The review includes guidance and oversight from both NCDR's Chief Science Officer (Frederick Masoudi) and chair of MRM (Jeptha Curtis).

Lastly the 16 member NCDR Management Board and 31member ACCF Board of Trustees approved these measures for submission to NQF.

In addition, the NCDR provides an open comment period (typically between 15 and 30 days) for: 1) all registry data set version changes, 2) new registry version measures and 3) significant changes/additions to registry version metrics/measures, including risk models and appropriate use criteria. The open comment period engages key registry shareholders (i.e., physicians and clinical care team members and hospital or practice representatives) as well as other external stakeholders (i.e., hospitals, physicians, payers, regulators, consumers, purchasers, etc.) Comments submitted are considered for modification of the version change. NCDR staff and members involved in developing the measures and reports receive all the comments submitted including the name of the individual and organization submitting comment. The NCDR determines which comments to incorporate into modifications and the internal timeline for any modifications. No formal response is provided back to individuals submitting comments through this process. The NCDR may choose to

provide a report of comments received and decisions made regarding the various feedback to a broader audience.

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). More recently, a 9-center study of providing pre-procedural bleeding risks demonstrated a fully-adjusted 44% lower odds of bleeding when the models were used (*BMJ*, 2015;350:h1302). The ultimate 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.

Predictive Validity:

To further underscore the importance of the bleeding measure, we examined the association of bleeding rates, by quintiles, with other clinically important outcomes, including mortality, complications of heart failure and stroke, length of stay and rates of same-day discharge. We hypothesized that patients experiencing a bleeding complication would also be at higher risk for longer post-procedure lengths of stay (because additional observation and treatment, such as transfusions and surgical repairs, would be needed to address the bleeding complication. Other complications that we hypothesized to be associated with bleeding events would be a greater risk for other complications, including stroke, heart failure and mortality. We postulated that the anemia and hypotension associated with severe bleeds could put patients at increased risk for stroke and death, while the fluid resuscitation and transfusions used to treat a bleeding event might be associated with heart failure exacerbations during the hospitalization. Because it is a hospital-based measure, we examined the risks of these other adverse outcomes across quintiles of bleeding rates throughout the NCDR registry. For the importance/predictive validity of this measure, we found important associations with all outcomes:

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

Predictive Validity:

	Total	**Risk Standardized Bleeding Rate**					**P-value**
	n = 717510	**Quintile 1 (0.0053980000 to 0.0210688560) n = 123567**	**Quintile 2 (0.0210688561 to 0.0259987236) n = 122031**	**Quintile 3 (0.0259987237 to 0.0324864439) n = 138056**	**Quintile 4 (0.0324864440 to 0.0414904733) n = 170144**	**Quintile 5 (0.0414904734 to 0.1278680000) n = 16372**	
Observed Bleed	23874 (3.3%)	1428 (1.2%)	2584 (2.1%)	4064 (2.9%)	6439 (3.8%)	9359 (5.7%)	<0.001
Discharge status							
Alive	707874 (98.7%)	122189 (98.9%)	120543 (98.8%)	136207 (98.7%)	167782 (98.6%)	161153 (98.4%)	<0.001
Deceased	9636 (1.3%)	1378 (1.1%)	1488 (1.2%)	1849 (1.3%)	2362 (1.4%)	2559 (1.6%)	<0.001
Los	3.0 ± 12.4	2.8 ± 12.9	2.8 ± 10.3	3.0 ± 11.9	3.1 ± 14.5	3.2 ± 11.4	<0.001
Heart Failure	10082 (1.4%)	874 (0.7%)	1283 (1.1%)	1805 (1.3%)	2470 (1.5%)	3650 (2.2%)	<0.001
Missing	8		3		4	1	
CVA/Stroke	2164 (0.3%)	232 (0.2%)	296 (0.2%)	418 (0.3%)	571 (0.3%)	647 (0.4%)	<0.001
Missing	11	1	3	2	3	2	
sameday_dc	65676 (11.0%)	12778 (12.6%)	12427 (12.3%)	12501 (11.0%)	14697 (10.3%)	13273 (9.7%)	<0.001
Missing	121466	21996	21349	23921	27170	27030	

Continuous variables compared using one-way analysis of variance.

Categorical variables compared using chi-square or Fisher's exact test.

The above table shows quintiles of performance based on hospitals' risk standardized bleeding rate from best performing (i.e. Quintile #1 – bleeding rate range of 0.5-2.1%) to worst performing hospitals (i.e. Quintile #2 – bleeding rate range of 4.1 to 12.8%) compared to other adverse event measures. For example, if mortality observed in the best performing hospitals had been observed in the worst performing hospitals, then 1826 deaths would have been averted. Similar patterns were observed for the complications of heart failure (3-fold increase in risk of heart failure from worst to best performing hospitals) and stroke (a doubling of risk from the hospitals with the lowest vs. the highest rates of bleeding). There was also evidence of more efficient care, with an observed LOS of 2.8 vs. 3.2 days and a same-day discharge rate of 12.6 vs. 9.7% between the best and worst performing hospitals. All of these associations support our hypotheses that bleeding, and its treatment, would be associated with other clinically important outcomes and support the predictive validity of the proposed bleeding measure.

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

Face-Validity:

As described above (2b1.2), we undertook an extensive effort to establish the definition and utility of risk-adjusted bleeding as a quality metric. These included an expert team developing the model, a group of experts, the Strategic Oversight Committee, overseeing the work and reporting of the measure – including ascertaining its alignment with both ACC/AHA PCI Guidelines and the Society of Coronary Angiography and Intervention's (SCAI's) 2016 Expert Consensus Statement – and an NCDR Oversight Group for NQF measures. It further underwent public comment and approval by the NCDR Management Board of the ACC's Board of Trustees. Beyond these traditional ascertainments of its face validity, we further leveraged evidence that the prospective use of the model was associated with a substantial reduction in bleeding after PCI, clearly demonstrating the model to serve as a means for improving the safety of PCI.

Predictive Validity:

The predictive validity of risk-adjusted bleeding being associated with mortality, post-PCI complications (stroke and heart failure exacerbations) and length of stay strongly underscores the importance of this adverse event and supported the hypothesized associations in conducting these analyses. Moreover, the actionability of the bleeding model suggests that bleeding rates can be improved by prospectively using this risk-adjustment model (*BMJ*, 2015;350:h1302). Given the broad range of bleeding outcomes in the US (range of unadjusted peri-procedural bleeding in 2016 across hospitals = 0-13%, with the adjusted rate 10th to 90th percentiles of hospitals bleeding rates = 1.7-5.0%) the use of this model to assess quality and inspire improvement is a critical step towards greater patient safety and outcomes. The fact that we have demonstrated that the model, when employed at the hospital level, can reduce the variation to those factors most under the locus of control of the operators/hospitals and that by providing pre-procedural risk estimates to providers we can improve the rational use of bleeding avoidance therapies and lower bleeding strongly support the validity of this performance measure not only in risk-adjusting bleeding, but also in improving care.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section* 2b3

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

The only exclusions from the bleeding model are patients undergoing CABG surgery, those who present to the hospital severely anemic (pre-procedure hemoglobin <8 g/dL) and do not have an obvious clinical bleed after their procedure, and patients who died during hospitalization (i.e. same day as their PCI 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. Lastly, patients who died at hospitalization would be captured under NQF measure 0133 which ACC believes complements this measure to ensure good care for patients who undergoing PCI.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

In 2016:

- CABG during the index hospitalization was performed in 8,137 (1/1%) patients. Beyond the small size of this excluded group, bleeding after a major operation is much more likely related to the operation than the preceding PCI.
- In 2016, 3,534 (0.47%) patients were excluded because they died the same day as their PCI procedure. They were excluded because they were not alive long enough to assess whether or not a bleeding event had occurred.
- There were 1,816 (0.25%) procedures in 2016 for which patients both had pre-procedural anemia (hgb<=8) and a transfusion. Of these, 1,344 (0.18%) had no bleeding evidence, and 472 had a bleed that was counted in the numerator of the bleeding measure. We do not believe that such a small rate for this exclusion would meaningfully impact the measure.
- In 2016, there were 22,406 procedures were excluded for being a second, non-index. Because it was not possible to separate which procedure, when multiple were performed in the same admission, was responsible for the bleeding event, the measure could be interpreted as bleeding events per admission in which PCI is performed, which is more clinically interpretable,

2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

We do not believe that the exclusions have any impact on the validity, accuracy or interpretability of the risk-adjusted bleeding outcome measure.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with** 32 **risk factors**

☐ **Stratification by risk categories**

☐ **Other,**

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

A hierarchical logistic regression model was created. The data definitions are available on the NCDR website (https://cvquality.acc.org/NCDR-Home/registries/hospital-registries/cathpci-registry). The beta coefficients and covariance matrix are available from NCDR upon request.

Variable	OR (95% CI)
Intercept	
STEMI	1.291 (1.269,1.314)
Age (<=70)	1.014 (1.014,1.015)
Age (>70)	1.01 (1.009,1.011)
BMI (<=30)	0.968 (0.966,0.969)
CVD	1.126 (1.112,1.141)
PVD	1.2 (1.184,1.216)
CLD	1.225 (1.21,1.24)
Prior PCI	0.767 (0.759,0.775)
Diabetes: Insulin	1.008 (0.994,1.022)
Dialysis	1.572 (1.537,1.608)
Mild CKD	1.308 (1.292,1.325)
Moderate CKD	1.66 (1.637,1.683)
Lytics	1.292 (1.246,1.34)
Shock at PCI Start	4.795 (4.58,5.02)
Shock at PCI start AND Salvage	6.319 (6.161,6.48
Shock w/in 24 hours	4.784 (4.659,4.911)
Emergent PCI	2.839 (2.783,2.897)
Urgent PCI	1.435 (1.416,1.454)
Thrombosis	1.565 (1.506,1.626)
SCAI II/III	1.203 (1.189,1.217)
SCAI IV	1.334 (1.308,1.36)
Prox LAD PCI	1.133 (1.12,1.416)
LM PCI	1.669 (1.633,1.707)
NYHA class IV	1.623 (1.594,1.653)
NYHA CLASS I/II/III	1.192 (1.174,1.21)
Prior Arrest	1.918 (1.881,1.956)
Pre Timi=0	1.248 (1.228,1.269)
Mult Vessel Disease	1.241 (1.229,1.253)
HGB (<=13)	0.724 (0.721,0.727)
HGB (>13)	1.115 (1.11,1.12)
Female	1.724 (1.707,1.741)

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g., potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care*) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

As described in Section 1.8, we did not believe that social factors needed to be included in risk-adjusting outcomes for peri-procedural bleeding after PCI. This was predicated on the feasibility (and current unavailability) of patient-level social factors. The belief that the consequence of adverse social factors (e.g. leading to greater rates of obesity, hypertension, smoking or other comorbidities) would be directly captured by our rich clinical data, and that the short duration of follow-up (72 hours, during which the patient was hospitalized), would negate potential barriers to healthcare access and treatment that might be more relevant with longer-term outcomes. Accordingly, we feel that in this model of in-hospital risk-adjusted bleeding rate, given the rich clinical data available through the NCDR CathPCI registry, that social risk factors, which are not readily available, would not likely improve this particular risk model.

As described in Section 2b.1.2, 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% of admissions. 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. A logistic regression model with backward selection and a retention criterion of $p<0.05$ was performed to develop the full risk model used for hospital comparisons. Of note, a more parsimonious model for clinical use was also developed by only using those variables with the strongest association (F-statistic >500) To further simplify prospective application of the simplified model the regression coefficients from the pre-procedure model were assigned 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 their bleeding rates and increase the safety of their PCI performance.

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).

In 2017, NCDR updated its reporting to sites using hierarchical models and one component of the outcome (changing the threshold of a bleeding event in the absence of overt bleeding from a Hb drop of 3g/dl to 4g/dl to align with the bleeding definition in other registries, such as the ACTION AMI registry). The same predictor variables from the published model were used although the beta weights and intercepts were inappropriately updated. Furthermore, the performance characteristics of the model was confirmed. The extensive data provided in this submission, all run with the new model and bleeding definition, justifies the updated model.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☐ **Published literature**

☐ **Internal data analysis**

☐ **Other (please describe)**

2b3.4a. What were the statistical results of the analyses used to select risk factors?

As described above, bivariate analyses were done to identify candidate variables that differed significantly between those with and without a clinically important bleeding event. Multivariable, hierarchical logistic regression analyses were then performed to retain those with a statistically significant association with bleeding ($p<0.05$ for each). Table 2 in Section 1.6 demonstrates the difference between those with and without bleeding events, based upon 2016 data.

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

As noted in Section 1.8 above, social risk factors are not included in this clinically-focused measure.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

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 2016 (see above). 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 2016 data.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

The c-statistic is 0.78 for the original model, 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:

Table 4. Derivation and Validation C-Statistic

	N		Full Model		Risk Score	
Group	**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 2016 data, the c-statistic was 0.79, slightly higher than that observed in the original data. Comparable performance was observed across all socio-demographic and clinical subsets, as shown below:

Group	Sample Size	C-statistic
Overall	717,510	0.790
STEMI	121,466	0.733
Women	223,562	0.742
Age >70 yrs	252,040	0.767
Diabetes	286,743	0.793
Caucasian	617,123	0.788
Non-Caucasian	100,387	0.796
No-Insurance	32,270	0.781

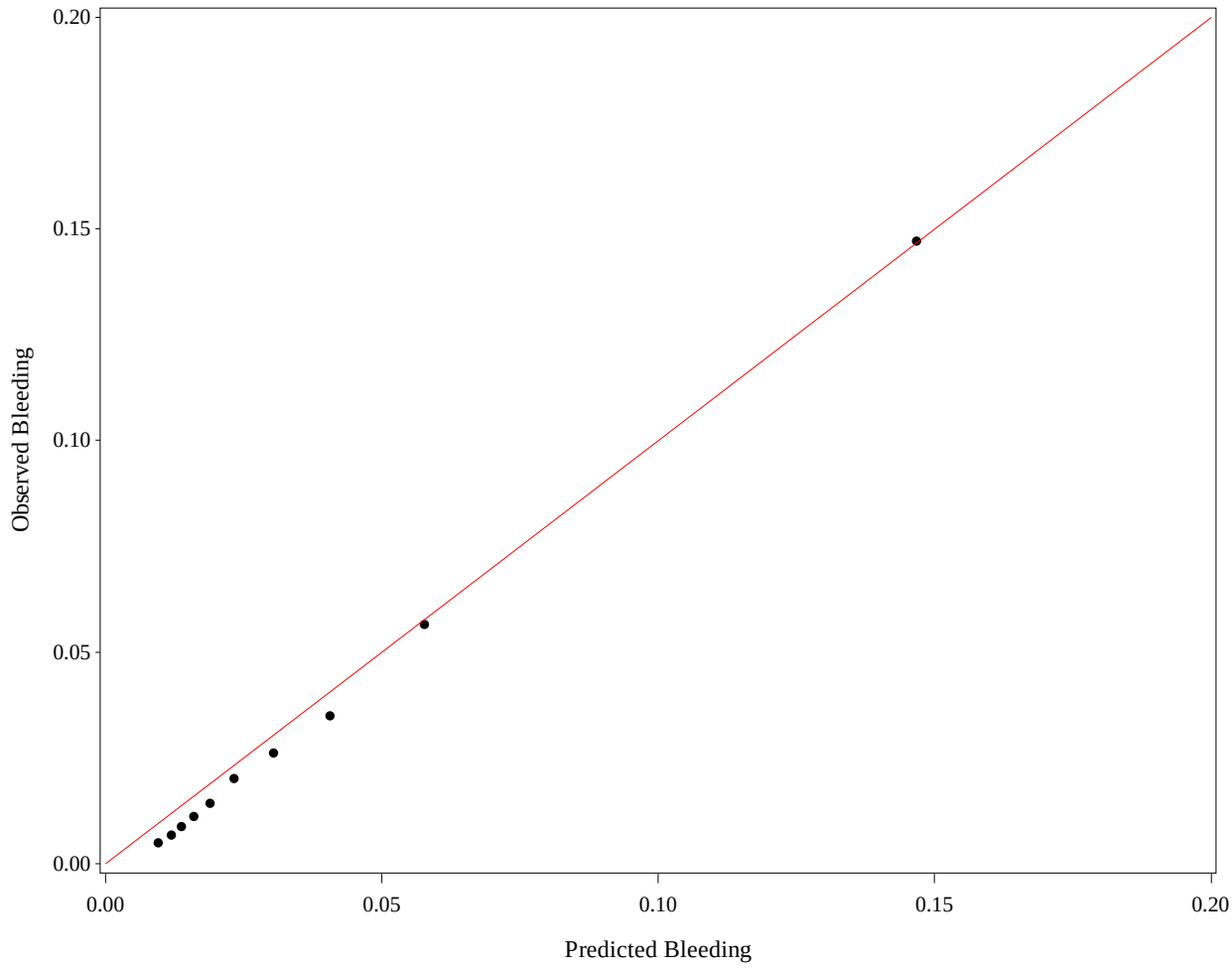
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

Before recalibrating the model to the 2016 data the slope of the calibration line was 1.059 (p<.001) indicating that the relationship between the independent variables in our model and the bleeding outcome slightly overpredicted bleeding in the lower risk patients, with a perfect slope being 1, and the intercept of the line was -0.1265 (p<.0001) indicating that the bleeding rate has decreased since the model was developed.

Due to the decreased bleeding rate from model development we recalibrated the model to the 2016 rates and obtained a slope and intercept of 1 and 0 respectively. See Figure 2.

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Figure 1. Calibration Curve Plot 2016



2b3.9. Results of Risk Stratification Analysis:

For 2016 data, the risk stratification adequately segregated 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 partly mitigated after adjusting for patient characteristics. The unadjusted distribution of bleeding is shown under Figure 3.

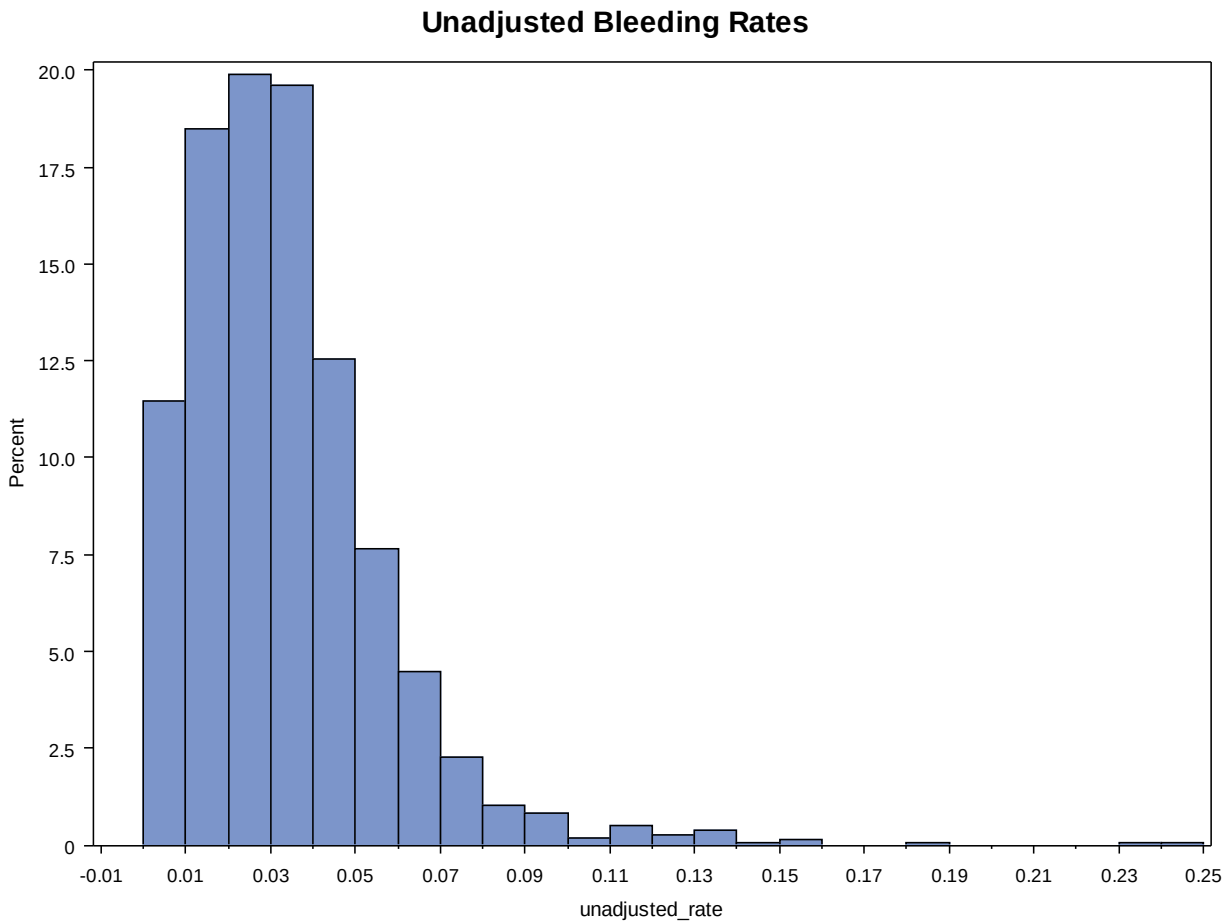


Figure 2. Unadjusted Distribution of Bleeding

The bleeding rates adjusted for patient characteristics is shown under Figure 4.

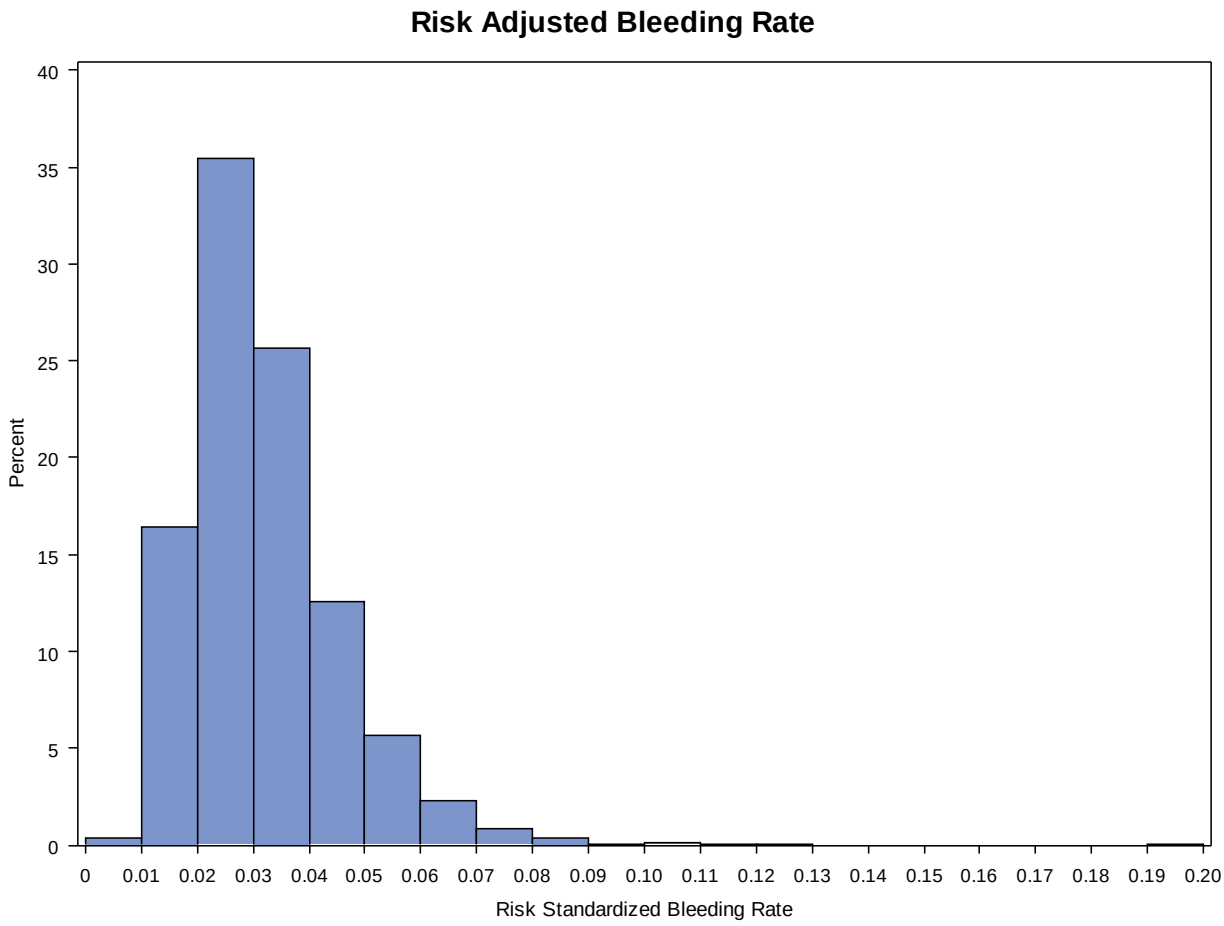
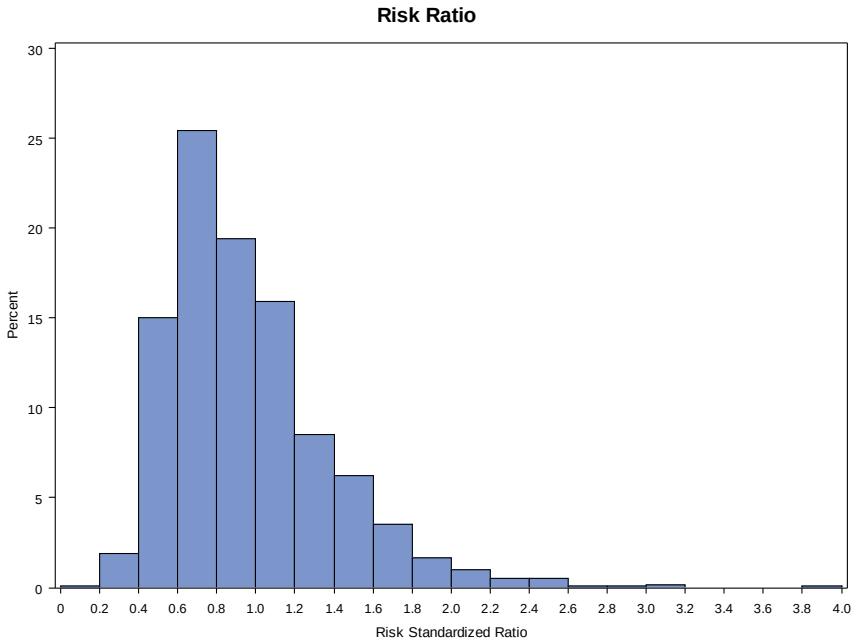


Figure 3. Adjusted Distribution of Bleeding

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

The distribution of sites' observed/expected ratios are shown under Figure 5.



2b3.10. 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 this model performs very well, accounting for patient characteristics present prior to the conduct of PCI and discriminating within important clinical subsets of patients. Moreover, there is substantial hospital variation before and after risk-adjustment. The distribution of institutional O/E ratios identifies 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.

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

We have provided extensive data about the model's performance in much more recent data from which the model was originally developed, further supporting its robustness.

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

As noted in the figures above, we found significant variability in bleeding rates across hospitals after adjusting for pre-procedural patient characteristics. Moreover, hospital performance on this measure is closely associated with risks for death, other complications and length of stay (Section 2b1.3).

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

A meaningful difference identifies the potential for improvement in comparison to others. Since all bleeding events are adverse outcomes, there are no absolute levels of bleeding risk that are significant as compared with others. To place the potential benefits of this model in context, it is helpful to compare the excess bleeds that could be avoided if the worst 25% of hospitals would have had the average bleeding rate of all hospitals. The average, adjusted bleeding rate was 3.2% and the upper quartile ranges from 3.9 to 13% bleeding rates. Given an average PCI volume of 410 cases/hospital, this suggests between 3 and 40 additional and potentially avoidable bleeding events per year among hospitals in the upper quartile as compared with the average hospital. This would be far larger if the worst performing hospitals were to achieve top-decile performance of a 1.8% bleeding rate. Clinically, these are a large number of excess events, particularly given that there are readily applied interventions, such as radial access, bivalirudin or the use of closure devices for femoral access, to mitigate bleeding.

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i.*e., what do the results mean in terms of statistical and meaningful differences?*)

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.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)
N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)
As noted above, there is minimal missing data due to the NCDR CathPCI submission requirements; missing data are imputed to include all cases in estimating performance.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Missing data used to derive the bleeding model are minimal. The most frequent missing variables are pre-procedure hemoglobin (missing in 4.1%) and pre-procedure creatinine (missing rate in 3.6%). Both of these were imputed as medians according to patient sex and MI type. The other variables have missing rates under 0.5% and were imputed using median or most frequent category.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

Given the low rates of missing data, we do not believe that the observed performance is systematically biased. Our efforts to impute data had little effect on site's performance on this measure.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims), Abstracted from a record by someone other than person obtaining original information (e.g., chart abstraction for quality measure or registry)

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in electronic clinical data (e.g., clinical registry, nursing home MDS, home health OASIS)

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

There were no difficulties noted with regard to data collection, availability of data, missing data, the frequency of data collection, patient confidentiality, time and cost of data collection, or other feasibility/implementation issues. In addition, the NCDR has a robust data collection process as outlined below.

Availability:

Participating hospitals report patient demographics, medical history, risk factors, hospital presentation, initial cardiac status, procedural details, medications, laboratory values and in-hospital outcomes. The majority of the 32 required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

Sampling:

There is no sampling of patient data allowed within the contractual terms of participation in the CathPCI Registry in NCDR. The registry is designed to include 100 percent of consecutive adult patients who undergo PCI at participating institutions. Section 2.b of the NCDR Master Agreement with participants includes 'Participant Responsibilities': "b. Use of ACCF Data Set and ACCF-Approved Software. Participant will submit a data record on each patient who receives medical care and who is eligible for inclusion in the Registries in which Participant is participating under this Agreement." Adult patients, ages 18 years and older, who undergo a diagnostic cardiac catheterization and/or PCI. Eligible diagnostic catheterizations are characterized by the passage of a catheter into the aortic root for pressure measurements and/or angiography, and can include Left Ventricle (LV) pressure measurements, LV angiography, coronary angiography, and coronary artery bypass angiography. Eligible PCI procedures include those that involve passage or attempted passage of a coronary device across one or more coronary lesions for purposes of increasing the intraluminal diameter of the vessel and/or restoring or improving circulation. Patients are selected for inclusion by reviewing existing medical records and no direct interaction with the patient will be required outside of the normal course of care. There will be no discrimination or bias with respect to inclusion on the basis of sex, race, or religion.

Patient confidentiality:

Patient confidentiality is preserved as the data are in aggregate form. The CathPCI Registry dataset, comprised of approximately 263, data elements was created by a panel of experts using available ACC-AHA guidelines, data elements and definitions, and other evidentiary sources. Private health information (PHI), such as social security number, is collected. The intent for collection of PHI is to allow for registry interoperability and the potential for future generation of patient-level drill downs in Quality and Outcomes Reports. Registry sites can opt out of transmitting direct identifiers to the NCDR, however, so inclusion of direct identifiers in the registry is at the discretion of the registry participants themselves. When using the NCDR web-based data collection tool, direct identifiers are entered but a partition between the data collection process and the data warehouse maintains the direct identifiers separate from the analysis datasets. The minimum level of PHI transmitted to the ACCF when a participant opts out of submitting direct identifiers meets the definition of a Limited Dataset as such term is defined by the Health Insurance Portability and Accountability Act of 1996.

Data collection within the NCDR conforms to laws regarding protected health information. Patient confidentiality is of utmost concern with all metrics. The proposed measure does not include a patient survey. Physician and/or institutional confidentiality is maintained by de-identified dashboard reports. There is no added procedural risk to patients through involvement in the CathPCI Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed. The primary risk associated with this measure is the potential for a breach of patient confidentiality. The ACCF has established a robust plan for ensuring appropriate and commercially reasonable physical, technical, and administrative safeguards are in place to mitigate such risks.

Data are maintained on secure servers with appropriate safeguards in place. The project team periodically reviews all activities involving protected health information to ensure that such safeguards including standard operating procedures are being followed. The procedure for notifying the ACCF of any breach of confidentiality and immediate mitigation standards that need to be followed is communicated to participants. ACCF limits

access to Protected Health Information, and to equipment, systems, and networks that contain, transmit, process or store Protected Health Information, to employees who need to access the PHI for purposes of performing ACCF's obligations to participants who are in a contractual relationship with the ACCF. All PHI are stored in a secure facility or secure area within ACCF's facilities which has separate physical controls to limit access, such as locks or physical tokens. The secured areas are monitored 24 hours per day, 7 days per week, either by employees or agents of ACCF by video surveillance, or by intrusion detection systems.

Each participant who has access to the NCDR website must have a unique identifier. The password protected webpages have implement inactivity time-outs. Encryption of wireless network data transmission and authentication of wireless devices containing NCDR Participant's information ACCF's network is required. Protected Health Information may only be transmitted off of ACCF's premises to approved parties, which shall mean: A subcontractor who has agreed to be bound by the terms of the Business Associate Agreement between the ACCF and the NCDR Participant.

Time of Data collection:

1 Full time employee can enter on average roughly 1200 patient records per year

(citation: ACC Marketing Intelligence Team)

Annual Fee:

See section 3c2

Overall there is no added procedural risk to patients through their hospital's involvement in the CathPCI Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed.

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

This measure was developed and designed to be used across other organizations and by other measure implementers. The fee and licensing information include below is specific to NCDR program requirements:

The ACCF's program the National Cardiovascular Data Registry (NCDR) provides evidence based solutions for cardiologists and other medical professionals committed to excellence in cardiovascular care. NCDR hospital participants receive confidential benchmark reports that include access to measure macro specifications and micro specifications, the eligible patient population, exclusions, and model variables (when applicable). In addition to hospital sites, NCDR Analytic and Reporting Services provides consenting hospitals' aggregated data reports to interested federal and state regulatory agencies, multi-system provider groups, third-party payers, and other organizations that have an identified quality improvement initiative that supports NCDR-participating facilities. Lastly, the ACCF also allows for licensing of the measure specifications outside of the Registry. For calendar year 2017the annual pricing for hospitals, NCDR Analytic and Reporting Services, and licensing of measure specifications ranges from $2900-$50,000.

Measures that are aggregated by ACCF and submitted to NQF are intended for public reporting and therefore there is no charge for a standard export package. However, on a case by case basis, requests for modifications to the standard export package will be available for a separate charge.

There is no added procedural risk to patients through their hospital's involvement in the CathPCI Registry. No testing, time, risk, or procedures beyond those required for routine care will be imposed.

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
	Public Reporting https://cvquality.acc.org/ncdr-home/acc-public-reporting NCDR Public Reporting Payment Program Quality Hospital Insight program for Anthem https://www.anthem.com/wps/portal/ahpmedprovider?content_path=shared/noapplication/f2/s3/t0/pw_b140403.htm&rootLevel=1&label=Hospital%20Quality%20and%20Safety Blue Distinction Centers for Cardiac Care http://www.bcbs.com/healthcare-partners/blue-distinction-forproviders/cardiacprogramcriteria.pdf Quality Improvement (external benchmarking to organizations) National Cardiovascular Data Registry https://cvquality.acc.org/NCDR-Home/registries/hospital-registries/cathpci-registry

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

Payment Program

Name of program and sponsor: Blue Distinction Centers for Cardiac Care; Sponsor: Blue Cross Blue Shield Association

Purpose:

The Blue Distinction Centers for Cardiac Care is a national designation program that recognizes hospitals that demonstrate expertise in delivering quality specialty care, safely and effectively. To earn the Blue Distinction Centers+ designation, hospitals must meet the same quality criteria as Blue Distinction Centers, and go an extra step to demonstrate that they do so cost efficiently. Quality is key: only those facilities that first meet Blue Distinction's nationally established, objective quality measures will be considered for designation as a Blue Distinction Center+. Blue Distinction Centers' goal is to help consumers find both quality and value for their specialty care needs, on a consistent basis, while encouraging healthcare professionals to improve the overall quality and delivery of care nationwide. [Retrieved from http://www.bcbs.com/healthcare-partners/blue-distinction-for-providers/cardiacprogramcriteria.pdf on 11/25/13]

Geographic area and number and percentage of accountable entities and patients included

Geographic Area: National program.

Number: Directory of Providers available at http://www.bcbs.com/why-bcbs/blue-distinction/blue-distinction-cardiac/bluedistinctioncardiac.pdf

% of accountable entities: Total of 414 hospitals

Alabama 10

Arizona 4

Arkansas 3

California 46
Colorado 6
Connecticut 5
Delaware 3
Florida 29
Georgia 4
Hawaii 1
Idaho 3
Illinois 29
Indiana 12
Iowa 8
Kansas 5
Kentucky 5
Louisiana 5
Maine 1
Massachusetts 8
Michigan 23
Minnesota 12
Missouri 12
Nebraska 5
New Hampshire 2
New Jersey 3
New York 12
Nevada 2
North Carolina 10
North Dakota 4
Ohio 26
Oklahoma 4
Patients included: information not available .

The measure is also used in the Quality Insight Hospital Program with Anthem, which overlaps with what is included above for Blue Distinction program

NCDR Public Reporting

ACC's National Cardiovascular Data Registry (NCDR) Voluntary Hospital Public Reporting Program: The ACC currently runs a program to give hospitals the opportunity to voluntarily publicly report their measure results based on data from the National Cardiovascular Data Registry (NCDR). Hospitals that choose to participate have their results displayed on ACC's CardioSmart. Currently Hospitals can report on the following three NQF-endorsed measures:

NQF #0965: Use of all recommended medications (ACEI or ARB and beta-blocker) to improve heart function and blood pressure after ICD implant.

NQF # 0964: Therapy with aspirin, P2Y12 inhibitor, and statin at discharge following PCI in eligible patients (composite measure)

NQF: 2377: Overall Defect Free Care Composite (labeled as "Complete Heart Attack Care" on the website)

NCDR CathPCI Registry:

The CathPCI Registry is sponsored by ACC in conjunction with the Society for Cardiovascular Angiographyand Interventions. The CathPCI Registry was designed to create a national surveillance system to assess the characteristics, treatments, and outcomes of patients with coronary heart disease who undergo procedures in cardiac catheterization laboratories. Eligible patients are adults (18 years of age and older) who undergo a diagnostic cardiac catheterization and/or PCI. More than 1,300 hospitals across the U.S submit data to the

CathPCI registry. Participation in the CathPCI Registry provides risk-adjusted quarterly benchmark reports that compares an institution's performance with that of volume-based peer groups and the national experience. The registry includes standardized, evidence-based data elements and definitions, a Dashboard tool that provides a custom query to control for variables (facility size, number of procedures, teaching vs. non-teaching sites, states and regions) to compare the participating facility data, metrics and volumes. ABIM Diplomates can also meet MOC recertification requirements by using CathPCI Registry data to earn up to 80 points toward evaluation of practice performance through the Clinical Quality Coach mobile app

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)
N/A

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

N/A

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

Performance results are distributed to all CathPCI registry participants as part of quarterly benchmark reports, which provide a detailed analysis of an institution´s individual performance in comparison to the entire registry population from participating hospitals across the nation. Reports include an executive summary dashboard, at-a-glance assessments, and patient level drill-downs. Registry participants also have access to an outcome report companion guide which provides common definitions and detailed metric specifications to assist with interpretation of performance rates.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

The majority of the required data elements are routinely generated and acquired during the delivery of standard cardiac care to this patient population. Electronic extraction of data recorded as part of the procedure expedites data collection. This strategy offers point of care collection and minimizes time and cost. Institutions can manually report using a free web-based tool or automate the reporting by using certified software developed by third-party vendors. The data elements required for this measure are readily available within the patient's medical record or can be attained without undue burden within the hospital. Most data elements exist in a structured format within patient's electronic health record.

There are a number of methods used to educate and provide general support to registry participants. This includes the following:

- Registry Site Manager Calls are available for all NCDR participants. RSM calls are provided as a source of communication between NCDR and participants to provide a live chat Q and A session on a continuous basis.
- New User Calls are available for NCDR participants, and are intended for assisting new users with their questions.
- NCDR Annual Conference

The NCDR Annual Conference is a well-attended and energetic two-day program at which participants from across the country come together to hear about new NCDR and registry-specific updates. During informative general sessions, attendees can learn about topics such as transcatheter therapies, the NCDR dashboard, risk

models, data quality and validation, and value-based purchasing. Attendees also receive registry updates and participate in advanced case studies covering such topics as Appropriate Use Criteria and outcomes report interpretation.

- Release notes (for outcomes reports)
- Clinical Support

The NCDR Product Support and Clinical Quality Consultant Teams are available to assist participating sites with questions Monday through Friday, 9:00 a.m. - 5:00 p.m. ET.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

Feedback is typically obtained through monthly registry site manager monthly calls, ad hoc phone calls tracked with salesforce software, and during registry –specific break-out sessions at the NCDR's annual meeting. Registry Steering Committee members may also provide feedback during regularly scheduled calls.

4a2.2.2. Summarize the feedback obtained from those being measured.

Users have not reported any difficulties with reporting this measure.

4a2.2.3. Summarize the feedback obtained from other users

No other feedback was received from other users

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

N/A

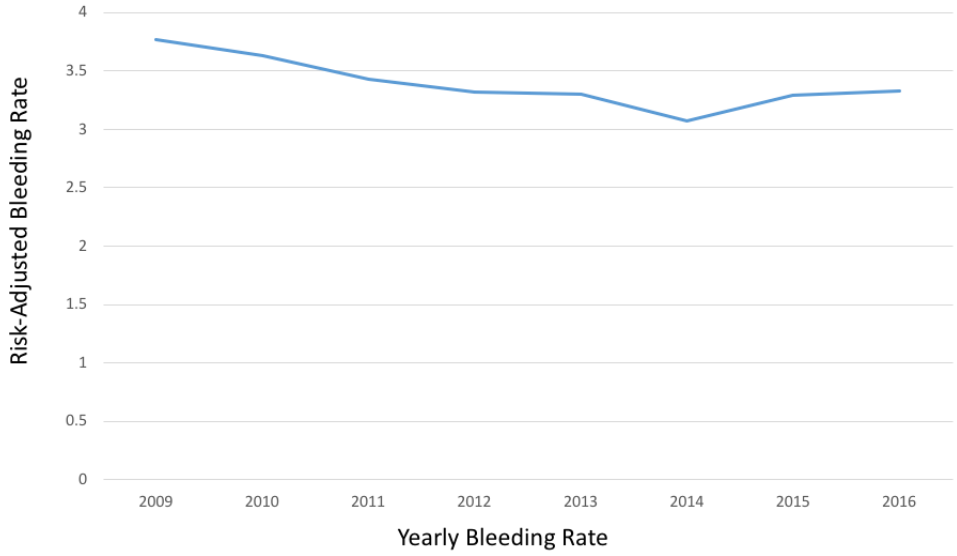
Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Over time, there has been improvement in the rates of peri-procedural bleeding, from 3.8% in 2009 to 3.3% in 2016 ($p<0.0001$). Then annual reduction in the odds of bleeding is 0.963 (95%CI=0.91, 0.965; $p<0.001$), suggesting a 4% reduction in the odds of bleeding per year, on average across all hospitals. The year over year results are shown in the figure below.



4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

N/A

4b2.2. Please explain any unexpected benefits from implementation of this measure.

The most vulnerable aspect of this measure pertains to physician transparency and willingness to report and record adverse events.

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

No

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

No

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

N/A

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);
OR
Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s):
Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

There are no bleeding related risk standardized measures endorsed by NQF currently for the PCI patient population.

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** 2013_JACC_Updated_Model_to_Predict_Risk_of_Post_PCI_Bleeding.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): American College of Cardiology

Co.2 Point of Contact: Sana, Gokak, sgokak@acc.org, 202-375-6596-

Co.3 Measure Developer if different from Measure Steward: American College of Cardiology

Co.4 Point of Contact: Esteban, Perla, eperla@acc.org

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

At the time of initial endorsement of this measure, the 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 and quality measurement. Serial phone calls were held to both define the eligible population and given process. These clinical leaders are noted below.

NCDR Clinical workgroup 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.

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2011

Ad.3 Month and Year of most recent revision: 12, 2017

Ad.4 What is your frequency for review/update of this measure? With dataset revisions and based on new evidence.

Ad.5 When is the next scheduled review/update for this measure? 12, 2019

Ad.6 Copyright statement: American College of Cardiology Foundation All Rights Reserved

Ad.7 Disclaimers: ACC realizes the various NCDR endorsed measures are not readily available on their own main webpage. However, ACCF plans to update their main webpage (cardiosource.org) to include the macrospecifications of the NQF endorsed measures. ACC hopes to work collaboratively with NQF to create a consistent and standard format would be helpful for various end users. In the interim, the supplemental materials include the details needed to understand this model.

Ad.8 Additional Information/Comments: ACC appreciates the opportunity to submit measures for this NQF endorsement maintenance project.