

AMI Mortality eMeasure Calculation Algorithm

The calculation algorithm consists of two steps. First, we used a logistic regression model to identify the model variables. Second, we used a hierarchical logistic regression model to calculate the risk-standardized mortality rates.

The logistic regression model links the outcome to the patient-level risk factors. Let Y_{ij} denote the outcome (in this measure, equal to 1 if patient dies, zero if patient lives) for the j^{th} patient who had an acute myocardial infarction (AMI) admission at the i^{th} hospital; \mathbf{Z}_{ij} denotes a set of risk factors based on the data. Let I denote the total number of hospitals and n_i the number of index patient stays in hospital i . We assume the outcome is related linearly to the covariates via a known linked function, h , where

Logistic regression model:

$$h(Y_{ij}) = \alpha_i + \beta \mathbf{Z}_{ij} \quad (1)$$

where $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \dots, Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link.

The hierarchical logistic regression model accounts for the natural clustering of observations within hospitals. It links the risk factors to the same outcome and a hospital-specific random effect,

Hierarchical logistic regression model:

$$h(Y_{ij}) = \alpha_i + \beta \mathbf{Z}_{ij} \quad (2)$$

$$\alpha_i = \mu + \omega_i; \quad \omega_i \sim N(0, \tau^2) \quad (3)$$

where h = the logit link, α_i represents the hospital-specific intercept, \mathbf{Z}_{ij} is defined as above, μ is the general intercept over all hospitals, and τ^2 the between-hospital variance component. This model separates within-hospital variation from between-hospital variation.

Both hierarchical logistic regression models and logistic regression models are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

See Figure 1 for a diagram of the analysis steps

Hospital Performance Reporting

Using the set of risk factors in the logistic regression model (1), we fit the hierarchical logistic regression model defined by Equations (2) - (3) and estimate the parameters $\hat{\mu}$, $\{\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_I\}$, $\hat{\beta}$ and $\hat{\tau}^2$. We calculate a standardized outcome, s_i , for each hospital by computing the ratio of the number of predicted deaths to the number of expected deaths, multiplied by the unadjusted overall mortality rate, \bar{y} . Specifically, we calculate

$$\text{Predicted} \quad \hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta} Z_{ij}) \quad (4)$$

$$\text{Expected} \quad \hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta} Z_{ij}) \quad (5)$$

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \quad (6)$$

If more (fewer) “predicted” cases than “expected” cases have the outcome in a hospital, then s_i will be higher (lower) than the unadjusted average. Higher (lower) s_i indicates worse (better) quality.

Creating Interval Estimates

Because the statistic described in Equation (6) is a complex function of parameter estimates, we use re-sampling and simulation techniques to derive an interval estimate. For each hospital, we compute an interval estimate of s_i to characterize the level of uncertainty around the point estimate using bootstrapping simulations. The point estimate and interval estimate can be used to characterize and compare hospital performance (e.g., higher than expected, as expected, or lower than expected). The bootstrapping simulation has the advantage of avoiding unnecessary distributional assumptions.

Calculation Algorithm

Let I denote the total number of hospitals in the sample. We repeat steps 1 – 4 below for $b = 1, 2, \dots, B$ times:

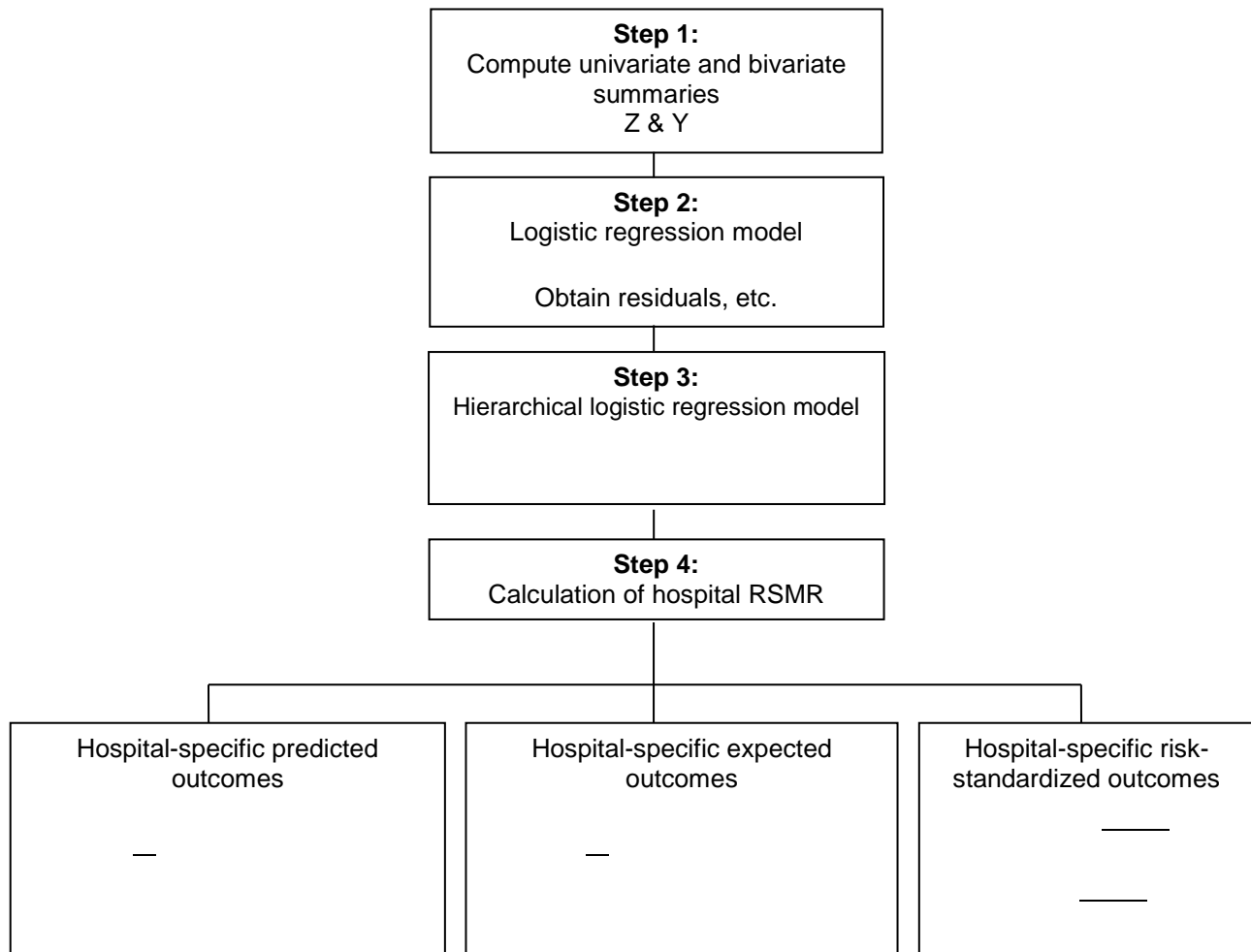
1. Sample I hospitals with replacement.
2. Fit the hierarchical logistic regression model using all patients within each sampled hospital. We use as starting values the parameter estimates obtained by fitting the model to all hospitals. If some hospitals are selected more than once in a bootstrapped sample, we

treat them as distinct so that we have I random effects to estimate the variance components. At the conclusion of Step 2, we have:

- a. $\hat{\beta}^{(b)}$ (the estimated regression coefficients of the risk factors).
 - b. The parameters governing the random effects, hospital-adjusted outcomes, distribution, $\hat{\mu}^{(b)}$ and $\hat{\tau}^{2(b)}$.
 - c. The set of hospital-specific intercepts and corresponding variances, $\{\hat{\alpha}_i^{(b)}, \text{var}(\hat{\alpha}_i^{(b)}); i = 1, 2, \dots, I\}$.
3. We generate a hospital random effect by sampling from the distribution of the hospital-specific distribution obtained in Step 2c. We approximate the distribution for each random effect by a normal distribution. Thus, we draw $\alpha_i^{(b*)} \sim N(\hat{\alpha}_i^{(b)}, \text{var}(\hat{\alpha}_i^{(b)}))$ for the unique set of hospitals sampled in Step 1.
 4. Within each unique hospital i sampled in Step 1, and for each case j in that hospital, we calculate $\hat{y}_{ij}^{(b)}$, $\hat{e}_{ij}^{(b)}$, and $\hat{s}_i(Z)^{(b)}$ where $\hat{\beta}^{(b)}$ and $\hat{\mu}^{(b)}$ are obtained from Step 2 and $\hat{\alpha}_i^{(b*)}$ is obtained from Step 3.

Ninety-five percent interval estimates (or alternative interval estimates) for the hospital-standardized outcome can be computed by identifying the 2.5th and 97.5th percentiles of randomly half of the B estimates (or the percentiles corresponding to the alternative desired intervals).

Figure 1. Analysis Steps



Hospital 30-day Risk-standardized Acute Myocardial Infarction (AMI) Mortality eMeasure

Measure Technical Report

**Submitted by Yale New Haven Health Services Corporation Center for Outcomes
Research & Evaluation (CORE):**

Robert L. McNamara, MD, MHS
Susannah Bernheim, MD, MHS
Purav Mody, MBBS
Julia Montague, MPH
Yongfei Wang, MS
Elizabeth Eddy, BA
Amena Keshawarz, MPH
Harlan M. Krumholz, MD, SM

Contract # HHSM-500-2008-000251/HHSM-500-T0001, Modification No. 000008

Prepared For:

Centers for Medicare & Medicaid Services (CMS)

September 2013

ACKNOWLEDGEMENTS

We would like to acknowledge the ongoing support from our colleagues affiliated with the American College of Cardiology and the National Cardiovascular Data Registry. Without their continued support, this project would not have been possible:

Frederick Masoudi, MD, MSPH
Gregg Fonarow, MD
Joanne Foody, MD
James Jollis, MD
James DeLemos, MD, PhD
Joseph Drozda, MD
George Dengas, MD, PhD
Vernon Anderson, MD
Lara Slattery, MHS

Special thanks are due to the group of electronic health records (EHR) system experts who communicated with us on a regular basis and provided useful insight and guidance. These individuals include:

Adam Landman, MD
Christopher Mast, MD, MS
Venkatesh Janakiraman

We would also like to acknowledge support from individuals at the Office of the National Coordinator for Health Information Technology (ONC) who provided their expertise regarding EHRs and eMeasures:

Kevin Larsen, MD
Jacob Reider, MD
Lauren Richie, MA

We would also like to thank members of Sentara Healthcare, Kaiser Permanente, Veterans Health Affairs, Mid America Heart Institute, Duke Clinical Research Institute, and Statewide Planning and Resource Cooperative System (SPARCS) for speaking to us regarding the availability of EHR data for development of our measure as well as input regarding model development. These members include:

Gabriel Escobar, MD
John Brush, MD
John Parker, MD
Marta Render, MD
David Magid, MD
Edward Hannan, PhD, MS, MS
John Spertus, MD, MPH
Mikhail Kosiborod, MD
James Tcheng, MD

In addition, we would like to acknowledge Jeremy Michel, MD, Postdoctoral Fellow at the Yale Center for Medical Informatics, for his valuable input during the eSpecification process.

Finally, we appreciate the efforts of our internal members here at the Center for Outcomes Research and Evaluation (CORE) and our consultants for their continual support in our project. These individuals include:

Elizabeth Drye, MD, SM

Zhenqiu Lin, PhD

Lori Geary, MPH

Jeptha Curtis, MD

Lisa Suter, MD

Sharon-Lise Normand, PhD (Department of Biostatistics, Harvard School of Public Health and
Department of Health Care Policy, Harvard Medical School)

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
1. EXECUTIVE SUMMARY	1
2. INTRODUCTION	3
2.1 Background	3
2.2 Rationale for AMI Mortality eMeasure.....	3
3. APPROACH TO <i>DE NOVO</i> DEVELOPMENT OF OUTCOME EMEASURE	5
3.1 <i>De Novo</i> Development	5
3.2 Registry Data Source	5
3.3 Establishment of Feasibility Criteria in the Current Clinical and EHR Environment	5
3.4 Working Group and Expert Input.....	6
3.5 eSpecification and eMeasure Testing	7
4. APPLICATION OF THE METHODS.....	8
4.1 Overview	8
4.2 Outcome	8
4.2.1 30-day Mortality	8
4.2.2 All-cause Mortality.....	8
4.3 Data Sources	8
4.3.1 2009 and 2010 NCDR® ACTION Registry®–GWTG™ (AR-G) Data	8
4.3.2 2009 and 2010 Medicare Data	11
4.4 Cohort Derivation	11
4.5 Model Development	17
4.5.1 Candidate Risk-adjustment Variables	17
4.5.2 Selection of Final Risk-adjustment Variables.....	22
4.6 Statistical Approach to Model Development.....	24
4.6.1 Logistic Regression Model and Hierarchical Logistic Regression Model	24
4.6.2 Calculation of Hospital-Specific RSMRs	25
4.7 Model Testing	26
4.7.1 Reliability	26
4.7.2 Validity	27
4.7.3 Disparities Assessment	27
4.7.4 Sensitivity Analysis – Assessment of Variables Deemed Clinically Relevant but Not Feasible for Use in eMeasures	27
4.8 eSpecification.....	28

5. RESULTS.....	29
5.1 Preliminary Model (Containing Variables with Questionable eMeasure Feasibility).....	29
5.1.1 Logistic Regression.....	29
5.2 Final Model (Containing only eMeasure-feasible Variables).....	29
5.2.1 Logistic Regression.....	29
5.2.2 Hierarchical Logistic Regression Model	30
5.2.3 30-day Mortality Rate Distribution.....	30
5.3 Model Assessment.....	32
5.3.1 Model Validation	32
5.3.2 Measure Score Validity Testing Results.....	33
5.3.3 Disparities Assessment	33
5.3.4 Sensitivity Analysis – Assessment of Variables Deemed Clinically Relevant but Not Feasible for Use in eMeasures.....	35
6. SUMMARY STATEMENT	37
7. REFERENCES	39
8. APPENDICES	41
Appendix A. eSpecification.....	42
Appendix B. eMeasure Testing.....	43
8.3.1 Overview.....	43
8.3.2 Alpha Testing	43
8.3.3 Beta Testing	46
8.3.4 Conclusions.....	49
Appendix C. Working Group Member Roster	51
Appendix D. Lists of EHR Vendors, Additional EHR Experts, and Hospitals Systems Consulted for Feedback on Data Sources and Model Development.....	52
Appendix E. Variables Excluded at Each Step of Variable Selection	54
Appendix F. Approach to Defining Continuous Candidate Variables	58

LIST OF TABLES

Table 1. Comparison of CMS Hospitals Participating and Not Participating in AR-G in 2009.....	10
Table 2. Selected Patient Characteristics and Outcomes in AR-G Data for Patients Unmatched and Matched to CMS Data.....	15
Table 3. Selected Patient Characteristics and Outcomes in CMS Data for Patients Unmatched and Matched to AR-G Data	16
Table 4. eMeasure Feasibility of Candidate Variables	20
Table 5. Model Candidate Variables	22
Table 6. Description of Preliminary and Final Risk-adjustment Models	24
Table 7. Preliminary Model: Logistic Regression Results (N=20,540 patients).....	29
Table 8. Final Model: Logistic Regression Results (N=20,540 patients).....	29
Table 9. Final Model: Hierarchical Logistic Regression Model Results (N=20,540 patients)	30
Table 10. Model Performance: Results Based on the Logistic Regression Model	32
Table 11. Final Model (Logistic Regression) Odds Ratios by Dataset	33
Appendix Table 1. Feasibility Survey Options and Feasibility Scores (Developed by Abt Associates)	44
Appendix Table 2. Feasibility Scores of Data Elements Included in the AMI Mortality eMeasure (Calculated by Abt Associates).....	45
Appendix Table 3. Percentage of Patients in the Electronically Extracted Cohorts that were Eligible based on Inclusion Criteria as Identified by Nurse Abstraction	48
Appendix Table 4. Identification of Transfer Patients by Electronic Extraction and Nurse Abstraction	48
Appendix Table 5. Agreement between Electronically Extracted and Manually Abstracted Risk-adjustment Variables	49

LIST OF FIGURES

Figure 1. Derivation of Cohort for Model Development.....	12
Figure 2. Candidate Variable Selection Process Flow Chart	18
Figure 3. Analysis Steps.....	26
Figure 4. Distribution of Hospital Unadjusted Mortality Rates (2009)	31
Figure 5. Distribution of Hospital Risk-standardized Mortality Rates (2009)	31
Figure 6. Correlation of RSMR based on the Currently Proposed Final Model with RSMR based on the Previously Developed, Publicly Reported, Claims-Based AMI Mortality Measure (Hospital Volume-weighted Pearson Correlation Coefficient=0.86).....	33
Figure 7. Hospital RSMR (2009) by Proportion of African-American Patients.....	34
Figure 8. Hospital RSMR (2009) by Proportion of Dual Eligible Patients	35
Figure 9. Correlation between RSMR based on the Final Model and RSMR based on the Final Model Plus ECG Results (Hospital Volume-weighted Correlation Coefficient=0.989)	36
Figure 10. Association between Age and Mortality: No Winsorization on Age.....	58
Figure 11. Association between Heart Rate and Mortality with Winsorization of Heart Rate: Low Values to 1 st Percentile (40 bpm) and High Values to 99 th Percentile (160 bpm).....	58
Figure 12. Association between Systolic Blood Pressure and Mortality with Winsorization of Systolic Blood Pressure: Lower Values to 1 st Percentile (67 mm Hg) and Higher Values to 99 th Percentile (224 mm Hg)	59
Figure 13. Association between Body Mass Index (BMI) and Mortality with Winsorization of BMI: Lower Values to 1 st Percentile (16.5 kg/m ²) and Upper Values to 99 th Percentile (48.0 kg/m ²)	60
Figure 14. Association between Troponin Ratio and Mortality with Winsorization of Troponin Ratio: High Values to 99 th Percentile (871). The 1 st Percentile is 0	61
Figure 15. Association between Troponin Ratio and Mortality with Winsorization of Troponin Ratio: High Values to 99 th Percentile (871). Only Range of Troponin Ratio between 0 and 60 Are Shown	61
Figure 16. Association between Creatinine and Mortality with Winsorization of Creatinine: Low Values to 1 st Percentile (0.6 mg/dL) and High Values to 99 th Percentile (6.1 mg/dL).....	62
Figure 17. Association between Creatinine Clearance and Mortality with Winsorization of Creatinine Clearance: Low Values to 1 st Percentile (9.1 mL/min) and High Values to 99 th Percentile (142 mL/min).....	62
Figure 18. Association between Hemoglobin and Mortality with Winsorization of Hemoglobin: Low Values to 1 st Percentile (7.7 g/dL) and High Values to 99 th Percentile (17.5 g/dL)	63

1. EXECUTIVE SUMMARY

The implementation of electronic health records (EHRs) offers opportunities for the advancement of quality measurement. Ideally, performance measures developed for use in EHRs will utilize detailed clinical data but not require the substantial resources involved in collecting registry data or abstracting medical records.

This report describes development of a hospital 30-day all-cause risk-standardized mortality eMeasure for acute myocardial infarction (AMI) admissions. To our knowledge, this is one of the first outcome eMeasures developed. Our objective was to build a measure that could feasibly be implemented in the near term in current EHR systems using data elements routinely entered in current clinical practice.

We developed this measure *de novo*, rather than “retooling” a previously developed measure, in order to best utilize the EHR data platform. Although the measure is intended for use with EHR data, we used clinical registry data for measure development due to a lack of an accessible multi-hospital or nationally representative EHR dataset. Therefore, as part of model development, we established a set of criteria (listed below) to evaluate and include only those clinical variables currently feasible for use in eMeasures. We later further tested the measure feasibility and data element validity in EHR data.

Outcome: We developed this measure with 30-day all-cause mortality after AMI as the outcome, in accordance with the claims-based AMI mortality measure that is currently publicly reported.¹

Data source: We used ACTION Registry®–GWTG™ (AR-G), designed and maintained by the National Cardiovascular Data Registry (NCDR®), for clinical data, merged with Centers for Medicare & Medicaid Services (CMS) claims and enrollment data for the mortality outcome for measure development. The final eMeasure is intended for use with EHR data.

Risk-adjustment modeling: To adequately account for relevant patient demographic and clinical characteristics present upon initial presentation to the hospital, we developed a risk-adjustment model.

- With input from the literature, EHR experts, and vendors, we developed three criteria to determine which variables within AR-G were feasible for use in an eMeasure:
 - Consistently obtained in the target population based on current clinical practice
 - Captured with a standard definition and recorded in a standard format
 - Entered in structured fields that are feasibly retrieved from current EHR systems
- We developed a risk model for 30-day all-cause mortality using logistic regression and estimated the hospital-level 30-day all-cause risk-standardized mortality rate (RSMR) using a hierarchical logistic regression model. Model development was consistent with the rationale articulated in the American Heart Association scientific statement “Standards for Statistical Models Used for Public Reporting of Health Outcomes”² and used to develop prior CMS mortality measures that are endorsed by the National Quality Forum (NQF) and which CMS now publicly reports on *Hospital Compare* (<http://www.hospitalcompare.hhs.gov>).
- The final model includes the following variables, assessed at presentation:
 - Age
 - Heart rate
 - Systolic blood pressure
 - Troponin ratio (initial troponin value / troponin upper range limit for hospital)

- Creatinine
- The overall performance of the model was comparable with or better than that of current publicly reported outcome measures.
- We tested for measure score validity by correlating the RSMR with that of the previously validated, publicly reported, claims-based AMI mortality measure.

eMeasure Testing: After eSpecification, the resulting eMeasure was evaluated through an information technology (IT)/quality expert survey (usability and feasibility), an EHR vendor survey (feasibility), XML Schema and Schematron testing (data element reliability), and field testing in a sample of hospitals with various EHR systems (data element validity). This testing supported the overall usability, feasibility, reliability, and validity of the eMeasure and fulfilled NQF guidelines for eMeasure testing. This testing also revealed that some aspects of near-term implementation of outcome eMeasures may be challenging; implementation of hybrid models that combine EHR data with information from other data sources may be required until these challenges are resolved.

In summary, we have built one of the first outcome eMeasures that produces estimates of hospital risk-standardized mortality rates for Medicare patients with AMI and that can be used to evaluate hospital quality of care using the EHR. The eMeasure is consistent with the consensus standards for publicly reported outcome measures, is parsimonious in risk adjustment, and performs well compared with the previously validated, publicly reported, claims-based AMI mortality measure. Feasibility and data element validity testing in the EHR environment demonstrate that, for the most part, the data elements for risk adjustment can be feasibly and accurately extracted. Near-term implementation of this measure will require input from other data sources as EHR implementation continues to evolve.

2. INTRODUCTION

2.1 Background

Since 2007, the Centers for Medicare & Medicaid Services (CMS) has publicly reported hospital 30-day risk-standardized mortality rates (RSMRs) for acute myocardial infarction (AMI).¹ This measure, developed by Yale New Haven Health Services Corporation Center for Outcomes Research and Evaluation (CORE) and endorsed by the National Quality Forum (NQF), is calculated using administrative claims data. The use of claims data allows CMS to measure and publicly report quality measures without any additional burden on hospitals for data collection.

The implementation of electronic health records (EHRs) offers an opportunity for the development of quality measures that utilize medical record data rather than claims, but without requiring the resources needed for manual medical record abstraction. The American Recovery and Reinvestment Act of 2009 established incentives for hospitals across the country to universally adopt EHR systems.³ In particular, the Health Information Technology for Economic and Clinical Health (HITECH) Act was enacted to promote the adoption and meaningful use of health information technology (IT) for the purpose of quality measurement and quality improvement (Appendix A).⁴ The Office of the National Coordinator for Health Information Technology (ONC) has established Meaningful Use criteria to ensure that EHRs support the collection of point-of-care, clinically relevant data to support quality improvement and the development of eMeasures (that is, performance measures that use EHR data). Benefits in quality improvement after the implementation of EHRs have been documented.⁵

Given the current expansion of EHR implementation and the expectation that quality measures will be increasingly able to draw off the rich clinical data resources furnished by EHRs, CMS contracted with CORE to develop an outcome eMeasure evaluating hospital 30-day mortality following admission for AMI.

2.2 Rationale for AMI Mortality eMeasure

We sought to build an eMeasure assessing quality for an important condition and outcome for which we had already developed a claims-based measure. AMI is a high-volume, high-severity, and high-cost condition. Each year, over 600,000 Americans will experience an AMI.⁶ Despite impressive improvements in treatments, 30-day mortality following AMI exceeds 7%.⁷ CMS pays approximately \$11.7 billion annually for in-hospital costs for Medicare beneficiaries with coronary heart disease, of which AMI is a major contributor.⁶ AMI is also a well-studied condition with a rich literature on important risk factors and risk models. Finally, as mentioned previously, an AMI mortality measure developed and calculated using administrative claims data is currently publicly reported.¹

Our goal was not to “retool” the previous measure (that is, simply create a crosswalk between data elements in claims and those in the EHR environment) but to develop a new eMeasure *de novo*. Nonetheless, developing an initial eMeasure for a condition and outcome with an existing claims measure allowed us to build on clinical and measurement expertise when making measure decisions about the included cohort. Therefore, we could focus on the methodology of eMeasure development and testing. The existing claims-based measure also provided a comparable measure

as a source of comparison for our final eMeasure. For all these reasons, AMI represents an excellent condition for which to develop an eMeasure.

3. APPROACH TO *DE NOVO* DEVELOPMENT OF OUTCOME eMEASURE

In order to develop an outcome measure for use with EHR data, we defined new approaches for measure development in this emerging area. In this section, we describe the key aspects of the approach we used to develop this *de novo* outcome eMeasure.

3.1 *De Novo* Development

This AMI mortality eMeasure was developed *de novo*; we did not seek to mirror a previously developed measure, but rather we made all methodology decisions and selected variables specifically for this measure. Many eMeasures are “retooled” measures, developed by creating a crosswalk between clinical data elements found in the original, paper-based measure and similar elements in EHR data. However, retooling a previously developed measure risks altering the measure in the process because the data elements in the two sources may not match precisely. Furthermore, a clinical data element that can be easily abstracted from a paper medical record may not be equally straightforward to extract from an EHR. By contrast, *de novo* development allowed us to target those data elements most reliably and feasibly extracted from EHRs. Through the process of *de novo* development of our eMeasure we established a roadmap for future outcome eMeasure development.

3.2 Registry Data Source

Outcome measures used to profile hospitals and assess relative performance need to be risk-adjusted to provide a fair assessment of quality. Development of a risk-adjusted outcome eMeasure, therefore, requires a data source with a broad array of clinical variables and a substantial number of hospitals for adequate risk model development. At the time of measure development, issues of data exchange and standardization limited the ability to aggregate EHR data from multiple hospitals. Moreover, many EHR vendors and health systems that have aggregated EHR data sources are not yet able to easily extract datasets to support measure development. Therefore, we opted to use a clinical registry for measure development.

The registry provided us with measure development data collected in a standard fashion from a large number of hospitals nationally which could be linked to patient outcomes. The variables collected by the registry included a wide array of data elements likely to be found in current EHRs. Moreover, through the process of the registry development, these variables had been thoroughly vetted to include important risk factors for AMI patients. In order to successfully use registry data to develop a measure for the EHR environment, we developed feasibility criteria to restrict our measure variables to only those that were currently available in EHR data at the time of development, as described below. A further advantage of using registry data was that it enabled us to test the importance of data elements that are clinically important but not feasibly extracted from many EHRs at the time of development. This testing would not have been possible in a data source limited to elements extracted from EHRs.

3.3 Establishment of Feasibility Criteria in the Current Clinical and EHR Environment

The EHR is primarily a tool for clinical practice; thus, optimal quality eMeasures consider current clinical practice and current EHR capability to avoid any disruption of clinical care. Furthermore, if quality measures rely on actions such as filling out additional checkboxes to collect data elements

that are not captured in the routine service of clinical care, there will be significant challenges to operationalizing functions across multiple health systems and vendors. **Therefore, our primary objective was to develop an eMeasure that could be implemented without changing standard clinical practice or requiring that EHRs be adapted.**

In order to meet this goal, early in the measure development process, we developed a set of criteria to ensure all data elements used in the measure, both in cohort identification and risk adjustment, could be feasibly obtained within current clinical practice and with current EHR systems. As Meaningful Use criteria provide standards for *future* EHR implementation, and the Quality Data Model (QDM) is not limited to current EHR capability, we needed to establish stricter criteria based on current EHR capability. In a series of calls with EHR experts, we developed the following criteria to assess the eMeasure feasibility of candidate model variables:

1. Consistently obtained in the target population based on current clinical practice
2. Captured with a standard definition and recorded in a standard format
3. Entered in structured fields that are feasibly retrieved from current EHR systems

The first criterion ensures that the measure will not rely on the adoption of new clinical practices, such as requiring medical staff to routinely collect a laboratory test they might not otherwise order. The second criterion confirms that data elements used in the measure have the same meaning across sites. The third aligns with our intention to build a measure that could be feasibly implemented in current EHRs.

Through discussions with the EHR experts and examination of the data, we assessed each potential candidate variable for the risk-adjustment model by these criteria. Variables satisfying all three criteria were deemed feasible for inclusion in an eMeasure given the current EHR environment at the time of development. This process was completed early in measure development so that only feasible variables were considered for the model. Further feasibility testing using EHR data was completed in later phases of development; see Appendix B for details.

3.4 Working Group and Expert Input

Development of the AMI mortality eMeasure involved input from a number of experts, including a working group from CORE, as well as external EHR and clinical experts. The working group consisted of clinical and methodological experts with extensive experience in both performance measure development and AMI; the group included cardiologists, health sciences researchers, and other professionals with expertise in biostatistics, measure methodology, and quality improvement (Appendix C). The working group provided regular input on all measure decisions, including data source identification, cohort derivation, outcome definition, model development, and model testing. Working group meetings were typically held once per week and addressed key issues to ensure the measure would be meaningful, useful, and well-designed.

Throughout measure development, we obtained expert and stakeholder input via discussions with EHR vendors and experts and clinical experts from the National Cardiovascular Data Registry (NCDR). The EHR vendors and experts provided key input regarding appropriate data sources for model development and the appropriateness of including certain clinical variables in an eMeasure. We solicited advice from representatives of the NCDR regarding the selected variables in the final model, the clinical value of the variables excluded from the model for eMeasure feasibility reasons, and the overall clinical face validity of the model.

3.5 eSpecification and eMeasure Testing

eSpecification is the process of converting a paper-based quality measure or implementing a measure specifically developed for EHR into a format usable in the EHR environment. This process includes encoding the measure specifications in a standard eMeasure format known as Health Quality Measures Format (HQMF).⁸ We collaborated with Abt Associates, another CMS contractor, to construct the human-readable file, the machine-readable file, and the specific value sets for the eMeasure. These files enable appropriate implementation of the eMeasure.

In addition to model testing performed using registry data, we also collaborated with Abt Associates to complete eMeasure testing of the eSpecified measure's usability, feasibility, reliability, and validity. The eMeasure testing included surveys with IT experts and EHR vendors and comparison of the electronic extraction from the EHR output vs. manual nurse abstraction of the electronic record. For detailed results of the eMeasure testing, refer to Appendix B.

4. APPLICATION OF THE METHODS

4.1 Overview

This section provides details about the development of the hospital risk-standardized AMI mortality eMeasure, including the identification of a relevant data source, the cohort definition, variable selection for the risk-adjustment model, and model testing. In developing the measure we followed the standards set forth in the development of prior outcome performance measures, specifically using guidance from NQF,⁹ the CMS Measures Management System, and the American Heart Association scientific statement, “Standards for Statistical Models Used for Public Reporting of Health Outcomes.”²

4.2 Outcome

4.2.1 30-day Mortality

As compared with in-hospital mortality, a 30-day outcome timeframe provides a standard period of assessment. Models with a fixed outcome period are preferable because they ensure hospital variation in length of stay does not affect performance and minimize the opportunity for misrepresentation (transferring of patients or other gaming mechanisms).¹⁰ In addition, the 30-day period may be a more clinically meaningful timeframe for patients, reflecting not only the outcomes of inpatient processes of care but also the transition of care to the outpatient setting. As such, a 30-day mortality measure may stimulate better collaboration between hospitals and their surrounding medical communities aimed at reducing mortality rates. These activities may include ensuring patients are clinically ready for discharge; improving communication among providers in transitions of care; and encouraging strategies that promote disease management principles and educate patients on what symptoms to monitor, whom to contact with questions, and where and when to seek follow-up care.

4.2.2 All-cause Mortality

We used all-cause mortality as opposed to cardiac-specific mortality for several reasons. First, from the patient perspective, mortality from any cause is the critical measure. Second, different causes of death may still be directly related to the quality of care. Third, making accurate determinations of specific causes of death is difficult and prone to error, particularly if the patient dies outside of the hospital setting.

4.3 Data Sources

4.3.1 2009 and 2010 NCDR® ACTION Registry®–GWTG™ (AR-G) Data

The NCDR AR-G serves as a national surveillance effort to improve the quality of care for AMI patients on a national level.¹¹ AR-G captures detailed data about patients aged 18 years or older undergoing management for AMI. The data include demographics, comorbid conditions, clinical status, laboratory values, diagnostic tests, management strategies adopted, complications, and outcomes. Clinical experts have extensively vetted the more than 300 data elements included in the registry. These data are collected by hospitals and submitted electronically on a quarterly basis to NCDR. (The data collection form and the complete list of variables collected and submitted by hospitals can be found at <http://www.ncdr.com/webncdr/ACTION/>.) The patient

records submitted to the registry focus on acute episodes of care, from admission to discharge. The NCDR does not currently link patient records longitudinally across episodes of care.

Admissions to participating hospitals were eligible for inclusion in AR-G if admitted patients had:

- 1) Ischemic symptoms at rest, lasting ≥ 10 minutes, occurring in the 24 hours before admission, or up to 72 hours for ST segment elevated myocardial infarction (STEMI);
- 2) Electrocardiogram (ECG) changes associated with STEMI (new left bundle-branch block [LBBB] or persistent STEMI ≥ 1 mm in two or more contiguous electrocardiographic leads); or
- 3) Positive cardiac markers associated with non-ST segment myocardial infarction (NSTEMI) (CK-MB or troponin I/T > local laboratory upper limit of normal values) within 24 hours after initial presentation

Of note, patients admitted for other clinical conditions but who develop qualifying symptoms for STEMI or NSTEMI during hospitalization are ineligible for inclusion in AR-G.

A wide spectrum of hospitals across the country participate in AR-G. We compared the characteristics of hospitals that participated in AR-G in 2009 with those of hospitals that did not using data from the American Hospital Association Survey. Compared with hospitals that did not participate in AR-G, hospitals that did participate were larger (had a greater number of beds), more likely to be teaching hospitals, and more likely to have the ability to provide coronary artery bypass graft (CABG) surgery. They were also more likely to be not-for-profit rather than government or for-profit hospitals and to be located in metropolitan rather than rural areas. Hospitals that participated in AR-G were less likely to be safety net hospitals (Table 1).

The NCDR has implemented a Data Quality Program (DQP) to ensure that data submitted to AR-G are complete, consistent, and accurate.¹² Under the DQP, data submitted from various sites are reviewed for overall completeness, and participating hospitals are provided with a confidential analysis. Additionally, each year participating sites are randomly selected to have the quality of their data audited.

Table 1. Comparison of CMS Hospitals Participating and Not Participating in AR-G in 2009

Description	Hospitals in AR-G (N=282)	Hospitals not in AR-G (N=3,897)
	%	%
Number of beds		
<100	6.4	46.9
100 to 300	41.8	36.9
>300	51.8	16.3
Mean (SD)	362 (234)	166 (182)
Ownership		
Government	11.4	23.3
Not-for-profit	77.0	60.5
For-profit	11.7	16.1
Region		
Associated area	0.4	1.2
New England	2.8	4.3
Middle Atlantic	6.7	9.5
South Atlantic	24.1	14.9
East North Central	20.6	15.5
East South Central	7.8	8.9
West North Central	12.1	13.5
West South Central	8.5	14.1
Mountain	6.0	7.3
Pacific	11.0	11.0
Teaching status		
Council of Teaching Hospitals	17.0	5.9
Other teaching	25.9	10.7
Non-teaching	57.1	83.4
Cardiac facility		
CABG surgery	78.7	32.3
Cath lab only	9.2	12.4
Other	12.1	55.3
Core-based statistical area**		
Division	14.9	14.7
Metro	74.8	41.2
Micro	9.2	19.3
Rural	1.1	24.9
Safety Net Hospital*		
No	83.7	69.3
Yes	16.3	30.7

* Defined as government hospitals or non-government hospitals with high Medicaid caseload

**Core-based statistical areas are defined on the basis of the population contained within them:

Division: >2.5 million inhabitants

Metro: 50,000 – 2.5 million inhabitants

Micro: 10,000 – 50,000 inhabitants

Rural: <10,000 inhabitants

4.3.2 2009 and 2010 Medicare Data

Part A inpatient data

Part A inpatient data include claims paid by Medicare for inpatient hospital care.

Medicare Enrollment Database (EDB)

This database contains Medicare beneficiary demographic and vital status information. These data have previously been shown to accurately reflect patient vital status.¹³

Mortality information in the Medicare EDB was linked to the Part A inpatient discharges with AMI using the unique patient identifier in the Medicare databases (health insurance claim [HIC] number).

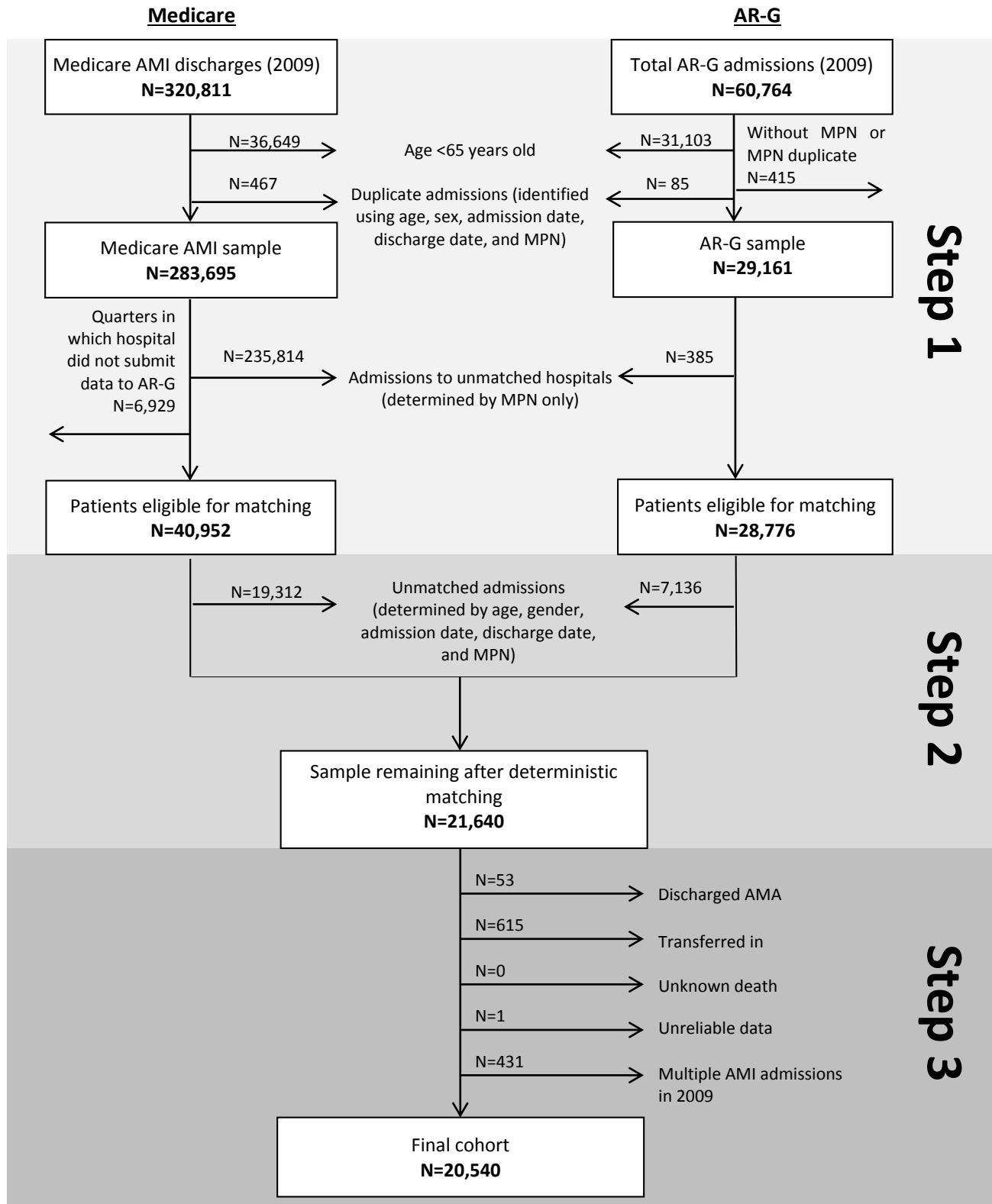
4.4 Cohort Derivation

For development of the model, we used discharges for AMI included in the AR-G dataset from January 1, 2009 through December 31, 2009, deterministically matched with discharges for AMI in CMS claims data from January 1, 2009 through December 31, 2009.

To derive the dataset for the deterministic match from AR-G data, AMI admissions were uniquely identified by hospital Medicare provider number (MPN), patient age, sex, admission date, and discharge date. Hospital MPNs were self-reported in the NCDR ICD Registry™ hospital profile. MPNs were manually verified through the American Hospital Association annual survey database or on the web using hospital name and address.

Similarly, we derived an appropriate cohort of discharges with AMI from the CMS dataset. We identified discharges with AMI by International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) principal discharge diagnosis code 410.xx (excluding 410.x2). We deterministically matched the derived datasets to obtain the final merged CMS-AR-G dataset. Figure 1 depicts the steps followed to derive the cohort, followed by a description of each step.

Figure 1. Derivation of Cohort for Model Development



Step 1: Preparation of datasets for deterministic matching

To derive datasets from AR-G and CMS claims data for the deterministic match, we applied a series of exclusion criteria to both datasets. This allowed us to obtain a comparable cohort of patients within each dataset in preparation for deterministic matching. The exclusion criteria applied were:

- Age <65 years (CMS claims and AR-G data): Admissions for patients aged <65 years at the time of admission were excluded.
Rationale: Patients younger than 65 in the Medicare dataset represent a distinct population that qualifies for Medicare due to disability. The characteristics and outcomes of these patients may not be representative of the larger population of AMI patients.
- Admissions to hospitals with missing or duplicate MPNs (AR-G data only): Any admissions to hospitals with a missing MPN or in hospitals that shared the same MPN were excluded.
Rationale: If the MPN is unreliable, we are unable to match patients in AR-G data to patients in CMS claims data or calculate hospital mortality rates with certainty.
- Duplicate admissions (CMS claims and AR-G data): Admissions for patients who have identical information in a single dataset indicated for age, sex, admission date, discharge date, and MPN are excluded.
Rationale: Admissions with identical demographics are excluded to avoid making matching errors upon merging of the two datasets.

We then excluded admissions for patients in certain hospitals:

- Admissions to hospitals that did not appear in the AR-G dataset
Rationale: Admissions to hospitals that do not submit data to AR-G would not be eligible for matching.
- Admissions occurring during quarters in which a hospital did not submit data to AR-G (CMS claims data only)
Rationale: Admissions occurring during a quarter in which a hospital did not submit data to AR-G would not be eligible for matching (e.g., if a hospital were to start submitting data to AR-G in July, patients in CMS data admitted during January through June would be excluded).

Step 2: Deterministic match of AR-G and CMS claims datasets

The remaining hospitalizations in both datasets were then merged using hospital MPN, patient age, sex, admission date, and discharge date as the linking fields. Admissions that did not match based on all five linking fields were excluded.

Among admissions eligible for matching in AR-G, 75% were successfully matched to CMS claims data. The observed characteristics of patients whose admissions did match were very similar to those of patients whose admissions did not match, including similar age, cardiac risk factors, and presentation heart rate and blood pressure (Table 2). Possible explanations for the failure of 25% of the admissions to match include admissions for patients ineligible for Medicare (e.g., non-U.S. citizens), admissions for patients in Medicare Advantage (not in fee-for-service Medicare) or with

non-governmental insurance, and inaccuracies within the CMS or AR-G data for linking fields (e.g., substituting age for date of birth).

Among admissions eligible for matching within the CMS claims dataset, 53% were successfully matched to AR-G data. Table 3 compares matched and unmatched admissions. Although age was similar between the two groups, fewer patients with subendocardial infarctions, history of congestive heart failure, and history of other comorbidities were found in the matched cohort. Possible explanations for mismatch include differences in selection criteria for the two databases, miscoding of principal discharge diagnoses in the CMS data, failure to include an eligible patient in AR-G, and data entry errors.

Table 2. Selected Patient Characteristics and Outcomes in AR-G Data for Patients Unmatched and Matched to CMS Data

Description	Unmatched (N=7,136) %	Matched (N=21,640) %
Demographics		
Age (y): Mean (SD)	76.2 (8.3)	77.0 (8.1)
Female	41.87	44.6
Race - White	84.9	90.3
Race - Black or African-American	8.9	6.8
Race - Other	12.9	8.4
History and Risk Factors		
Weight (kg): Mean (SD)	79.6 (20.1)	79.5 (20.0)
Current/Recent Smoker (w/in 1 year)	16.5	15.7
Hypertension	80.7	80.1
Dyslipidemia	62.8	62.9
Currently on Dialysis	3.2	2.5
Chronic Lung Disease	19.7	17.7
Diabetes Mellitus	37.2	34.5
Prior MI	29.5	27.9
Prior Heart Failure	20.2	18.4
Prior PCI	24.8	23.7
Prior CABG	19.9	20.1
Cerebrovascular Disease	17.7	17.6
Prior Stroke	12.0	11.4
Peripheral Arterial Disease	13.7	14.1
Cardiac Status on First Medical Contact		
STEMI or STEMI Equivalent	28.1	32.3
Heart Failure	24.1	23.1
Cardiogenic Shock	4.9	4.7
Heart Rate (beat/min): Mean (SD)	87.1 (25.7)	85.0 (24.3)
Systolic Blood Pressure (mm Hg): Mean (SD)	143.1 (34.7)	143.3 (33.6)
Baseline Creatinine (mg/dL): Mean (SD)	1.4 (1.1)	1.4 (1.0)
Baseline CrCl* (mL/min): Mean (SD)	58.4 (29.5)	57.9 (30.2)
Baseline Hemoglobin (g/dL): Mean (SD)	13.0 (2.1)	13.1 (2.0)
Baseline Troponin Ratio (×ULN): Mean (SD)	33.0 (200.5)	45.7 (292.9)

*Cockcroft-Gault formula

Table 3. Selected Patient Characteristics and Outcomes in CMS Data for Patients Unmatched and Matched to AR-G Data

Description	Unmatched (N=19,312) %	Matched (N=21,640) %
Demographics		
Age: Mean (SD)	78.2 (8.2)	77.0 (8.0)
Female	48.3	44.6
Principal discharge diagnosis		
410.0 (Anterolateral wall)	1.2	2.5
410.1 (Other anterior wall)	5.6	9.7
410.2 (Inferolateral)	1.0	2.3
410.3 (Inferoposterior)	0.7	1.5
410.4 (Other inferior)	6.4	14.2
410.5 (Other lateral)	0.8	1.5
410.6 (Posterior)	0.3	0.5
410.7 (Subendocardial)	78.8	63.8
410.8 (Other)	0.6	0.5
410.9 (Unspecified)	4.6	3.5
History and Risk Factors		
Percutaneous intervention	8.8	9.8
CABG surgery	5.5	5.7
Congestive heart failure	31.1	21.4
AMI	27.2	13.8
Unstable angina	17.1	10.7
Anterior myocardial infarction	6.8	12.2
Other location of myocardial infarction	9.1	20.0
Chronic atherosclerosis	79.4	84.0
Cardio-respiratory failure and shock	10.1	6.7
Valvular or rheumatic heart disease	25.3	21.0
Comorbidity		
Hypertension	80.5	79.1
Stroke	6.9	4.9
Cerebrovascular disease	16.8	14.7
Renal failure	22.4	16.5
Chronic obstructive pulmonary disease	26.8	22.2
Pneumonia	23.2	16.1
Diabetes and DM complications	42.1	38.7
Protein-calorie malnutrition	5.3	3.5
Dementia and senility	15.4	11.3
Hemiplegia, paralysis, functional disability	5.4	4.0
Vascular or circulatory disease	22.3	19.1
Metastatic cancer and acute leukemia	3.7	2.9
Trauma	24.1	20.8
Major psych disorders	5.4	4.3
Liver and biliary disease	1.1	0.6

Step 3: Exclusion criteria applied to the merged dataset

After performing the deterministic match, we applied exclusion criteria to the matched cohort to derive the final cohort of patients for building the risk-adjustment model. These exclusion criteria are similar to those in the currently publicly reported claims-based AMI mortality measure.¹

The following exclusions were applied to the merged dataset:

- 1) Discharged against medical advice (AMA): Admissions in which the patient was discharged AMA were removed from the matched dataset.
Rationale: Patients who leave AMA do not allow the hospital to provide the entire spectrum of necessary care for management of AMI.
- 2) Transfer-in admissions: Among patients transferred from one acute care institution to another, the second admission with an AMI was not eligible as an index admission. We used the CMS data to define transfers as two admissions that occur within one day of each other.
Rationale: We assign the outcome for the acute episode of care to the first admitting hospital because the first hospital initiates patient management and is responsible for any decision to transfer the patient. Therefore, the first admission in an acute episode of care is eligible to be an index admission in the measure. The second admission and any subsequent admissions in the same acute episode are excluded from the measure.
- 3) Admissions with missing death: Records with missing vital status were excluded.
Rationale: Records with no vital status information would prevent ascertainment of the mortality outcome.
- 4) Admissions with unreliable/missing data: Records with unreliable or missing data for age or sex were excluded.
Rationale: Unreliable or missing data limit the validity of the risk-adjustment model.
- 5) Multiple AMI admissions in 2009: We randomly selected one admission to retain and excluded the other admissions for patients in the merged AR-G-CMS dataset who had multiple admissions for AMI within the year.
Rationale: Episodes of care must be mutually independent, each with the same probability of the outcome. For patients with multiple admissions in a year, the probability of death increases with each subsequent admission, and therefore the episodes of care are not mutually independent. We therefore randomly select one admission for inclusion in the measure.

Each exclusion criterion was evaluated for EHR feasibility using the feasibility criteria detailed in Section 3.3.

4.5 Model Development

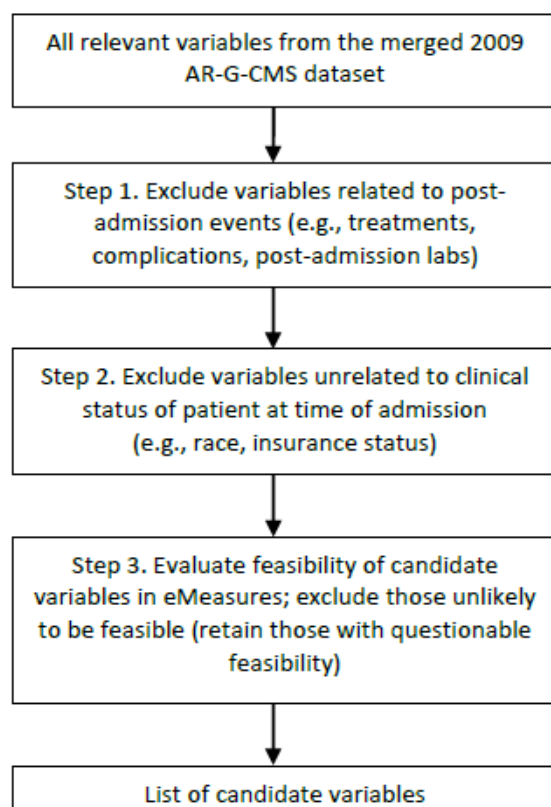
4.5.1 Candidate Risk-adjustment Variables

We sought to develop a model that included key variables that are clinically relevant, demonstrate a strong statistical association with 30-day mortality, and are feasible for use in an eMeasure. Although EHRs likely will ultimately link across clinical episodes of care and contain historical

patient data, given the EHR environment at the time of measure development and inability to reliably obtain data from the outpatient setting prior to admission, we only considered for inclusion in the measure variables that would be available and consistently collected at admission. To select candidate variables, the members of the working group reviewed the entire list of variables in the AR-G registry database. (The complete list of variables can be found at: <http://www.ncdr.com/WebNCDR/Action/Elements.aspx>.) These variables have undergone extensive vetting by clinical and methodological experts during development of the AR-G registry.^{11,14} To identify clinically meaningful variables to review for the candidate variable selection process, we excluded irrelevant variables not suitable for use in risk adjustment (e.g., patient name, physician name, etc.). In addition, we combined certain variables to derive other clinically meaningful variables; for example, we derived body mass index (BMI) from height and weight.

We applied a series of exclusion criteria to the remaining 193 variables to obtain a list of candidate variables for building the model. Refer to Figure 2 for the variable selection strategy. Refer to Appendix E for a list of variables excluded at each step.

Figure 2. Candidate Variable Selection Process Flow Chart



Step 1: Exclusion of variables related to post-admission events

We excluded all variables pertaining to post-admission events such as treatments, complications, and post-admission labs. This resulted in the exclusion of 121 variables.

Step 2: Exclusion of variables unrelated to the clinical status of the patient at the time of admission

Next, we excluded remaining variables that were unrelated to the clinical status of the patient at the time of admission, such as insurance status, patient ZIP code, means of transfer to the first facility, race, etc. This resulted in the exclusion of an additional 30 variables.

Step 3. Exclusion of variables not feasible for use in an eMeasure

As described in Section 3.3, we sought to develop a measure that was feasible for use in current EHR systems at the time of development. We developed the following criteria to assess eMeasure feasibility of candidate variables:

1. Consistently obtained in the target population based on current clinical practice
2. Captured with a standard definition and recorded in a standard format
3. Entered in structured fields that are feasibly retrieved from current EHR systems

Through discussions with the EHR experts and examination of the data, we assessed each variable by these criteria. Variables satisfying all three criteria were deemed feasible for use in an eMeasure given the current EHR environment (Table 4).

Variables clearly not fulfilling one or more of the criteria were deemed not feasible for use in an eMeasure given the current EHR environment. For example, brain natriuretic peptide (BNP) is not consistently obtained for patients with AMI. Thus, although when it is obtained BNP is captured using a standard definition, recorded in a standard format, and entered in a structured field, BNP was not considered feasible for this eMeasure. As another example, heart failure on presentation is consistently obtained in patients with AMI; however, the definition of what constitutes heart failure varies among providers. As a final example, ECGs are consistently obtained in patients with AMI but are not entered in a structured field that is feasibly retrieved from current EHR systems.

In some cases, our review determined that certain variables questionably fulfilled one or more criteria and were thus deemed “questionably feasible” in the current EHR environment. For example, it is unclear how frequently history of peripheral arterial disease is captured using a standard definition or recorded in structured fields in current EHRs. To maximize inclusiveness at this stage, we retained the candidate variables deemed “questionably feasible” in the candidate variable selection process.

Table 4. eMeasure Feasibility of Candidate Variables

Variable	Consistently obtained in target population based on current clinical practice	Captured with a standard definition and recorded in a standard format	Entered in structured fields that are feasibly retrieved from current EHR systems
1. Candidate variables deemed to fulfill all three criteria required for eMeasure feasibility			
Age	✓	✓	✓
Sex	✓	✓	✓
Heart Rate at First Medical Contact (bpm)	✓	✓	✓
Systolic Blood Pressure at First Medical Contact (mm Hg)	✓	✓	✓
Body Mass Index (BMI) (kg/m ²)	✓	✓	✓
Initial Troponin Ratio*	✓	✓	✓
Initial Creatinine Clearance (mL/min)	✓	✓	✓
Initial Creatinine Value (mg/dL)	✓	✓	✓
Initial Hemoglobin Value (g/dL)	✓	✓	✓
2. Candidate variables deemed to have questionable feasibility in current EHR environment			
History of Hypertension (No/Yes)	✓	✓	?
History of Dyslipidemia (No/Yes)	✓	✓	?
Currently on Dialysis (No/Yes)	✓	✓	?
History of Chronic Lung Disease (No/Yes)	✓	✓	?
History of Diabetes Mellitus (No/Yes)	✓	✓	?
Prior MI (No/Yes)	✓	✓	?
Prior Stroke (No/Yes)	✓	?	?
History of Peripheral Arterial Disease (No/Yes)	✓	?	?
Prior Percutaneous Coronary Intervention (No/Yes)	✓	?	?
Prior CABG (No/Yes)	✓	?	?
Prior Heart Failure (No/Yes)	✓	?	?
Current/Recent Smoker (w/in 1 year) (No/Yes)	✓	?	?
Atrial Fibrillation or Flutter in the Past 2 Weeks (No/Yes)	?	?	?

Variable	Consistently obtained in target population based on current clinical practice	Captured with a standard definition and recorded in a standard format	Entered in structured fields that are feasibly retrieved from current EHR systems
3. Candidate variables deemed not feasible for use in eMeasures given current EHR environment			
ST Segment Elevated Myocardial Infarction (STEMI) or STEMI Equivalent (No/Yes)	✓	×	×
ECG Findings for STEMI Equivalent (Selections: ST Elevation; LBBB; Isolated Posterior MI)	✓	×	×
Other ECG Findings (Selections: New or Presumed New ST Depression, New or Presumed New T-Wave Inversion, Transient ST Elevation Lasting <20 Minutes, None)	✓	×	×
Heart Failure at First Medical Contact (No/Yes)	✓	×	×
Cardiogenic Shock at First Medical Contact (No/Yes)	✓	×	×
Diabetes Therapy (Selections: None, Diet, Oral, Insulin, Other)	×	×	×
Most Recent Percutaneous Coronary Intervention Date	×	×	×
Most Recent CABG Date	×	×	×
Brain Natriuretic Peptide (BNP) (pg/mL)	×	✓	✓
Initial N-Terminal –proBNP Value (pg/mL)	×	✓	✓
History of Cerebrovascular disease (No/Yes) (Includes history of stroke, transient ischemic attack, >79% occlusion by imaging, or prior carotid artery surgery or intervention)	×	×	×
Initial CK-MB Value	×	✓	✓
Initial CK-MB ULN	×	✓	✓
Initial Hemoglobin A1c Value	×	✓	✓
INR Value	×	✓	✓
Total Cholesterol (mg/dL)	×	✓	✓
HDL Cholesterol (mg/dL)	×	✓	✓
LDL Cholesterol (mg/dL)	×	✓	✓
Triglycerides (mg/dL)	×	✓	✓
*Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)			

After these three steps were applied, 22 variables remained candidates for inclusion in the final model (see Table 5).

Table 5. Model Candidate Variables

Description
Demographics
Age
Sex
Cardiac Status On First Medical Contact
Heart Rate at First Medical Contact (bpm)
Systolic Blood Pressure at First Medical Contact (mm Hg)
History and Risk Factors
BMI
Current/Recent Smoker (w/in 1 year) (No/Yes)*
History of Hypertension (No/Yes)*
History of Dyslipidemia (No/Yes)*
Currently on Dialysis (No/Yes)*
History of Chronic Lung Disease (No/Yes)*
History of Diabetes Mellitus (No/Yes)*
Prior MI (No/Yes)*
Prior Heart Failure (No/Yes)*
Prior PCI (No/Yes)*
Prior CABG (No/Yes)*
Atrial Fibrillation or Flutter Past 2 Weeks (No/Yes)*
Prior Stroke (No/Yes)*
History of Peripheral Arterial Disease (No/Yes)*
Laboratory Results
Initial Creatinine Value (mg/dL)
Initial Hemoglobin Value (g/dL)
Troponin Ratio** (ng/mL)
Creatinine Clearance (mL/min)

*Variables with questionable eMeasure feasibility given current EHR environment

**Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)

4.5.2 Selection of Final Risk-adjustment Variables

We examined distributions of the 22 candidate variables. For missing “Yes/No” categorical variables, we assumed a “No” response. For all continuous variables, to reduce the effect of spurious outliers, we transformed extreme values by replacing them with a value at the outer limit of a designated range by a process called Winsorization.^{15,16} All continuous variables were initially Winsorized to the 1st and 99th percentiles (i.e., values less than the 1st percentile were assigned to the value of the 1st percentile, and values greater than the 99th percentile were assigned to the value of the 99th percentile). The variables were then plotted against 30-day mortality rates and further Winsorized as appropriate to the clinically meaningful values or derived as simple regression splines^{17,18} (see Appendix F). For missing values for BMI, we imputed sex-specific

median values. For all other continuous variables, we imputed the median value of the entire group.¹⁹

After Winsorization of the continuous variables, with the pre-selected candidate variables and the outcome of 30-day mortality, we performed a bootstrap simulation with 1,000 iterations by allowing patients to be selected repeatedly. In each iteration, a bootstrap data sample was constructed and a logistic regression model with stepwise selection (entry variables with $p < 0.05$; retained variables with $p < 0.01$) was performed over all the candidate variables. Lastly, we summarized the model information of all 1,000 iterations on the following: number and frequency of times that a variable is selected (e.g., 70% would mean that the candidate variable was selected as significant at $p < 0.05$ in 70% of the iterations), minimum, maximum, and the range of the standardized coefficient for a selected variable. We also assessed the direction and magnitude of the distribution of regression coefficients.

The working group reviewed the results of the bootstrap simulation and decided to retain all risk-adjustment variables above a 90% cutoff (i.e., the variables were selected as significant at $p < 0.05$ in 90% of the iterations), which was thought to demonstrate a consistently strong association with mortality. All variables selected less than 90% of the time in 1,000 iterations were excluded except heart rate < 70 bpm, which was included based on integrity of a variable, as its counterpart, heart rate > 70 bpm, remained in the model. The resulting preliminary risk-adjustment model consisted of nine variables, including five variables deemed feasible for use in eMeasures and four variables with questionable eMeasure feasibility.

To create a model with increased usability while retaining excellent model performance, we tested the performance of the model without those variables considered to be questionably feasible and compared it with that of the model containing the variables considered to be questionably feasible. Based on the results of that testing, the final parsimonious risk-adjustment model consisted of five variables that were clinically relevant and deemed to be eMeasure feasible (Table 6).

Table 6. Description of Preliminary and Final Risk-adjustment Models

	Preliminary model (Contains variables with questionable eMeasure feasibility)	Final model (Contains only variables deemed feasible in eMeasures)
Age (years)	✓	✓
Heart Rate: HR<70 (10 bpm)	✓	✓
Heart Rate: HR≥70 (10 bpm)	✓	✓
Systolic Blood Pressure (10 mm Hg)	✓	✓
Troponin Ratio** (ng/mL) (per 10 units)	✓	✓
Creatinine (mg/dL)	✓	✓
History of Dyslipidemia* (No/Yes)	✓	
Prior PCI* (No/Yes)	✓	
Prior Heart Failure* (No/Yes)	✓	
Prior Stroke* (No/Yes)	✓	

*Variables have questionable feasibility in eMeasures given current EHR environment

**Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)

4.6 Statistical Approach to Model Development

4.6.1 Logistic Regression Model and Hierarchical Logistic Regression Model

For model development and calculation of the hospital RSMR, we estimated two types of regression models using the combined CMS-AR-G dataset. First, we fit a generalized logistic regression model linking the outcome to the risk factors.²⁰ Let Y_{ij} denote the outcome (equal to 1 if patient dies within 30 days, zero otherwise) for the j th patient who presented with an AMI at the i th hospital; \mathbf{Z}_{ij} denotes a set of risk factors based on the administrative data. Let I denote the total number of hospitals and n_i the number of index admissions to hospital i . We assume the outcome is related linearly to the covariates via a known linked function, h , where

$$\text{LRM} \quad h(Y_{ij}) = \alpha + \beta \mathbf{Z}_{ij} \quad (1)$$

and $\mathbf{Z}_{ij} = (Z_{1ij}, Z_{2ij}, \dots, Z_{pij})$ is a set of p patient-specific covariates. In our case, h = the logit link, which is the logistic regression model.

To account for the natural clustering of observations within hospitals, we estimated a hierarchical logistic regression model that links the risk factors to the same outcome and a hospital-specific random effect,

$$h(Y_{ij}) = \alpha_i + \beta \mathbf{Z}_{ij} \quad (2)$$

$$\alpha_i = \mu + \omega_i; \quad \omega_i \sim N(0, \tau^2) \quad (3)$$

where h = the logit link, α_i represents the hospital-specific intercept, \mathbf{Z}_{ij} is defined as above, μ is the adjusted average outcome over all hospitals in the sample, and τ^2 is the between-hospital variance component.²¹ This model separates within-hospital variation from between-hospital variation. Both hierarchical logistic regression models and logistic regression models are estimated using the SAS software system (GLIMMIX and LOGISTIC procedures, respectively).

We first fit the logistic regression model described in Equation (1) using the logit link for the model development and model performance.

Having identified the covariates that remained, we next fit the hierarchical logistic regression models described in Equations (2) and (3), again using the logit link function; i.e.,

$$\text{Logit } (P(Y_{ij} = 1)) = \alpha_i + \beta \mathbf{Z}_{ij}$$

$$\alpha_i = \mu + \omega_i; \quad \omega_i \sim N(0, \tau^2)$$

where \mathbf{Z}_{ij} consisted of the covariates retained in the logistic regression model. As before, $Y_{ij} = 1$ if patient j treated at hospital i had the event; 0 otherwise.

4.6.2 Calculation of Hospital-Specific RSMRs

With the hierarchical logistic regression model, we calculated hospital-specific RSMRs. These rates were calculated as the ratio of predicted to expected mortality, multiplied by the overall unadjusted mortality rate. The expected number of deaths in each hospital was estimated using its patient mix and the average hospital-specific intercept. The predicted number of deaths in each hospital was estimated given the same patient mix but an estimated hospital-specific intercept. Operationally, the expected number of deaths for each hospital was obtained by regressing the risk factors on the mortality outcome using all hospitals in our sample, applying the subsequent estimated regression coefficients to the patient characteristics observed in the hospital, adding the average of the hospital-specific intercepts, transforming, and then summing over all patients in the hospital to get a value. This is a form of indirect standardization. The predicted hospital outcome is the number of deaths in the specific hospital estimated given its performance and case mix. Operationally, this was accomplished by estimating a hospital-specific intercept that herein represents baseline mortality risk within the hospital, applying the estimated regression coefficients to the patient characteristics in the hospital, transforming, and then summing over all patients in the hospital to get a value.

Using the set of risk factors in the logistic regression model, we fitted the hierarchical generalized logistic regression models defined by Equations (2) and (3) and estimated the corresponding parameters. We calculated a standardized outcome, s_i , for each hospital by computing the ratio of the predicted to expected mean outcomes, multiplied by the unadjusted mean mortality rate. Specifically, we calculated:

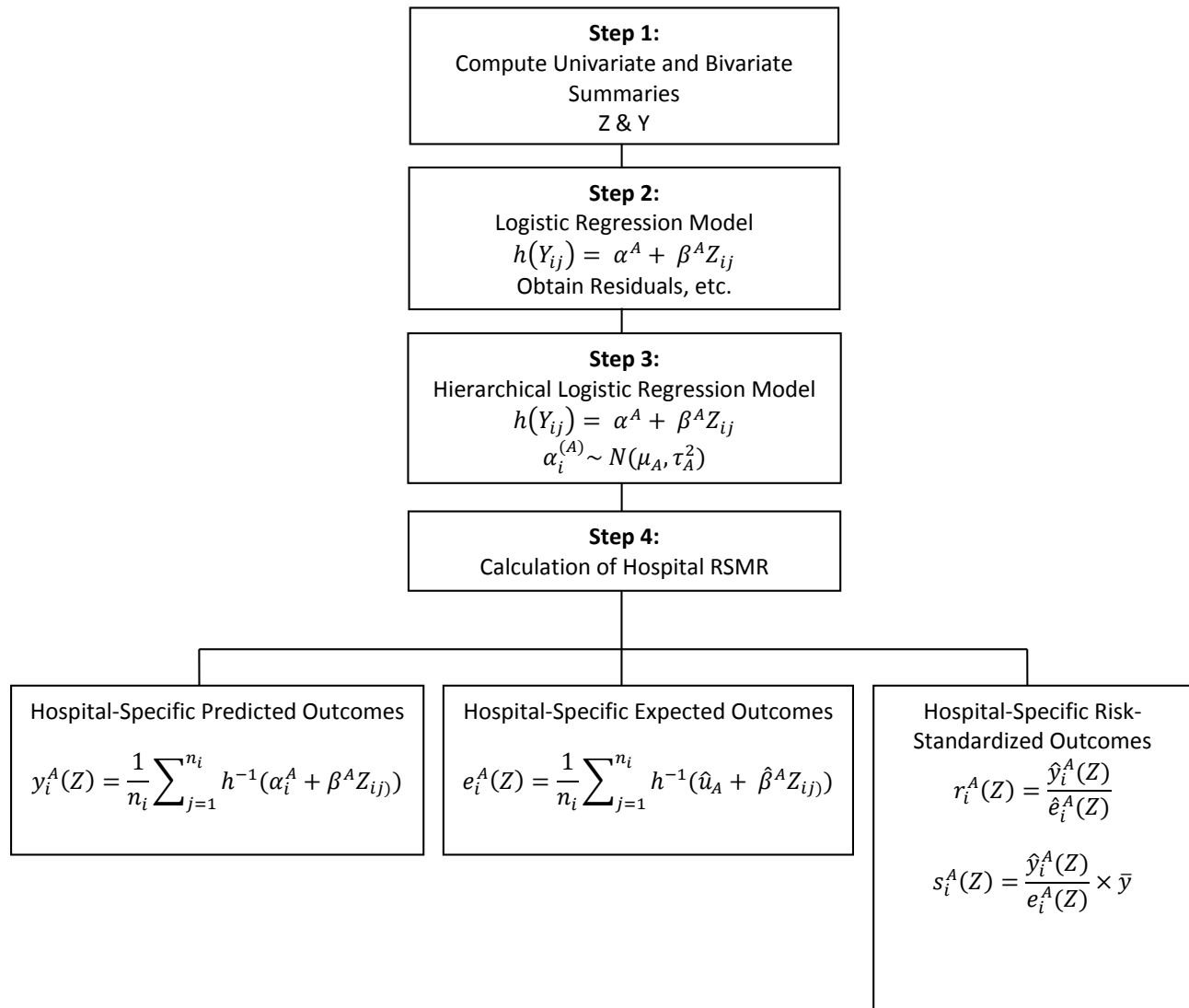
$$\text{Predicted} \quad \hat{y}_{ij}(Z) = h^{-1}(\hat{\alpha}_i + \hat{\beta} \mathbf{Z}_{ij}) \quad (4)$$

$$\text{Expected} \quad \hat{e}_{ij}(Z) = h^{-1}(\hat{\mu} + \hat{\beta} \mathbf{Z}_{ij}) \quad (5)$$

$$\hat{s}_i(Z) = \frac{\sum_{j=1}^{n_i} \hat{y}_{ij}(Z)}{\sum_{j=1}^{n_i} \hat{e}_{ij}(Z)} \times \bar{y} \quad (6)$$

If more (fewer) cases than “expected” have the outcome in a hospital, then the hospital risk-standardized outcome will be higher (lower) than the unadjusted average. See Figure 3 for analysis steps.

Figure 3. Analysis Steps



4.7 Model Testing

This section describes testing performed using the AR-G registry data to assess the reliability and validity of the risk-adjustment model. Further testing performed on the eSpecified eMeasure is referred to throughout this report as “eMeasure testing” and is described in Appendix B; this includes additional validity, feasibility, and usability testing.

4.7.1 Reliability

Reliability of data elements

We conducted data element reliability testing using the fully eSpecified eMeasure, as described in Appendix B.

4.7.2 Validity

Model validation

To assess the validity of the model, we constructed a dataset as described in Section 4.4, except using hospital discharges from the AR-G registry and Medicare claims files from January 1, 2010 through December 31, 2010 (as opposed to 2009 for the derivation cohort). A validation model was created using the same five final model risk-adjustment variables. Summary characteristics were compared for the 2009 and 2010 models (see Section 5.3.1). We also examined the temporal variation of the odds ratios and 95% confidence intervals of the model variables in the 2009 dataset vs. the 2010 dataset.

Validity of data elements

Data element validity testing was performed in the fully eSpecified eMeasure by comparing the output of various EHRs with visual inspection of the EHR (see Appendix B).

Validity of measure score

To assess the validity of the measure score, we applied the model in the publicly reported claims-based AMI mortality measure to the study sample and calculated hospital RSMRs. Then we calculated the weighted Pearson correlation between the hospital RSMR based on the claims-based model and the hospital RSMR based on our final model.

The publicly reported claims-based AMI mortality measure was also previously validated with a comprehensive medical record model from an earlier time period. Specifically, claims-based model validation was conducted by building comparable models using abstracted medical record data for risk adjustment using Cooperative Cardiovascular Project data. When both models were applied to the same patient population, the hospital risk-standardized rates estimated using the claims-based risk-adjustment models had a high level of agreement with the results based on the medical record model, thus supporting the use of the claims-based model for public reporting (see Section 5.3.2).²² Thus, the claims-based AMI mortality model is suitable for validation of the measure score of the current model.

4.7.3 Disparities Assessment

We conducted analyses to explore disparities in AMI mortality by socioeconomic status (SES) and race at the hospital level. We used Medicaid eligibility status as identified in the Medicare EDB as a proxy for SES. This approach is consistent with prior research as well as NQF recommendations.²³ Hospitals were categorized into quintiles based on their proportion of patients eligible for both Medicaid and Medicare (dual-eligible patients). Similar analyses were conducted for the proportion of African-American patients in hospitals.

4.7.4 Sensitivity Analysis – Assessment of Variables Deemed Clinically Relevant but Not Feasible for Use in eMeasures

Individual variables' eMeasure feasibility may change over time, particularly with increasing adoption of and improving technology in EHRs. For the current measure, clinical experts assessed the clinical importance of those variables deemed not currently feasible for use in eMeasures. Although not feasible for inclusion in the current model, these variables may warrant additional

consideration for future models as EHRs evolve.

4.8 eSpecification

eSpecification is the process of converting a paper-based quality measure, or implementing a measure specifically developed for an EHR, into a format appropriate for the EHR environment. This process includes encoding the measure specifications in a standard eMeasure format known as HQMF. Refer to Appendix A for more information.

5. RESULTS

5.1 Preliminary Model (Containing Variables with Questionable eMeasure Feasibility)

5.1.1 Logistic Regression

The preliminary logistic regression model performed very well, with a C-statistic of 0.79 and an adjusted R-square of 0.22. The variable descriptions, estimates, and standard errors for the logistic regression model are shown in Table 7.

Table 7. Preliminary Model: Logistic Regression Results (N=20,540 patients)

Description	Estimate	SE	Chi Sq	Pr>Chi Sq	OR	95% CI
Intercept	-5.29	0.349	229	0.00		
Age (per year)	0.06	0.003	351	0.00	1.06	1.05, 1.07
Heart Rate: HR<70 (per 10 bpm)	-0.06	0.040	2	0.17	0.94	0.87, 1.03
Heart Rate: HR≥70 (per 10 bpm)	0.14	0.010	124	0.00	1.15	1.12, 1.18
Systolic Blood Pressure (per 10 mm Hg)	-0.25	0.010	545	0.00	0.78	0.76, 0.80
Troponin Ratio** (ng/mL) (per 10 units)	0.11	0.001	107	0.00	1.12	1.10, 1.15
Creatinine (per mg/dL)	0.63	0.038	282	0.00	1.88	1.75, 2.02
History of Dyslipidemia* (No/Yes)	-0.29	0.051	32	0.00	0.75	0.68, 0.83
Prior PCI* (No/Yes)	-0.27	0.064	17	0.00	0.77	0.68, 0.87
Prior Heart Failure* (No/Yes)	0.45	0.057	62	0.00	1.56	1.40, 1.74
Prior Stroke* (No/Yes)	0.30	0.068	19	0.00	1.35	1.18, 1.55

*Variables with questionable eMeasure feasibility given current EHR environment

**Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)

5.2 Final Model (Containing only eMeasure-feasible Variables)

5.2.1 Logistic Regression

The final logistic regression model performed very well, with a C-statistic of 0.78 and an adjusted R-square of 0.20. The variable descriptions, estimates, and standard errors for the logistic regression model using the final model are shown in Table 8.

Table 8. Final Model: Logistic Regression Results (N=20,540 patients)

Description	Estimate	SE	Chi Sq	Pr>Chi Sq	OR	95% CI
Intercept	-6.045	0.342	312	0.000		
Age (years)	0.063	0.003	453	0.000	1.07	1.06, 1.07
Heart Rate: HR<70 (10 bpm)	-0.051	0.042	2	0.217	0.95	0.88, 1.03
Heart Rate: HR≥70 (10 bpm)	0.150	0.013	140	0.000	1.16	1.13, 1.19
Systolic Blood Pressure (10 mm Hg)	-0.249	0.011	555	0.000	0.78	0.76, 0.77
Troponin Ratio** (ng/mL) (per 10 units)	0.118	0.011	117	0.000	1.13	1.10, 1.15
Creatinine (mg/dL)	0.671	0.037	336	0.000	1.96	1.82, 2.10

**Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)

5.2.2 Hierarchical Logistic Regression Model

In the final hierarchical logistic regression model, the estimated between-hospital variance in the log-odds of mortality was 0.0248 (standard error=0.0143). This result implies that the odds of mortality for a high-mortality hospital (+1 standard deviation) were 1.37 times those for a low-mortality hospital (-1 standard deviation). Model variable descriptions, estimates, standard errors, and odds ratios are shown in Table 9.

Table 9. Final Model: Hierarchical Logistic Regression Model Results (N=20,540 patients)

Description	Estimate	SE	T Value	Pr > t	OR	95% CI
Intercept	-6.050	0.333	-18.151	0.000		
Age (years)	0.063	0.003	21.826	0.000	1.07	1.06, 1.07
Heart Rate: HR<70 (10 bpm)	-0.050	0.040	-1.243	0.214	0.95	0.88, 1.03
Heart Rate: HR≥70 (10 bpm)	0.149	0.012	12.135	0.000	1.16	1.13, 1.19
Systolic Blood Pressure (10 mm Hg)	-0.249	0.010	-24.244	0.000	0.78	0.76, 0.80
Troponin Ratio** (ng/mL) (per 10 units)	0.121	0.011	11.285	0.000	1.13	1.11, 1.15
Creatinine (mg/dL)	0.670	0.036	18.852	0.000	1.95	1.82, 2.10

280 hospitals with between-hospital variance=0.0248, standard error=0.0143.

**Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)

5.2.3 30-day Mortality Rate Distribution

The hospital unadjusted 30-day mortality rate in 2009 data ranged from 0% to 60% across 280 hospitals with a median (interquartile range) of 10.5% (8.2%, 13.3%) (Figure 4). After adjusting for patient characteristics and clustering within hospitals, RSMRs at the hospital level were found to be more normally distributed, ranging from 9.6% to 13.1% across 280 hospitals. The median (interquartile range) RSMR was 10.7% (10.3%, 11.1%) (Figure 5).

Figure 4. Distribution of Hospital Unadjusted Mortality Rates (2009)

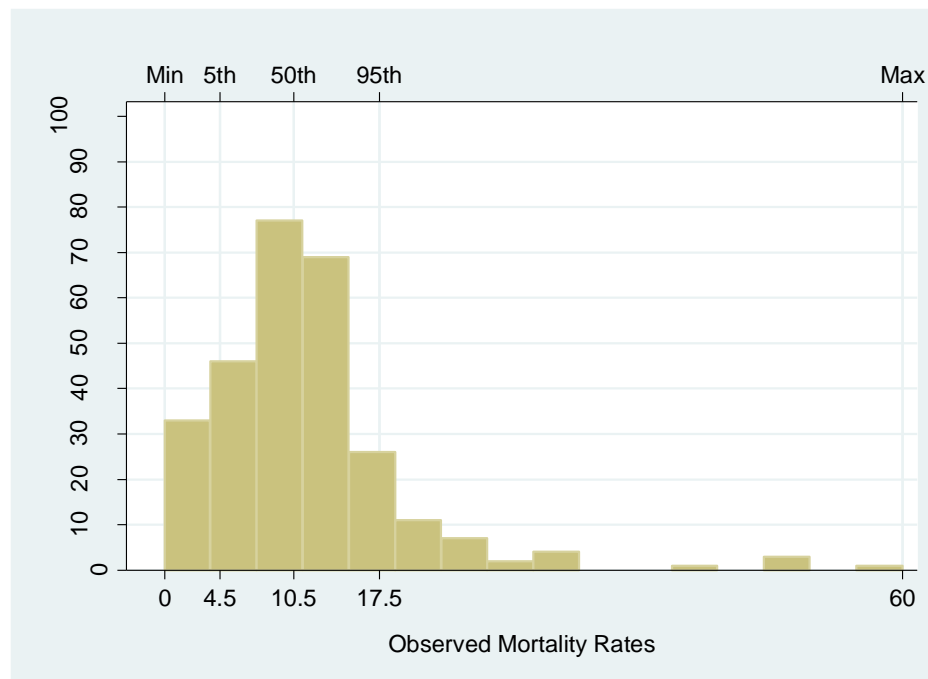
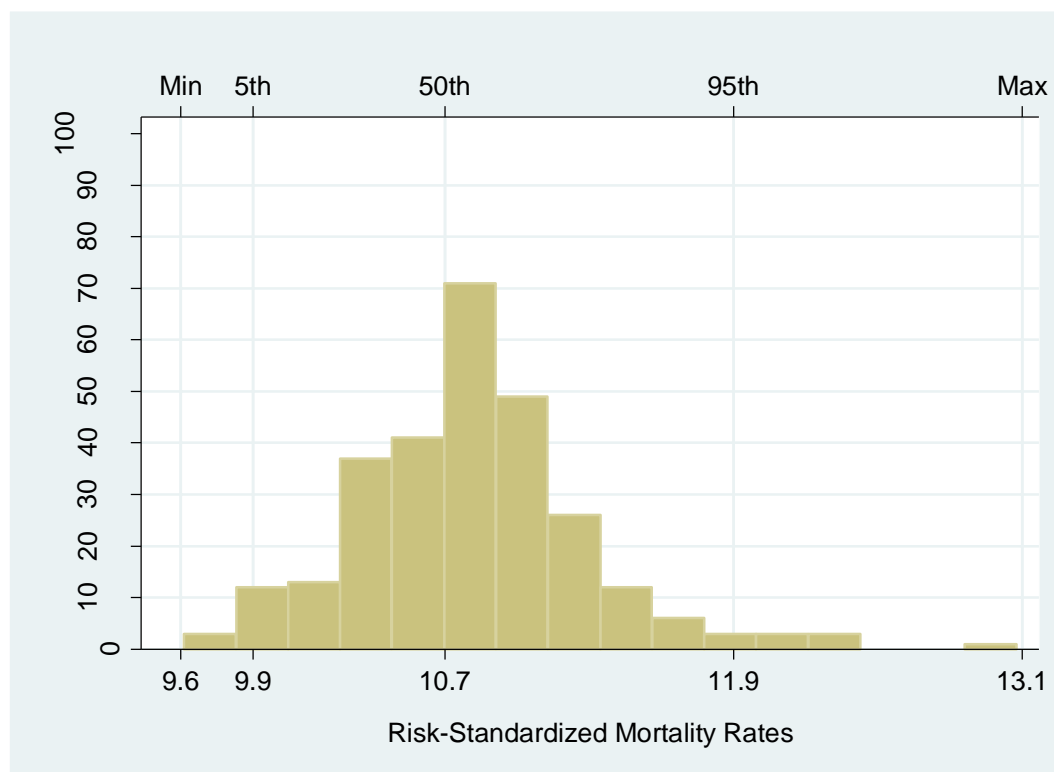


Figure 5. Distribution of Hospital Risk-standardized Mortality Rates (2009)



5.3 Model Assessment

5.3.1 Model Validation

We computed five summary statistics for assessing model performance²⁴: over-fitting indices,* predictive ability, area under the receiver operating characteristic (ROC) curve, distribution of residuals, and model chi-square.[†] The final model, originally developed with 2009 data, was validated using 2010 data. Due to the low sample size, we did not split the development sample. Model performance was similar in each dataset, with strong model discrimination and fit. Predictive ability was also similar across datasets. The C-statistic (area under the ROC curve) was 0.78 for both datasets (Table 10).

Table 10. Model Performance: Results Based on the Logistic Regression Model

Indices	2009 Derivation Sample	2010 Validation Sample
Number of Admissions	20,540	34,196
Mortality Rate	10.80	10.98
Calibration		
γ_0, γ_1	0.000, 1.000	-0.013, 0.979
Adjusted R-square	0.204	0.194
Discrimination		
Predictive Ability (lowest decile %, highest decile %)	0.012, 0.375	0.012, 0.374
C-statistic	0.78	0.78
Residuals Lack of Fit (Pearson Residual Fall %)		
<-2	0.015	0.000
[-2, 0)	89.187	89.019
[0, 2)	4.869	4.849
[2+	5.930	6.132
Model χ^2 (number of covariates)	1880.576 (6)	3029.846 (6)

We also examined the temporal variation of the odds ratios (95% confidence intervals) of the model variables. The odds ratios are consistent over the two years of data (Table 11).

* Over-fitting refers to the phenomenon in which a model describes the relationship between predictive variables and outcome in the development dataset well, but fails to provide valid predictions in new patients.

† Chi-square – A test of statistical significance usually employed for categorical data to determine whether there is a good fit between the observed data and expected values; i.e., whether the differences between observed and expected values are attributable to true differences in characteristics or instead are the result of chance variation. The formula for computing the chi-square is as follows:

$$\sum \frac{(O-E)^2}{E}$$

where O = observed value

E = expected value, and

degrees of freedom (df) = (rows-1)(columns-1)

Table 11. Final Model (Logistic Regression) Odds Ratios by Dataset

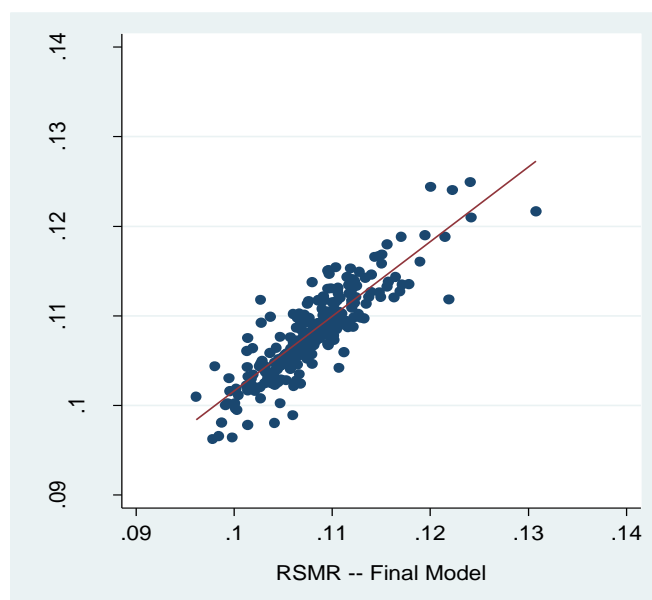
Description	2009 Development Sample	2010 Validation Sample
Age (years)	1.07 (1.06, 1.07)	1.07 (1.06, 1.07)
Heart Rate: HR<70 (10 bpm)	0.95 (0.88, 1.03)	0.99 (0.93, 1.05)
Heart Rate: HR≥70 (10 bpm)	1.16 (1.13, 1.19)	1.14 (1.12, 1.17)
Systolic Blood Pressure (10 mm Hg)	0.78 (0.76, 0.77)	0.78 (0.76, 0.79)
Troponin Ratio** (ng/mL) (per 10 units)	1.13 (1.10, 1.15)	1.12 (1.10, 1.14)
Creatinine (mg/dL)	1.96 (1.82, 2.10)	1.85 (1.75, 1.95)

**Troponin Ratio = Initial troponin value (ng/mL) / Troponin upper range limit (ng/mL)

5.3.2 Measure Score Validity Testing Results

We calculated the correlation of the RSMR from our final model with that of the previously validated, publicly reported, claims-based AMI mortality measure, using data from 2009. The correlation coefficient of 0.86 demonstrates excellent correlation (Figure 6).

Figure 6. Correlation of RSMR based on the Currently Proposed Final Model with RSMR based on the Previously Developed, Publicly Reported, Claims-Based AMI Mortality Measure (Hospital Volume-weighted Pearson Correlation Coefficient=0.86)

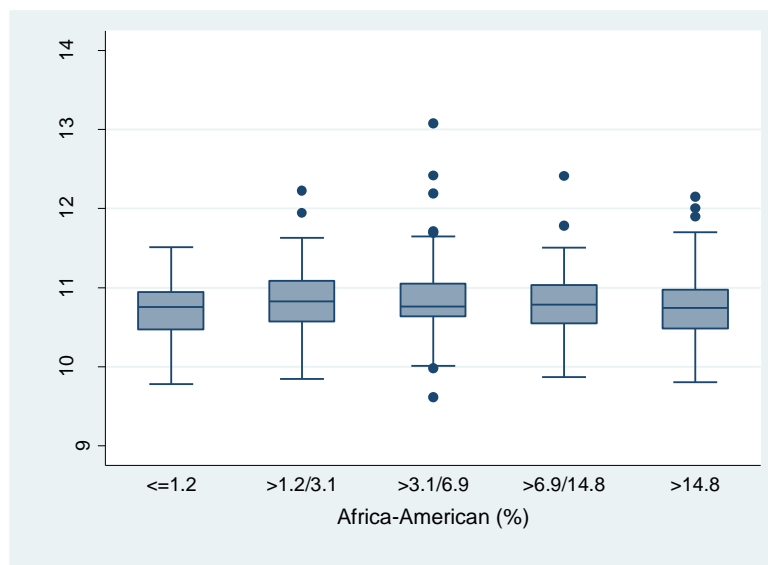


5.3.3 Disparities Assessment

RSMRs in the 2009 data were consistent across quintiles of hospitals based on the hospital proportion of African-American patients. Thus, hospitals with high proportions of African-

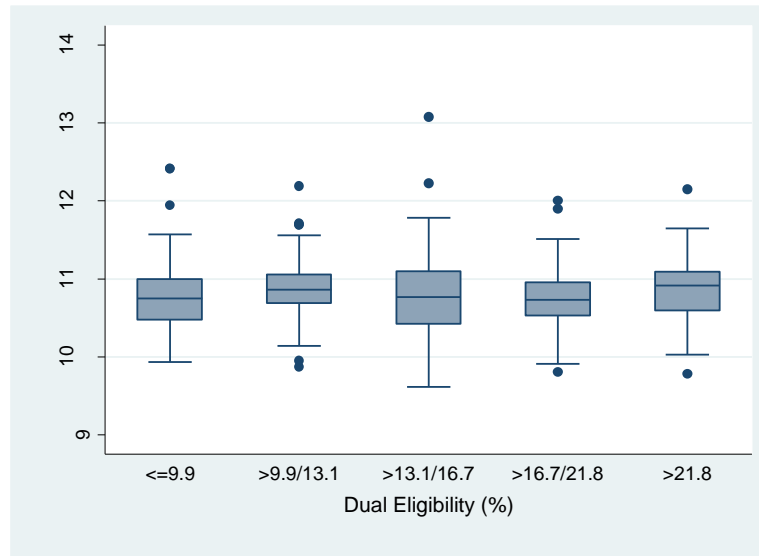
American patients generally performed as well on the measure as hospitals with lower proportions of African-American patients (Figure 7).

Figure 7. Hospital RSMR (2009) by Proportion of African-American Patients



Similarly, RSMRs in 2009 data were consistent across quintiles of hospitals based on the hospital proportion of dual eligible patients. This analysis suggests that that many hospitals with a high proportion of dual eligible patients performed well on the measure (Figure 8).

Figure 8. Hospital RSMR (2009) by Proportion of Dual Eligible Patients

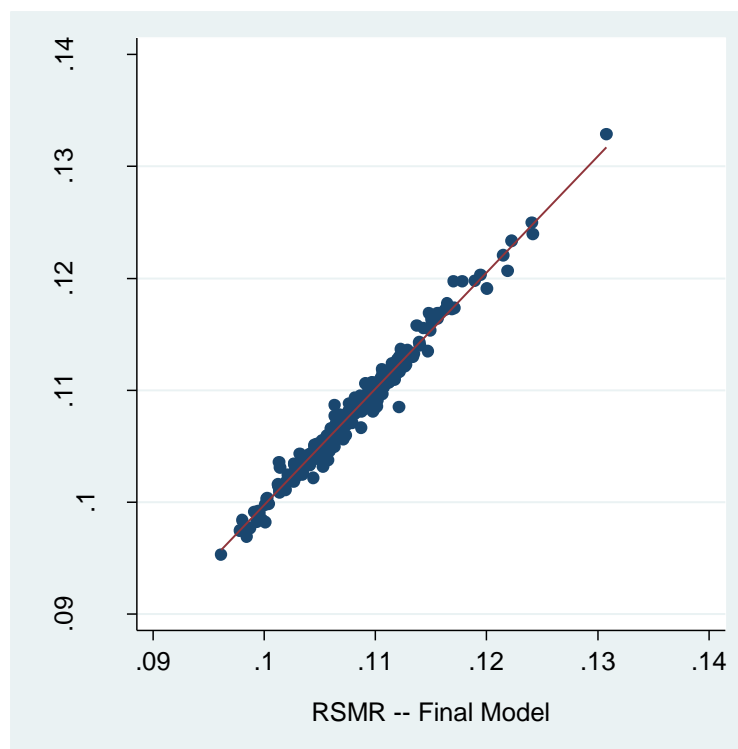


5.3.4 Sensitivity Analysis – Assessment of Variables Deemed Clinically Relevant but Not Feasible for Use in eMeasures

Clinical experts identified three variables – STEMI on the ECG, heart failure on admission, and cardiogenic shock on admission – as clinically important despite not being feasible for use in an eMeasure given the current EHR environment.

STEMI is identified on the ECG on presentation. While an ECG is consistently obtained on presentation in the target population based on current clinical practice, the results are not reliably recorded in a standard format nor entered in structured fields that are feasibly retrieved from current EHR systems. However, within the AR-G dataset, the ECG results are recorded in a standard format and entered in structured fields. Thus, although ECG findings did not meet the eMeasure feasibility criteria, we were able to evaluate the effect of including this variable in the model. The addition of ECG results to the final model in the 2009 data only increased the C-statistic from 0.78 to 0.80. This increase suggests that future models may be improved if the ECG results become more eMeasure-feasible (e.g., was captured in a structured format within EHRs); however, the resulting improvement in model performance will likely be modest. In addition, the correlation of RSMRs between the final model and the final model with ECG results was 0.989 (Figure 9). This high correlation confirms the low likelihood of substantial improvement in the eMeasure with the addition of ECG results.

Figure 9. Correlation between RSMR based on the Final Model and RSMR based on the Final Model Plus ECG Results (Hospital Volume-weighted Correlation Coefficient=0.989)



Heart failure on admission and cardiogenic shock on admission are also consistently obtained in current clinical practice. However, definitions of these variables are inconsistent, and their reliability is limited^{25,26}; thus, the criteria for being captured in a standard format and entered in structured fields are not met. Given this questionable reliability of the data elements, assessment of the incremental value of including these variables would not be helpful.

6. SUMMARY STATEMENT

We developed a hospital 30-day all-cause risk-standardized mortality eMeasure for AMI admissions. This measure was developed *de novo* using clinical registry data through a deliberate process to select only those variables feasible for use in an eMeasure.

- The measure was developed using clinical registry data from the NCDR AR-G merged with administrative claims data from CMS. The measure was developed with extensive input from clinical, EHR, and methodological experts with knowledge and experience relevant to quality measurement of AMI.
- The cohort consists of hospitalizations for patients admitted to a short-term acute care facility with a principal diagnosis of AMI.
- The outcome is all-cause mortality within 30 days of admission.
- In the model, we included only those risk-adjustment variables deemed currently “eMeasure-feasible” at the time of development – meeting all three of the following requirements:
 - Consistently obtained in the target population based on current clinical practice
 - Captured with a standard definition and recorded in a standard format
 - Entered in structured fields that are feasibly retrieved from current EHR systems
- The hierarchical modeling accounts for hospital case mix, hospital sample size, and the clustering of patients within hospitals, thereby making the measure suitable for public reporting.
- The final model consists of five clinical variables that are present on admission and eMeasure-feasible:
 - Age
 - Heart rate
 - Systolic blood pressure
 - Troponin ratio
 - Creatinine
- Of note, three variables that did not meet the requirements for eMeasure feasibility were identified to be particularly important to the clinical community – ST segment elevation myocardial infarction on the ECG, presence of heart failure, and presence of cardiogenic shock, all on admission. The clinical importance of these variables may warrant efforts to improve their eMeasure feasibility for consideration in future models.
- The final model performed very well, with a C-statistic of 0.78. In addition, we confirmed measure score validity by testing the correlation of RSMR from our final model with that of the previously validated, publicly reported, claims-based AMI mortality measure. The correlation coefficient of 0.86 demonstrated excellent correlation.
- Testing of the eMeasure, described in Appendix B, demonstrated the overall feasibility and usability of the eSpecified eMeasure and the reliability of the data elements, and field testing indicated the overall validity of data elements used in the eMeasure. The eMeasure testing results indicated that, given the current EHR environment, implementation of hybrid measures

that use EHR data and data from other sources is likely more feasible in the short-term than implementing outcome eMeasures that use EHR data alone.

In summary, we have built one of the first outcome eMeasures that produces estimates of hospital risk-standardized mortality rates for Medicare patients with AMI and that can be used to evaluate hospital quality of care using the EHR. The eMeasure is consistent with the consensus standards for publicly reported outcome measures, is parsimonious in risk adjustment, and performs well compared with the previously validated, publicly reported, claims-based AMI mortality measure.

7. REFERENCES

1. Bernheim S, Wang Y, Grady J, et al. Measures Maintenance Technical Report: Acute Myocardial Infarction, Heart Failure, and Pneumonia 30-day Risk Standardized Mortality Measures. 2011; Available at: <http://qualitynet.org/>.
2. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes - An American Heart Association scientific statement from the quality of care and outcomes research interdisciplinary writing group - Cosponsored by the Council on Epidemiology and Prevention and the Stroke Council - Endorsed by the American College of Cardiology Foundation. *Circulation*. 2006;113(3):456-462.
3. American Recovery and Reinvestment Act of 2009. Available at: www.ama-assn.org/ama1/pub/upload/mm/.../arra-hit-provisions.pdf.
4. CMS EHR Meaningful Use Overview. Available at: https://www.cms.gov/EHRIncentivePrograms/30_Meaningful_Use.asp.
5. Chaudhry B, Wang J, Wu S, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med*. May 16 2006;144(10):742-752.
6. Roger VL, Go AS, Lloyd-Jones DM, et al. Heart Disease and Stroke Statistics--2012 Update: A Report From the American Heart Association. *Circulation*. Jan 3 2012;125(1):e2-e220.
7. Yeh RW, Sidney S, Chandra M, Sorel M, Selby JV, Go AS. Population trends in the incidence and outcomes of acute myocardial infarction. *N Engl J Med*. Jun 10;362(23):2155-2165.
8. Health Level 7. Available at: <https://www.hl7.org>.
9. National Quality Forum. National voluntary consensus standards for patient outcomes, first report for phases 1 and 2: A consensus report. 2010; http://www.nysna.org/images/pdfs/practice/nfq_ana_outcomes_draft10.pdf. . Accessed August 19, 2010.
10. Drye EE, Normand SL, Wang Y, et al. Comparison of hospital risk-standardized mortality rates calculated by using in-hospital and 30-day models: an observational study with implications for hospital profiling. *Ann Intern Med*. Jan 3 2012;156(1 Pt 1):19-26.
11. Peterson ED, Roe MT, Chen AY, et al. The NCDR ACTION Registry-GWTG: transforming contemporary acute myocardial infarction clinical care. *Heart*. Nov 2010;96(22):1798-1802.
12. Messenger JC, Ho KKL, Young CH, et al. The National Cardiovascular Data Registry (NCDR) Data Quality Brief: The NCDR Data Quality Program in 2012. *J Am Coll Cardiol*. September 19, 2012; Published online. doi:10.1016/j.jacc.2012.07.020.
13. Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly. The advantages of a merged data base for Medicare and Veterans Affairs hospitals. *Med Care*. May 1992;30(5):377-391.
14. Peterson ED, Roe MT, Rumsfeld JS, et al. A call to ACTION (acute coronary treatment and intervention outcomes network): a national effort to promote timely clinical feedback and support continuous quality improvement for acute myocardial infarction. *Circ Cardiovasc Qual Outcomes*. Sep 2009;2(5):491-499.
15. Attenburg HP. Estimation of radioimmunoassay data using robust nonlinear regression methods. *Proceedings of the 10th Symposium on Computational Statistics*. 1992:367-372.
16. Dixon WJ, Yuen KK. Trimming and winsorization: A review. *Statistische Hefte*. 1974;15(2-3):157-170.
17. Bollaerts K, Eilers PHC, Mechelen Iv. Simple and multiple P-splines regression with shape constraints. *British Journal of Mathematical and Statistical Psychology*. 2006;59:451-469.
18. Keele L. Splines. *Semiparametric Regression for the Social Sciences*: John Wiley & Sons, Ltd.;

- 2008:49-83.
19. Chin CT, Chen AY, Wang TY, et al. Risk adjustment for in-hospital mortality of contemporary patients with acute myocardial infarction: the acute coronary treatment and intervention outcomes network (ACTION) registry-get with the guidelines (GWTG) acute myocardial infarction mortality model and risk score. *American heart journal*. Jan 2011;161(1):113-122 e112.
 20. McCullagh P, N.J. *Generalized Linear Models*: Chapman and Hall; 1989.
 21. Daniels MJ, Gatsonis C. Hierarchical Generalized Linear Models in the Analysis of Variations in Health Care Utilization. *J. Am. Stat. Assoc.* 1999;94(445):29-42.
 22. Krumholz HM, Wang Y, Mattera JA, et al. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*. Apr 4 2006;113(13):1683-1692.
 23. Kindig D, Stoddart G. What Is Population Health? *American Journal of Public Health*. March, 2003;93(3):380-383.
 24. Harrell F, Shih Y. Using full probability models to compute probabilities of actual interest to decision makers. *Int J Technol Assess Health Care*. Winter 2001;17(1):10.
 25. Barringhaus KG, Zelevinsky K, Lovett A, Normand SL, Ho KK. Impact of independent data adjudication on hospital-specific estimates of risk-adjusted mortality following percutaneous coronary interventions in massachusetts. *Circ Cardiovasc Qual Outcomes*. Jan 1 2011;4(1):92-98.
 26. Curtis JP, Rathore SS, Wang Y, Chen J, Nallamothu BK, Krumholz HM. Use and Effectiveness of Intra-Aortic Balloon Pumps Among Patients Undergoing High Risk Percutaneous Coronary Intervention: Insights From the National Cardiovascular Data Registry. *Circ Cardiovasc Qual Outcomes*. Jan 1 2012;5(1):21-30.
 27. Quality Data Model. Available at:
http://www.qualityforum.org/Projects/h/QDS_Model/Quality_Data_Model.aspx.
 28. Health Information Technology Standards Committee. Available at:
<http://healthit.hhs.gov/portal/server.pt?open=512&objID=1271&parentname=CommunityPage&parentid=6&mode=2>. .
 29. ONC HITSC Vocabulary Standards. Available at:
http://healthit.hhs.gov/portal/server.pt?open=512&objID=1818&parentname=CommunityPage&parentid=34&mode=2&in_hi_userid=11673&cached=true.

8. APPENDICES

Appendix A. eSpecification

The process of converting existing paper-based quality measures or encoding measures developed *de novo* for implementation in an electronic platform is known as “eSpecification.” The resulting measures, known as eMeasures, are encoded in the Health Quality Measure Format (HQMF), which is a Health Level 7 (HL7) standard for representing health quality measures in a machine-readable form (in this case, XML programming). HL7 is a standards development organization responsible for communication of healthcare data among applications.⁸ Encoding of measures in HQMF enables consistency and standardization of different health quality measure structures being developed for the electronic platform.

The Health Information Technology Expert Panel (HITEP), consisting of content experts convened by NQF, developed the Quality Data Model (QDM) in collaboration with the Office of the National Coordinator for Health Information Technology (ONC), American Health Information Community (AHIC), and the Agency for Healthcare Research and Quality (AHRQ). The QDM provides a standard structure and grammar to represent health quality measures precisely and accurately in an electronic environment that can be used across electronic patient care systems.²⁷

The Health Information Technology Standards Committee (HITSC) was formed to make recommendations to ONC on standards, implementation specifications, and certification criteria for the electronic exchange and use of health information.²⁸ The HITSC evaluated and recommended a minimum set of different vocabulary standards for defining an eMeasure. These vocabulary standards consist of different coding systems used for defining different types of data elements. For example, the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT) is used to define the QDM concept of “Condition/Diagnosis/Problem,” while RxNorm provides identifiers for clinical drugs.²⁹

Following development of our final model, we created value sets consisting of codes from the appropriate code taxonomies in preparation for eSpecification of the model. We collaborated with Abt Associates to construct the human-readable file, the machine-readable file, and the specific value sets for the eMeasure.

8.3.1 Overview

Following the development and eSpecification of our eMeasure, we worked with Abt Associates, another CMS contractor, to perform eMeasure testing. This testing consisted of two major components: alpha testing and beta testing. The alpha testing component assessed the usability, feasibility, and reliability of the eMeasure; the beta testing component assessed the validity of the data elements included in the eMeasure.

8.3.2 Alpha Testing

Alpha testing of the eMeasure consisted of three parts. Two were qualitative surveys: the hospital information technology (IT)/quality expert survey, and the electronic health record (EHR) vendor survey. The third was the quantitative data element reliability (XML Schema and Schematron) testing, which assessed how well the eMeasure specification conforms to technical standards.

Hospital IT/quality expert survey

Overview

The purpose of the hospital IT/quality expert survey was to determine the overall usability and feasibility of the eMeasure. We received survey responses from experts representing seven hospitals. These experts are the individuals who would be responsible for writing the specific queries that would extract our eMeasure from the hospitals' EHRs.

The survey asked experts to describe the level of ease with which they understood the human-readable form of the AMI mortality eMeasure and how the data elements included in the eMeasure were stored in their hospital's EHR (e.g., structured field vs. free text). The experts also had the opportunity to provide comments on the eMeasure's usability and feasibility and the specific aspects of the eMeasure that they found confusing or difficult.

Results

Five of the seven experts indicated that they could understand the AMI mortality eMeasure well enough to write or run reports or extract the data necessary to calculate the eMeasure. Of the remaining two experts, one indicated difficulty understanding the measure and one did not complete this section of the survey.

The IT/quality experts indicated that most data elements included in the AMI mortality eMeasure were stored as structured data, allowing data elements to be extracted easily. However, they noted two exceptions that may pose challenges to implementation of the eMeasure. First, they noted a potential difficulty in extracting a hospital's upper limit of normal for troponin, a variable used in the risk adjustment to calculate the patient-to-hospital troponin ratio. Second, they foresaw difficulties in linking emergency department (ED) data to inpatient data, which is necessary to identify the first collected value of the risk-adjustment variables. Additionally, they indicated that electrocardiogram (ECG) results are not stored as structured data in current EHRs. This last finding confirmed our decision not to include this element in our final model.

The results of the IT/quality expert survey indicated that overall, the experts understood the eMeasure well and that the majority of the data elements could be extracted from the EHR easily.

EHR vendor survey

Overview

The purpose of the EHR vendor survey was to further determine the feasibility of the eMeasure. Individuals from nine EHR vendors representing 85% of the current EHR market completed the survey. This survey was not specific to the AMI mortality eMeasure but was a generic survey whose results could be applied to all eMeasures. The survey asked the vendors to assess whether 122 data types and their attributes, as defined by the Quality Data Model (QDM), could feasibly be retrieved as structured data from their existing hospital EHRs. Vendors also indicated which data types in each QDM category were currently stored in their EHRs as structured data, which data types they could develop as structured data within 18 months, and the potential burden associated with the development of these data types within the EHR.

Abt Associates analyzed the results of the EHR vendor survey and, based on these results, assigned each QDM data type a feasibility score that ranged from 1 to 4 (Appendix Table 1). They also compiled a report specific to the data types included in the AMI mortality eMeasure.

Appendix Table 1. Feasibility Survey Options and Feasibility Scores (Developed by Abt Associates)

Response	Score
No, we do not provide it now, and could not or would not develop it within 18 months	1
No, we do not provide it now but could develop it within 18 months with moderate to major burden	2
No, we do not provide it now, but could develop it within 18 months with relatively minor burden	3
Yes, we provide it now or are already working on it	4

A response assigned a score of 1 indicated that the data type was not feasible for use in eMeasures, whereas a score of 4 indicated that the data type was highly feasible for use in eMeasures.

Results

Vendors calculated a feasibility score for each data type, and Abt Associates averaged feasibility scores from each vendor for each data type. The feasibility scores of the QDM data types used in the AMI mortality eMeasure are shown in Appendix Table 2.

Appendix Table 2. Feasibility Scores of Data Elements Included in the AMI Mortality eMeasure (Calculated by Abt Associates)

Data Item	Mean Score	Highly Feasible?
Physical exam findings	3.56	No
Date/time of physical exam	3.56	No
Diagnostic study result	3.75	No
Primary encounter diagnosis?	3.88	Yes
Date/time of start of condition	3.89	Yes
Transfer from (used when receiving a patient, as in “where did the patient transfer in from?”)	3.89	Yes
Date/time of transfer	3.89	Yes
Condition	4.00	Yes
Condition status (e.g., active, inactive, resolved)	4.00	Yes
Date/time of resolution of condition	4.00	Yes
Diagnostic study type - lab (e.g., CBC, chemistry panel)	4.00	Yes
Date/time - ordered	4.00	Yes
Date/time - performed	4.00	Yes
Encounter type (e.g., inpatient, ambulatory)	4.00	Yes
Discharge status (e.g., alive and well, expired, left against medical advice)	4.00	Yes
Date/time of encounter	4.00	Yes
Birth date	4.00	Yes
Patient expired	4.00	Yes
Date/time patient expired	4.00	Yes

As shown in Appendix Table 2, the average scores for the data types in the AMI mortality eMeasure ranged from 3.56 to 4.0, with the majority of data types scoring a 4.0. Sixteen of the 19 data types in the AMI mortality eMeasure were identified as highly feasible (i.e., vendors responded that the item was already provided, they are working on it now, or could provide it with relatively minor burden). The vendors indicated that these 16 data types were either already included in their EHR, or could be developed as a structured field within 18 months with relatively little burden. Based on the vendor responses to the survey, we determined the AMI mortality eMeasure to be overall feasible.

Data element reliability testing

Overview

Abt Associates conducted data element reliability testing to determine how well individual data elements conform to standardized formats such as Extensible Markup Language (XML), as well as syntactic validation of the eMeasure logic. The syntactic validation was conducted by validating the data elements against the XML Schema encoded definitions. The QDM Schematron testing was conducted by validating the measure XML against QDM conformance constraints using Schematron. Finally, Abt conducted limited validity testing of the eMeasure value sets. Overall, this testing would ensure that the eMeasure’s expression in XML is written

correctly and is technically adequate and that the data elements and value sets used in the eMeasure are reliable.

Results

Overall, the AMI mortality eMeasure passed the XML Schema testing, the Schematron testing, and the value set validity testing. Minor errors identified during the Schematron testing were addressed iteratively and testing was re-conducted using the updated eSpecification. The final eMeasure passed all aspects of the data element reliability testing with no additional issues.

8.3.3 Beta Testing

Overview

The purpose of beta testing was to assess the validity of the data elements included in the AMI mortality eMeasure. Testing was conducted at three hospitals, referred to as hospitals A, B, and C. Each hospital used a different EHR system at the time of testing: Epic, Cerner, and McKesson, respectively. We planned to conduct testing at a fourth hospital, a critical access hospital that used an EHR system developed by Healthcare Management Systems, Inc. However, the hospital was unable to extract laboratory values from its EHR system and thus beta testing could not be completed at this hospital. The eMeasure was thus evaluated within the three EHR systems implemented at hospitals A, B, and C.

Of note, hospital EHR systems do not contain post-discharge mortality. Thus, we did not aim to assess data element validity of the outcome in this step of testing.

We provided programmers at each hospital with the human-readable and XML forms of our eMeasure. The programmers wrote code specific to their EHR systems to identify the cohort and extract the data elements (inclusion criteria, exclusion criteria, and risk-adjustment variables) necessary for the eMeasure. Using these queries, each hospital electronically extracted the relevant data elements from the EHR records of 40-60 of their AMI patients. Nurse abstractors then manually abstracted the same data elements from the same patients' medical records. The results of the electronic extraction and nurse abstraction were compared for consistency, with the manual nurse abstraction as the "gold standard."

Methods

Electronic extraction

Hospitals first used electronic extraction to identify patients for review based on the inclusion criteria; namely, a principal discharge diagnosis of AMI and an inpatient admission. The diagnosis was pulled from the EHR's standard reporting interface. Once patients were identified, electronic extraction pulled all remaining data elements, including those related to the exclusion criteria and risk-adjustment variables. The included hospitals reported that in some cases the data elements could be extracted with standard reporting tools, but at least one hospital reported developing a custom report to extract the data elements.

Of note, the AMI mortality eMeasure is designed to risk-adjust using the first collected value of the risk-adjustment variables in order to reflect the clinical status of the patient at presentation. However, due to limitations of the eSpecification, electronic extraction pulled all values for the risk-adjustment variables. For example, if a patient's heart rate was recorded four times

throughout the visit, all four values were electronically extracted. Following electronic extraction, the first collected value was manually identified for comparison with the nurse abstractor's results.

Manual abstraction

Nurse abstractors manually reviewed the EHR records for the cohort of patients identified for inclusion by electronic extraction. The nurse abstractors fully abstracted only those medical records for patients whom they found to be eligible based on the eMeasure's inclusion and exclusion criteria. If a nurse abstractor found that a patient was not eligible for inclusion in the measure based on the manual review, the nurse abstractor would then stop reviewing the record. For example, if the nurse abstractor found a patient not to have had a principal discharge diagnosis of AMI, the nurse abstractor would not continue abstracting to determine whether the patient met the remaining inclusion and exclusion criteria. Nurse abstractors also identified first collected values for the risk-adjustment variables included in the eMeasure.

Abt Associates compared the values pulled by the electronic extraction with those identified by manual abstraction to assess data element validity. The inter-rater reliability (IRR) among the nurse abstractors was computed by having a second abstractor review 10% of a given nurse abstractor's records and comparing the results of the second abstraction with those of the first abstraction. The extent to which these abstractions agreed determined the IRR score.

Results

Hospital A identified 60 patients for beta testing. From these 60 patients identified by the electronic extraction, the nurse abstractors fully abstracted 53 patient charts. The remaining seven patients were not fully abstracted because they were identified as not having had a principal discharge diagnosis of AMI (n=1) or as having been transferred into the hospital (n=6).

Hospital B identified 40 patients for beta testing. From these 40 patients, nurse abstractors completely abstracted 26 patient charts. The remaining 14 patients were not fully abstracted because they were identified as having been transferred into the hospital (n=12) or as having left AMA (n=2).

Hospital C identified 40 patients for beta testing. From these 40 patients, nurse abstractors completely abstracted 34 patient charts. The remaining six patients were not fully abstracted because they were identified as not having been admitted as inpatients (n=3), not having had a principal discharge diagnosis of AMI (n=2), or as having been transferred into the hospital (n=1).

As mentioned previously, the initial patient cohort was identified through electronic identification of patients with a principal discharge diagnosis of AMI. Electronic extraction suggested all patients identified as having had a principal discharge diagnosis of AMI also had inpatient admissions. Appendix Table 3 shows the percentage of patients the nurse abstraction also identified as having met the inclusion criteria.

Appendix Table 3. Percentage of Patients in the Electronically Extracted Cohorts that were Eligible based on Inclusion Criteria as Identified by Nurse Abstraction

	Hospital A (n=60)	Hospital B (n=40)	Hospital C (n=40)
Patients with a principal discharge diagnosis of AMI (%)	98.3	95	95
Patients with an inpatient admission (%)	100	100	92.5

Appendix Table 4 shows the percentage of patients who were identified as transfers into the hospital by electronic extraction and nurse abstraction.

Appendix Table 4. Identification of Transfer Patients by Electronic Extraction and Nurse Abstraction

	Hospital A (n=60)	Hospital B (n=40)	Hospital C (n=40)
Identified as transfers in by both electronic extraction and nurse abstraction (%)	1.7	25.0	0.0
Identified as transfers in by electronic extraction only (%)	23.3	25.0	0.0
Identified as transfers in by nurse abstraction only (%)	10.0	5.0	2.5
Not identified as transfers in by either electronic or nurse abstraction (%)	65.0	45.0	97.5

As shown in Appendix Table 4, transfer-in patients were not reliably identified in the EHR through electronic extraction. In all three hospitals, some patients identified as transfers by manual abstraction were not flagged as such in the electronically extracted data. Although the manual medical record abstraction was considered the “gold standard,” the cause of the reverse situation, in which patients were flagged as transfer patients in the electronic extraction and not by the manual abstraction, is unclear.

Additionally, there were rare problems identifying patients who had left AMA. Hospital B was the only hospital to have discrepancies between the electronically extracted and nurse abstracted data element for “left AMA.” At hospital B, two patients were identified by nurse abstraction only as having left AMA; electronic extraction did not identify any patients as having left AMA. It appears that too few patients overall left AMA for adequate assessment.

Because risk-adjustment variables were not manually abstracted for patients who did not qualify for inclusion in the eMeasure, such patients are excluded from the following comparison of the risk-adjustment variables. Risk-adjustment variables were fully abstracted for 53 patients at hospital A, 26 patients at hospital B, and 34 patients at hospital C. We compared the risk-adjustment variables that were electronically extracted against the manually abstracted risk-adjustment variables for these patients. Appendix Table 5 shows the percentage of agreement between the electronic extraction and the manual abstraction for each risk-adjustment data element.

Appendix Table 5. Agreement between Electronically Extracted and Manually Abstracted Risk-adjustment Variables

	Hospital A (n=53)	Hospital B (n=26)	Hospital C (n=34)
Age	100.0	100.0	100.0
Heart Rate	98.1	100.0	91.2
Systolic Blood Pressure	98.1	100.0	88.2
Patient Troponin	98.1	100.0	94.1
Hospital Upper Limit for Troponin	0.0	100.0	0.0
Creatinine	100.0	96.2	82.4

We investigated the reasons for the discrepancies at hospitals A and C. At hospital A, the patient troponin level discrepancy was due to an error in rounding. At hospital C, the discrepancies in the heart rate and systolic blood pressure values were due to errors in the electronic identification of the first collected value. The discrepancies in patient troponin at hospital C were due to errors in both the electronic extraction and the nurse abstraction. For hospital upper limit for troponin, hospital A did not provide any output, and hospital C provided a range for indeterminate values. Finally, the discrepancies in creatinine values at hospital C were due to incorrect documentation and typographical errors on the part of the nurse abstractor.

As mentioned previously, the manually abstracted results were considered the gold standard against which electronically extracted values were compared. IRR testing assessed the consistency of these abstracted results. IRR values for hospitals A, B, and C were 95.7%, 99.1%, and 92.9%, respectively, indicating a high level of agreement between the nurse abstractors' reviews at all hospitals.

8.3.4 Conclusions

Alpha testing of the AMI mortality eMeasure showed that the eMeasure was overall usable, feasible, and reliable. Hospital and vendor IT experts expressed few concerns about running the eMeasure in hospital EHRs and confirmed our decision to exclude ECG results from the eMeasure.

Beta testing of the AMI mortality eMeasure supported the overall validity of nearly all of the data elements included in the eMeasure. There were notable issues with identifying patients who had been transferred into the hospital. All other data elements for cohort identification and risk adjustment were consistently found for all patients and were extractable and accurate.

The testing process confirmed a few issues for measure implementation unrelated to data element validity that we expect to be resolved in the near future for the majority of EHRs. These issues include identifying the initiation of an episode of inpatient care and the ability to incorporate logic to extract the first collected values for the risk-adjustment variables into the eMeasure specification. Furthermore, as expected, beta testing revealed the difficulty of extracting the hospital upper limit of troponin, reinforcing the IT and quality experts' concerns over extracting this value correctly. To address this problem, our eMeasure metadata requests that hospitals provide this value manually.

Thus, the majority of data elements needed to calculate this measure can be feasibly and

accurately retrieved from current hospital EHRs. Remaining challenges identified during alpha and beta testing are likely to be resolved in near future (identification of the first collected vital sign or lab value) or can be handled by collection of supplemental data from other sources (e.g., transfer status or upper limit of normal for a lab value). For the near-term implementation of outcome eMeasures, hybrid models combining EHR data with information from other data sources may be required until these challenges are resolved. Overall, the model was found to be usable, scientifically sound, and valid.

Appendix C. Working Group Member Roster

Name	Title/Affiliation
Yale-CORE Members	
Harlan Krumholz, MD, SM	Co-Lead; Cardiologist; Professor of Medicine (Cardiovascular Medicine), Yale School of Medicine
Robert McNamara, MD, MHS	Co-Lead; Cardiologist; Associate Professor of Medicine (Cardiovascular Medicine), Yale School of Medicine
Susannah Bernheim, MD, MHS	Director, Quality Measures
Lori Geary, MPH	Senior Project Manager, Quality Measures
Yongfei Wang, MS	Lead Analyst
Zhenqiu Lin, PhD	Supporting Project Analyst
Julia Montague, MPH	Project Coordinator
Purav Mody, MBBS	Research Assistant
Elizabeth Eddy, BA	Research Assistant
Amena Keshawar, MPH	Research Assistant

Appendix D. Lists of EHR Vendors, Additional EHR Experts, and Hospitals Systems Consulted for
Feedback on Data Sources and Model Development

EHR vendor

Epic Corporation
Christopher Mast, MD, MS
Venkatesh Janakiraman

Consultants regarding electronic databases

Edward Hannan, PhD, MS, MS
Distinguished Professor
School of Public Health
University at Albany – State University of New York

James Tcheng, MD
Director, DTMI Biomedical Informatics Core
Professor of Medicine
Professor of Community and Family Medicine
Duke University Medical Center

Mikhail Kosiborod, MD
Associate Professor
Mid America Heart Institute

John Spertus, MD, MPH
Professor/ Daniel J. Lauer Missouri Endowed Chair in Metabolism and Vascular Disease Research
Mid America Heart Institute
University of Missouri – Kansas City

Hospital systems

Sentara Cardiovascular Research Institute
John Brush, MD
John Parker, MD

Kaiser Permanente Northern California
Gabriel Escobar, MD
Research Scientist

Kaiser Permanente, Institute for Health Research Colorado
David Magid, MD
Director of Research for CPMG

Veterans Health Affairs
Marta Render, MD
Director, VA Inpatient Evaluation Center (IPEC)

Additional EHR experts

Adam Landman, MD
Brigham and Women's Hospital

Jeremy Michel, MD
Yale Center for Medical Informatics

Jacob Reider, MD
Office of the National Coordinator for Health Information Technology (ONC)

Lauren Richie, MA
Office of the National Coordinator for Health Information Technology (ONC)

Appendix E. Variables Excluded at Each Step of Variable Selection

Step 1: Exclude variables related to post-admission events

Variable category	ACTION Premier Form Variables
E. Medications	
Aspirin post admission	Aspirin in First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	Aspirin at Discharge (Selections: No; Yes; Contraindicated; Blinded)
	Aspirin at Discharge - Dose
Clopidogrel post admission	Clopidogrel in First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	Clopidogrel in First 24 Hours - Dose
	Clopidogrel at Discharge (Selections: No; Yes; Contraindicated; Blinded)
	Clopidogrel at Discharge - Dose
	Clopidogrel at Discharge - Recommended Duration of Therapy (months)
Ticlopidine post admission	Ticlopidine in First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	Ticlopidine in First 24 Hours - Dose
	Ticlopidine at Discharge (Selections: No; Yes; Contraindicated; Blinded)
	Ticlopidine at Discharge - Dose
	Ticlopidine at Discharge - Recommended Duration of Therapy (months)
Prasugrel post admission	Prasugrel in First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	Prasugrel in First 24 Hours – Dose
	Prasugrel at Discharge (Selections: No; Yes; Contraindicated; Blinded)
	Prasugrel at Discharge - Dose
	Prasugrel at Discharge - Recommended Duration of Therapy (months)
Warfarin at discharge	Warfarin at Discharge (Selections: No; Yes; Contraindicated; Blinded)
Beta blocker post admission	Beta blocker First 24 Hrs (Selections: No; Yes; Contraindicated; Blinded)
	Beta Blocker at Discharge (Selections: No; Yes; Contraindicated; Blinded)
ACE Inhibitor post admission	ACE Inhibitor First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	ACE Inhibitor at Discharge (Selections: No; Yes; Contraindicated; Blinded)
Angiotensin Receptor Blocker post admission	Angiotensin Receptor Blocker First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	Angiotensin Receptor Blocker at Discharge (Selections: No; Yes; Contraindicated; Blinded)
Aldosterone Blocking post admission	Aldosterone Blocking Agent First 24 Hours (Selections: No; Yes; Contraindicated; Blinded)
	Aldosterone Blocking Agent at Discharge (Selections: No; Yes; Contraindicated; Blinded)
Statin post admission	Statin First 24 Hrs (Selections: No; Yes; Contraindicated; Blinded)
	Statin at Discharge (Selections: No; Yes; Contraindicated; Blinded)
Non-Statin Lipid-lowering Agent post admission	Non-Statin Lipid-lowering Agent First 24 Hr (Selections: No; Yes; Contraindicated; Blinded)
	Non-Statin Lipid-lowering Agent at Discharge (Selections: No; Yes; Contraindicated; Blinded)
GP IIb/IIIa	GP IIb/IIIa Inhibitor Administered (Selections: No; Yes; Contraindicated; Blinded)
	GP IIb/IIIa Inhibitor Type (Selections: Eptifibatide; Tirofiban; Abciximab)
	GP IIb/IIIa Dose
Anticoagulants	Anticoagulants Administered (Selections: No; Yes; Contraindicated; Blinded)
Unfractionated Heparin	IV Unfractionated Heparin (No/Yes)
	Unfractionated Heparin Initial Bolus (No/Yes)
	Unfractionated Heparin Dose of Initial Bolus
	Unfractionated Heparin Initial Infusion (No/Yes)

Variable category	ACTION Premier Form Variables
	Unfractionated Heparin Dose of Initial Infusion
Enoxaparin	Enoxaparin (No/Yes)
	Enoxaparin Initial Subcutaneous Dose
	Enoxaparin Initial IV Bolus (No/Yes)
	Enoxaparin Frequency of Injections Per Day (Selections: q12h; q24h; None)
Dalteparin	Dalteparin (No/Yes)
	Dalteparin Dose
Bivalirudin	Bivalirudin (No/Yes)
Fondaparinux	Fondaparinux (No/Yes)
Argatroban	Argatroban (No/Yes)
Lepirudin	Lepirudin (No/Yes)
F. Procedures and Tests	
Positive Cardiac Markers	Positive Cardiac Markers w/in First 24 hours (No/Yes)
Non-invasive Stress Testing	Non-invasive Stress Testing (No/Yes)
LVEF	LVEF (%)
	LVEF Not Assessed (No/Yes)
Diagnostic Coronary Angiography	Diagnostic Coronary Angiography (No/Yes)
Angiography Findings	Left Main Stenosis Percent (%)
	Left Main Not Available (No/Yes)
	Proximal LAD Stenosis Percent (%)
	Proximal LAD Not Available (No/Yes)
	Mid/Distal LAD, Diag Branches Stenosis Percent (%)
	Mid/Distal LAD, Diag Branches Not Available (No/Yes)
	CIRC, OMs, LPDA and LPL Branches Stenosis Percent (%)
	CIRC, OMs, LPDA and LPL Branches Not Available (No/Yes)
	RCA, RPDA, RPL, AM Branches Stenosis Percent (%)
	RCA, RPDA, RPL, AM Branches Not Available (No/Yes)
	Ramus Stenosis Percent (%)
	Ramus Not Available (No/Yes)
Diagnostic Cath Contraindication	Diagnostic Cath Contraindication (No/Yes)
PCI	PCI (No/Yes)
	Stent(s) Placed (No/Yes)
	Bare Metal Stent Implanted (No/Yes)
	Drug Eluting Stent Implanted (No/Yes)
	Other Stents Implanted (No/Yes)
	PCI Indication (Selections: Immediate primary PCI for STEMI; Rescue PCI (after failed full-dose lytics for STEMI); PCI for NSTEMI; Stable, successful reperfusion for STEMI, or completed infarction post-STEMI; Other)
	Non-system Reason for Delay in PCI (Selections: Difficult vascular access; Cardiac arrest and/or need for intubation before PCI; Patient delays in providing consent for the procedure; Difficulty crossing the culprit lesion during the PCI procedure; Other; None)
CABG	CABG (No/Yes)
G. Reperfusion Strategy	
Reperfusion	Reperfusion Candidate (No/Yes)
	Primary Reason Not Indicated (Selections (~30) not listed; refer to coder's data dictionary)

Variable category	ACTION Premier Form Variables
Thrombolytic therapy	Thrombolytics (No/Yes)
	Strength of Thrombolytic Dose
	Type of Thrombolytics (Selections: Tenecteplase; Alteplase; Reteplase; Streptokinase; Other)
Delay in Reperfusion	Non-System Reason for Delay (No/Yes)
H. In-hospital Clinical Events	
Reinfarction	Reinfarction (No/Yes)
Cardiogenic Shock	Cardiogenic Shock (No/Yes)
Heart Failure	Heart Failure (No/Yes)
	Heart Failure Date
CVA/Stroke	CVA/Stroke (No/Yes)
	Hemorrhagic Stroke (No/Yes)
Suspected Bleeding Event	Suspected Bleeding Event (No/Yes)
	Suspected Bleeding Event Location - Access Site (No/Yes)
	Suspected Bleeding Event Location - Retroperitoneal (No/Yes)
	Suspected Bleeding Event Location - GI (No/Yes)
	Suspected Bleeding Event Location - GU (No/Yes)
	Suspected Bleeding Event Location - Other (No/Yes)
Surgical Procedure or Intervention	Surgical Procedure or Intervention Required (No/Yes)
Blood Transfusion	RBC/Whole Blood Transfusion (No/Yes)
	Transfusion Related to CABG (No/Yes)
Peak Troponin	Peak Troponin Collected (Selections: No; Yes – I; Yes – T)
	Peak Troponin Value (ng/mL)
	Peak Troponin URL (ng/mL)
Peak CK-MB	Peak CK-MB Collected (No/Yes)
	Peak CK-MB Value
	Peak CK-MB Unit (Selections: IU/L; %; (mg/mL)/IU; ng/mL)
	Peak CK-MB ULN
Peak Creatinine	Peak Creatinine Collected (No/Yes)
	Peak Creatinine Value (mg/dL)
Lowest Recorded Hemoglobin	Lowest Recorded Hemoglobin Collected (No/Yes)
	Lowest Recorded Hemoglobin Value (g/dL)
J. Discharge	
Discharge	Comfort Measures Only (No/Yes)
	Clinical Trial (No/Yes)
	Discharge Status (Selections: Alive; Deceased)
	Smoking Counseling (No/Yes)
	Dietary Modification Counseling (Selections: No; Yes; N/A)
	Exercise Counseling (Selections: No; Yes; Ineligible)
	Cardiac Rehabilitation Referral (Selections: No; Yes; Ineligible)
	Discharge Location (Selections: Home; Extended Care/Transitional Unit; Other Hospital; Nursing Home; Hospice; Other)
	Transfer Time
	Transfer for PCI (No/Yes)
	Transfer for CABG (No/Yes)
	Cause of Death (Selections: Cardiac; Non-Cardiac)
	Time of Death

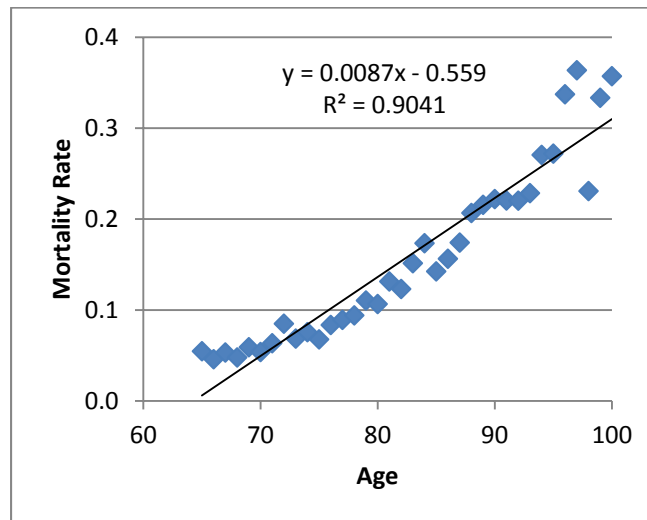
Step 2: Exclude variables unrelated to clinical status of patient at time of admission

Variable category	ACTION Premier Form Variables
A. Demographics	
Race	White
	Black/African-American
	Asian
	American Indian/Alaskan Native
	Native Hawaiian/Pacific Islander
Hispano or Latino Ethnicity	Yes/No
B. Admission	
Patient Zip Code	Patient Zip Code
	Zip Code N/A
Means of Transport to First Facility/ Arrival Time	Means of transport to First Facility (Selections: Self/Family; Ambulance; Mobile ICU; Air)
	Pre-arrival First Medical Contact Time Estimated (No/Yes)
Insurance Payer	Insurance Payer - Private Health Insurance (No/Yes)
	Insurance Payer - Medicare (No/Yes)
	Insurance Payer - Medicaid (No/Yes)
	Insurance Payer - Military Health Care (No/Yes)
	Insurance Payer - State-Specific Plan (No/Yes)
	Insurance Payer - Indian Health Service
	Insurance Payer - Non-US Insurance
	Insurance Payer - None
Cocaine Use	Cocaine Use (No/Yes)
Aspirin at Home	Aspirin at Home (No/Yes)
Clopidogrel at Home	Clopidogrel at Home (No/Yes)
Ticlopidine at Home	Ticlopidine at Home (No/Yes)
Prasugrel at Home	Prasugrel at Home (No/Yes)
Warfarin at Home	Warfarin at Home (No/Yes)
Beta Blocker at Home	Beta Blocker at Home (No/Yes)
ACE Inhibitor at Home	ACE Inhibitor at Home (No/Yes)
Angiotensin Receptor Blocker at Home	Angiotensin Receptor Blocker at Home (No/Yes)
Aldosterone Blocking Agent at Home	Aldosterone Blocking Agent at Home (No/Yes)
Statin at Home	Statin at Home (No/Yes)
Non-Statin Lipid-lowering Agent at Home	Non-Statin Lipid-lowering Agent at Home (No/Yes)

Step 3: Exclude variables that are deemed not feasible for use in eMeasures

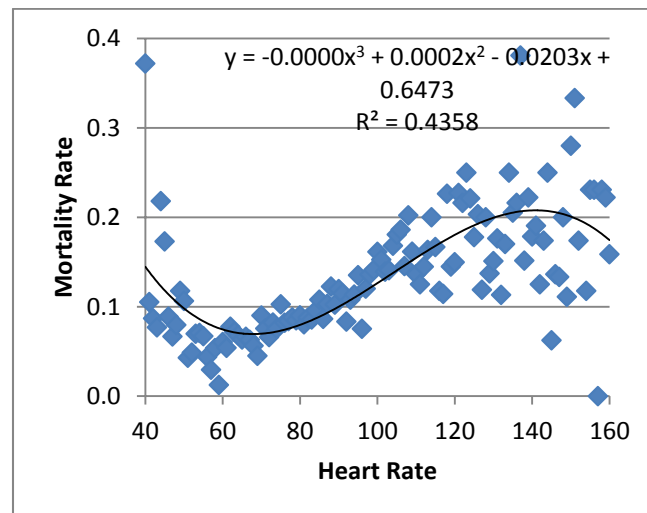
See Section 4.5.1.

Figure 10. Association between Age and Mortality: No Winsorization on Age



Decision: Variable to be kept unchanged.

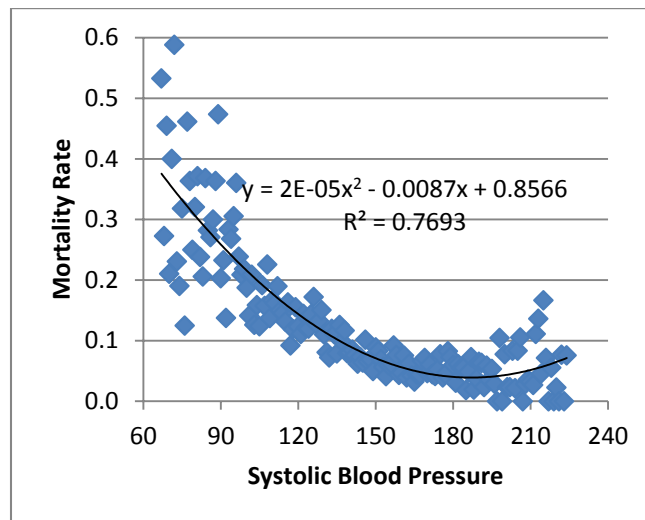
Figure 11. Association between Heart Rate and Mortality with Winsorization of Heart Rate: Low Values to 1st Percentile (40 bpm) and High Values to 99th Percentile (160 bpm)



Decision: Winsorize lower limit to 40 bpm and upper limit to 140 bpm; use splines with a knot at 70 bpm.

Rationale: Presence of a clear linear relationship between heart rate and mortality in the region 40-70 and 70-140 bpm. Thus, it is more appropriate to consider the two linear relationships separately rather than as a single linear relationship in the model. In addition, the use of splines with a knot at 70 bpm is the same approach that was used by the Duke Clinical Research Institute for their AR-G risk model.¹⁹

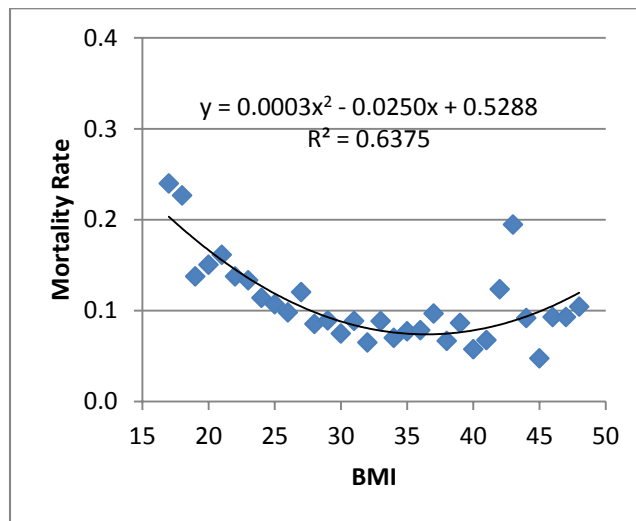
Figure 12. Association between Systolic Blood Pressure and Mortality with Winsorization of Systolic Blood Pressure: Lower Values to 1st Percentile (67 mm Hg) and Higher Values to 99th Percentile (224 mm Hg)



Decision: Winsorize lower limit to 70 mm Hg and upper limit to 150 mm Hg.

Rationale: Risk at systolic blood pressure >150 mm Hg is not clear; risk at 150 mm Hg appears to approximate risk thereafter. Although the risk between 70 mm Hg and 90 mm Hg is variable, the risk appears to decrease as the blood pressure increases. Additionally, 70 mm Hg is clinically more meaningful than 67 mm Hg (1st percentile) as the lower endpoint.

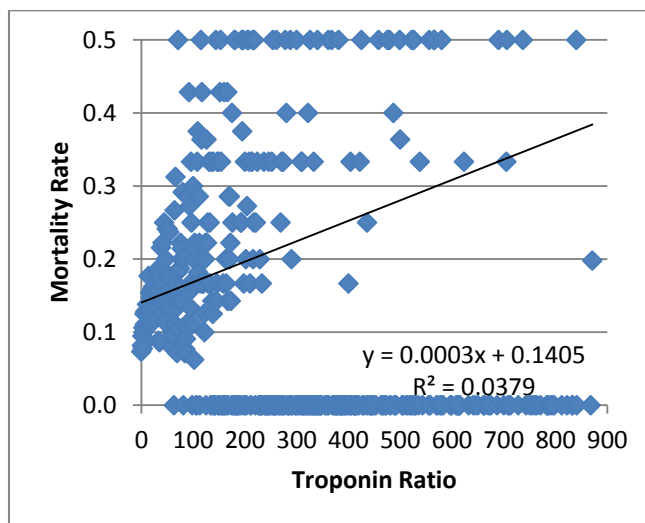
Figure 13. Association between Body Mass Index (BMI) and Mortality with Winsorization of BMI: Lower Values to 1st Percentile (16.5 kg/m²) and Upper Values to 99th Percentile (48.0 kg/m²)



Decision: Winsorize lower limit to 18.5 kg/m² and upper limit to 30 kg/m².

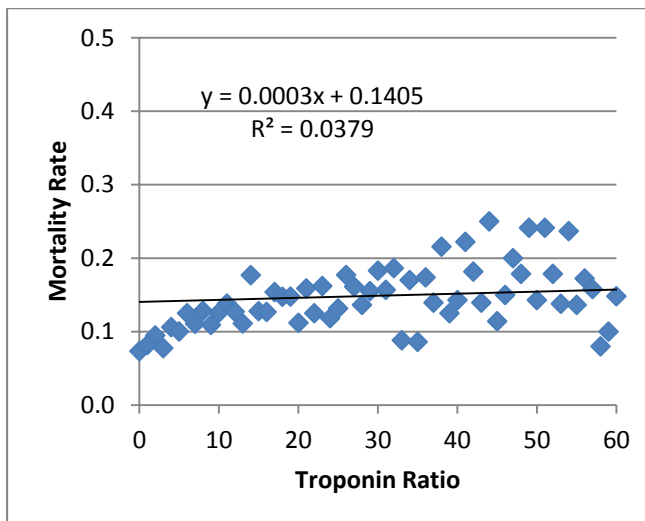
Rationale: Winsorization cutpoint selected as per NCDR CathPCI mortality models. In addition, risk after BMI 30 kg/m² is unclear but appears relatively constant. Decision to Winsorize the lower end point to 18.5 kg/m² compared with 16.5 kg/m² as risk appears unclear under this value and it is the lower end of the normal range (18.5-24.99) for BMI.

Figure 14. Association between Troponin Ratio and Mortality with Winsorization of Troponin Ratio: High Values to 99th Percentile (871). The 1st Percentile is 0



(Note: <10% of values were >60)

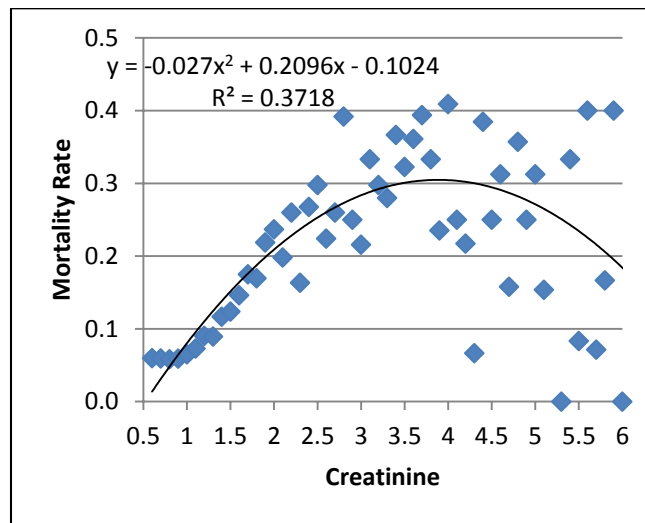
Figure 15. Association between Troponin Ratio and Mortality with Winsorization of Troponin Ratio: High Values to 99th Percentile (871). Only Range of Troponin Ratio between 0 and 60 Are Shown



Decision: Winsorize upper values to 60.

Rationale: Troponin ratio covers a large range of values from 0 to 871; however, the 90th percentile is 60. Risk above 60 is unstable and with relatively few data points.

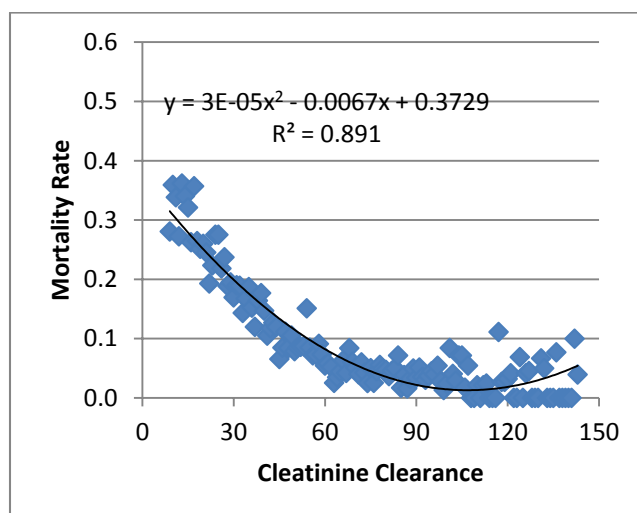
Figure 16. Association between Creatinine and Mortality with Winsorization of Creatinine: Low Values to 1st Percentile (0.6 mg/dL) and High Values to 99th Percentile (6.1 mg/dL)



Decision: Winsorize lower limit to 0.6 mg/dL and upper limit to 3 mg/dL.

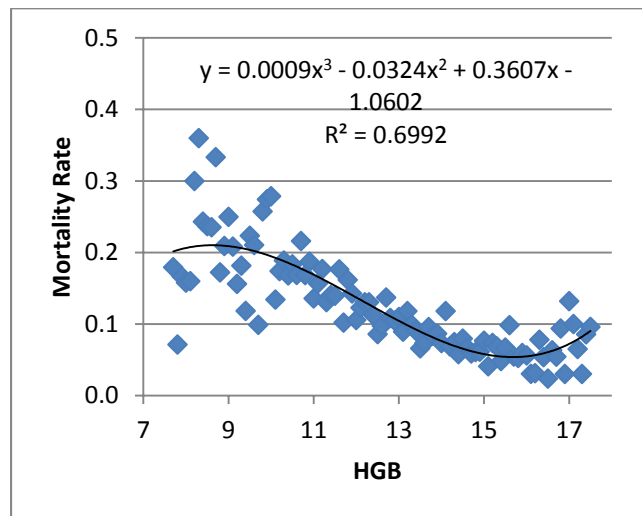
Rationale: The 95th percentile of the data points is 2.7 mg/dL and there is an increasing linear trend in risk prior to 3 mg/dL. Risk >3.0 mg/dL is unclear.

Figure 17. Association between Creatinine Clearance and Mortality with Winsorization of Creatinine Clearance: Low Values to 1st Percentile (9.1 mL/min) and High Values to 99th Percentile (142 mL/min)



Decision: Winsorize lower limit to 9.1 mL/min and upper limit to 90 mL/min.

Figure 18. Association between Hemoglobin and Mortality with Winsorization of Hemoglobin: Low Values to 1st Percentile (7.7 g/dL) and High Values to 99th Percentile (17.5 g/dL)



Decision: Winsorize lower limit to 9 g/dL and upper limit to 16 g/dL.

Rationale: Risk below and above these values is unclear.