Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure with Electronic Health Record Extracted Risk Factors Methodology Report
Version 2.0

Submitted By:

Yale New Haven Health Services Corporation – Center for Outcomes Research & Evaluation (YNHHSC/CORE)

Prepared For:

Centers for Medicare & Medicaid Services (CMS)

March 2020

Table of Contents

List of Figures

List of Figures

Center for Outcomes Research & Evaluation Reevaluation Team

Grace Glennon, MS – Reevaluation Lead
Erica Norton, BS – Project Coordinator II
Elizabeth Triche, PhD – Reevaluation Division Director
Yongfei Wang, MS – Lead Analyst, Senior Statistician
Shuling Liu, PhD – Lead Analyst
Sonam Lama, MPH – Project Coordinator
Rajvi Shah, MPH – Research Associate
Amanda Audette, BS – Research Associate
Alex Ferrante, BS – Research Assistant II
Xin, Xin, MS, MA – Statistician II
Jeph Herrin, PhD** – Statistical Consultant
Doris Peter, PhD – Reevaluation Consultant
Jacqueline Grady, MS – Statistical Consultant and Associate Director of Data Management and Analytics
Zhenqiu Lin, PhD –Director of Data Management and Analytics
Lisa Suter, MD** – Contract Director
Karen Dorsey, MD, PhD** –Director, Quality Measurement
Susannah Bernheim, MD, MHS – Senior Director, Quality Measurement
Harlan M. Krumholz, MD, SM** – Principal Investigator
**Yale School of Medicine

Measure Reevaluation Team Contributors

Chandni Vasisht, MPH – Project Manager
Darinka Djordjevic, PhD –Project Manager
Shu-Xia Lin, PhD –Research Scientist and Associate Director of Data Management and Analytics
Erica Spatz, MD** – Clinical Consultant
Vrunda Desi, MD**– Clinical Consultant
Amy Salerno, MD* – Clinical Consultant
*University of Virginia
**Yale School of Medicine

Acknowledgements

This work is a collaborative effort and the authors gratefully acknowledge and thank our many colleagues and collaborators for their thoughtful and instructive input, especially those former team members on this project: Amy Salerno, MD, MHS Development lead, Lesli Ott, MA, MA Lead Analyst; Maggie Bierlein, MS Supporting Analyst; Nicole Cormier, MPH, Project Manager; Kerry McCole, M.Phil.Ed., M.S.Ed Project Coordinator II; Kumar Dharmarjan, MD, MBA Clinical Consultant; Carolyn Presley, MD Clinical Consultant; and Theodore Long, MD, MPH Clinical Consultant.

Special thanks to Melissa Miller, MPH, and Lori Schroeder, LLM, JD from Yale-New Haven Health Services Corporation/Center for Outcomes Research & Evaluation; and Sharon-Lise Normand, PhD from Harvard School of Public Health.

We would like to give special thanks to Christine Broderick at the National Partnership of Women and Family, along with the participants of our Patient and Family Caregiver Work Group. All of your feedback was invaluable to this work.

Additionally, we would like to thank the Technical Expert Panel members (Appendix A), who provided helpful insight and feedback on key measure development and re-specification decisions. We would also like to acknowledge the members of our Technical Work Group (Appendix A) who generously gave their time to provide guidance on key clinical and statistical decisions.

We would like to especially acknowledge the contribution of data and ongoing support from our colleagues affiliated with Kaiser Permanente Northern California, Division of Research, Gabriel Escobar, MD and John Greene, MA.

Finally, we would like to thank our Contracting Officer Representatives at the Centers for Medicare & Medicaid Services, Lein Han, PhD, and Vinitha Meyyur, PhD, for their continued support of our work.

1. EXECUTIVE SUMMARY

Goal of Measure

The goal of developing a Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure with Electronic Health Record-Extracted Risk Factors, or Hybrid HWM Measure, was to broadly measure the quality of care across hospitals, including smaller-volume hospitals. This measure will provide information to hospitals that can facilitate targeted quality improvement, provide more transparent information for the public, and allow policymakers to monitor a very important outcome. In addition, the goal of this Hybrid HWM Measure, which employs a combination of administrative claims data and clinical electronic health record (EHR) data, is to minimize provider burden while enhancing clinical case mix adjustment with clinical data.

Background and Rationale

Mortality is an important health outcome that is meaningful to patients and providers, and estimates suggest that more than 400,000 patients die each year from preventable harm in hospitals.[1] The vast majority of patients admitted to the hospital have survival as a primary goal. Existing condition-specific mortality measures support targeted quality improvement work and may have contributed to national declines in hospital mortality rates for measured conditions and/or procedures.[2] They do not, however, allow for measurement of a hospital's broader performance, nor do they meaningfully capture performance for smaller volume hospitals. While we do not ever expect mortality rates to be zero, studies have shown that, for selected conditions and diagnoses, mortality within 30 days of hospital admission is related to quality of care and that high and variable mortality rates across hospitals indicate opportunities for improvement[3,4]. Therefore, it is reasonable to consider an all-condition, all-procedure, risk-standardized 30-day mortality rate as a quality measure.

Development of a hospital-wide mortality measure that includes EHR data addressed stakeholder preference for the use of patient-level clinical EHR data to support risk adjustment in assessing hospital performance by using data from claims as well as clinical data elements pulled from the EHR for risk adjustment.

Measure Development Process

This measure aims to report the hospital-level, risk-standardized rate of mortality within 30 days of hospital admission for most conditions or procedures. The Center for Outcomes Research and Evaluation (CORE) initially developed a Claims-only HWM measure, and then built upon the Claims-only HWM measure by utilizing the same concept, outcome, and cohort, and adding clinical data elements extracted from EHR to augment the risk adjustment models. This report focuses on the Hybrid HWM measure, which was developed in a small sample of hospitals with EHR data. We refer to Claims-only HWM development and results where necessary, for example, where specialty division results were not able to be produced in the Hybrid HWM measure.

CORE initiated development of the measure by conducting an extensive literature review and environmental scan to inform measure development. We also engaged with several stakeholder groups throughout the development process, including providers, technical experts, and patients and family caregivers. We solicited feedback on the measure concept, outcome, cohort, risk model variables (including claims variables and clinical EHR variables), and how to develop and report measure results in

a meaningful way. These engagements have included two advisory groups in the form of a Technical Work Group and a Patient and Family Caregiver Work Group. We also convened a national Technical Expert Panel (TEP) consisting of a diverse set of stakeholders, including providers and patients who reviewed final measure results and voted on face validity. We also hosted two public comment periods to solicit input from the general public. The first public comment period was hosted in 2016, where we specifically requested input on the cohort, outcome, and approach to grouping patients by condition and procedure for risk adjustment. The second public comment period was hosted in 2018 and presented the measure specifications and results utilizing International Classification of Disease, Ninth Revision (ICD-9) data. Subsequently the measure was re-specified in ICD-10 data and National Quality Forum (NQF) endorsed in 2019, which included an additional opportunity for public comment. This report focuses on results from the final measure specified in ICD-10 data, and refers to initial development results where applicable.

Measure Specifications Overview

Our cohort definition attempts to capture as many admissions as possible for which survival would be a reasonable indicator of quality and for which adequate risk adjustment is possible. We assumed survival would be a reasonable indicator of quality for admissions fulfilling two criteria: 1) survival is most likely the primary goal of the patient when they enter the hospital; and 2) the hospital can reasonably influence the chance of survival through quality of care. We determined the adequacy of risk adjustment using clinical judgement and by examining survival patterns and model performance. Therefore, we included in the measure all admissions except those for which 30-day mortality cannot reasonably be considered a signal of quality care.

The outcome for this measure is all-cause 30-day mortality. We define all-cause mortality as death from any cause within 30 days of the index hospital admission date.

To compare mortality performance across hospitals, the measure accounts for differences in patient characteristics (patient case mix) as well as differences in mixes of services and procedures offered by hospitals (hospital service mix). We account for differences in patient case mix using patient clinical comorbidity variables from claims and patient clinical data derived from the EHR. We account for differences in hospital service mix using the patient's principal discharge diagnosis.

Rather than assume that the effects of risk variables are homogeneous across all discharge condition and procedure categories, we separated the cohort into 15 different service-line divisions and estimated separate risk models within each. We then derived a single summary score from the results of the 15 models by combining separate standardized mortality ratios to calculate one hospital-wide mortality rate for each hospital. Using 15 models rather than a single model allows for better risk adjustment for diverse patient groups and improves the usability of the measure. The service-line divisions allow hospitals and consumers to have more detailed information on hospital performance. The 15 service-line divisions include the non-surgical divisions: Cancer, Cardiac, Gastrointestinal, Infectious Disease, Neurology, Orthopedics, Pulmonary, Renal, Other; and the surgical divisions: Cancer, Cardiothoracic, General, Neurosurgery, Orthopedics, Other. While the measure is intended to include all 15 service-line divisions, the dataset used to develop and test the Hybrid HWM Measure did not contain enough patients in the non-surgical Cancer and surgical Cardiothoracic, Cancer, Neurology, Orthopedics, and Other service-line divisions, so most of the Hybrid Measure testing results only capture 9 service-line divisions. We provide results from the Claims-only HWM measure for these other service-line divisions

because the Hybrid HWM Measure uses the same concept, cohort, outcome and claims-based risk adjustment variables as the Claims-only HWM measure; therefore, there is no conceptual reason that the results from the Claims-only HWM measure would be substantially dissimilar to results from the Hybrid HWM Measure.

The Hybrid HWM measure utilizes some data derived from hospital EHR systems. To use these data to calculate the measure, CMS provides electronic specifications in the form of the Measure Authoring Tool (MAT) Output. The Hybrid HWM Measure uses much of the same measure logic as is used in the electronic specifications for the Hybrid Hospital-Wide Readmission (HWR) Measure. The Hybrid HWM Measure electronic specifications can be found at the Electronic Clinical Quality Improvement (eCQI) Resource Center here: https://ecqi.healthit.gov/pre-rulemaking-eh-cah-ecqms.

This report serves as a summary of the measure development, stakeholder input, measure specifications, and measure testing for the Hybrid HWM Measure.

2. INTRODUCTION

2.1. Overview of Report

The Centers for Medicare & Medicaid Services (CMS) contracted with Yale New Haven Health System/Center for Outcomes Research and Evaluation (YNHHS/CORE) to develop a Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Mortality Measure with Electronic Health Record Extracted Risk Factors based on administrative claims data and clinical electronic health record (EHR) data. Throughout this report, we refer to this measure as the Hybrid HWM Measure.

Mortality is an important outcome that is meaningful to patients and providers. The vast majority of patients admitted to the hospital have survival as a primary goal. This important outcome is already the focus of existing CMS condition- and procedure-specific mortality quality measures; hospital-level risk-standardized mortality rates (RSMRs) are reported for patients admitted for heart failure, pneumonia, acute myocardial infarction, chronic obstructive pulmonary disease, stroke, and coronary artery bypass graft surgery.[5,6] Existing mortality measures support targeted quality improvement work around specific conditions and may have contributed to national declines in hospital mortality rates for measured conditions and/or procedures.[2] They do not, however, capture admissions for patients admitted for a majority of the conditions or procedures for which a patient may use the hospital or allow for measurement of a hospital's broader performance. In addition, the condition and procedure-specific mortality measures do not always allow for performance measurement for smaller volume hospitals.

In Medicare data from July 2016 through June 2017, there were approximately ten million inpatient admissions among Medicare fee-for-service (FFS) beneficiaries between ages 65 and 94 across 4,700 United States (US) hospitals. The observed 30-day mortality rate was 8.17%. A HWM measure is likely to capture about 6.5 million of those admissions across 4,700 hospitals, allowing for broad performance measurement and quality improvement efforts.

In addition to the obvious harm to individuals and their families and caregivers that results from preventable death, there are also significant financial costs to the healthcare system. Capturing monetary savings for preventable mortality events is challenging, as patients who die may incur fewer expenses than those who survive. Further, distinguishing between truly preventable hospital deaths and those deaths that are truly not preventable is challenging. However, using two estimates of the number of deaths due to preventable medical errors, and assuming an average of ten lost years of life per death (valued at $75,000 per year in lost quality adjusted life years), the annual direct and indirect cost of potentially preventable deaths could be as much as $73.5 to $735 billion.[7-9].

Under contract with CMS, CORE had previously identified a set of core clinical data elements (CCDE) that are routinely collected on hospitalized adults, feasibly extracted from hospital EHR systems, and are related to patients' clinical status at the start of an inpatient encounter. The CCDE are the first captured vital signs and laboratory results. The CCDE have been utilized in conjunction with administrative claims data to create hybrid outcome measures, which are quality measures that utilize more than one source of data. The Hybrid HWM Measure builds upon the initial CCDE work and the Hybrid Hospital-Wide (All-Condition, All-Procedure) Risk-Standardized Readmission (HWR) Measure by using the CCDE as additional candidate risk variables to test various risk models and develop the Hybrid HWM Measure. For more information on how the CCDE were originally developed, please refer to the Core Clinical Data Elements Technical Report and the Hybrid HWR Measure with Electronic Health Record Extracted Risk Factors

report posted at https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html under "Core Clinical Data Elements and Hybrid Measures.zip". For testing results of the electronic specifications of the CCDE, we refer readers to the Hybrid HWR Measure with Electronic Health Record Extracted Risk Factors Data posted on Quality Positioning System section of the National Quality Forum (NQF) website.

In this technical report, we provide detailed information on the development and specifications of the Hybrid HWM Measure. This includes details on the major decisions to form the cohort, the outcome, and the divisions. It also includes information on risk adjustment, measure testing, and reporting considerations. The Hybrid HWM Measure complies with accepted standards for outcomes measure development, including appropriate risk adjustment and transparency of specifications. Our goal is to include admissions for patients for whom mortality is likely to present a quality signal and those where the hospital has the ability to influence the outcome for the patient. The performance metric, RSMRs, are derived from the combined results of multiple statistical models built for groups of admissions that are clinically related and share similar risk profiles. This report reflects specifications that have been developed with close input from patients, caregivers, clinicians and methodological experts. In addition, the measure reflects input from a nationally convened Technical Expert Panel (TEP) representing a diverse set of stakeholders as well as input from public comment periods.

Following initial development with International Classification of Disease, Ninth Revision (ICD-9) codes, this Hybrid HWM Measure received conditional support from the Measure Application Partnership in December 2017, and CMS signaled the possibility of including one or both of the HWM measures (Claims-only and Hybrid measures) within the Inpatient Quality Reporting (IQR) program through the 2019 Inpatient Prospective Payment System (IPPS) Proposed Rule. Following ICD-10 re-specification, measure results and testing for both measures were presented to the TEP and Technical Work Group. The Hybrid HWM Measure was endorsed by the NQF in 2019.

2.2. Hospital-Wide Mortality as a Quality Indicator

Importance

Mortality is an unwanted outcome for the overwhelming majority of patients admitted to US hospitals. Although mortality within 30 days of hospitalization is uncommon, this outcome provides a concrete signal of care quality across conditions and procedures when assessed among appropriate patients. It captures the result of care processes, such as peri-operative management protocols, and the impact of both optimal care and adverse events resulting from medical care.

For some conditions and diagnoses, evidence supports that optimal medical care reduces mortality[4,3]. We know from ongoing improvements in condition- and procedure-specific mortality rates that interventions to improve these outcomes are feasible.[2] Multiple organizations, including the Institute for Healthcare Improvement (IHI), promote a range of evidence-based strategies to reduce hospital mortality[10]. These strategies include:

- Adoption of strategies shown to reduce ventilator-associated pneumonia;[11-14]
- Delivery of reliable, evidence-based care for acute myocardial infarction;[14,15]
- Prevention of adverse drug events though medication reconciliation;[16]
- Prevention of central line infections through evidence-based guideline-concordant care;[17] and

- Prevention of surgical site infections through evidence-based guideline-concordant care.[18,19]

To reduce mortality, the IHI further encourages hospitals to use multidisciplinary rounds to improve communication, employ Rapid Response Teams to attend to patients at the first sign of clinical decline, identify high-risk patients on admission and increase nursing care and physician contact accordingly, standardize patient handoffs to avoid miscommunication or gaps in care, and establish partnerships with community providers to promote evidenced-based practices to reduce hospitalizations before patients become critically ill.[20] The IHI's 100,000 Lives Campaign, which was created to enlist hospitals in a coordinated effort to adopt the above interventions, led to an estimated more than 120,000 lives saved over the first 18 months of the campaign.[21]

Some of the evidence-based recommendations above apply to specific diagnoses. While condition- and procedure-specific initiatives to reduce mortality may broadly impact mortality rates across other conditions and procedures, there is likely more to be gained by a measure of hospital-wide mortality that can inform and encourage quality improvement efforts for patients not currently captured by existing CMS mortality measures. In addition, there is evidence that a hospital's organizational culture is linked to key measures of hospital quality performance.[22] Since these cultural and leadership qualities affect the entire hospital, the Hybrid HWM Measure may provide important incentives for hospitals not only to examine their care processes and improve care for individual conditions, but may also provide incentives to encourage care transformation and improve overall organizational culture.

In fact, because of its importance, hospital-wide mortality has been the focus of a number of previous quality reporting initiatives in the US and other countries. Prior efforts have met with some success and a number of challenges.[23] Despite these challenges, countries such as the United Kingdom, Scotland, and Australia, continue to report measures of hospital-wide mortality.[24]

From 1986 through 1993, the Health Care Financing Administration (now CMS) measured hospital-wide mortality. Hospitals used this information to reduce avoidable deaths and closely examine hospital care processes. However, this effort was stopped partly due to concerns over the adequacy of the case-mix adjustment in the measure that was used, which was based on administrative claims data. The Hybrid HWM Measure described in this report aims to address the limitations[23,25,26] of the earlier measure specifications, which led to the removal of the measure.[27-29]

Other hospital-wide mortality measures have been reported in the United Kingdom and Canada. These prior efforts to measure hospital-wide mortality similarly faced a number of challenges including concerns about adequate exclusion of patients for whom survival is not the primary goal, such as hospice and palliative care patients; risk adjustment for disease severity; ability to distinguish between conditions present on admission and events occurring after admission; and addressing imbalances in both case mix and capability (for example, coronary artery bypass graft surgery performed or not) across hospitals.[26,30-32] In developing the current Hybrid HWM Measure, we aimed to take advantage of advances in coding and design of the measure to address prior challenges.

While we do not expect optimal mortality rates to be zero, we know, as stated above, that studies have shown that mortality within 30 days is related to quality of care; that interventions have been able to reduce 30-day mortality rates for a variety of specific conditions; and that high and variable mortality rates indicate opportunity for improvement. Therefore, it is reasonable to consider an all-condition, all-procedure risk-standardized 30-day mortality rate as an important quality performance measure for hospitals.

Feasibility

Since the initial CMS hospital-wide mortality effort, much has changed to improve potential feasibility. Since 2015, administrative claims coding has advanced significantly. Advancements include the transition from ICD-9 to ICD-10 with much more specificity in both diagnosis and procedure codes, allowing up to 25 diagnostic codes per admission encounter (previously there were only 10 available diagnostic codes) and expanding the use of present on admission codes to signify conditions that were present prior to admission. CMS also has the benefit of years of experience successfully calculating and reporting the claims-based condition- and procedure-specific mortality measures, including performing chart-based validation of a number of these measures. Additionally, CMS has reported results for the claims-based Hospital-Wide Readmission (HWR) Measure since July 2013, which utilizes novel methods to aggregate readmission rates across diverse patient cohorts, to adjust more accurately for service mix and begun voluntary reporting of a Hybrid HWR measure in 2018. Moreover, CMS has further evolved its measure development approach to expand stakeholder engagement across all phases of measure development and to specifically include patients' perspectives and input to ensure more patient-centered measures. Therefore, it is now feasible to construct a measure which will be scientifically sound and acceptable to stakeholders.

Finally, the use of electronic clinical data in this Hybrid HWM Measure will better account for more critical clinical information about the patient's health status at the time of arrival to the hospital. This information can be incorporated into risk adjustment for more detailed clinical risk adjustment. This electronic clinical information is now more broadly available, due to national incentives aimed at increasing EHR adoption[*], related work to standardize data element definitions across providers[†], and specific work by our team to develop and test CCDEs, some of which are used in this risk adjustment model. The clinical data required in the risk adjustment model will be derived electronically from hospital EHRs. We have previously tested the feasibility and validity of each of these data elements empirically and have shown them to be consistently captured for nearly all adults hospitalized for acute care and extractable from hospital EHRs. Since the EHR system used by these 22 hospitals (Epic) is widely used in the US, we can make the reasoned inference that these data are representative. For testing results of the CCDE included in the Hybrid HWR Measure, we refer readers to the specifications posted on the on Quality Positioning System section of the NQF website.

Usability

A primary motivation for this measure was to provide policymakers with a summary performance assessment of patient survival, particularly for lower volume hospitals that care for insufficient numbers of patients to produce stable, reportable performance estimates for condition- and procedure-specific measures. In addition, from the outset, CMS and CORE sought to make this measure broadly usable by both patients and providers, as well as policymakers. Therefore, we approached this measure development from three distinct perspectives – policymakers; providers; and patient and family

[*] EHR Incentive Program. http://www.cms.gov/EHRIncentivePrograms/

[†] Health Information Technology (IT) for Economic and Clinical Health (HITECH) Act of 2009 provided Health and Human Services with authority to establish programs to improve healthcare quality, safety, and efficiency through the promotion of health IT, establishing the Office of National Coordinator to set standards, implementation specifications, and certification criteria for electronic exchange and use of health information. https://www.healthit.gov/policy-researchers-implementers/health-it-legislation-and-regulations

caregivers – in order to create a measure that provides meaningful, scientifically acceptable hospital performance information for all of these user groups.

The multiple model approach, which uses results for each of the service-line division models to create the overall hospital-wide mortality measure score, could increase the practical utility of the measure by providing information on differences in performance among divisions (service-line areas) within hospitals. This aspect of the measure will allow hospitals to better target quality improvement efforts. The patient and family caregivers identified that information on the division level would be helpful for consumers. Further consideration is needed to identify an approach that provides transparency and granularity for consumers and ensures the division information displayed is a reliable representation of quality within each service line for each hospital.

2.3. Approach to Measure Development

We developed this measure in consultation with national guidelines for publicly reported outcome measures, following the technical approach to outcomes measurement set forth in NQF guidance for outcome measures, CMS Measure Management System guidance, and the guidance articulated in the American Heart Association scientific statement, "Standards for Statistical Models Used for Public Reporting of Health Outcomes."[33,34] This measure was originally developed in ICD-9 data, and later re-specified in ICD-10. We have engaged with several stakeholder groups continuously during the development process, eliciting feedback on the measure concept, outcome, cohort, risk model variables, measure results, and how to present the measure results in a meaningful way for patients, family caregivers, and providers. These have included two formal advisory groups:

- A Technical Work Group, comprised of clinicians and a statistician; and
- A Patient and Family Caregiver Work Group comprised of patients, family members, and caregivers for patients who have had multiple encounters with the healthcare system.

We also convened a national TEP of diverse stakeholders, including providers and patients, and hosted two public comment periods in 2016 and 2018. The Hybrid HWM Measure was submitted to NQF and endorsed in 2019.

3. METHODS

3.1. Overview

This document aims to report the development and specifications of the measurement of hospital-level, risk-standardized mortality within 30 days of hospital admission for most conditions or procedures using administrative claims and EHR data. The measure produces a single score, derived from the results of risk-adjustment models for 15 mutually exclusive divisions (admissions grouped based on categories of discharge diagnoses or procedures). Hospitalizations are eligible for inclusion in the measure if the patient was hospitalized at a non-Federal short-stay acute care hospital or critical access hospital. To compare mortality performance across hospitals, the measure accounts for differences in patient characteristics (patient case mix) as well as differences in mixes of services and procedures offered by hospitals (hospital service mix).

The measure cohort, outcome, divisions, and approach to risk factors were initially developed in CMS administrative claims data, with key decisions described below. Because there is currently no large national dataset that includes patient-level EHR data to develop, test, and validate various risk models using clinical EHR data, we used Kaiser Permanente Northern California (KPNC) data from their EHR data warehouses; these data contain patient-level clinical variables (for example, laboratory test results, vital signs, care directives) used to develop the risk-adjustment models for the Hybrid HWM Measure. KPNC serves more than 4.4 million members at its 21 acute care hospitals. While the risk model was developed and tested using these hospitals, the Hybrid HWM Measure is designed to be implemented in all non-Federal short-stay acute care hospitals in the US. In this report, we have described the decisions and final measure specifications as they would be implemented in the Medicare FFS population. However, for development purposes, throughout this report we note where we used slightly modified specifications that were necessary for testing purposes due to the smaller number of hospitals in the dataset provided by KPNC, referred to as the Clinical Hybrid Re-specification Dataset (Section 3.2 Data Sources). Throughout this Methods section, we focus on and outline the final measure specifications (as developed for Medicare FFS population), and note the differences used for Hybrid HWM measure development only. In the Results section of this report, we report results based on the modified measure specifications using the Clinical Hybrid Re-specification Dataset. We refer to the Claims-only HWM measure for select results that use the Medicare FFS data source and the Claims-only HWM measure specifications. To note, this measure was originally developed using ICD-9 data (developed with claims data from July 2013-June 2014) and later re-specified for use with ICD-10 data. This report focuses on ICD-10 results and supplements with original ICD-9 results where necessary.

This section provides details about the measure development and final measure specifications of the Hybrid HWM Measure. Below we detail the data sources used, the measure cohort inclusion and exclusion criteria, the outcome definition and attribution, the approach to risk adjustment, final risk models, and approach to measure calculation.

3.2. Data Sources

To develop and re-specify the Hybrid HWM Measure including the cohort, outcome, service-line divisions, and, for most of testing, we constructed multiple datasets, listed below.

1. Medicare Claims-Only Measure Re-specification Dataset. Several sources of data derived from Medicare FFS claims were used to define the measure specifications:
    a. Cohort and outcome:
        i. An index dataset containing administrative inpatient hospitalization data, enrollment data, and post-discharge mortality status for Medicare FFS beneficiaries, 65 and older on admission, hospitalized from July 1, 2016 – June 30, 2017.
        ii. Enrollment and mortality status were obtained from the Medicare Enrollment Database, which contains beneficiary demographic, benefit, coverage, and vital status between July 2016 to June 2017.
        iii. Hospice enrollment data was also obtained from the Medicare Enrollment Database, for patients enrolled into hospice between July 2016 to June 2017.

    b. Case-mix risk adjustment:
        i. A history dataset including inpatient hospitalization data on each patient for the 12 months prior to the index admission.

2. Clinical Hybrid Re-specification Dataset. Constructed KPNC matched administrative claims and EHR data, admissions with discharge date from October 1, 2015 to December 30, 2016.
    a. Note: We expanded our measurement period to have enough data for measure development purposes, from 12 months to 15 months. The final Hybrid HWM Measure specifications would be a one-year measure, similar to the claims and Hybrid HWR measures.
3. Surgical Mortality Hierarchy. A Medicare FFS claims dataset with mortality rates for surgical admissions from July 1, 2013 to June 30, 2014, used to assign admissions to a surgical division for patients with more than one major surgical procedure on the earliest date of their index hospitalization. The admission is assigned to the division that includes the procedure (the "defining surgical procedure") with the highest mortality rate. For more details, please see Section 3.4 Service-Line Divisions.

4. Original Development Data Sources (ICD-9 data):
    a. Medicare Claims-Only Development Dataset. Provides the same data as the Claims-Only Re-specification Dataset but includes administrative inpatient hospitalization data, enrollment data, and post-discharge mortality status for FFS Medicare beneficiaries, 65 and older on admission, hospitalized from July 1, 2014 – June 30, 2015.
    b. Clinical Hybrid Development Dataset. Provides similar data as the Clinical Hybrid Re-specification Dataset with matched administrative claims and EHR data, admission dates from January 1, 2009 – June 30, 2015.

Clinical Hybrid Development Dataset and Clinical Hybrid Re-specification Dataset Description

Data used to develop the Hybrid HWM Measure were provided by KPNC from their EHR data warehouses. KPNC is an integrated healthcare delivery system that serves over 4.4 million members at its 21 acute care hospitals. Although the number of KPNC hospitals is much smaller than the number of hospitals in the nation that will be ultimately included in the implemented measure, the patients within the KPNC hospitals represent an adequate sample for measure development. Comparison of similarly

aged patients (65 years and older) in the Clinical Hybrid Development Dataset and Medicare Claims-Only Development Dataset demonstrated similar prevalence of those comorbidities included in the Claims-only HWM measure risk model; see Appendix B Comorbidity Comparison: Claims vs Clinical Hybrid Datasets. All KPNC hospitals use an integrated EHR system that runs Epic software to capture and store patient management, administrative, and clinical data in their outpatient and inpatient healthcare settings. The Systems Research Initiative within the Kaiser Permanente Northern California Division of Research has worked to develop an extensive clinical risk-adjustment methodology for internal benchmarking and quality assurance and is in the process of developing the capability to use these clinical data in real time for clinical decision support and quality measurement. Their work has required mapping specific clinical data elements within their databases, extracting data, and validating their source and accuracy.

Additionally, members enrolled in the KPNC health system receive comprehensive care through the KPNC network of outpatient and inpatient providers. In the rare instance that a member is admitted to an acute care facility outside of the network, KPNC will receive a claim for those services unless the patient decides to pay out-of-pocket. Thus, almost all hospital admissions in this patient population are captured by KPNC databases, which facilitates the observation of mortality outcomes.

During original Hybrid HWM Measure development in ICD-9 data, and re-specification in ICD-10 data we partnered with KPNC to provide datasets that include all admissions for adult patients to any of their member hospitals between January 1, 2009 and June 30, 2015, and October 1, 2015 to December 30, 2016. These datasets contained both the claims data as well as the clinical data that were used to derive the cohort, outcome, comorbidities, and CCDE. The clinical data included values for the 21 data elements in the CCDE from which we derived first-captured vital signs and laboratory test results from all hospital entry locations including the emergency department, operating rooms, inpatient floors, and units. Specifically, they provided:

- Hospital identifier and hospital entry location;
- Time and date stamps for patients' arrival at the hospital for care;
- Principal discharge diagnosis;
- Secondary diagnoses;
- The patients' vital signs and laboratory test results from each admission (including data values, time and date stamps);
- Variables related to cohort exclusion criteria (discharged against medical advice, transferred from another acute care facility, discharge status); and
- Whether the patient died for any reason within 30-days from admission (from their linked administrative claims).

In addition, they provided the following information from claims submitted by their members for admissions to out-of-network hospitals: admission dates, discharge dates, and principal discharge diagnoses. In this dataset, all of these data elements were linked to a single hospital admission using a unique encounter identification number. Individual patients may have had one or more admissions in the database and were linked using unique patient identifiers assigned by KPNC.

3.3. Cohort

Index Admissions Included in the Measure

Our guiding principle for defining eligible admissions was that the measure should appropriately reflect a meaningful quality signal across a large number of acute care hospitals. Therefore, our cohort should capture as many admissions as possible for which survival would be a reasonable indicator of quality. We defined an admission as having a reasonable indicator of quality if it fulfilled two criteria: 1) survival was most likely the primary goal of the patient when they entered the hospital (for example, a patient admitted at the end of their life under hospice care for comfort measures likely does not have 30-day survival as their primary goal); and 2) the hospital could be reasonably expected to impact the chance of the patient's survival with improved quality of care (for example, the hospital does not have the ability to meaningfully impact the chance of survival for a patient admitted with brain death). Inclusion and exclusion criteria described below were based on these principles.

Grouping Patients into Clinically Coherent Categories

Because of the large and diverse number of admissions considered and tens of thousands of included ICD-10 codes, the Hybrid HWM Measure uses the Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS) to group the numerous ICD-10 codes into clinically meaningful categories. The HWM Measure use the AHRQ CCS to group the principal discharge diagnoses and major procedures, with slight modifications specific to mortality risk, described below.

CCS is a software tool developed as part of the Healthcare Cost and Utilization Project (HCUP), a Federal-State-Industry partnership sponsored by the AHRQ. It collapses ICD-10 condition and procedure codes into a smaller number of clinically meaningful condition and procedure categories.

Rationale for using CCS:

- Using ICD-10 codes would have been impractical because there are tens of thousands of ICD-10 codes, some of which occur so infrequently that using this level of detail in statistical modeling would produce unreliable results.
- AHRQ CCS categories are grouped specifically for the purpose of clinical coherence. They have been deployed in many other policy and research projects to analyze outcomes and utilization of services in hospitals.
- By using a categorization taxonomy that is widely known, publicly available, and clinically coherent, the methods are more transparent and the results are more easily interpreted.
- The AHRQ CCS categorization is consistent with the methods used in the existing CMS claims-only and hybrid HWR measures, which the Hybrid HWM Measure was designed to complement.

We have tested for, and made modifications to, highly heterogeneous CCS and low mortality CCS, as outlined in Section 3.6 Risk Adjustment, to ensure that each CCS will be a robust and accurate risk adjuster. The specifications for each clinical division can be found in the CCS Modifications tab in the data dictionary.

Inclusion Criteria

The inclusion and exclusion criteria were vetted by clinical and measurement experts on the Technical Work Group and TEP. During both development and re-specification with ICD-10 codes, codes were reviewed by a team of three clinicians who independently reviewed discharge diagnosis codes within CCSs that had high mortality rates or high variation in mortality rates for potential non-inclusion or exclusion from the measure. Any discrepancies between the three clinicians were finalized based on consensus. Final inclusion criteria are as follows:

1. **Enrolled in Medicare FFS Part A for at least 12 months prior to the date of admission and during the index admission**
    - Rationale: Claims data are consistently available only for Medicare FFS beneficiaries. The 12-month prior enrollment criterion ensures a full year of administrative data is available for risk adjustment. Medicare Part A is required at the time of admission to ensure no Medicare Advantage patients are included in the measure.
    - Note that for testing purposes only, this inclusion criterion was not applied because the KPNC data is not Medicare Claims.

2. **Not transferred from another acute care facility**
    - Rationale: Admissions to an acute care hospital within one day of discharge from another acute care hospital are considered transfers. Transferred patients are included in the measure cohort, but it is the initial hospitalization rather than any "transfer-in" hospitalization(s), that is included as the hospitalization to which the mortality outcome is attributed (the index admission).

3. **Aged between 65 and 94 years**
    - Rationale: Medicare patients younger than 65 usually qualify for the program due to severe disability. They are not included in the measure because Medicare patients younger than 65 are considered to be too clinically distinct from Medicare patients 65 and over. Patients over the age of 94 are not included to avoid holding hospitals responsible for the survival of the oldest elderly patients, who may be less likely to have survival as a primary goal. With the guidance of our work groups and TEP, we decided to only include patients between 65 and 94 years of age. While we acknowledge that many elderly patients do have survival beyond 30 days as a primary goal for their hospitalization, we also understand that, on average, very old patients may be less likely to have survival as a primary goal and that the hospital may not always be able to impact the chance of survival in the oldest elderly patients.
    - Note that for testing purposes only, we include patients between the ages of 50 and 94 because the Hybrid HWM Measure used a sample of only 21 hospitals with claims and EHR data for both development and re-specification.

4. **Not admitted for primary psychiatric diagnoses**
    - Rationale: Patients admitted for psychiatric treatment are typically cared for in separate psychiatric facilities that are not comparable to short-term acute care hospitals (see data dictionary, Non-Acute Care NonInclusion tab).

5. **Not admitted for rehabilitation**

- Rationale: These admissions are not typically to a short-term acute care hospital and are not for acute care (see data dictionary, Non-Acute Care NonInclusion tab).

6. **Not enrolled in hospice at the time of, or 12 months prior to their index admission**
   - Rationale: Patients enrolled in hospice in the prior 12 months or at the time of admission are unlikely to have 30-day survival as a primary goal.

7. **Not enrolled in hospice within two days of admission**
   - Rationale: This exclusion reflects input from our TEP and working groups and analyses performed in response to their feedback. There is not a single, correct approach regarding patients enrolled in hospice during admission or upon discharge – mortality may or may not represent a quality signal for this group of patients and hospice enrollment is inadequate to differentiate this issue. However, for most patients and/or families who had the discussion and agreed to enroll in hospice within two days of admission, 30-day survival is not likely the primary goal.

8. **Not with a principal diagnosis of cancer and enrolled in hospice during their index admission**
   - Rationale: Patients admitted primarily for cancer who are enrolled in hospice during admission are unlikely to have 30-day survival as a primary goal of care (see data dictionary, Cancer Hospice NonInclusion tabs).

9. **Without any diagnosis of metastatic cancer**
   - Rationale: Although some patients admitted with a diagnosis of metastatic cancer will have 30-day survival as a primary goal of care, for many such patients admitted to the hospital, death may be a clinically reasonable and patient-centered outcome. Therefore, this is a group of patients that may not have 30-day survival as a primary goal of care (see data dictionary, Metastatic Cancer NonInclusion tab).

10. **Not with a principal discharge diagnosis, or a secondary diagnosis that is present on admission (POA) for a condition for which hospitals have limited ability to influence survival**
    - Rationale: Hospitals have little ability to impact mortality for these conditions. This list of conditions (see data dictionary, Survival NonInclusion tab) was determined by three independent clinicians who reviewed high mortality conditions (mortality rates greater than 40% within CCSs). Using a consensus process, the three clinicians identified clinical conditions (defined by primary discharge diagnoses) where hospitals have limited ability to influence survival, and therefore, death would not be a quality signal. The list was then reviewed with our TEP and Technical Work Group. During re-specification in ICD-10, codes were again reviewed and additional codes were added. Admissions are not included in the cohort if the patient had a primary diagnosis code that is on this list, or a secondary code that is POA.

See Section 4.1 Index Cohort: Inclusion and Exclusion, Figure 3, for a cohort flowchart, including the percent of admissions that did not meet the inclusion criteria described below.

Exclusion Criteria

The final cohort that includes the percent of admissions that were excluded using the below criteria, can be found in Section 4.1 Index Cohort: Inclusion and Exclusion, Figure 3.

This measure excludes index admission for patients:

1. **With inconsistent or unknown vital status;**
    - Rationale: We do not include stays for patients where the admission date is after the date of death in the Medicare Enrollment Database, or where the date of death occurs before the date of discharge, but the patient was discharged alive.

2. **Discharged against medical advice (AMA);**
    - Rationale**:** Hospitals had limited opportunity to implement high-quality care and is not responsible for events that follow a discharge AMA.

3. **With an admission for crush injury (CCS 234), burn (CCS 240), intracranial injury (CCS 233), spinal cord injury (CCS 227), skull and face fractures (CCS 228), or open wounds of head, neck, and trunk (CCS 235)**;
    - Rationale: Even though a hospital likely can influence the outcome of some of these conditions, we felt that there were specific challenges to risk adjustment using claims data. These conditions are less frequent events that are unlikely to be uniformly distributed across hospitals and may entail distinct risk profiles (see the data dictionary, Exclusions tab).

4. **With an admission in a low volume CCS**, defined as less than or equal to 100 patients with that principal discharge diagnosis per service-line division across all hospitals.
    - Rationale: To calculate a stable and precise risk model, there are a minimum number of admissions that are needed. In addition, a minimum number of admissions and/or outcome events are required to group admissions into larger categories. These admissions present challenges to both accurate risk prediction and coherent risk grouping and are therefore excluded.
    - Note: Due to the smaller number of hospitals in the Clinical Hybrid Re-specification Dataset, we used a smaller cut-off for low volume diagnoses (25 instead of 100) to include more patients for development. As we implement this nationwide, we intend to use a 100-admission cut-off.

Addressing Patients with Multiple Admissions

The risk of mortality is not independent of the number of admissions a patient has had in a given time period, as a patient with multiple admissions can have at most one negative outcome (death). In addition, we know that the overall mortality rate for patients admitted more than once is higher than for those patients with only one admission. We also know that the percent of patients with multiple admissions that a hospital cares for varies. While patients do not always go back to the same hospital for repeat admissions, empiric analyses of Medicare data demonstrate that the majority of patients return to the same hospital. Other condition-specific hospital mortality measures reported by CMS address this issue by randomly selecting only one admission per patient per year.

As this measure includes all conditions and procedures, we systemically investigated different approaches to handling the issue of patients with multiple admissions within the measurement period. There was no practical statistical modeling approach that could account or adjust for the complex relationship between the number of admissions and risk of mortality in the context of a hospital-wide

mortality measure. Therefore, in order to provide a scientifically rigorous, statistically appropriate, and technically feasible measure that provides transparency, and where appropriate, emphasizes simplicity, we used the approach currently employed in existing CMS mortality measures of including only one randomly selected admission per patient in the one-year measurement period. This reduces the number of admissions, but does not exclude any patients from the measure.

Rationale: Random selection better reflects that the results of their hospitalizations can be death or survival when patients enter the hospital. Selecting the last admission would not be as accurate a reflection of the risk of death as random selection, as the last admission is inherently associated with higher mortality risk.

Other Cohort Considerations

With the approval of our TEP, the measure does not currently utilize codes that identify patients that have do-not-resuscitate (DNR) orders, as this is not a reliable method for determining a patient's wishes at the time of or during the admission. [**Note**: We will continue to explore clinically relevant variables related to patient care preferences for end-of-life care during measure reevaluation.]

3.4. Service-Line Divisions

It is unlikely that the effect of risk variables (such as diabetes) is homogeneous across all discharge condition categories. Therefore, we chose to group the cohort into clinically-related service-line divisions where the prevalence and effect of risk factors would likely be less heterogeneous, and then estimate separate risk adjustment regression models within each service-line division. For this multiple model approach, we create 15 different risk models for 15 different service-line divisions and then derive a summary score from the results of the models, representing an overall hospital-wide mortality rate for each hospital. This approach allows risk variables to have different effects for different conditions. For example, the effect of the comorbid risk factor of having diabetes may be different for a patient who is admitted for pneumonia than for a patient who had knee replacement surgery.

In particular, service-line divisions allow the measure to account for differences in mortality risk between surgical and non-surgical patients. Our analyses found that even within the same discharge condition, patient risk of death was strongly affected by whether a major surgical procedure was performed during hospitalization. Patients undergoing major surgical procedures are typically clinically different than those that are admitted with the same discharge condition but do not undergo a major surgical procedure. For example, a patient admitted for a hip fracture that undergoes a major surgical procedure such as hip replacement to treat their fracture is likely considered healthy enough to have the surgery, compared to patients who are so ill that they either would not survive or choose not to risk surgery. In this example, surgery is associated with a lower observed mortality rate. In other examples, surgery is likely an indicator of more severe disease. For example, patients with a principal discharge diagnosis gastrointestinal ulcer (except hemorrhage) that undergo a major surgery are generally those that have ulcers causing perforation or obstruction, which are markers of more severe disease compared to patients without perforation and obstruction requiring only medical therapy or minor surgical interventions.

In theory, estimating more models, such as a separate model for each of the diagnostic condition categories, could provide greater discrimination of mortality risk at the patient level. However, such an

approach is not feasible; many hospitals would not have an index admission in many of the condition categories rendering the measure less useful. We use distinct service-line divisions to balance the desire for more models to maximize discrimination of mortality risk with the need to minimize the number of models to ensure reliable results would be obtainable for most hospitals.

Finally, through input from the TEP and all of the work groups, we heard the importance of providing more detailed information than a single summary score for the usability of this measure for both clinicians and patients. The multiple model approach, which uses results for each of the service-line division models to create the overall hospital-wide mortality measure score, could increase the practical utility of the measure by providing information on differences in performance among divisions (service-line areas) within hospitals. This aspect of the measure could allow hospitals to better target quality improvement efforts and was supported by patients, family caregivers, and our TEP. However, further consideration is needed to identify an approach that provides transparency and granularity for consumers and ensures the division information displayed is a reliable representation of quality within each service line for each hospital.

In summary, using 15 models rather than a single model may allow for better risk adjustment for diverse patient groups, and will likely improve the usability of the measure. Using many more risk models (service-line divisions) may not be feasible given the number of cases per hospital in each condition.

Because we had matched administrative claims-EHR data for only 21 hospitals we had sufficient data for only nine service-line divisions with which to calculate the overall RSMR for the Hybrid HWM Measure. The Hybrid HWM measure uses the same concept, cohort, outcome and claims-based risk adjustment variables as the Claims-only HWM measure, therefore Claims-only HWM measure results are provided where necessary. There is no conceptual reason that the results from the Claims-only HWM measure would be substantially dissimilar to results from the Hybrid HWM Measure.

Defining Service-Line Divisions

We expect the hospital component of mortality risk to be in part related to the care provided by a team of doctors, nurses, care coordinators, pharmacists, etcetera. Conditions typically cared for by the same team of clinicians would therefore be expected to experience similar added (or reduced) levels of mortality risk. Each eligible admission is assigned to one of 15 mutually exclusive clinical divisions (see Table 1 below). The divisions reflect how care for patients is organized within hospitals. Organizing results by care team in this way will allow hospitals to identify areas of strength and weakness if their hospital performance varies across service-line divisions. This approach also addresses the strong preference of patients and caregivers to have a better understanding of hospital performance for certain conditions or procedures.

These 15 service-line divisions were created through a detailed process, led by clinicians and vetted by all of the work groups and TEP (see Appendix C Creating the Final Service-Line Divisions for details). The process consisted of the following steps:

1. Identified surgical versus non-surgical admissions;
2. Grouped admissions into 10 surgical sub-divisions and 23 non-surgical subdivisions based on clinical coherence and similar care teams;
3. Combined subdivisions into six surgical divisions and eight non-surgical divisions based on clinical coherence and risk variable performance; and

4. Presented results to work groups and TEP and, in response to feedback, added an additional surgical division of surgical cancer, creating the final 15 service-line divisions.

Table 1. List of Non-Surgical and Surgical Service-Line Divisions

Non-Surgical Divisions	Surgical Divisions
Cancer	**General***
Cardiac*	**Cancer**
Gastrointestinal*	**Cardiothoracic**
Infectious Disease*	**Neurosurgery**
Neurology*	**Orthopedic**
Orthopedics*	**Other**
Pulmonary*	**-**
Renal*	**-**
Other*	**-**

***Division included in the overall RSMR calculation (for Hybrid HWM Measure testing).**

Surgical vs. Non-Surgical Assignment

Admissions were first screened for the presence of an eligible surgical procedure category. These were defined as "major surgical procedures," representing procedures for which a patient is likely to be cared for primarily by a surgical service and identified using the approach used by the HWR measure to identify surgical admissions. Admissions with any such major surgical procedures were assigned to a surgical division, regardless of the principal discharge diagnosis code for the admission (see data dictionary, Surgical Cohort Divisions CCS tab). All remaining admissions are assigned to divisions based on the AHRQ CCS diagnosis category of the principal discharge diagnosis, modified as described below.

Identifying the Defining Surgical Procedure

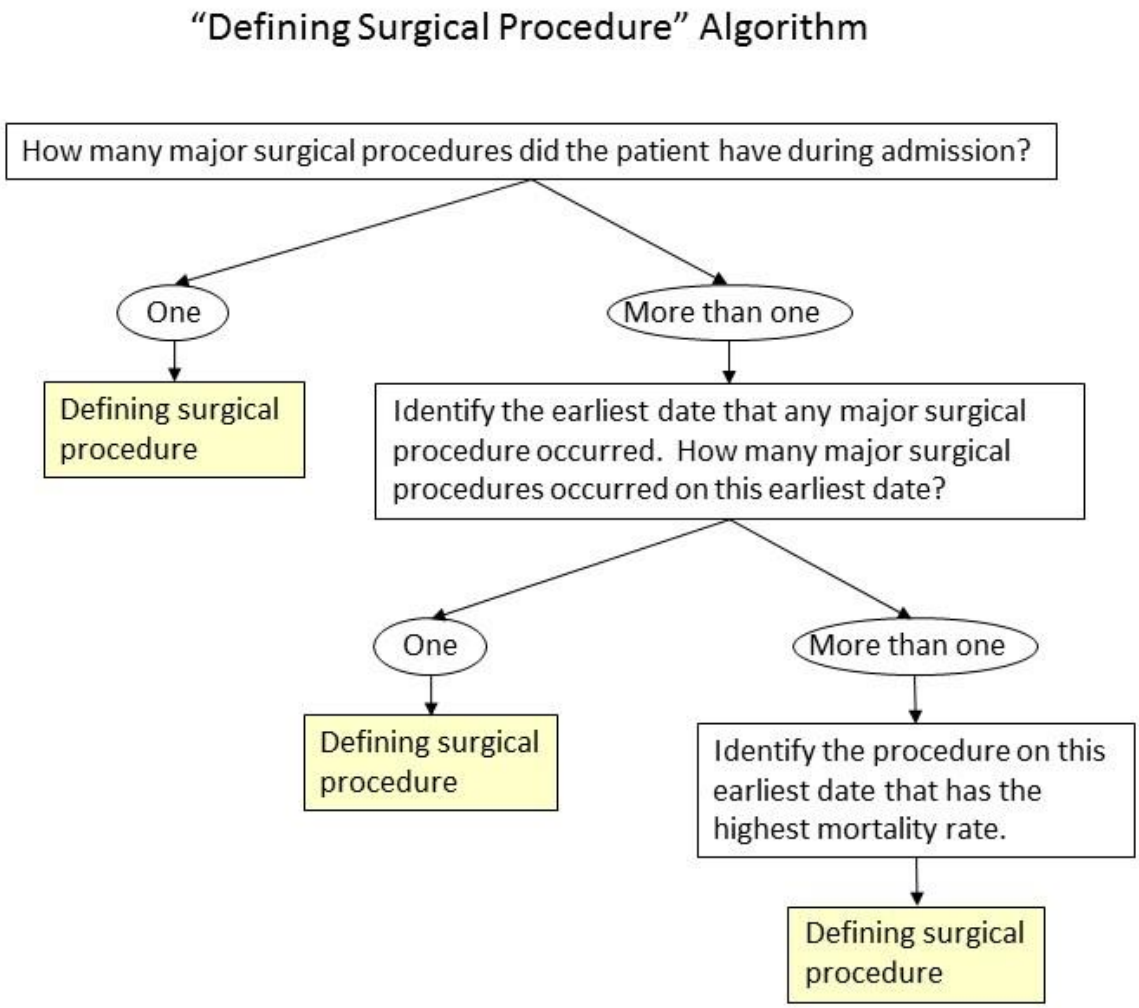
Unlike principal discharge diagnoses, of which there can only be one per admission, patients can undergo multiple surgical procedures during a hospital stay, and it is not possible in claims data to determine which, if any, procedure was related to the reason for admission. In order to report on service-line divisions that are more granular than a single division containing all surgical patients, we created an algorithm to assign a "defining surgical procedure" (Figure 1). If a patient only has one major surgical procedure, that procedure will be the "defining surgical procedure." However, if a patient has more than one major surgical procedure within a single hospitalization, the first dated major surgical procedure will be assigned as the "defining surgical procedure." If there is more than one major surgical procedure that occurs on that earliest date, the procedure with the highest mortality rate (defined by unadjusted mortality rates for all admissions with major surgical procedures from Medicare FFS data including admissions from July 1, 2013 – June 30, 2014) is the "defining surgical procedure."

The surgical cancer division is defined as an admission for a patient that undergoes any of the "major surgical procedures" and also has a principal discharge diagnosis of cancer.

Modified CCS Groupings

As described in Section 3.6 Risk Adjustment, the CCS groupings that define the clinical divisions were modified to facilitate improved risk adjustment. The specifications for each clinical division are defined in the data dictionary (CSS Modifications tab) provided with this methodology report.

Figure 1. "Defining Surgical Procedure" Algorithm



3.5. Outcome

All-Cause 30-Day Mortality

The outcome for this measure is all-cause 30-day mortality. We define mortality as death from any cause within 30 days of the index hospital admission date.

Thirty-Day Timeframe

We combined input from clinical experts with empirical analyses, published literature, and consistency with existing CMS mortality measures to define the 30-day timeframe for capturing mortality.

It is imperative to have a standard period of assessment so that the outcome for each patient is measured consistently from the date of admission. Without a standard period, variation in length of stay would have an undue influence on mortality rates, and hospitals would have an incentive to adopt strategies to shift deaths out of the hospital without improving quality. Most prior all-condition mortality measures that assess a standard time frame and all existing CMS condition- and procedure-specific hospital mortality measures utilize a 30-day timeframe, starting the day of admission, for assessing mortality.

To evaluate the appropriateness of the 30-day time frame across the HWM cohort, we reviewed survival curves for Medicare beneficiaries 65 years and older across all diagnostic CCS groupings up to 90 days following admission. We found that diagnostic CCS groups with the highest mortality rates had their steepest declines in the first few days and the curves continued to decline but at a slower rate after that time. In general, few diagnostic CCS groups showed complete leveling off of mortality, even at 90 days. However, the 30-day period does capture the largest declines in mortality.

Additional support for the 30-day time frame stemmed from evidence that mortality can be influenced by hospital care and the early transition to the outpatient setting during this time. Finally, we reviewed the 30-day timeframe with our Technical Work Group, Patient and Family Caregiver Work Group and TEP, and they supported the 30-day timeframe. In summary, we chose a post-admission observation period of 30-days balancing considerations of empirical data findings, actionability, cross-measure consistency, and fairness of attribution.

All-Cause Mortality

We defined the outcome as “all-cause” mortality rather than related to the index hospitalization for multiple reasons. First, from the patient perspective, mortality for any reason is an undesirable outcome of care. In defining the measure cohort, we worked with clinical experts and patients to only include patients for whom it is reasonable to assume that 30-day survival is a primary goal of care. Second, there is no reliable way to determine whether mortality is related to the index hospitalization based on the documented cause of mortality. As with readmissions, many deaths that might not be deemed related are in fact influenced by the care received during hospitalization. For example, a heart failure patient who is discharged with inappropriately dosed medications may develop renal failure from over-diuresis and die. It would be inappropriate to treat this death as unrelated to the care the patient received for heart failure. Third, all existing CMS mortality measures report all-cause mortality, making this approach consistent with existing measures. Finally, defining the outcome as all-cause mortality may encourage hospitals to implement broader initiatives aimed at improving the overall care within the hospital and

transitions from the hospital setting instead of limiting the focus to a narrow set of condition- or procedure-specific approaches.

Outcome Attribution

Outcomes for surgical and non-surgical patients are attributed to the admitting hospital. In cases of transfers, the sequence of hospitalizations is treated as one episode of care and the admission and associated outcome are attributed to the first admitting hospital. For example, if a patient is admitted to acute care Hospital A, and then transferred to acute care Hospital B, the admission and associated outcome (survival or death within 30-days) is attributed only to Hospital A.

A surgical transfer patient is defined as a patient who is originally admitted to one hospital where no major surgical procedure is performed and is then transferred to a different hospital where they receive a major surgical procedure. Given that surgical transfer patients are more likely to have risks that are similar to other surgical patients (rather than non-surgical patients), we assigned surgical transfer patients to a surgical division for risk adjustment and reporting (rather than a non-surgical division). However, the mortality outcome remains attributed to the original admitting hospital that made the decision to both admit and transfer the patient.

3.6. Risk Adjustment

Overview

The goal of risk adjustment is to account for differences across hospitals in patient demographic and clinical characteristics that might be related to the outcome but are unrelated to quality of care. The Hybrid HWM measures adjust for case mix differences (clinical status of the patient on admission, accounted for by adjusting for comorbidities and diagnoses POA), and service mix differences (the types of conditions/procedures cared for by the hospital, accounted for by adjusting for the discharge condition category). In addition, the Hybrid HWM measure adds EHR data to the case mix risk adjustment in the form of 10 clinical risk variables extracted from EHRs. Please note that the case-mix variables are the same but their coefficients can vary by between divisions.

We do not adjust for patients' admission source or discharge disposition (for example, skilled nursing facilities) because these factors are associated with the structure of the health care system, and may reflect the quality of care delivered by the system. We do not adjust for socioeconomic status, gender, race, or ethnicity because hospitals should not be held to different standards of care based on the demographics of their patients; This measure was endorsed by NQF without adjustment for patient-level socioeconomic status factors.

The risk adjustment variables included in the Hybrid HWM measure include the following three types of risk variables:

1. Case Mix (claims-derived comorbidities): comorbidity risk variables derived from administrative claims data. Comorbidities for inclusion were identified during the 12 months prior to and including the index admission. To assemble the ICD-10 codes into clinically coherent variables for risk adjustment, the measure employs the publicly available CMS condition categories (CMS-CCs) to group codes into CMS-CCs, and selects comorbidities on the basis of clinical relevance and statistical significance;[35]

2. Case Mix (clinical EHR data): clinical data outlined in Case Mix Risk Adjustment (EHR-Based Clinical Status Risk Variables), derived from the Clinical Hybrid Re-specification Dataset; and
3. Service Mix (principal discharge diagnoses): the AHRQ CCS categories for the principal discharge diagnosis associated with each index admission derived from ICD-10 codes in administrative claims data from the index admission. These are also the codes that are used to define the service-line divisions for the non-surgical divisions.

The final risk adjustment model can be found in the data dictionary, RiskVariable ParameterEstimates. Below we explain our general approach to capturing patient-level case mix in the risk model, followed by an explanation of service-line risk adjustment. These sections are followed by a description of the division-level and overall hospitals-level statistical models in detail.

Case-Mix Adjustment

To account for differences in case mix among hospitals, the measure adjusts for variables (that is, age and comorbid diseases) that are clinically relevant and have relationships with the outcome. Case mix differences among hospitals are based on the clinical status of the patient at the time of the index admission. Accordingly, only comorbidities that convey information about the patient at the time of the index admission, or any time within the preceding 12 months, are included in risk adjustment. Complications that arise during the course of the hospitalization are not used in risk adjustment.

Refer to the Complications tab in the data dictionary for the list of comorbidity risk-adjustment variables common to all divisions and the list of potential complications that are excluded from risk adjustment (detailed below in Complications of Hospitalization). The Condition Categories (CCs) outlined in this table are used to identify risk variables in claims for discharges on or after October 1, 2015 for ICD-10-CM diagnosis codes as well as discharges prior to October 1, 2015 for ICD-9-CM diagnosis codes.

Initial Development of Comorbid Risk Variables

Our goal was to develop parsimonious models that include clinically relevant variables strongly associated with the risk of mortality in the 30 days following an index admission. For candidate variable selection, using the Medicare Claims-Only Development Dataset, we started with the CMS-CC grouper, used in previous CMS risk-standardized outcome measures, to group ICD-9-CM codes into comorbid risk-adjustment variables.

To select candidate variables, a team of clinicians reviewed all CMS-CCs and combined some of these CMS-CCs into clinically coherent groups to ensure adequate case volume. Any combined CMS-CCs were combined using both clinical coherence and consistent direction of mortality risk prediction across the CMS-CC groups in the majority of the 15 divisions. All original candidate risk variables are listed in Appendix C, Table C.1 Candidate Comorbid (Claims-Based) Risk Variables.

To select final risk variables, we used the claims-only development sample to create 500 bootstrap samples for each of the service-line divisions. For each sample, we ran a standard logistic regression model that included all candidate variables. The results were summarized to show the percentage of times that each of the candidate variables was significantly associated with 30-day mortality (at the $p<=0.05$ level) in the 500 bootstrap samples (for example, 70% would mean that the candidate variable was significant at $p<=0.05$ in 70% of the bootstrap samples). We also assessed the direction and magnitude of the regression coefficients.

We found that models containing all risk factors performed similarly to models containing a more limited set of "significant" risk factors, described below. We therefore used a fixed, common set of comorbidity variables in all of our models for simplicity and ease of implementation and analysis. Please note that the case-mix variables are the same but their coefficients can vary by between divisions.

We describe below the steps for variable selection.

1. The CORE Project Team reviewed the bootstrapping results and decided to provisionally examine risk adjustment variables at or above a 90% cutoff in one of the 15 service-line division models (in other words, retain variables that were significant at the p<=0.05 level in at least 90% of the bootstrap samples for each division). We chose the 90% cutoff because this threshold has been used across other measures and produced a model with adequate discrimination.
2. In order to develop a statistically robust and parsimonious set of comorbid risk variables, we then chose to limit the variables to those that met a 90% threshold in at least 13/15 divisions. This step resulted in the retention of 20 risk factors, including age and 19 comorbid risk variables. This resulted in C-statistics that did not change by more than 0.02 in any of the 15 divisions compared to models that contained all possible risk variables.

To re-specify these risk variables in ICD-10, we used established ICD-9 to ICD-10 cross-walks with the Fiscal Year (FY) 2017 version of the V22 CMS-HCC. We then compared the volume of admissions included in each division using ICD-10 codes in a set of claims submitted between July 2016 and June 2017, with volume of admissions in each division using previously-defined ICD-9 codes in a set of claims submitted between July 2014 and June 2015. We also then re-ran the risk-adjustment models. Final comorbid risk variables are listed in Table 2.

Table 2. Final Comorbid Risk Variables

Comorbid Risk Variables
Age
Other Infectious Diseases (CC 7)
coma/brain compression/ anoxic injury and severe head injury (CC 80, 166)
Metastatic & Severe Cancers (CC 8,9)
Protein-Calorie Malnutrition (CC 21)
Disorders of Fluid/Electrolyte/Acid-Base Balance (CC 24)
Disorders of Lipoid Metabolism (CC 25)
Liver Failure (CC 27,30)
Other GI Disorders (CC 34, 35, 3738)
Other Musculoskeletal and Connective Tissue Disorders (CC 44,45)
Hematologic or Immunity Disorders (CC 46-48)
Dementia and Other Nonpsychotic Organic Brain Syndromes (CC 51-53)
Respiratory Failure, Respirator Dependence, Shock (CC 82-84)
Congestive Heart Failure (CC 85)
Hypertension and hypertensive heart disease (CC 94,95)
Pneumonia (CC 114-116)

Comorbid Risk Variables
Dialysis or Severe Chronic Kidney Disease (CC 134,136,137)
Acute or Unspecified Renal Failure (CC 135,140)
Poisonings and Allergic and Inflammatory Reactions (CC 175)
Minor Symptoms, Signs, Findings (CC 179)

Complications of Hospitalization

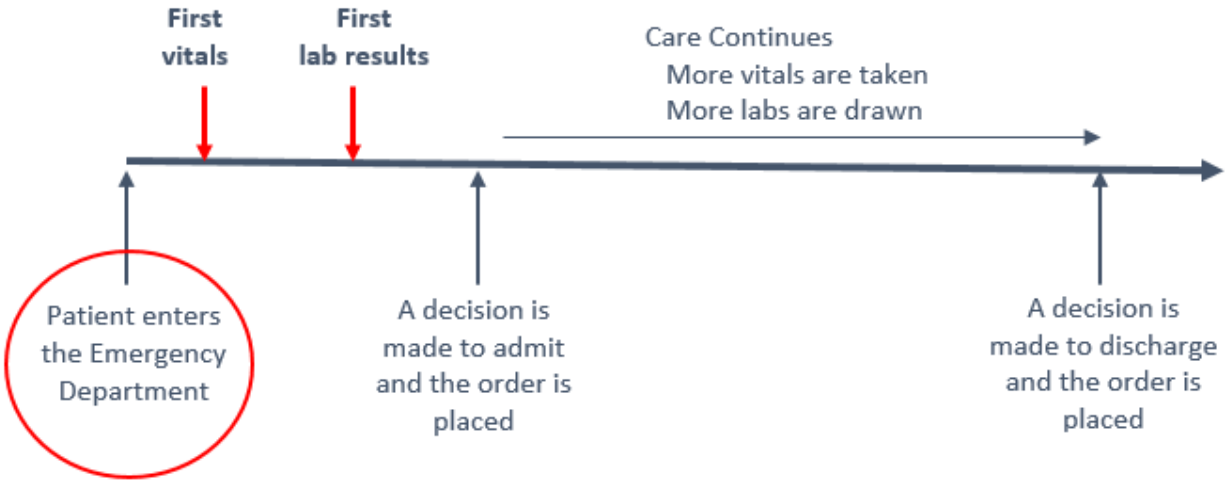
Complications occurring during hospitalization are not comorbid illnesses and do not reflect the health status of patients upon presentation. In addition, they likely reflect hospital quality of care, and, for these reasons, should **not** be used for risk adjustment. Although adverse events occurring during hospitalization may increase the risk of mortality, including them as risk factors in a risk-adjusted model could lessen the measure's ability to characterize the quality of care delivered by hospitals. We have previously reviewed every CMS-CC and identified those which, if they were to occur only during the index hospitalization, are more likely than not to represent potential complications rather than pre-existing comorbidities. For example: fluid, electrolyte, or base disorders; sepsis; and acute liver failure are all examples of CMS-CCs that could potentially be complications of care (see Complications tab in the data dictionary for the list).

We took a two-step approach to identifying complications of care. First, we searched the secondary diagnosis codes in the index admission claim for all patients in the measure and identified the presence of any ICD-10 code associated with a CMS-CC in in the Potential Complications of Care (see data dictionary, Complications tab). If these codes appeared only in the index admission claim, we flagged them because they are potential complications of care. Next, we determined if these potential complications of care were associated with a "present on admission" code. Any potential complication of care with an associated "present on admission" code was kept in the risk model; any potential complication of care without an associated "present on admission" code was removed under the assumption that it represented a complication of care. In this way, we supplemented the existing approach to identifying potential complications of care used in CMS's publicly reported mortality measures by incorporating "present on admission" codes. Our analyses demonstrate that a majority of hospitals currently use "present on admission" codes across a majority of conditions. Therefore, we felt that a combined approach to excluding complications of care from the risk model that used both the existing methodology and "present of admission" codes allow the measure to capture as many clinically appropriate risk variables as possible while simultaneously removing complications of care from the risk model.

Case Mix Risk Adjustment (EHR-Based Clinical Status Risk Variables)

The electronic clinical risk variables reflect a patient's clinical status upon arrival to the hospital as the first captured value or first set of vital signs and laboratory results. For example, as shown in Figure 2 below, we would incorporate only the first set of vital statistics (for example, blood pressure) and laboratory results (for example, glucose) on the patient once they arrive at the hospital.

Figure 2. Identifying First Captured Values for the Core Clinical Data Elements



To be able to use electronic clinical data for a measure of hospitals nationally, we must collect accurate data from all hospitals. Because of this, any electronic clinical data that we use must meet the following criteria:

1. Consistently captured on all adult hospitalized inpatients;
2. Captured with a standard definition; and
3. Entered into the electronic health record in a structured field and feasibly extracted.

Example of usable data elements: Blood Pressure

- ✓ Captured on all patients upon arrival at the hospital in any setting (hospital outpatient, inpatient, emergency department);
- ✓ Captured using the same units of measurement across the country (mmHg); and
- ✓ Entered into a structured field (numeric) in the EHR that can be extracted.

Example of unusable data element: Medication history or adherence

- × Inconsistently or not reliably collected on all patients by clinicians;
- × Units of measurement could range;
- × name of medication could differ; or
- × Possibly captured in clinical notes, and not a structured field.

The CCDE are a standard "set" of clinical data consistently obtained on hospital inpatients and feasibly extracted from EHRs, as shown in Table 3. We have shown that these variables are consistently captured with a standard definition, entered in a structured field, and feasibly extracted.[36,37] Therefore, they represent a feasible set of candidate variables from which to select our risk model. The CCDE were designed to be a dynamic list that can be modified for specific measures, and potentially expanded as the use of EHRs evolves and clinical practice changes over time.

Table 3. Currently Specified CCDE Variables

Clinical Data Elements	Units of Measurement	Window for First Captured Values
Patient Characteristics		
Age	Years	-
First-Captured Vital Signs		
Diastolic Blood Pressure	mmHg	0-2 hours
Heart Rate	Beats per minute	0-2 hours
Oxygen Saturation	Percent	0-2 hours
Respiratory Rate	Breath per minute	0-2 hours
Systolic Blood Pressure	mmHg	0-2 hours
Temperature	Degrees Fahrenheit	0-2 hours
Weight	Pounds	0-24 hours
First-Captured Laboratory Results		
Anion Gap	mEq/L	0-24 hours
Bicarbonate	mmol/L	0-24 hours
BUN	mg/dL	0-24 hours
Chloride	mEq/L	0-24 hours
Creatinine	mg/dL	0-24 hours
Glucose	mg/dL	0-24 hours
Hematocrit	Percent	0-24 hours
Hemoglobin	g/dL	0-24 hours
Platelet	Count	0-24 hours
Potassium	mEq/L	0-24 hours
Sodium	mmol/L	0-24 hours
WBC Count	10^9 per liter (X10E+09/L)	0-24 hours

CCDE Risk Variable Selection

CCDE risk variable selection occurred in the initial development of the measure. To select candidate clinical EHR variables, we began with the list of CCDE variables, listed above in Table 3 Currently Specified CCDE Variables.

First, we looked at how many admissions in our Clinical Hybrid Development Dataset were missing values for each CCDE. The non-surgical divisions had fewer than 10% of admissions that were missing values. However, in the surgical divisions, while vitals were missing in fewer than 10% of admissions, the laboratory result values were missing in 15% - 50% of admissions, depending upon division. For **development purposes only**, we imputed values for missing labs or vital signs, as described below:

- For all admissions missing any vital signs and for admissions within the non-surgical divisions missing any laboratory result values, we used multiple imputation (imposing limits to ensure the imputed values were within clinical possibilities) with 5 copies of data with different imputations based on a multi-normal distribution.
- For admissions within the surgical divisions missing any laboratory results, we randomly imputed a value within the normal range for that lab. For the normal ranges, see Table 4 below.

- **Rationale:** Surgical patients that are missing initial labs are most likely elective surgical admissions that had the labs collected within 30 days prior to admission. It is less likely that a patient with an extremely abnormal lab value would undergo an elective surgery without having the labs checked again on admission. **This approach is for development purpose only**.

Second, we selected which CCDE would be the most appropriate to include in the Hybrid HWM measure. We approached risk variable selection from the perspective of ensuring a parsimonious list of clinical EHR variables that would minimize hospital burden to report the data and provide face validity from a clinical perspective.

Therefore, we first sought to ensure that each candidate variable was modeled in a clinically appropriate way. For example, the laboratory value sodium has a U-shaped predictive association with mortality: Normal sodium levels are associated with a low risk of mortality, while both abnormally high and abnormally low levels are associated with an increased risk of mortality. The association between each CCDE variable and mortality was reviewed by four clinicians and selected based on the best association. See Table 4 for the approach used for each risk variable. In addition, we report the normal values used for imputing missing laboratory results within the surgical divisions.

Table 4. Candidate Clinical EHR Risk Variable (CCDE) Mortality Association Modelling Approaches

Candidate EHR Risk Variables	Normal Range	Modelling Approach
Age	-	linear
Diastolic Blood Pressure	-	splined, knot at 80
Heart Rate	-	linear
Oxygen Saturation	-	linear
Respiratory Rate	-	splined, knot at 16
Systolic Blood Pressure	-	splined, knot at 140
Temperature	-	linear
Weight	-	splined, knot at 180
Anion Gap	7-17	splined, knot at 10
Bicarbonate	22-30	splined, knot at 26
BUN	8-18	splined, knot at 14 and 40
Chloride	96-106	linear
Creatinine	0.5-1.2	linear but winsorized at 5
Glucose	70-100	splined, knot at 180
Hematocrit	37-52	linear
Hemoglobin	12-18	linear
Platelets	140-440	splined, knot at 200
Potassium	3.3-5.0	splined knot at 4.0
Sodium	135-145	splined, knot at 140
White Blood Count	4.0-10.0	splined, knot at 7.0

Next, we examined the strength of different clinical variables in the context of a multivariable model. We performed bootstrapping with 1,000 iterations allowing patient admissions to be repeatedly selected and produced 1,000 bootstrapping samples for each of the 5 multiple imputations (for the missing data). We used logistic regression with stepwise selection to create risk models for each division in each bootstrapping sample in each imputation run, identifying the variables most significantly associated with

mortality for that division (present in 80% or more of runs). This approach produced risk models that might be missing important clinical variables. For example, the selected model for the Surgical Cancer Division contained only age and blood urea nitrogen (BUN) and the model discrimination (judged by using the c-statistic) was not as strong as compared to the model that only used administrative claims (comorbidities, principal discharge diagnosis).

Based upon this information, we selected a standard set of clinically coherent risk variables in order to ensure that each division-level risk model included key laboratory results and vital signs data. As with prior hybrid measures that use EHR data in their risk model, we did not include risk variables if they were strongly correlated with another variable. For example, we selected systolic blood pressure but not diastolic blood pressure, as these variables were highly correlated and provide very similar risk prediction. Using a standard set of clinically selected variables produced improved c-statistics compared to the models based purely upon stepwise selection. We also tested allowing the risk variables to vary across the 15 divisions (using stepwise selection) but still forcing in clinical variables and found that the model discrimination (c-statistic) was very similar, in some cases identical, to using a standard set of variables. Therefore, we proceeded with a common set of 10 clinical risk variables plus age across all divisions (Table 5).

Table 5. Final CCDE Risk Variables

Final CCDE	Units of Measurement	Window for First Captured Values
First-Captured Vital Signs		
Heart Rate	Beats per minute	0-2 hours
Oxygen Saturation	Percent	0-2 hours
Systolic Blood Pressure	mmHg	0-2 hours
Temperature	Degrees Fahrenheit	0-2 hours
First-Captured Laboratory Results		
Bicarbonate	mmol/L	0-24 hours
Creatinine	mg/dL	0-24 hours
Hematocrit[‡]	g/dL	0-24 hours
Platelet	Count	0-24 hours
Sodium	mmol/L	0-24 hours
WBC Count	10^9 per liter (X10E+09/L)	0-24 hours

Since the Hybrid HWM Measure was developed, we have updated the core clinical data elements to include hematocrit instead of hemoglobin to better align the Hybrid HWM with the Hybrid HWR measure that is currently reported by hospitals to CMS on a voluntary basis. This change was made to better align the two hybrid measures and reduce the reporting burden for hospitals. Hemoglobin and hematocrit values are highly correlated (correlation ranged from 0.88—0.99 by service-line division, according to testing results from the Clinical Hybrid Re-specification Dataset, and risk model performance was not impacted by the switch to hematocrit.

[‡] Hemoglobin will be replaced by Hematocrit for future iterations of this measure

Service-Mix Adjustment

To account for differences in service mix among hospitals, the measure adjusts for the principal discharge diagnosis of the index admission (grouped into AHRQ CCS diagnosis categories, modified as described below). Thus, for the service-line divisions, the AHRQ CCS diagnosis categories used for risk adjustment are the same as those used to define each of the divisions (see data dictionary tab, Non-Surgical Cohort Divisions CCS tab).

Rationale: Principal discharge diagnoses differ in their baseline mortality risks and hospitals will differ in their relative distribution of these principal discharge diagnoses (service mix) within each division. Therefore, adjusting for these principal discharge diagnoses levels the playing field across hospitals with different service mixes.

Highly Heterogeneous CCSs

For some of the CCS groups, risk of mortality varied significantly across the different primary discharge diagnoses within the CCS. During measure development, using ICD-9-coded data, there was concern voiced by our Technical Work Group and TEP that we may not be adequately risk-adjusting using these heterogeneous CCSs. To identify heterogenous CCSs, we calculated the correlation between mortality rate and inpatient admissions grouped by principal discharge diagnosis ICD-9 code within each CCS. We identified any CCS with an intraclass correlation (ICC) score >0.05 as having high heterogeneity. The ICC is used in this context to identify heterogeneity of mortality risk across primary discharge diagnosis codes within the ICC. The value of 0.05, or 5%, is a conventional threshold for accounting for between group heterogeneity. For ICD-10 re-specification, we revisited this approach, but calculated the ICC after patients were separated into surgical and non-surgical divisions.

To address the heterogeneity, three clinicians independently, and through consensus, modified the highly heterogeneous CCSs using clinically informed recategorizations. We modified these highly heterogenous CCSs by either:

1. Splitting the CCSs into more than one CCS,
2. Moving ICD-10 codes to more clinically coherent CCSs,
3. Making no change to the CCS based on the CCS's clinical cohesiveness, or based on the number of patients affected; or
4. Identifying and removing for purposes of the cohort definition, admissions with primary ICD-10 codes are clinically different from others in the CCS and for which care unlikely impacts survival (where mortality is not a quality signal).

During ICD-10 re-specification, we identified 44 highly heterogeneous CCSs and made modifications to 20 of them, as described in the data dictionary, tab CCS Modifications. The changes to the CCSs resulted in more homogenous CCS risk variable groups and increased the face validity of the risk models.

Therefore, CCSs for risk adjustment and cohort have been slightly modified from the AHRQ definitions. For consistency, these changes were also applied to the service-line division definitions for the non-surgical divisions and are reflected in the final division definitions as specified in the data dictionary, Non-Surgical Cohort Divisions CCS.

Low-mortality CCSs

During initial measure development, the patient-level risk models for two divisions (the "Other" surgical and "Other" non-surgical divisions) did not converge due to the large number of CCS category codes in these divisions, and due to low mortality rates associated with some of the CCSs in these divisions (which are used for service-line risk adjustment). However, the TEP and Patient and Family Caregiver Work Group had a strong interest in retaining these admissions (more than half a million admissions) in the measure. To address this issue, within each division, CCSs with low mortality rates (those less than or equal to 1%) are combined into one independent group, which reduces the total number of risk variables (CCS category codes) in the model.

Final Risk Model Selection

During initial development, three different risk models were tested in the Clinical Hybrid Development Dataset to select the best risk model based on statistical performance and face validity as determined by our TEP. We tested the following risk models:

1. "Clinical-Only Risk Model": Uses only EHR-based clinical variables in risk model (no claims comorbidity OR principal discharge diagnoses)
    a. Service mix: None
    b. Case mix: age plus 10 clinical variables captured from EHR data
2. "Clinical + Principal Discharge Diagnoses Risk Model": Uses EHR-based clinical variables with claims-based principal discharge diagnoses in risk model (no claims comorbidity)
    a. Service mix: AHRQ CCS categories for patients' principal discharge diagnoses captured from claims data
    b. Case mix: age plus 10 clinical variables captured from EHR data
3. "Clinical + Claims Risk Model": Uses EHR-based clinical variables + claims-based comorbidity and principal discharge diagnosis variables in risk model:
    a. Service mix: AHRQ CCS categories for patients' principal discharge diagnoses captured from claims data
    b. Case mix: Both the age plus 10 clinical variables captured from EHR data and the CCs for patients' comorbidities captured from claims data during hospitalizations in the 12 months prior to and including the index admission (19 CC risk variables and age plus 10 clinical variables for each division risk model)

After reviewing the results with our TEP and based upon their preference for higher discrimination over other features (parsimony, not requiring 12 months of history data), we selected Clinical + Claims Risk Model for the Hybrid HWM measure. See Appendix D Final Risk Model Selection for more testing details.

Rationale: The Clinical + Claims Risk Model, which includes the broadest set of risk variables, had the best statistical performance and the highest face validity per the majority of the TEP by accounting for clinical EHR variables, principal discharge diagnoses, and comorbidities identified using claims-data. While it does require the exclusion of patients not enrolled in Medicare for 12 months prior to admission, this was the preferred model by the majority of the TEP.

3.7. Measure Calculation

To calculate an overall hospital-wide mortality rate, we needed to combine the results of the 15 risk models (divisions) into one overall score. We envisioned a HWM measure that will provide a broad indication of a hospital's performance and capture cross-cutting hospital-wide characteristics that contribute to quality of care. As with CMS's other claims-based performance measures, the measure result will be a point estimate (the RSMR). While there are multiple approaches to calculate this overall RSMR through combining the results of the 15 models, after consultation with multiple statisticians and review with our Technical Work Group, our patient and family caregiver working groups, and our TEP during initial development, we are using a weighted mean with a hierarchical general logistic model (HGLM) approach. In the future when a national sample is available, an empirical correlation approach such as Markov Chain Monte Carlo (MCMC), which produces a statistically precise and conservative estimate of better and worse outliers, may be applied.

Weighted Mean with Volume

This approach requires first calculating a Standardized Mortality Ratio (SMR) for each hospital for each service-line division and then combines the SMRs for each hospital's divisions by taking the weighted average of the performance in each of the divisions, taking into account how precisely we were able to predict the outcome for that division. In technical terms, to calculate the point estimate for each hospital, we used the point estimates of all 15 SMRs (one from each division) and took the volume-weighted mean to create an overall hospital-wide combined SMR, similar to the HWR measure methodology. To calculate the RSMR for each hospital, we multiplied the overall hospital-wide SMR by the national observed mortality rate. The statistical approach is described in greater detail below.

Statistical Approach to Calculating Division-Level and Overall Standardized Mortality Ratios

This section provides further detail on the specific technical information for the statistical modeling for creating the final measure results. This includes information on the statistical models for each of the 15 divisions, how the results are calculated for each of the divisions, and then how those results are combined to form the overall mortality rate.

Models for Each Service-Line Division

We created 15 service-line division patient-level risk-adjustment models using logistic regression, with outcome Y_{idj} for the i-th patient in d division at the j-th hospital equal to 1 if the patient died within 30 days of admission and 0 otherwise. The patient-level risk-adjustment models allowed us to assess risk factors and model performance without reference to the variation in performance across hospitals.

For the hospital-level results among each of the 15 service-line divisions, we used hierarchical logistic regression models where death within 30 days is modeled as a function of patient-level demographic and clinical characteristics and a random hospital-level intercept. This accounted both for the natural clustering of observations within hospitals and captured a hospital-specific signal. We used the results of each hierarchical logistic regression model to calculate a standardized mortality ratio (SMR) for each hospital. The SMR was computed as the predicted mortality rate divided by the expected mortality rate at each hospital for each division. These contributing SMRs were then pooled for each hospital to create an overall hospital-wide SMR using the volume-weighted mean approach. To aid interpretation, this ratio was then multiplied by the overall national observed mortality rate for all index admissions in all cohorts, to produce the risk-standardized mortality rate or RSMR.

Specifically, for a given service-line division, we estimated a hierarchical logistic regression model as follows. Let Y_{idj} denote the outcome (equal to 1 if patient *i* in *d* division at hospital *j* dies within 30 days, 0 otherwise) for a patient in a specified division $d \subseteq \{1,...,15\}$, at hospital *j*; $\mathbf{Z_{idj}}$ denotes a set of risk factors. Let *M* denote the total number of hospitals and V_{dj} the number of index patient stays among *d* division in hospital *j*. We assume the outcome is related linearly to the covariates via a logit function:

$$\text{logit}(\Pr(Y_{idj} = 1)) = \alpha_{dj} + \boldsymbol{\beta} * \mathbf{Z_{idj}} \quad (1)$$

$$\alpha_{dj} = \mu_d + \omega_{dl}$$

$$\omega_{dj} \sim N(0, \tau_d^2)$$

where $\mathbf{Z_{idj}} = (Z_{idj1}, Z_{idj2}, \ldots Z_{idjk})$ is a set of *k* patient-level covariates. α_{dj} represents the hospital-specific intercept in d division; μ_d is the adjusted average outcome over all hospitals in *d* division; and τ_d^2 is the between hospital variance component. The hierarchical logistic regression model for each cohort was estimated using the SAS software system (GLIMMIX procedure).

Standardized Mortality Ratio for Each Service-Line Division

We used the results of each hierarchical logistic regression model to calculate standardized mortality ratio as the predicted number of deaths over the expected number of deaths for each service-line division at each hospital. The predicted mortality rate in each division was calculated, using the corresponding hierarchical logistic regression model, as the sum of the predicted probability of death for each patient, including the hospital-specific (random) effect. The expected number of deaths in each division for each hospital were similarly calculated as the sum of the predicted probability of death for each patient, setting the hospital-specific (random) effect to be zero. Using the notation of the previous section, the model specific risk-standardized mortality ratio was calculated as follows. To calculate the predicted mortality rate $pred_{dj}$ for index admissions in each division d=1,...,15 at hospital *j*, we use:

$$pred_{dj} = \sum \text{logit}^{-1}(\alpha_{dj} + \boldsymbol{\beta} * \mathbf{Z_{idj}}) \quad (2)$$

where the sum is over all m_{Dj} index admissions in division d with index admissions at hospital *j*. To calculate the expected number exp_{dj} we use:

$$exp_{dj} = \sum \text{logit}^{-1}(\mu_d + \boldsymbol{\beta} * \mathbf{Z_{idj}}) \quad (3)$$

Then, as a measure of excess or reduced mortality rate among index admissions in cohort D at hospital *j*, we calculate the standardized mortality ratio SMR_{dj} as:

$$SMR_{dj} = pred_{dj}/exp_{dj} \quad (4)$$

Hospital-Wide Risk-Standardized Mortality Rate

To report a single mortality score, the separate service-line division SMRs are combined into a single value.

For a given hospital, *j*, which has patients in some subset of divisions $d \subseteq \{1,...,15\}$, we calculate the SMR as described above for each division for which the hospital discharged patients. If the hospital does not have index admissions in a given division *d*, then the weight $w_{dj} = 0$. Then, calculate the variance-weighted logarithmic mean:

$$SMR_j = \exp((\sum w_{dj} \log(SMR_{dj})) / \sum w_{dj}) \quad (5)$$

where the sums are over all service-line divisions and w_{dj} is the hospital volume V_{dj}; note that if a hospital does not have index admissions in a given division ($w_{dj} = 0$) then that cohort contributes nothing to the overall score SMR_j. This value, SMR_j, is the hospital-wide standardized mortality ratio for hospital *j*. To aid interpretation, this ratio is then multiplied by the overall national observed mortality rate for all index admissions in all cohorts, $\bar{Y}$, to produce the risk-standardized hospital-wide mortality rate ($RSMR_j$).

$$RSMR_j = SMR_j * \bar{Y} \quad (6)$$

Creating Interval Estimates

For Hybrid HWM Measure development, confidence interval estimates were not calculated due to the smaller sample size in the development and re-specification datasets. In the future, the below approach can be used to create confidence interval estimates.

We will first estimate the mean and variance for each $\log(SMR)_{dj}$ based on the MCMC posterior distribution of the $\log(SMR_{dj})$. We let $\log(SMR_d)$ denote the vector of $\log(SMR_{dj})$, where j=1,2,…,J. We will then utilize all posterior means of $\log(SMR_{dj})$ from each division and each hospital, if it exists, to construct the covariance matrix of $\log(SMR_d)$, where d=1,2,…,15. This covariance matrix estimates the dependency of SMRs between divisions and will be same for all the hospitals. We then will construct our confidence interval for SMR_j by considering all possible variances and covariances. Let f(.) denotes the equation (5). According to the delta method, we have:[38]

Because the $\log(SMR_{dj})$ are estimates rather than observations we will account for the measure errors using $\sum_{d=1}^{D}(Var(\log(SMR_{dj}))$, which will be estimated from the posterior distribution. Because we will not assume the $\log(SMR_{dj})$ from different divisions are independent we cannot set the covariances to zero; instead as an approximation we will sum over all the empirical variances and covariances of $\log(SMR_{dj})$ using $\sum_{d=1}^{D}\sum_{d'=1}^{D} Cov(\log(SMR_d), \log(SMR_{d'}))$, which will be from the covariance matrix. Assuming a normal distribution for each SMR_j, the confidence interval estimates will be calculated as $SMR_j \pm Z_{0.975} \times SD(SMR_j)$ where $Z_{0.975}$ is the 97.5% quantile for a standard normal distribution.

Given $RSMR_{j=} SMR_j * \bar{Y}$, we will calculate the lower and upper bound of the confidence interval for $RSMR_j$ by multiply $\bar{Y}$ to the corresponding estimates of the lower and upper bound of the SMR_j.

4. MEASURE TESTING

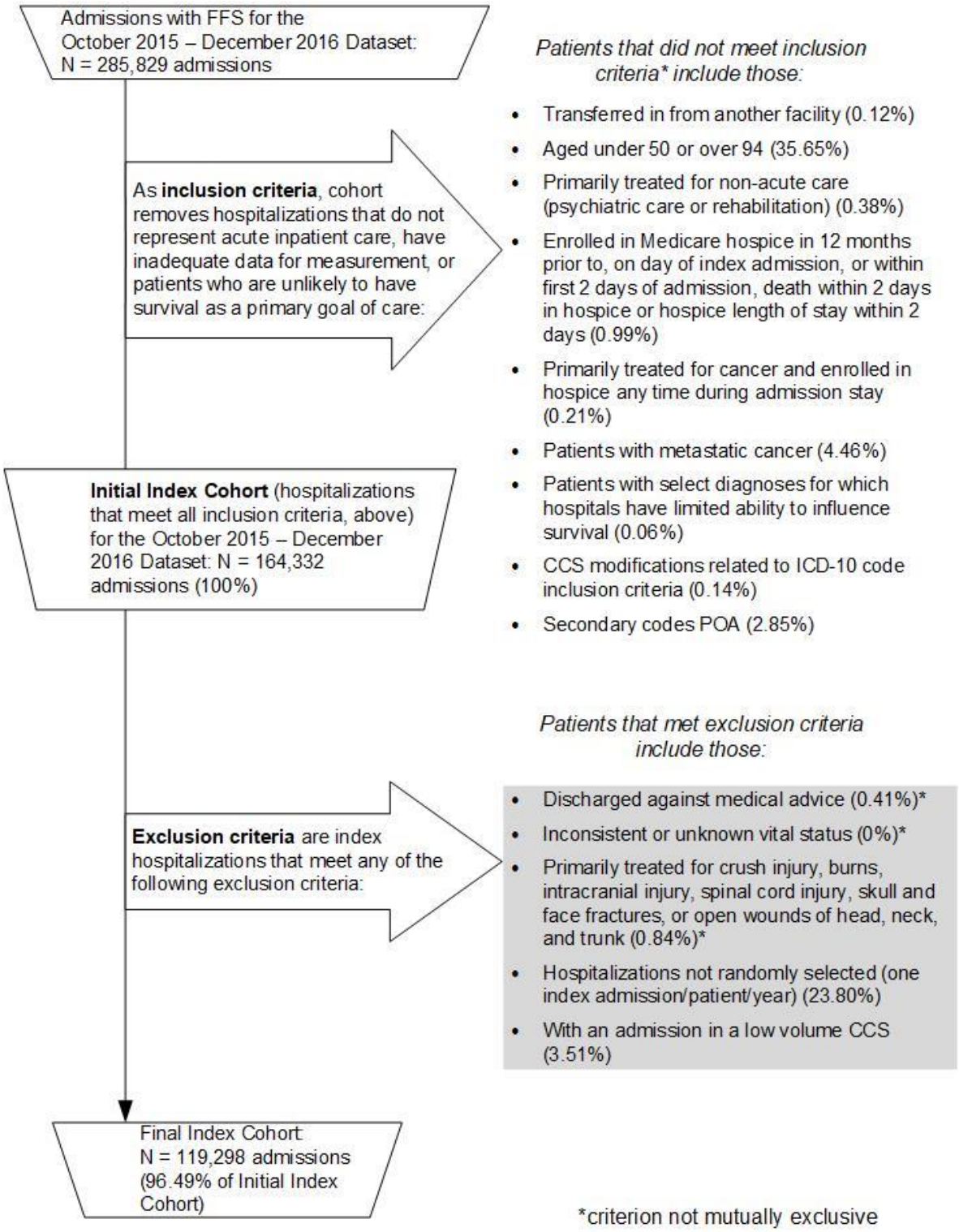
The Hybrid HWM Measure estimates hospital-specific 30-day, all-cause RSMRs using hierarchical logistic regression models. We calculated division-level measure results, using October 1, 2015 – December 30, 2016 data. The results of these analyses are presented below.

4.1. Index Cohort: Inclusion and Exclusion

The exclusion and inclusion criteria for this measure are presented in Section 3.3 Cohort and in the data dictionary tabs. A flow chart of the cohort construction, including the percentage of admissions that met each exclusion criterion in the October 2015 – December 2016 dataset is presented in Figure 3.

As shown in Figure 3, our original dataset with admissions from October 1, 2015-December 30, 2016 contained 285,829 admissions. After applying our inclusion and exclusion criteria as well as random selection for one admission per patient, our index cohort contained 119,298 admissions.

Figure 3. Cohort Inclusions and Exclusions in the October 2015 - December 2016 Dataset



4.2. Service-line Divisions

Results for each division for the patient-level logistic regression models, including the number of admissions, unadjusted 30-day mortality rate, and the c-statistic are shown in Table 6.

The c-statistic is a measure of how accurately a statistical model is able to distinguish between a patient with and without an outcome. While a higher c-statistic is desirable, we do not want to maximize it by adjusting for factors that should not be adjusted for. The range of c-statistic results is 0.82 to 0.95 which is better than results we have seen for other 30-day mortality measures.

Table 6. Index Hospitalizations, Observed Mortality Rates, and C-statistics by Division (October 2015-December 2016) (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

Service-Line Division	Index Hospitalizations (# Patients)	Observed Mortality Rate	C-Statistics
Non-Surgical: Cardiac	16,845	3.24%	0.89
Non-Surgical: Gastrointestinal	9,854	2.25%	0.91
Non-Surgical: Infectious Disease	16,190	9.80%	0.85
Non-Surgical: Neurology	5,113	4.28%	0.87
Non-Surgical: Orthopedics	3,512	1.94%	0.90
Non-Surgical: Pulmonary	7,280	6.94%	0.82
Non-Surgical: Renal	4,063	6.45%	0.87
Non-Surgical: Other	12,827	2.71%	0.89
Surgical: General	9,141	2.17%	0.95

For the service lines with insufficient data in the Clinical Hybrid Re-specification Dataset to calculate results, we present results for the Claims-only HWM measure (using the Medicare Claims Re-specification Dataset). The Hybrid HWM Measure uses the same concept, cohort, outcome and claims-based risk adjustment variables as the Claims-only HWM measure, therefore there is no conceptual reason that results from the Claims-only HWM measure would differ substantially from the Hybrid HWM Measure results. Table 7 below shows high C-statistics for the service-line divisions ranging from 0.75 to 0.91.

Table 7. Index Hospitalizations, Observed Mortality Rates, and C-statistics by Division (July 2016-June 2017) (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

Service-Line Division	Index Hospitalizations (# Patients)	Total # Hospitals	# Hospitals ≥ 25 patients	Observed Mortality Rate	C-Statistics
Non-Surgical: Cancer	**34,662**	**3,089**	**346**	**14.60%**	**0.75**
Non-Surgical: Cardiac	536,468	4,429	2,775	6.43%	0.84
Non-Surgical: Gastrointestinal	321,708	4,401	2,573	4.85%	0.83
Non-Surgical: Infectious Disease	559,106	4,510	3,187	13.01%	0.83
Non-Surgical: Neurology	229,784	4,189	1,983	7.94%	0.82
Non-Surgical: Orthopedics	133,023	4,323	1,576	4.91%	0.81
Non-Surgical: Pulmonary	483,033	4,502	3,522	9.50%	0.80
Non-Surgical: Renal	330,453	4,418	2,509	8.73%	0.77
Non-Surgical: Other	400,634	4,537	2,635	5.54%	0.81
Surgical: Cancer	**85,825**	**3,108**	**904**	**2.30%**	**0.83**
Surgical: Cardiothoracic	**132,613**	**3,134**	**1,084**	**6.37%**	**0.82**
Surgical: Neurosurgery	**30,816**	**2,042**	**382**	**3.01%**	**0.91**
Surgical: Orthopedics	**676,192**	**3,710**	**2,812**	**1.46%**	**0.90**
Surgical: Other	**156,833**	**3,597**	**1,588**	**4.06%**	**0.87**
Surgical: General	205,562	4,008	2,065	6.56%	0.87

4.3. Measure Results

Hospital-Level Overall Risk-Standardized Mortality Rates

The Hybrid HWM Measure score, or RSMR, for the 21 hospitals, ranged from a minimum of 3.98% to a maximum of 5.43%. This narrow range is due to the small sample size and lack of a nationally representative database. When tested in a full set of Medicare claims data we expect the results to be similar to those for the Claims-only HWM measure; the distribution of the RSMR for the Claims-only measure was a minimum of 3.95% to maximum of 8.70%. The mean RSMR was 6.85%. The distribution can be found in Appendix E, Figure E.1.

The final Hybrid HWM Measure hierarchical logistic regression model results can be found in data dictionary, RiskVariable ParameterEstimates.

Hospital-level Service-line Division Level SMRs

Table 8 below shows the distribution of division-level SMRs, and the between hospital variance for each of the divisions for which we were able to calculate the SMR for the Hybrid HWM Measure. Table 9 includes service-line division results for the Claims-only HWM measure using the Claims Re-specification Dataset for the other service-line divisions.

Table 8. Standardized Mortality Ratio (SMR) Distribution by Service-Line Division (October 2015-December 2016) (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

Divisions	Index Hospitalizations (# Patients)	Between Hospital variance	Mean SMR	Min SMR	Max SMR
Non-Surgical: Cardiac	16,845	0.0314	0.9997	0.8530	1.2433
Non-Surgical: Gastrointestinal	9,854	0.0321	1.0058	0.8542	1.1233
Non-Surgical: Infectious Disease	16,190	0.0296	1.0140	0.8273	1.2415
Non-Surgical: Neurology	5,113	0.1286	1.0397	0.7277	1.4785
Non-Surgical: Orthopedics	3,512	0.0085	1.0002	0.9823	1.0247
Non-Surgical: Pulmonary	7,280	0.0831	1.0373	0.7415	1.3691
Non-Surgical: Renal	4,063	0.0324	1.0028	0.8976	1.1421
Non-Surgical: Other	12,827	0.0609	1.0095	0.7676	1.3016
Surgical: General	9,141	0.1149	1.0211	0.7495	1.4238

Table 9. Standardized Mortality Ratio (SMR) Distribution by Service-Line Division (July 2016-June 2017) (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

Division	Between Hospital Variance (SE)	Mean SMR	Mean SMR SE	Min SMR	Max SMR
Non-Surgical: Cancer	**0.1601 (0.0237)**	**1.0393**	**0.1172**	**0.5144**	**1.8943**
Non-Surgical: Cardiac	0.0776 (0.0058)	1.0235	0.1134	0.6251	1.5963
Non-Surgical: Gastrointestinal	0.0869 (0.0087)	1.0316	0.1061	0.6440	1.5304
Non-Surgical: Infectious Disease	0.1031 (0.0052)	1.0287	0.1463	0.3877	1.7613
Non-Surgical: Neurology	0.0726 (0.0078)	1.0219	0.0871	0.6951	1.4718
Non-Surgical: Orthopedics	0.0874 (0.0156)	1.0320	0.0782	0.6025	1.6871
Non-Surgical: Pulmonary	0.1082 (0.0063)	1.0345	0.1696	0.5304	2.2827
Non-Surgical: Renal	0.0625 (0.0054)	1.0207	0.0937	0.5807	1.4609
Non-Surgical: Other	0.0898 (0.00710	1.0323	0.1201	0.5833	1.7853
Surgical: Cancer	**0.2399 (0.0427)**	**1.1030**	**0.1585**	**0.5502**	**2.3198**
Surgical: Cardiothoracic	**0.1181 (0.0142)**	**1.0345**	**0.1142**	**0.5415**	**1.7790**
Surgical: Neurosurgery	**0.2644 (0.0847)**	**1.1032**	**0.1313**	**0.6676**	**1.9969**
Surgical: Orthopedics	**0.0819 (0.0109)**	**1.0323**	**0.0994**	**0.6151**	**1.6125**
Surgical: Other	**0.0567 (0.0128)**	**1.0202**	**0.1201**	**0.6934**	**1.2765**
Surgical: General	0.0942 (0.0102)	1.0289	0.0975	0.6056	1.6774

Predictive Ability

Method

Discrimination in predictive ability measures the ability to distinguish high-risk subjects from low-risk subjects; therefore, for a model with good predictive ability we would expect to see a wide range in mortality rates between the lowest decile and highest decile. To calculate the predictive ability, we

randomly split the Clinical Hybrid Re-specification Dataset into two parts, one part was used as development cohort and the other half was used validation. We examined the range of observed mortality rates between the lowest and highest predicted deciles. We then observed mortality rate in deciles of the predicted mortality rate and constructed calibration plots based on the validation data.

Results

The Hybrid HWM predictive ability results demonstrate a wide range between the lowest decile and highest decile for each of the models, showing that that each model can distinguish between high and low-risk subjects (Table 10).

Table 10. Division-Level Model Discrimination (Predictive Ability) (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

Division	Predictive Ability	
	Lowest Decile (%)	Highest Decile (%)
Non-Surgical: Cardiac	0.00	20.57
Non-Surgical: Gastrointestinal	0.00	16.22
Non-Surgical: Infectious Disease	0.25	42.92
Non-Surgical: Neurology	0.16	26.04
Non-Surgical: Orthopedics	0.00	12.69
Non-Surgical: Pulmonary	0.27	28.96
Non-Surgical: Renal	0.00	33.60
Non-Surgical: Other	0.00	17.19
Surgical: General	0.00	19.12

Additionally, the risk-decile plots (Appendix F, Figures F.1-F.15) show that the predicted risk closely approximated the observed risk in most deciles. Higher observed mortality rates are associated with higher predicted mortality rates in deciles, suggesting good calibration.

Measure Score Reliability

Method

The reliability of a measurement is the degree to which repeated measurements of the same entity agree with each other. We estimated the overall measure score reliability by calculating the intra-class correlation coefficient (ICC) using a split-sample method, where hospital performance is measured once using a random subset of patients, then measured again using a second random subset exclusive of the first; the agreement between the two resulting performance measures across hospitals is then compared.[39]

Fifteen months of ICD-10 claims and EHR data (Clinical Hybrid Re-specification Dataset, October 1, 2015 – December 30, 2016) was used to calculate split-sample reliability. Admissions were randomly and evenly split into the two split samples within each individual hospital. For each sample, we fit a hierarchical generalized linear model for each service-line division and then aggregated the results into an overall RSMR. The ICC estimated was ICC [2, 1], described in Landis and Koch, and assessed using conventional standards.[40] In total, 84,825 admissions and 21 hospitals were split randomly into two datasets. All 21 hospitals met the 25-case minimum.

Results

The agreement between the two independent assessments of the RSMR for each hospital in the split samples was 0.6826, and the adjusted ICC (which estimates the ICC if we had been able to use one full year of data in each split sample, or 24 months in total), is 0.7748 (Table 11). Both results show that the measure demonstrates high reliability, according to conventional standards.[40]

Table 11. Split Sample Reliability (All Hospitals, Hybrid HWM Measure, Splitting Samples of the Clinical Hybrid Re-specification Dataset)

Statistic	Split-sample reliability (all hospitals)
Number of hospitals	21
ICC [2,1]	0.6826
Adjusted ICC [2,1]	0.7748

Empiric Validity

The external empiric validity was not directly tested in the Hybrid HWM Measure due to lack of availability of EHR data from a nationally representative set of hospitals. Instead, we report results of testing done in the Claims-only HWM measure, using the Medicare Claims-Only Measure Re-specification Dataset. Because of the homology between the two measures, there is no reason to suspect that the results of analyses done for the claims-only measure would differ in any significant way from results of analyses for a nationally representative hybrid measure. Also, the measure scores based on the claims-only model in the hybrid data are highly correlated to the measure scores based on the hybrid model (correlation coefficient = 0.96).

Methods

To test the validity of the HWM measure score, we examined whether better performance on the Claims-only HWM measure was related to better performance for other relevant structural and outcome measures. However, together with our Technical Work Group, which consists of nationally recognized experts in measure development, as well as other measurement experts, we have concluded that there is no single recognized and accepted "gold standard" measure that specifically measures factors most relevant to such a broad measure as Hospital-Wide Mortality (HWM). We did, however, identify three relevant metrics against which we could compare the measure score with the hypothesis that a trend toward correlation with these external assessments would support a conclusion of high measure score validity.

1. Nurse-to-bed ratio: Several studies have found that higher levels of nurse staffing are associated with improved patient outcomes and lower mortality rates.[41-44] We used a nurse-to-bed ratio calculated using two fields from the American Hospital Association's (AHA) annual survey. The AHA surveys all hospitals in the US and the response rate averages 85–95 percent annually,[45] from about 6,000 hospitals. Staffing is measured as the numbers of full-time and part-time registered nurses, and licensed practical nurses. Within the AHA's annual survey from 2016, we used the fields "FTEN" and "HOSPBD", which are self- reported fields that are defined in the

AHA data dictionary as: number of reported full-time registered nurse and number of hospital beds.

2. Overall Hospital Quality Star Rating mortality group score: CMS's Overall Hospital Quality Star Rating assesses hospitals' overall performance (expressed on *Hospital Compare* graphically, as stars) based on a weighted average of group scores from different domains of quality (mortality, readmissions, safety, patient experience, imaging, effectiveness of care, timeliness of care). The mortality group is comprised of mortality measures that are publicly reported on *Hospital Compare*. The mortality group score is derived from a latent-variable model that identifies an underlying quality trait for that group. We used mortality group scores from 4,106 Medicare FFS hospitals from July 2018. The full methodology for the Overall Hospital Quality Star Rating can be found at: https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228775957165.
3. Overall Hospital Quality Star Rating: CMS's Overall Hospital Quality Star Rating assesses hospitals' overall performance (expressed on *Hospital Compare* graphically, as stars) based on a weighted average of "group scores" from different domains of quality (mortality, readmissions, safety, patient experience, imaging, effectiveness of care, timeliness of care). Each group has within it, measures that are reported on *Hospital Compare*. Group scores for each individual group are derived from latent-variable models that identify an underlying quality trait for each group. Group scores are combined into an overall hospital score using fixed weights; overall hospital scores are then clustered, using k-means clustering, into five groups and are assigned one-to-five stars (the hospital's star rating). We used hospital's star ratings from 3,715 Medicare FFS hospitals from July 2018. The full methodology for the Overall Hospital Quality Star Rating can be found at https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier3&cid=1228775957165.

We examined the relationship of performance on the claims only measure scores (RSMR) with each of the three external measures of hospital quality. For the external measures, the comparison was against performance within quartiles for nurse-to-bed ratio and mortality group score, or in the case of Star Ratings, to the Star Rating category (one to five Stars).

We also compared performance on these external measures with categories of performance on the HWM measure by determining "outliers" of performance for the RSMR. Specifically, we identified outliers by estimating an interval estimate (similar to a confidence interval) around each hospital score and identified those facilities that had a 95% interval estimate entirely above or entirely below the national average. We then assigned scores to one of three performance categories: 1) "no different than national average," 2) "better than the national average," or 3) "worse than the national average." – with 95% confidence. Hospitals categorized as outliers ("better" or "worse" than national average) on the HWM measure were identified within the quartiles of performance on the comparator measure.

Results

For each external measure of quality, the comparison showed a trend toward better performance on the HWM measure with better performance on the comparator measure (Figure 4, Figure 5, and Figure 6). For example, in Figure 4, when comparing the Claims-only HWM measure scores to the nurse-to-bed-

ratio, as the number of nurses per bed increases (more nurses in the hospital) across quartiles of nurse-to-bed ratio (from left to right on the graph), the median overall HWM mortality rate is lower (better). Likewise, in Figure 5, better performance on the HWM measure is associated with better Overall Hospital Quality Star Rating mortality group scores across quartiles of mortality group score performance. Finally, in Figure 6, we show that HWM performance improves across the star rating category in the expected direction: HWM scores are better (lower) as the star rating category improves (increases from one to five stars).

Note that the Overall Hospital Quality Star Rating includes quality measures that are much broader than the HWM Measure, such as patient experience. This is consistent with the stronger relationship between HWM and the Overall Hospital Quality Star Rating mortality measure group score (Figure 5), compared to the Overall Hospital Quality Star Rating (Figure 6).

Within the graphs in Figure 4, Figure 5, and Figure 6, we also overlay "better" and "worse" outliers (95% confidence interval, as described in Methods above) on the HWM measure, with performance on the external measure. The overlay results are consistent with the trend toward better performance on the HWM measure with better performance on the quality measure; there are more high outliers (shown in total as "better" at the bottom of the graph, and as blue squares in the graph) with higher performance for each comparator measure (moving left to right on the graphs below); there are also fewer "worse" outliers (shown in total as "worse" at the bottom of the graph, and as red triangles in the graph). The inverse is also observed: fewer "better" outliers and more "worse" outliers are present in quartiles of worse performance on the comparator measure. For example, in Figure 4 below, which compares RSMR to nurse-to-bed ratio, there are 19 HWM "better than national average" outliers in the third quartile (and five "worse" outliers), and 49 HWM "better" outliers in the fourth quartile (and zero "worse" outliers). In addition, in Figure 5, which compares RSMR to mortality group score, there are 96 HWM "better than national average" outliers (and zero "worse" outlies) in the fourth (best performing) quartile, and 13 "worse than national average" (and zero "better") outliers in the first (worst performing) quartile. A similar relationship can be seen in Figure 6, in comparison to the Overall Hospital Quality Star Rating.

Figure 4. Claims-only HWM RSMR: Relationship to Nurse-to-Bed Ratio

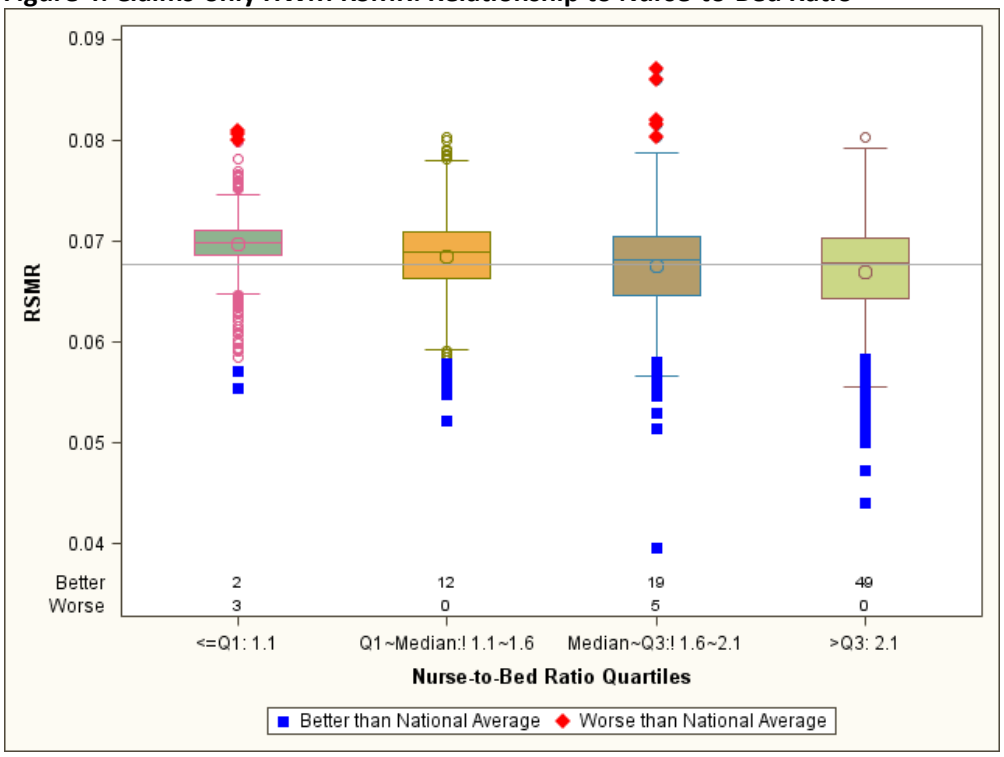


Figure 5. Claims-only HWM RSMR: Relationship to Star Ratings Mortality Group Score

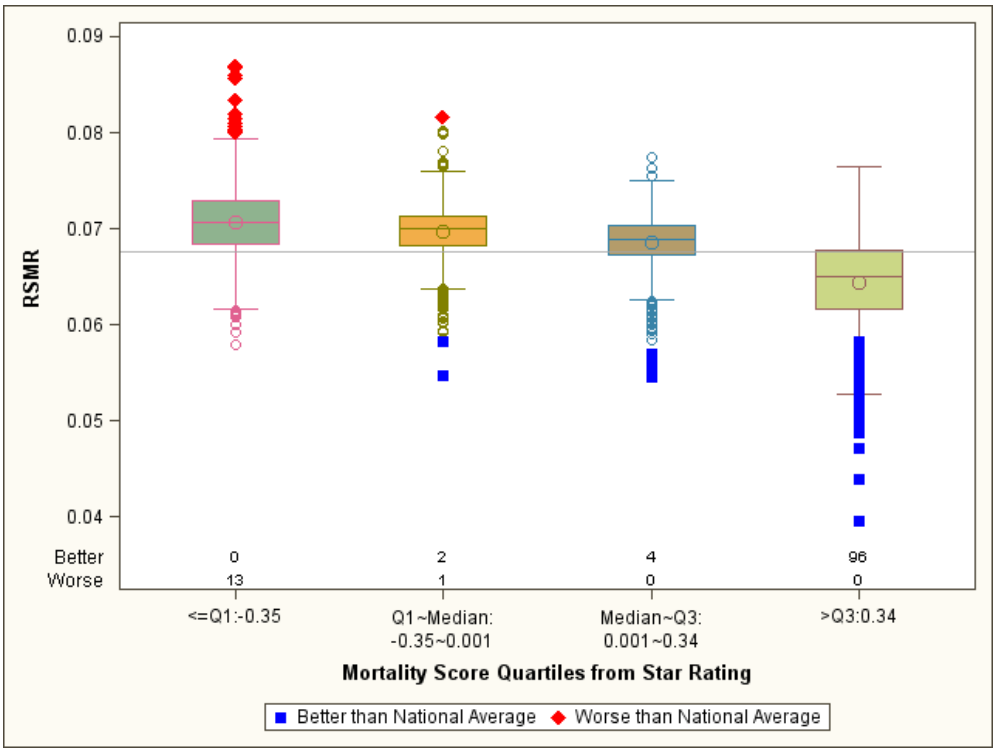
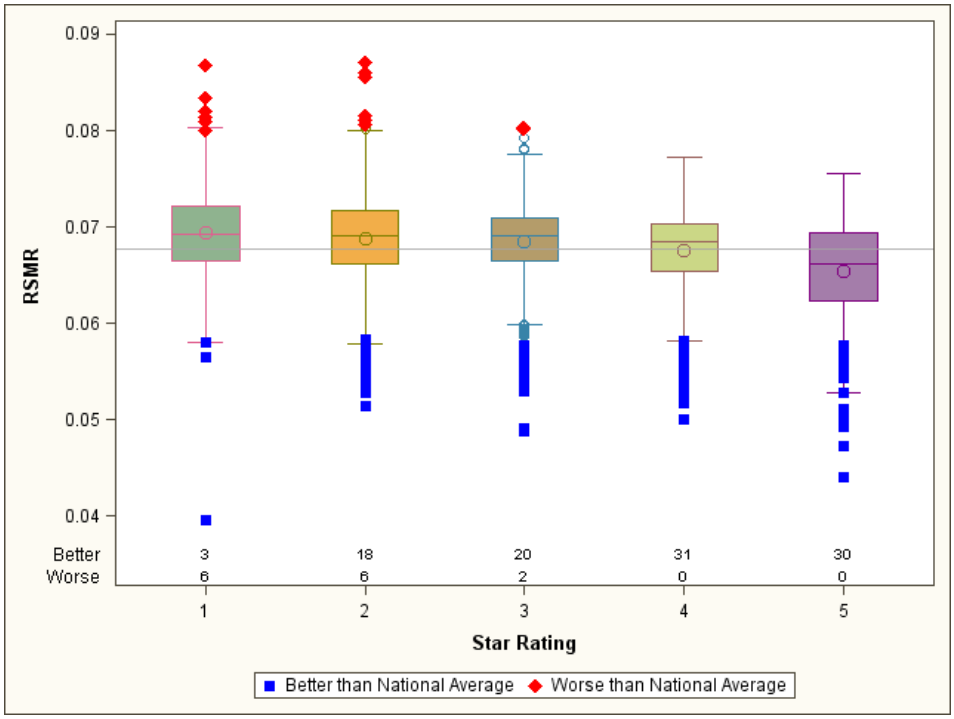


Figure 6. Claims-only RSMR: Relationship to Overall Hospital Star Rating



There is no single analysis that is sufficient to validate the measure because there is no gold standard that exists for the validation of a hospital-wide quality measure. With this limitation in mind, the three empiric external analyses support measure score validity based on trends in correlation with different hospital quality metrics.

Face Validity

Methods

We systematically assessed the face validity of the Hybrid HWM Measure score as an indicator of quality by confidentially soliciting the TEP members' agreement with the following statement via an online survey following the final TEP meeting: "The risk-standardized mortality rates obtained from the Hybrid Hospital-Wide Mortality Measure as specified can be used to distinguish between better and worse quality facilities." The survey offered participants response options on a six-point scale (1=Strongly Disagree, 2=Moderately Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5= Moderately Agree, and 6=Strongly Agree).

Results

A total of six of the eight TEP members completed the face validity survey. Of the six respondents, five respondents (83%) indicated that they somewhat, moderately, or strongly agreed and one somewhat disagreed with the following statement: "The risk-standardized mortality rates obtained from the Hybrid Hospital-Wide Mortality Measure as specified can be used to distinguish between better and worse quality facilities." Survey results from the TEP indicated high agreement (83%) regarding the face validity of the Hybrid HWM measure.

5. SUMMARY

This report summarizes the development, specifications, and testing of a hospital-level all-cause hospital-wide 30-day mortality measure based on administrative claims data enhanced with clinical data elements from the EHR for risk adjustment. We used a standard, accepted, and transparent approach to develop the measure, and relied on close input from patients and clinicians throughout the development process.

The Hybrid HWM Measure offers several important benefits. First, it allows CMS, the public, patients, and providers to monitor an important, patient-centered outcome. Second, it provides CMS with a tool for broad performance assessment across a wide span of hospitals, complementing their existing Claims-only and Hybrid HWR Measures. It does so with minimal burden to hospitals and no burden to patients by leveraging reliably captured and valid clinical data elements that have been shown to be feasibly extracted from the EHR without changes to standard clinical workflow, which improves the measure's risk models. Finally, measure score results with a full set of Medicare claims show a large range of hospital performance, and the measure can provide more granular division-level performance information that is of interest to both patients and clinicians.

However, one limitation of the testing results for the Hybrid HWM Measure is that due to the lack of availability of a nationally representative dataset containing the needed EHR data elements, the current report does not include testing results for the following clinical divisions: Non-surgical Cancer, and Surgical Cardiothoracic, Cancer, Neurology, Orthopedics, and Other. Instead, we provide results from the Claims-only HWM measure run with a one-year of Medicare claims. The claims only HWM measure uses the same concept, cohort, outcome and claims-based risk adjustment variables as the Hybrid HWM Measure; there is no conceptual reason that the results from the Claims-only HWM measure would be substantially dissimilar to results from the Hybrid HWM Measure.

Measuring hospital-wide mortality is challenging. Earlier attempts did not exclude patients for whom mortality is likely not a quality signal nor did they have the benefit of close patient and clinician engagement in measure design. Throughout our discussions with stakeholders, including our TEP, we heard support for the concept of measuring hospital-wide mortality and a strong desire for a measure that offers patients and providers meaningful, detailed, and statistically valid performance data. This measure was developed with expert and patient input throughout the process aimed at addressing previous concerns, and is technically sound. The measure also offers the extra benefit of clinical risk variables without adding to the overall measure burden.

REFFERENCES

1. James JT. A new, evidence-based estimate of patient harms associated with hospital care. *Journal of patient safety.* 2013;9(3):122-128.
2. Suter LG, Li SX, Grady JN, et al. National patterns of risk-standardized mortality and readmission after hospitalization for acute myocardial infarction, heart failure, and pneumonia: update on publicly reported outcomes measures based on the 2013 release. *Journal of general internal medicine.* 2014;29(10):1333-1340.
3. Peterson ED, Roe MT, Mulgund J, et al. Association between hospital process performance and outcomes among patients with acute coronary syndromes. *Jama.* 2006;295(16):1912-1920.
4. Writing Group for the Checklist- I.C.U. Investigators, Brazilian Research in Intensive Care Network. Effect of a quality improvement intervention with daily round checklists, goal setting, and clinician prompting on mortality of critically ill patients: A randomized clinical trial. *JAMA.* 2016;315(14):1480-1490.
5. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE). Condition-Specific Measures Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Mortality Mesures. 2016; https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed October 28, 2016.
6. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE). Procedure-Specific Measure Updates and Specificatiosn Report Hospital-Level 30-Day Risk-Standardized Mortality Measure. 2016; https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830. Accessed October 28, 2016.
7. Classen DC, Resar R, Griffin F, et al. 'Global trigger tool'shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs.* 2011;30(4):581-589.
8. Andel C, Davidow SL, Hollander M, Moreno DA. The economics of health care quality and medical errors. *Journal of Health Care Finance.* 2012;39(1):39-50.
9. Institute of Medicine. To Err is Human: Building a Safer Health System. 1999; https://iom.nationalacademies.org/~/media/Files/Report%20Files/1999/To-Err-is-Human/To%20Err%20is%20Human%201999%20%20report%20brief.pdf. Accessed January 20, 2016.
10. Berwick DM, Calkins DR, McCannon CJ, Hackbarth AD. The 100,000 lives campaign: setting a goal and a deadline for improving health care quality. *Jama.* 2006;295(3):324-327.
11. Tablan O, Anderson L, Besser R, Bridges C, Hajjeh R. CDC; Healthcare Infection Control Practices Advisory Committee. Guidelines for preventing health-care-associated pneumonia, 2003: recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee. *MMWR Recomm Rep.* 2004;53(RR-3):1-36.
12. Resar R, Pronovost P, Haraden C, Simmonds T, Rainey T, Nolan T. Using a bundle approach to improve ventilator care processes and reduce ventilator-associated pneumonia. *Joint Commission Journal on Quality and Patient Safety.* 2005;31(5):243-248.
13. Kalil AC, Metersky ML, Klompas M, et al. Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clinical Infectious Diseases.* 2016;63(5):e61-e111.
14. O'Gara PT, Kushner FG, Ascheim DD, et al. 2013 ACCF/AHA Guideline for the Management of ST-Elevation Myocardial Infarction. *A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines.* 2013;61(4):e78-e140.

15. Centers for Medicare & Medicaid Services. Hospital Quality Initiative Overview. 2005; https://www.hsdl.org/?view&did=464393. Accessed February 28, 2020.
16. Joint Commission. 2005 Joint Commission National Patient Safety Goals: Practical Strategies and Helpful Solutions for Meeting These Goals. 2005; https://www.jointcommission.org/en/standards/national-patient-safety-goals/hospital-2020-national-patient-safety-goals/. Accessed January 20, 2016.
17. O'grady NP, Alexander M, Burns LA, et al. Guidelines for the prevention of intravascular catheter-related infections. *Clinical infectious diseases.* 2011;52(9):e162-e193.
18. Berríos-Torres SI, Umscheid CA, Bratzler DW, et al. Centers for Disease Control and Prevention Guideline for the Prevention of Surgical Site Infection, 2017. *JAMA Surgery.* 2017;152(8):784-791.
19. Rosenberger LH, Politano AD, Sawyer RG. The surgical care improvement project and prevention of post-operative infection, including surgical site infection. *Surgical infections.* 2011;12(3):163-168.
20. Whittington J, Simmonds T, Jacobsen D. *Reducing hospital mortality rates.* Institute for Healthcare Improvement; 2005.
21. Emmi Poteliakhoff. Update on IHI's 100k Lives Campaign. October 2006; https://pdfs.semanticscholar.org/2e08/e791708bd1745caf2e4f8b3ce72baa1c51b8.pdf. Accessed February 28, 2020.
22. Curry LA, Linnander EL, Brewster AL, Ting H, Krumholz HM, Bradley EH. Organizational culture change in U.S. hospitals: a mixed methods longitudinal intervention study. *Implementation science : IS.* 2015;10:29.
23. Shahian DM, Iezzoni LI, Meyer GS, Kirle L, Normand SL. Hospital-wide mortality as a quality metric: conceptual and methodological challenges. *Am J Med Qual.* 2012;27(2):112-123.
24. Health and Social Care Information Centre. Summary Hospital-Level Mortality Indicator. 2015; https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/448742/SHMI_specification.pdf. Accessed February 28, 2020.
25. Green J, Wintfeld N, Sharkey P, Passman LJ. The importance of severity of illness in assessing hospital mortality. *Jama.* 1990;263(2):241-246.
26. Blumberg MS. Comments on HCFA hospital death rate statistical outliers. Health Care Financing Administration. *Health services research.* 1987;21(6):715-739.
27. Hawkes N. Patient coding and the ratings game. *BMJ.* 2010;340.
28. Wen E, Sandoval C, Zelmer J, Webster G. Understanding and using the hospital standardized mortality ratio in Canada: challenges and opportunities. *Healthc Pap.* 2008;8(4):26-36; discussion 69-75.
29. Brien SE, Ghali WA. CIHI's hospital standardized mortality ratio: friend or foe. *Healthc Pap.* 2008;8(4):57-61.
30. Jarman B, Gault S, Alves B, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ.* 1999;318(7197):1515-1520.
31. Black N. Assessing the quality of hospitals. *BMJ.* 2010;340.
32. Lilford R, Pronovost P. Using hospital mortality rates to judge hospital performance: a bad idea that just won't go away. *BMJ.* 2010;340.
33. Centers for Medicare & Medicaid Services. Measures Management System. 2016; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo. Accessed October 28, 2016.
34. 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.
35. Pope GC, Ellis RP, Ash AS, et al. Principal inpatient diagnostic cost group model for Medicare risk adjustment. 2001.
36. (YNHHSC/CORE) YNHHSCCfORaE. 2013 Core Clinical Data Elements Technical Report (Version 1.1). 2015; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed November 10, 2017.
37. (YNHHSC/CORE) YNHHSCCfORaE. Hybrid Hospital-Wide Readmission Measures with Electronic Health Record Extracted Risk Factors (Version 1.1). 2015; https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/Measure-Methodology.html. Accessed November 10, 2017.
38. Klein LR. Textbook of econometrics. 1953.
39. Rousson V, Gasser T, Seifert B. Assessing intrarater, interrater and test–retest reliability of continuous measurements. *Statistics in medicine.* 2002;21(22):3431-3446.
40. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *biometrics.* 1977:159-174.
41. Aiken LH, Clarke SP, Sloane DM, Sochalski J, Silber JH. Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Jama.* 2002;288(16):1987-1993.
42. Griffiths P, Ball J, Murrells T, Jones S, Rafferty AM. Registered nurse, healthcare support worker, medical staffing levels and mortality in English hospital trusts: a cross-sectional study. *BMJ open.* 2016;6(2):e008751.
43. Needleman J, Buerhaus P, Mattke S, Stewart M, Zelevinsky K. Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine.* 2002;346(22):1715-1722.
44. Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M. Nurse staffing and inpatient hospital mortality. *New England Journal of Medicine.* 2011;364(11):1037-1045.
45. (AHA) AHA. *The AHA Annual Survey Database Fiscal Year 1997 Documentation.* Chicago: Health Forum; 1999.

GLOSSARY

C-statistic: An indicator of the model's discriminant ability or ability to correctly classify those who have and have not died within 30 days of the start of the admission. Potential values range from 0.5, meaning no better than chance, to 1.0, an indication of perfect prediction. Perfect prediction implies that patients' outcomes can be predicted completely by their risk factors, and physicians and hospitals play no role in their patients' outcomes.

Case mix: The particular illness severity and age characteristics of patients with index admissions at a given hospital.

Cohort: The index admissions used to calculate the measure after inclusion and exclusion criteria have been applied.

Comorbidities: Medical conditions the patient had in addition to his/her primary reason for admission to the hospital.

Complications: Medical conditions that may have occurred as a consequence of care rendered during hospitalization.

Condition categories (CMS-CCs): Groupings of ICD-9-CM or ICD-10-CM diagnosis codes in clinically relevant categories, from the Hierarchical Condition Categories (HCCs) system. CMS uses the grouping but not the hierarchical logic of the system to create risk factor variables. Description of the Condition Categories can be found at http://www.cms.hhs.gov/Reports/downloads/pope_2000_2.pdf.

Confidence interval (CI): A CI is a range of probable values for an estimate that characterizes the amount of associated uncertainty. For example, the 95% CI for the ORs associated with risk-adjustment variables in the model indicates there is 95% confidence that the odds ratio (OR) lies between the lower and the upper limit of the interval. The 95% CI serves as a proxy for statistical significance for ORs; if the CI does not contain the value of 1.0, the association is considered significant.

Core clinical data elements (CCDE): A standardized set of clinical data that are consistently obtained on adult hospital inpatients that could be feasibly extracted from electronic health records, to be used in risk adjustment for hospital quality outcome measures.

Discharge condition category: A group of related discharge diagnosis ICD-9 or ICD-10 codes (principal diagnoses), as grouped by the Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS).

Electronic health record (EHR): A record in digital format that allows for systematic collection of electronic health information about individual patients or populations. It theoretically allows for sharing information across different healthcare settings.

Electronic health record data: Data derived specifically from the hospital EHR. In this report, in most cases we are referring to the clinical data on patients, which are the core clinical data elements (CCDE).

Electronic specification: Refers to measure specifications derived from EHRs and contain four main components, which are contained within the Measure Authoring Tool (MAT) Output: measure overview/description, measure logic, measure code lists, and quality datasets elements.

Expected mortality*:* The number of deaths expected based on average hospital performance with a given hospital's case mix and service mix.

First captured values: The first value for a data element recorded in the EHR after a patient arrives at the facility for care. Identification of the first value requires a time and date stamp for the first interaction a patient has with facility staff which results in a time or date stamp being entered in the Patient Management System. This is most often the time and date of registration when basic demographic and insurance information are provided and confirmed by non-clinical staff. An arrival location is also required because patients can arrive in various locations including the emergency department, pre-operative area, or to an inpatient unit or floor. The time and date stamps associated with the specific data elements are then compared against the time of arrival to identify the first captured value.

Hierarchical model: A widely accepted statistical method that enables fair evaluation of relative hospital performance by accounting for patient risk factors as well as the number of patients a hospital treats. This statistical model accounts for the structure of the data (patients clustered within hospitals) and calculates (1) how much variation in hospital mortality rates overall is accounted for by patients' individual risk factors (such as age and other medical conditions); and (2) how much variation is accounted for by hospital contribution to mortality risk.

Hybrid measure: A measure that uses two separate data sources. Specifically, the Hybrid HWM measure uses Medicare claims data to derive the cohort, outcome, and comorbidities, and EHR-derived data to add patient-level clinical data into the risk adjustment. This is in comparison to only using Medicare claims as a single source of data for measure development and implementation.

Index admission: Any admission included in the measure calculation as the initial admission for an episode of care to which the outcome is attributed.

Medicare fee-for-service (FFS): Original Medicare plan in which providers receive a fee or payment for each individual service provided directly from Medicare. Only beneficiaries in Medicare FFS, not in managed care (Medicare Advantage), are included in this measure.

National observed mortality rate: All included hospitalizations with the outcome divided by all included hospitalizations.

Odds ratio (OR): The ORs express the relative odds of the outcome for each of the predictor variables. For example, the OR for Protein-calorie malnutrition (CC 21) represents the odds of the outcome for patients with that risk variable present relative to those without the risk variable present. The model coefficient for each risk variable is the log (odds) for that variable.

Outcome: The result of a broad set of healthcare activities that affect patients' well-being. For this measure, the outcome is mortality within 30 days of admission.

Predicted mortality*:* The number of deaths within 30 days, predicted based on the hospital's performance with its observed case mix and service mix.

Risk-adjustment variables: Patient demographics and comorbidities used to adjust for differences in case mix and service mix across hospitals.

Risk-standardized mortality rate (RSMR): The risk-standardized mortality rate is the standardized mortality ratio (SMR) (see definition below), multiplied by the national observed mortality rate.

Service-line divisions: A group of index admissions for patients with related conditions or procedures categories that are likely treated by similar care teams. There were 15 defined cohorts in this report. Each service-line division has its own risk model. The service-line divisions are, Non-Surgical: Cancer, Cardiac, Gastrointestinal, Infectious Disease, Neurology, Orthopedics, Pulmonary, Renal, Other; Surgical: Cancer, Cardiothoracic, General, Neurosurgery, Orthopedics, Other.

Service mix: The particular conditions and procedures of the patients with index admissions at a given hospital.

Standardized mortality ratio (SMR): For each hospital, the numerator of the ratio is the number of deaths predicted for the hospital's patients, accounting for its observed mortality rate, the number of patients, and the hospital's case- and service-line mix. The denominator is the number of deaths expected nationally for that hospital's case/service-line mix. A ratio greater than one indicates that more patients died at that hospital than expected, compared to an average hospital with similar case/service-line mix. A ratio less than one indicates that the hospital's patients have fewer deaths than expected, compared to an average hospital with a similar case/service-line mix.

APPENDIX A. ACKOWLEDGEMENT DETAILS

We would like to thank the members of the Technical Expert Panel (TEP). The TEP members provided important insight and feedback on key measure decisions for the development and re-specification of the Hybrid Hospital-Wide Mortality (HWM) Measure.

Disclaimer: The views, thoughts, and opinions expressed in this report belong solely to the author and do not represent endorsement by any entity or individual, including the Technical Work Group and TEP members and the organizations those members are affiliated with, as well as other contributors and consultants. Acknowledgment of input does not imply endorsement of the methodology and policy decisions.

TEP Members:

Jonathan Bae, MD (2016-2018) – Associate Chief Medical Officer for Patient Safety and Clinical Quality, Duke University Health System, Durham, NC

Michelle Beck (2018-2019)– Patient/Family Caregiver Representative, University of Maryland Upper Chesapeake Medical Center, Bel Air, MD

Jeanne Black, PhD, MBA – Director of Health Services Research, Department of Orthopedics, Cedars-Sinai Health System, Los Angeles, CA

John Bott, MBA, MSSW– Independent Consultant

Roger Dmochowski, MD, MMHC, FACS – Associate Surgeon-in-Chief, Vanderbilt University Medical Center, Nashville, TN

Richard Dutton, MD, MBA – Chief Quality Officer, United States Anesthesia Partners, Dallas, TX

Chris Ghaemmaghami, MD (2016-2018) – Chief Medical Officer and Senior Associate Dean for Clinical Affairs, University of Virginia Health System and University of Virginia School of Medicine, Charlottesville, VA

Gaye Hyre – Patient/Family Caregiver Representative, West Haven, CT

Irene Katzan, MD, MS –Neurologist, Cleveland Clinic, Cleveland, OH

Amy Kelley, MD, MSHS – Associate Professor and Staff Physician of Geriatrics and Palliative Medicine, Icahn School of Medicine at Mt Sinai, New York, NY

Brenda Matti-Orozco, MD, FACP – Chief of Division of General Internal Medicine and Palliative Medicine, Morristown Medical Center, Morristown, NJ

Colleen O'Leary, MSN, RN, AOCNS (2016-2018) – Associate Director Nursing Education and Evidence-based Practice, and Director at Large, The Ohio State University James Cancer Hospital and Solove Research Institute, Westerville, OH

Jyotirmay Sharma, MD, FACS, FACE – Chief Quality Officer, Emory University School of Medicine, Atlanta, GA

Fredda Valdeck, LSMW (2016-2018) – Patient Advocate, New York, NY

We would also like to give thanks to the members of our Technical Work Group who generously gave their time to provide guidance on key clinical and statistical decisions.

Technical Work Group Members:

Dr. Lee Fleisher, MD – Chair, Department of Anesthesiology and Critical Care, University of Pennsylvania Health System; and Vice-Chair of the Consensus Standards Advisory Committee (CSAC) and co-chair of the Surgery Standing Committee of the NQF.

Dr. Cary P. Gross, MD – Yale Professor of Medicine (General Medicine), in the Institute for Social and Policy Studies and of Epidemiology (Chronic Diseases); and Director and founder of Yale's Cancer Outcomes Public Policy and Effectiveness Research (COPPER). Dr. Gross served from March 2016 to July 2016.

Dr. Leora Horwitz, MD, MHS – Associate Professor in the Departments of Population Health and Medicine at New York University School of Medicine; founding director of the Center for Healthcare Innovation and Delivery Science, New York University Langone Medical Center; and of the Division of Healthcare Delivery Science, Department of Population Health, New York University School of Medicine.

Mr. Kristopher Huffman, MS – Senior Statistician, American College of Surgeons.

Dr. David M. Shahian, MD – Professor of Surgery at Harvard Medical School; Vice President of the Massachusetts General Hospital (MGH) Center for Quality and Safety; and Associate Director of the MGH Codman Center for Clinical Effectiveness in Surgery; Vice Chair of the NQF Health Professionals Council; Chair of The Society of Thoracic Surgeons (STS) Workforce on National Databases and its Quality Measurement Task Force

Appendix B. COMOBIDITY COMPARISON: CLAIMS vs CLINICAL HYBRID DATASET

Table B.1. below compares patients 65 years and older in both the [Medicare] Claims-Only Development Dataset and the Clinical Hybrid Development Dataset. The mean age and standard deviation of the population is very similar. Comorbidity burden is relatively similar across the two datasets, although some specific diagnoses are more common in the Claims-Only Development Dataset (such as Disorders of Fluid/Electrolyte/Acid-Base Balance and Congestive Heart Failure), while others (such as Disorders of lipid Metabolism and Septicemia) are more common in the Clinical Hybrid Development Dataset.

Table B.1. Risk Variable Frequencies Comparing Medicare Claims-Only Development Dataset and Clinical Hybrid Development Dataset

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Age (Mean/SD)	77.87	7.90	77.45	7.93
Other Infectious Diseases (CC 7)	539171	13.93	12458	5.35
Metastatic & Severe Cancers (CC 8,9)	103144	2.66	6259	2.69
Protein-Calorie Malnutrition (CC 21)	296449	7.66	24746	10.62
Disorders of Fluid/Electrolyte/Acid-Base Balance (CC 24)	1388492	35.87	23226	9.97
Disorders of Lipoid Metabolism (CC 25)	2117182	54.70	158747	68.13
Liver Failure (CC 27,30)	51192	1.32	2540	1.09
Other GI Disorders (CC 34-38)	1822504	47.09	142036	60.96
Other Musculoskeletal and Connective Tissue Disorders (CC 44,45)	1333561	34.45	107530	46.15
Hematologic or Immunity Disorders (CC 46-48)	355945	9.20	11400	4.89
Dementia and Other Nonpsychotic Organic Brain Syndromes (CC 51-53)	686894	17.75	39206	16.83
Coma/Brain Compression/Anoxic Injury and Severe Head Injury (CC 80,166)	42028	1.09	908	0.39
Respiratory Failure, Respirator Dependence, Shock (CC 82-84)	524093	13.54	15202	6.52
Congestive Heart Failure (CC 85)	1112605	28.75	24276	10.42
Hypertension and hypertensive heart disease (CC 94,95)	2448768	63.27	145258	62.34
Pneumonia (CC 114-116)	593118	15.32	44056	18.91
Dialysis or Severe Chronic Kidney Disease (CC 134,136,137)	223271	5.77	12231	5.25
Acute or Unspecified Renal Failure (CC 135,140)	710072	18.35	11172	4.79
Poisonings and Allergic and Inflammatory Reactions (CC 175)	200537	5.18	9937	4.26
Minor Symptoms, Signs, Findings (CC 179)	1626182	42.01	152483	65.44
Principal Discharge Diagnosis CCS				
Tuberculosis (CCS 1)	286	0.01	32	0.01

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Septicemia (except in labor) (CCS 2)	296918	7.67	34017	14.60
Bacterial infection; unspecified site (CCS 3)	588	0.02	52	0.02
Mycoses (CCS 4)	3158	0.08	104	0.04
HIV infection (CCS 5)	407	0.01	32	0.01
Hepatitis (CCS 6)	1764	0.05	169	0.07
Viral infection (CCS 7)	6800	0.18	244	0.10
Other infections; including parasitic (CCS 8)	1726	0.04	--	--
Sexually transmitted infections (not HIV or hepatitis) (CCS 9)	193	<0.00	--	--
Cancer of head and neck (CCS 11)	3552	0.09	316	0.14
Cancer of esophagus (CCS 12)	1410	0.04	132	0.06
Cancer of stomach (CCS 13)	2626	0.07	204	0.09
Cancer of colon (CCS 14)	16978	0.44	1300	0.56
Cancer of rectum and anus (CCS 15)	4942	0.13	332	0.14
Cancer of liver and intrahepatic bile duct (CCS 16)	2438	0.06	363	0.16
Cancer of pancreas (CCS 17)	3630	0.09	306	0.13
Cancer of other GI organs; peritoneum (CCS 18)	2400	0.06	185	0.08
Cancer of bronchus; lung (CCS 19)	18505	0.48	1234	0.53
Cancer; other respiratory and intrathoracic (CCS 20)	326	0.01	31	0.01
Cancer of bone and connective tissue (CCS 21)	1589	0.04	131	0.06
Melanomas of skin (CCS 22)	330	0.01	--	--
Other non-epithelial cancer of skin (CCS 23)	1214	0.03	--	--
Cancer of breast (CCS 24)	5497	0.14	1954	0.84
Cancer of uterus (CCS 25)	4481	0.12	649	0.28
Cancer of cervix (CCS 26)	411	0.01	--	--
Cancer of ovary (CCS 27)	1698	0.04	--	--
Cancer of other female genital organs (CCS 28)	925	0.02	--	--
Cancer of prostate (CCS 29)	12301	0.32	1871	0.80
Cancer of other male genital organs (CCS 31)	110	<0.00	--	--
Cancer of bladder (CCS 32)	6266	0.16	898	0.39
Cancer of kidney and renal pelvis (CCS 33)	8416	0.22	602	0.26
Cancer of other urinary organs (CCS 34)	974	0.03	--	--
Cancer of brain and nervous system (CCS 35)	3605	0.09	259	0.11

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Cancer of thyroid (CCS 36)	1042	0.03		
Non-Hodgkin`s lymphoma (CCS 38)	4873	0.13	354	0.15
Leukemias (CCS 39)	4078	0.11	181	0.08
Multiple myeloma (CCS 40)	2646	0.07	91	0.04
Cancer; other and unspecified primary (CCS 41)	656	0.02	75	0.03
Malignant neoplasm without specification of site (CCS 43)	995	0.03	109	0.05
Neoplasms of unspecified nature or uncertain behavior (CCS 44)	6918	0.18	389	0.17
Maintenance chemotherapy; radiotherapy (CCS 45)	4511	0.12	188	0.08
Benign neoplasm of uterus (CCS 46)	197	0.01	--	--
Other and unspecified benign neoplasm (CCS 47)	12349	0.32	1677	0.72
Diabetes mellitus with complications (CCS 50)	9035	0.23	626	0.27
Gout and other crystal arthropathies (CCS 54)	167	<0.00	--	--
Fluid and electrolyte disorders (CCS 55)	72337	1.87	3259	1.40
Deficiency and other anemia (CCS 59)	211	0.01	--	--
Coagulation and hemorrhagic disorders (CCS 62)	165	<0.00	15	0.01
Other hematologic conditions (CCS 64)	135	<0.00	--	--
Meningitis (except that caused by tuberculosis or sexually transmitted disease) (CCS 76)	1270	0.03	66	0.03
Encephalitis (except that caused by tuberculosis or sexually transmitted disease) (CCS 77)	1120	0.03	61	0.03
Other CNS infection and poliomyelitis (CCS 78)	707	0.02	56	0.02
Parkinson`s disease (CCS 79)	4006	0.10	92	0.04
Multiple sclerosis (CCS 80)	834	0.02	41	0.02
Other hereditary and degenerative nervous system conditions (CCS 81)	8496	0.22	211	0.09
Paralysis (CCS 82)	552	0.01	--	--
Epilepsy; convulsions (CCS 83)	21684	0.56	1014	0.44
Coma; stupor; and brain damage (CCS 85)	2066	0.05	620	0.27
Heart valve disorders (CCS 96)	36247	0.94	1714	0.74
Peri-; endo-; and myocarditis; cardiomyopathy (except that caused by tuberculosis or sexually transmitted disease) (CCS 97)	7576	0.20	503	0.22
Essential hypertension (CCS 98)	8910	0.23	213	0.09

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Hypertension with complications and secondary hypertension (CCS 99)	42013	1.09	1516	0.65
Acute myocardial infarction (CCS 100)	131410	3.40	7036	3.02
Coronary atherosclerosis and other heart disease (CCS 101)	88744	2.29	6162	2.64
Nonspecific chest pain (CCS 102)	41726	1.08	3923	1.68
Pulmonary heart disease (CCS 103)	34976	0.90	2057	0.88
Conduction disorders (CCS 105)	18702	0.48	1838	0.79
Cardiac dysrhythmias (CCS 106)	183804	4.75	7764	3.33
Congestive heart failure; non-hypertensive (CCS 108)	203230	5.25	11071	4.75
Occlusion or stenosis of precerebral arteries (CCS 110)	3690	0.10	639	0.27
Other and ill-defined cerebrovascular disease (CCS 111)	2752	0.07	120	0.05
Transient cerebral ischemia (CCS 112)	39882	1.03	1574	0.68
Late effects of cerebrovascular disease (CCS 113)	3027	0.08	269	0.12
Peripheral and visceral atherosclerosis (CCS 114)	3459	0.09	235	0.10
Aortic and peripheral arterial embolism or thrombosis (CCS 116)	184	<0.00	--	--
Phlebitis; thrombophlebitis and thromboembolism (CCS 118)	57	<0.00	--	--
Hemorrhoids (CCS 120)	5785	0.15	325	0.14
Pneumonia (except that caused by tuberculosis or sexually transmitted disease) (CCS 122)	199409	5.15	6198	2.66
Influenza (CCS 123)	20888	0.54	275	0.12
Acute bronchitis (CCS 125)	12931	0.33	272	0.12
Other upper respiratory infections (CCS 126)	3942	0.10	194	0.08
Chronic obstructive pulmonary disease and bronchiectasis (CCS 127)	132341	3.42	3020	1.30
Asthma (CCS 128)	28473	0.74	2003	0.86
Aspiration pneumonitis; food/vomitus (CCS 129)	39964	1.03	1507	0.65
Pleurisy; pneumothorax; pulmonary collapse (CCS 130)	16124	0.42	772	0.33
Respiratory failure; insufficiency; arrest (adult) (CCS 131)	68985	1.78	4388	1.88
Lung disease due to external agents (CCS 132)	1010	0.03	35	0.02
Other lower respiratory disease (CCS 133)	18308	0.47	1395	0.60
Other upper respiratory disease (CCS 134)	277	0.01	-	-

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Intestinal infection (CCS 135)	35578	0.92	1570	0.67
Esophageal disorders (CCS 138)	15068	0.39	955	0.41
Gastritis and duodenitis (CCS 140)	12060	0.31	515	0.22
Other disorders of stomach and duodenum (CCS 141)	9945	0.26	540	0.23
Appendicitis and other appendiceal conditions (CCS 142)	10224	0.26	1205	0.52
Abdominal hernia (CCS 143)	34344	0.89	3006	1.29
Regional enteritis and ulcerative colitis (CCS 144)	5428	0.14	281	0.12
Intestinal obstruction without hernia (CCS 145)	68942	1.78	4091	1.76
Diverticulosis and diverticulitis (CCS 146)	63076	1.63	2867	1.23
Anal and rectal conditions (CCS 147)	6193	0.16	566	0.24
Biliary tract disease (CCS 149)	56274	1.45	4170	1.79
Pancreatic disorders (not diabetes) (CCS 152)	26989	0.70	1798	0.77
Gastrointestinal hemorrhage (CCS 153)	82245	2.12	4851	2.08
Noninfectious gastroenteritis (CCS 154)	19058	0.49	553	0.24
Other gastrointestinal disorders (CCS 155)	27867	0.72	2093	0.90
Nephritis; nephrosis; renal sclerosis (CCS 156)	712	0.02	26	0.01
Acute and unspecified renal failure (CCS 157)	112224	2.90	3813	1.64
Chronic kidney disease (CCS 158)	2108	0.05	246	0.11
Urinary tract infections (CCS 159)	125457	3.24	3801	1.63
Other diseases of kidney and ureters (CCS 161)	3433	0.09	261	0.11
Other diseases of bladder and urethra (CCS 162)	842	0.02	--	--
Hyperplasia of prostate (CCS 164)	127	<0.00	--	--
Nonmalignant breast conditions (CCS 167)	160	<0.00	--	--
Inflammatory diseases of female pelvic organs (CCS 168)	114	<0.00	--	--
Prolapse of female genital organs (CCS 170)	936	0.02	--	--
Ovarian cyst (CCS 172)	231	0.01	--	--
Menopausal disorders (CCS 173)	46	<0.00	--	--
Other female genital disorders (CCS 175)	937	0.02	68	0.03
Skin and subcutaneous tissue infections (CCS 197)	72797	1.88	2608	1.12
Chronic ulcer of skin (CCS 199)	778	0.02	--	--

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease) (CCS 201)	8630	0.22	538	0.23
Rheumatoid arthritis and related disease (CCS 202)	890	0.02	--	--
Osteoarthritis (CCS 203)	319802	8.26	25820	11.08
Other non-traumatic joint disorders (CCS 204)	9679	0.25	303	0.13
Spondylosis; intervertebral disc disorders; other back problems (CCS 205)	104694	2.70	5258	2.26
Pathological fracture (CCS 207)	14886	0.38	871	0.37
Acquired foot deformities (CCS 208)	218	0.01	--	--
Other acquired deformities (CCS 209)	11662	0.30	390	0.17
Other connective tissue disease (CCS 211)	2639	0.07	392	0.17
Other bone disease and musculoskeletal deformities (CCS 212)	9090	0.23	530	0.23
Cardiac and circulatory congenital anomalies (CCS 213)	970	0.03	20	0.01
Digestive congenital anomalies (CCS 214)	362	0.01	--	--
Nervous system congenital anomalies (CCS 216)	99	<0.00	--	--
Other congenital anomalies (CCS 217)	3162	0.08	--	--
Joint disorders and dislocations; trauma-related (CCS 225)	3105	0.08	190	0.08
Fracture of neck of femur (hip) (CCS 226)	121231	3.13	6536	2.81
Skull and face fractures (CCS 228)	2932	0.08	105	0.05
Fracture of upper limb (CCS 229)	25228	0.65	1412	0.61
Fracture of lower limb (CCS 230)	34873	0.90	1983	0.85
Other fractures (CCS 231)	56917	1.47	1881	0.81
Sprains and strains (CCS 232)	3256	0.08	145	0.06
Open wounds of head; neck; and trunk (CCS 235)	2058	0.05	135	0.06
Open wounds of extremities (CCS 236)	1657	0.04	124	0.05
Complication of device; implant or graft (CCS 237)	46649	1.21	2880	1.24
Complications of surgical procedures or medical care (CCS 238)	10409	0.27	1171	0.50
Superficial injury; contusion (CCS 239)	7477	0.19	419	0.18
Syncope (CCS 245)	36058	0.93	2989	1.28
Fever of unknown origin (CCS 246)	5620	0.15	327	0.14
Lymphadenitis (CCS 247)	239	0.01	33	0.01
Gangrene (CCS 248)	3082	0.08	58	0.02
Shock (CCS 249)	343	0.01	40	0.02

Risk Variable	Claims-Only Development Dataset		Clinical Hybrid Development Dataset	
	Frequency #	Percentage (%)	Frequency #	Percentage (%)
Nausea and vomiting (CCS 250)	3987	0.10	389	0.17
Abdominal pain (CCS 251)	10576	0.27	849	0.36
Other aftercare (CCS 257)	532	0.01	3140	1.35
Residual codes; unclassified (CCS 259)	115	<0.00	--	--
Other and ill-defined heart disease (CCS 104_2)	2222	0.06	153	0.07
Cardiac arrest and ventricular fibrillation (CCS 107_1)	178	<0.00	18	0.01
Cardiac arrest and ventricular fibrillation (CCS 107_2)	2200	0.06	109	0.05
Acute cerebrovascular disease (CCS 109_1)	28019	0.72	1700	0.73
Acute cerebrovascular disease (CCS 109_2)	128914	3.33	7262	3.12
Aortic; peripheral; and visceral artery aneurysms (CCS 115_1)	182	<0.00	--	--
Aortic; peripheral; and visceral artery aneurysms (CCS 115_2)	501	0.01	--	--
Aortic; peripheral; and visceral artery aneurysms (CCS 115_3)	1015	0.03	35	0.02
Other circulatory disease (CCS 117_2)	179	<0.00	--	--
Gastroduodenal ulcer (except hemorrhage) (CCS 139_1)	3865	0.10	224	0.10
Gastroduodenal ulcer (except hemorrhage) (CCS 139_2)	3768	0.10	195	0.08
Peritonitis and intestinal abscess (CCS 148_1)	853	0.02	47	0.02
Peritonitis and intestinal abscess (CCS 148_2)	1563	0.04	88	0.04
Other liver diseases (CCS 151_1)	10153	0.26	634	0.27
Other liver diseases (CCS 151_2)	2461	0.06	216	0.09
Other injuries and conditions due to external causes (CCS 244_1)	978	0.03	45	0.02
Other injuries and conditions due to external causes (CCS 244_2)	6752	0.17	398	0.17
Other nutritional; endocrine; and metabolic disorders (CCS 58_2)	4474	0.12	--	--
Other nervous system disorders (CCS 95_1)	19952	0.52	176	0.08
Other nervous system disorders (CCS 95_2)	18617	0.48	1330	0.57

APPENDIX C. CREATING THE FINAL SERVICE-LINE DIVISIONS

Grouping of Sub-Divisions

For surgical admissions, we used work done previously for the Hospital-Wide Readmission (HWR) measure, which identified and then classified each major surgical procedure Clinical Classification Software (CCS) into one of 10 surgical sub-divisions based on surgical service-line with clinician input; these groupings were re-reviewed by five physicians on our team as well as our Technical Expert Panel (TEP).

For the non-surgical admissions, two practicing physicians at CORE reviewed the CCS categories for principal discharge diagnoses and grouped them into 23 clinically coherent non-surgical sub-divisions based upon service-line. These sub-divisions were reviewed by three additional physicians and any discrepancies were resolved by consensus among all physicians. The final sub-divisions were then reviewed and endorsed by our TEP.

Combining Sub-Divisions into Service-Line Divisions

For each of the 23 non-surgical and 10 surgical sub-divisions, we then calculated the odds ratios (OR) for risk of 30-day mortality with 95% confidence intervals (CI) for all of the candidate comorbidity variables (see Table C.1. below), and, for each of the surgical sub-divisions, we also calculated the OR for risk of 30-day mortality with 95% CI for all of the principal discharge diagnosis CCSs. This ensured that the reason for admission for the surgical patients (the principal discharge diagnosis) was also considered for combining sub-divisions. This was not necessary for non-surgical divisions, as the non-surgical divisions were defined using the principal discharge diagnosis CCS. We also calculated the number of patients within each sub-division to understand possible case volume limitations across the sub-divisions. We used this information to further combine sub-divisions into divisions based on clinical coherence as well as similar directionality across the majority of the comorbid conditions, while still trying to ensure adequate case volume.

Using this approach, we combined the 23 non-surgical sub-divisions into nine service-line divisions (eight more homogeneous divisions, and one “Other Condition” division that included admissions across multiple specialties), and the 10 surgical sub-divisions into six surgical divisions (five more homogeneous divisions, and one “Other Procedures” division that included admissions across multiple types of procedures). This created a total of 15 clinical or “service-line” divisions that are used both to organize the cohort and for service-line risk adjustment. The 15 final service-line divisions are: Non-surgical: Cancer, Cardiac, Gastrointestinal, Infectious Disease, Neurology, Orthopedics, Pulmonary, Renal, Other; and Surgical: Cancer, Cardiothoracic, General, Neurosurgery, Orthopedics, Other.

Table C.1. Candidate Claims-Based Risk Variables and Associated Condition Category (CC) (Original Development)

Risk Adjustment Variable	Condition Category (CC)
Age	N/A
Transfer from Outside ED	N/A
Opportunistic/Chronic Infections	CC 1, 3-6, 39
Lymphoma & Other Cancers	CC 10
TIA and Other Cerebrovascular Disease	CC 101, 102
Vascular Disease with Complications	CC 106, 107
Vascular Disease	CC 108
Other Circulatory Disease	CC 109
Other Cancers & Heart or Respiratory Tumors	CC 11-13
Fibrosis of Lung and Other Chronic Lung Disorders	CC 110, 112
Chronic Obstructive Pulmonary Disease	CC 111
Asthma	CC 113
Pneumonia	CC 114-116
Pleural Effusion/Pneumothorax	CC 117
Other Respiratory Disorders	CC 118
Eye Infections and Retinal Disorders	CC 120-122, 124, 125
Glaucoma	CC 126
Other Eye Disorders	CC 128
Other ENT and Mouth Disorders	CC 129, 131
Hearing Loss	CC 130
Transplant Status	CC 132, 186, 187
Dialysis or Severe Chronic Kidney Disease	CC 134, 136, 137
Acute or Unspecified Renal Failure	CC 135, 140
Mild to Moderate Chronic Kidney Disease	CC 138, 139
Other Benign Tumors	CC 14-16
Other Renal or Urinary Tract Disorders	CC 141, 145
Urinary Obstruction and Retention	CC 142
Urinary Incontinence	CC 143
Urinary Tract Infection	CC 144
Female Genital Disorders	CC 147, 148
Male Genital Disorders	CC 149
Pressure Ulcer	CC 157-160
Burns, Non-pressure Ulcers	CC 161-163
Cellulitis, Local Skin Infection	CC 164
Other Dermatological Disorders	CC 165
Other Head Injuries or Concussion	CC 167, 168
Amputation Status and Major Fractures Including Vertebral, Hip, and Other	CC 169-171, 173, 189, 190

Risk Adjustment Variable	Condition Category (CC)
Diabetes	CC 17-19
Other Injuries	CC 172, 174
Poisonings and Allergic and Inflammatory Reactions	CC 175
Complications of Care	CC 176, 177
Major Symptoms, Abnormalities	CC 178
Minor Symptoms, Signs, Findings	CC 179
Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock	CC 2
Protein-Calorie Malnutrition	CC 21
Morbid Obesity	CC 22
Other Significant Endocrine and Metabolic Disorders	CC 23
Disorders of Fluid/Electrolyte/Acid-Base Balance	CC 24
Disorders of Lipoid Metabolism	CC 25
Other Endocrine/Metabolic/Nutritional Disorders	CC 26
Liver Failure	CC 27, 30
Cirrhosis & Chronic Hepatitis	CC 28, 29
Other Liver & Biliary Disease	CC 31, 32
Intestinal Obstruction/Perforation, Peptic Ulcer, Hemorrhage, and Other Specified GI Disorders	CC 33, 36
Other GI Disorders	CC 34, 35, 37, 38
Rheumatoid Arthritis and Inflammatory Connective Tissue Disease	CC 40
Disorders of the Vertebrae and Spinal Discs	CC 41
Osteoarthritis of Hip or Knee	CC 42
Osteoporosis and Other Bone/Cartilage Disorders	CC 43
Other Musculoskeletal and Connective Tissue Disorders	CC 44, 45
Hematologic or Immunity Disorders	CC 46-48
Iron Deficiency and Other/Unspecified Anemias and Blood Disease	CC 49
Delirium and Encephalopathy	CC 50
Dementia and Other Nonpsychotic Organic Brain Syndromes	CC 51-53
Drug/Alcohol Dependence or Psychosis	CC 54, 55
Drug/Alcohol Abuse, Without Dependence	CC 56
Psychosis: Schizophrenia, Reactive, and Unspecified	CC 57, 59
Major Depressive, Bipolar, and Paranoid Disorders	CC 58
Other Psychiatric Disorders	CC 60, 63

Risk Adjustment Variable	Condition Category (CC)
Depression	CC 61
Anxiety Disorders	CC 62
Other Developmental Disorders	CC 64-68
Other Infectious Diseases	CC 7
Paralytic Syndromes	CC 70-72, 103, 104
Neuromuscular Disorders	CC 73-76, CC78
Seizure Disorders and Convulsions	CC 79
Metastatic & Severe Cancers	CC 8, 9
Coma/Brain Compression/Anoxic Injury and Severe Head Injury	CC 80, 166
Polyneuropathy, Mononeuropathy, and Other Neurological Conditions/Injuries	CC 81
Respiratory Failure, Respirator Dependence, Shock	CC 82-84
Congestive Heart Failure	CC 85
Acute Myocardial Infarction	CC 86
Angina and Unstable Angina	CC 87, 88
Coronary Atherosclerosis/Other Chronic Ischemic Heart Disease	CC 89
Other and Unspecified Heart Disease	CC 90, 92, 93, 98
Valvular and Rheumatic Heart Disease	CC 91
Hypertension and Hypertensive Heart Disease	CC 94, 95
Heart Rhythm and Conduction Disorders	CC 96, 97
Cerebral Hemorrhage, Stroke, Late Effects of Stroke	CC 99, 100, 105

Note: Descriptions of the Condition Categories can be found at http://www.cms.hhs.gov/Reports/downloads/pope_2000_2.pdf

APPENDIX D. FINAL RISK MODEL SELECTION (CLINICAL HYBRID DEVELOPMENT DATASET)

After we finalized the risk variables during initial development (including age, the 10 clinical electronic health record (HER)-based risk variables, the claims-based comorbid risk variables, and the principal discharge diagnosis variables), we tested three different risk models within the Clinical Hybrid Development Dataset. We directly compared the claims-only risk model calculated in the Clinical Hybrid Dataset to multiple variants that included clinical EHR-based risk variables and selected the best risk model based upon statistical performance and face validity as determined by our Technical Expert Panel (TEP). We tested the following risk models:

1. Baseline: "Claims-Only Risk Model": Uses only claims-based variables in risk model
    a. Service mix: Agency for Healthcare Research and Quality (AHRQ) Clinical Classification Software (CCS) categories for patients' principal discharge diagnoses captured from claims data
    b. Case mix: CMS Condition Categories (CCs) for patients' comorbidities captured from claims data during hospitalizations in the 12 months prior to and including the index admission (age plus 19 CC risk variables for each service-line division risk model from Claims-only HWM measure)
2. "Clinical-Only Risk Model": Uses only EHR-based clinical variables in risk model (no claims comorbidity OR principal discharge diagnoses)
    a. Service mix: None
    b. Case mix: age plus 10 clinical variables captured from EHR data
3. "Clinical + Principal Discharge Diagnoses Risk Model": Uses EHR-based clinical variables with claims-based principal discharge diagnoses in risk model (no claims comorbidity)
    a. Service mix: AHRQ CCS categories for patients' principal discharge diagnoses captured from claims data
    b. Case mix: age plus 10 clinical variables captured from EHR data
4. "Clinical + Claims Risk Model": Uses EHR-based clinical variables + claims-based comorbidity and principal discharge diagnosis variables in risk model:
    a. Service mix: AHRQ CCS categories for patients' principal discharge diagnoses captured from claims data
    b. Case mix: Both the age plus 10 clinical variables captured from EHR data and the CCs for patients' comorbidities captured from claims data during hospitalizations in the 12 months prior to and including the index admission (19 CC risk variables and age plus 10 clinical variables for each division risk model)

Table D.1 shows the c-statistics produced by each of the four models for each division and demonstrates that all four models provide similar discrimination. Each risk model offers slightly different advantages. The risk models with clinical data offer greater face validity and capture data reflecting the status of patients upon presentation. The Clinical + Principal Discharge Diagnoses Risk Model, without claims-based comorbidity data, would allow inclusion of patients who do not have 12 months of history data available (approximately 700,000 more potential admissions in the measure cohort when applied to the entire Medicare FFS population). However, this model performed slightly worse than other models. After reviewing the results with our TEP and based upon their preference for higher discrimination over other

features (parsimony, not requiring 12 months of history data), we selected Clinical + Claims Risk Model for the Hybrid HWM Measure.

Rationale: The Clinical + Claims Risk Model, which includes the broadest set of risk variables, had the best statistical performance and the highest face validity per the majority of the TEP by accounting for clinical EHR variables, principal discharge diagnoses, and comorbidities identified using claims-data. While it does require the exclusion of patients not enrolled in Medicare for 12 months prior to admission, this was the preferred model by the majority of the TEP.

Table D.1. Comparison of C-Statistics by Division of Clinical-Only Model, Claims-Only Model, Clinical + Principal Discharge Diagnoses Model, and Final Hybrid (Clinical + Claims) Model, Using Clinical Hybrid Dataset (January 1, 2010 – December 31, 2015)

Division	Clinical-Only Model C-Statistic	Claims-Only Model C-Statistic	Clinical + Principal Discharge Diagnoses Model C-Statistic	Clinical + Claims (Final Hybrid) Model C-Statistic
Non-Surgical: Cancer	0.79	0.83	0.84	0.87
Non-Surgical: Cardiac	0.84	0.84	0.86	0.88
Non-Surgical: Gastrointestinal	0.81	0.87	0.85	0.89
Non-Surgical: Infectious Disease	0.79	0.78	0.79	0.83
Non-Surgical: Neurology	0.74	0.81	0.80	0.83
Non-Surgical: Orthopedics	0.82	0.86	0.85	0.88
Non-Surgical: Pulmonary	0.75	0.76	0.78	0.80
Non-Surgical: Renal	0.82	0.83	0.83	0.86
Surgical: Cardiothoracic	0.80	0.83	0.85	0.85
Surgical: General	0.89	0.92	0.93	0.94
Surgical: Neurosurgery	0.85	--	--	--
Surgical: Orthopedics	0.89	0.92	0.92	0.93

APPENDIX E. ADDITIONAL CLAIMS-ONLY HWM RESULTS (USING MEDICARE CLAIMS RE-SPECIFICATION DATASET)

Table E.1 through Table E.4 and Figure E.1 provide additional results for the Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset, for reference as to what may be expected for the Hybrid HWM Measure in a national dataset.

Table E. 1. Hospital Volume Distribution by Number of Divisions (July 2016-June 2017) (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

# Divisions	# Hospitals	Mean # Patients	SD	Median # Patients	Min # Patients	25th Quartile	75th Quartile	Max # Patients
1-5 Divisions	209	73	136	14	1	6	55	674
6-10 Divisions	974	100	210	63	9	38	103	5,043
11-15 Divisions	3,509	1,204	1,292	747	35	283	1,686	14,256

Table E. 2. Number of Hospitals by Volume and Divisions (July 2016-June 2017) (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

Hospital Volume	# Hospitals	1-5 Divisions	6-10 Divisions	11-15 Divisions
All hospitals	4,692	209	974	3,509
<25 Patients	237	136	101	0
25-144 Patients	1,111	39	747	325
144-456 Patients	1,115	26	102	987
456-1,384 Patients	1,115	8	21	1,086
≥1,384 Patients	1,114	0	3	1,111

Figure E. 1. Distribution of Hospital 30-Day HWM RSMRs between July 2016 and June 2017 (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

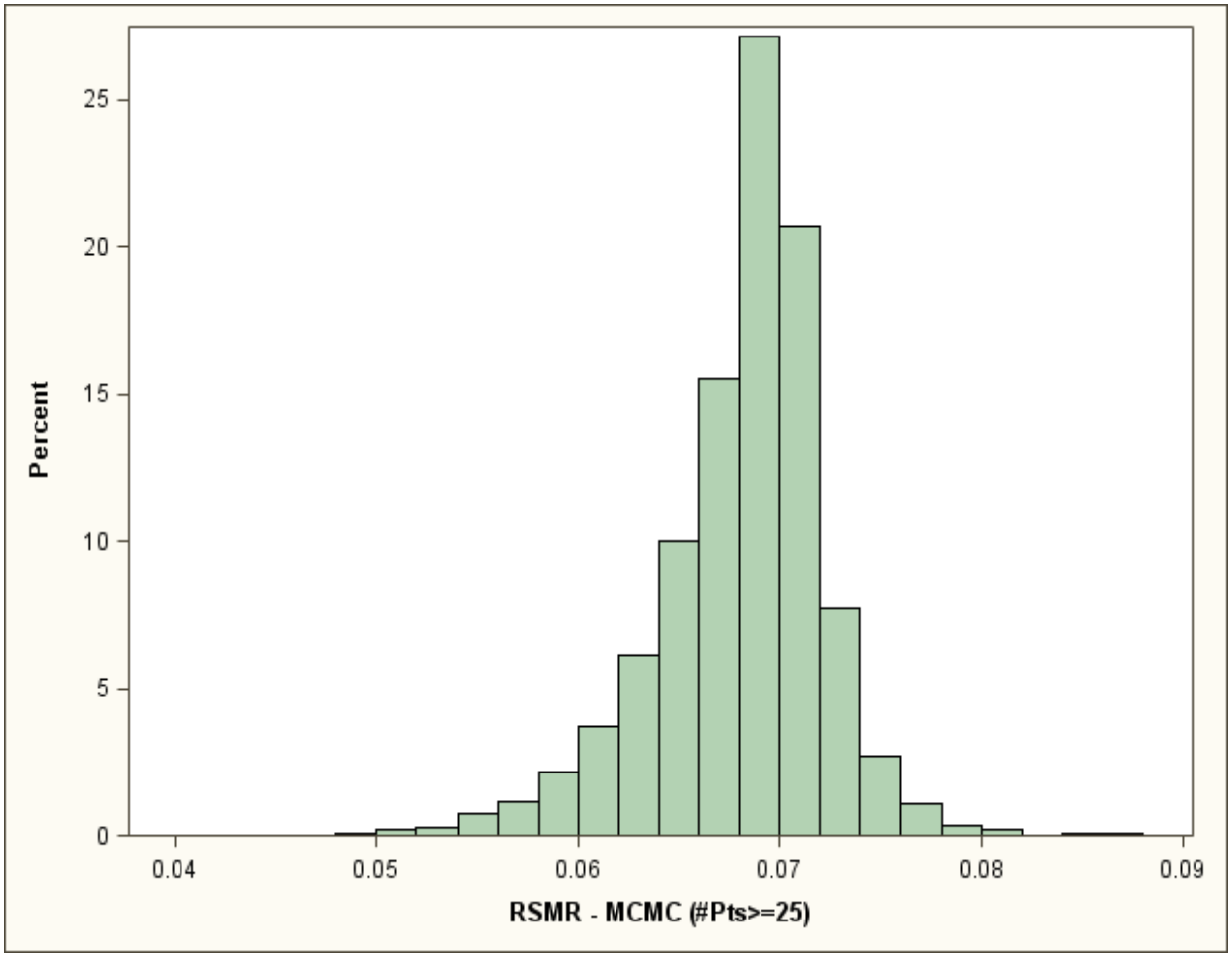


Table E. 3. Distribution of RSMRs (July 2016-June 2017) (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

Hospitals	# Hospitals	Mean	SD	Median	Minimum	25% Quartile	75% Quartile	Maximum
All Hospitals	4,692	6.85%	0.41%	6.93%	3.95%	6.66%	7.09%	8.70%
Hospitals with >=25 Patients	4,455	6.84%	0.42%	6.91%	3.95%	6.64%	7.09%	8.70%

Table E. 4. Distribution of RSMRs by Number of Divisions (July 2016-June 2017) (Claims-only HWM measure, using the Medicare Claims Re-specification Dataset)

# Divisions	# Hospitals	Mean	SD	Median	Minimum	25% Quartile	75% Quartile	Maximum
Division:1-5	209	7.03%	0.16%	7.02%	6.29%	6.96%	7.08%	7.65%
Division:6-10	974	7.01%	0.18%	7.01%	5.51%	6.91%	7.11%	7.60%
Division:11-15	3,509	6.80%	0.46%	6.86%	3.95%	6.56%	7.08%	8.70%

APPENDIX F. RISK DECILE PLOTS

Figure F.1 through Figure F.9 below show the risk-decile plots for each of the nine divisions with sufficient data for calculation in the Hybrid HWM Measure (using the Clinical Hybrid Re-specification Dataset).

Figures F.10 through Figure F.15 show the risk-decile plots for the for the other six divisions in the Claims-only HWM measure, using the Medicare Claims Re-specification Dataset.

Figure F.1. Risk Decile Plot for the Non-Surgical Cardiac Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

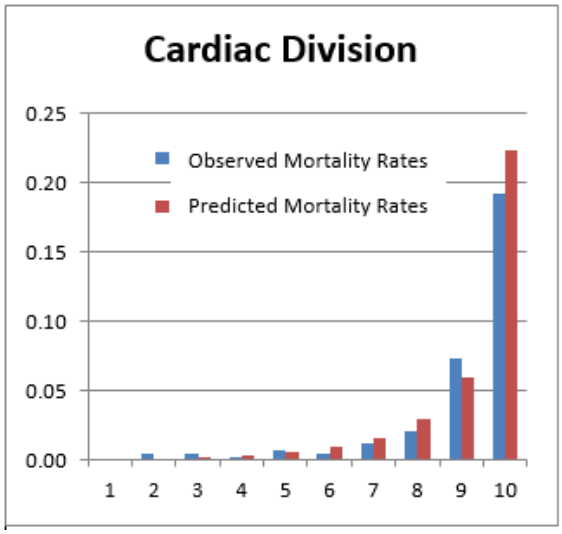


Figure F.2. Risk Decile Plot for the Non-Surgical GI Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

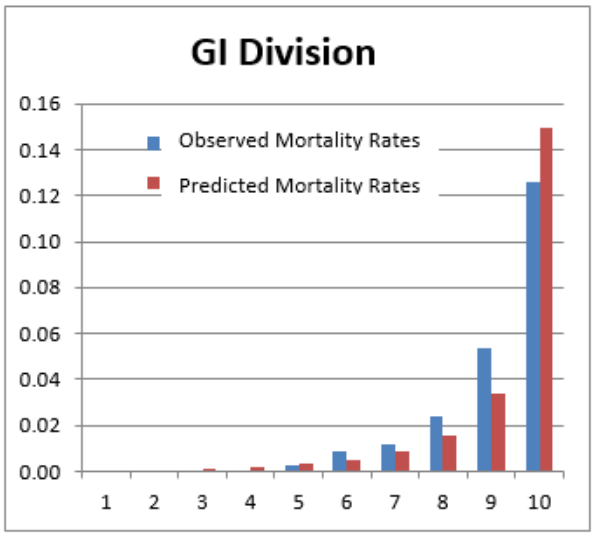


Figure F.3. Risk Decile Plot for the Non-Surgical Infectious Disease Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

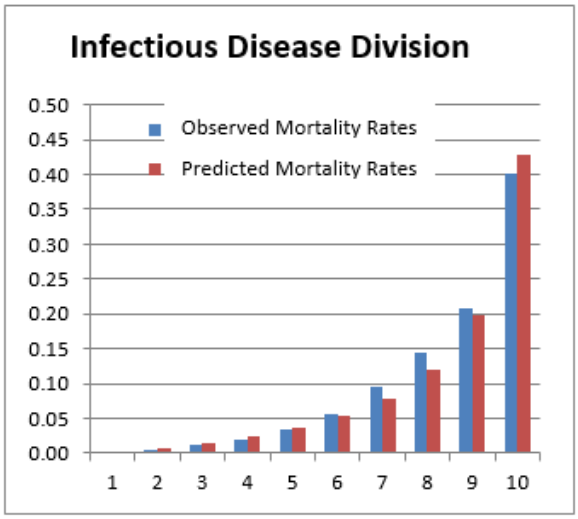


Figure F.4. Risk Decile Plot for the Non-Surgical Neurology Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

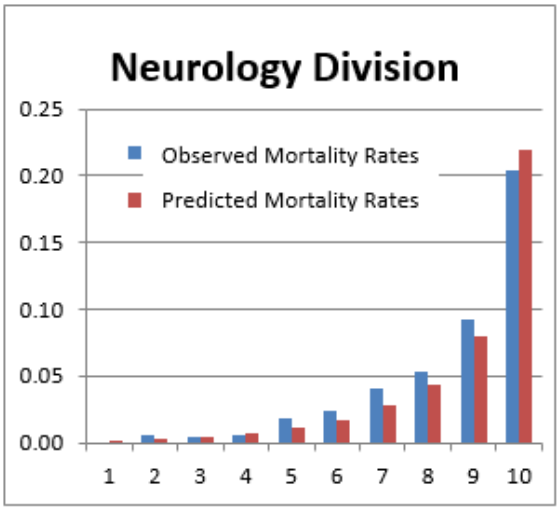


Figure F.5.Risk Decile Plot for the Non-Surgical Orthopedic Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

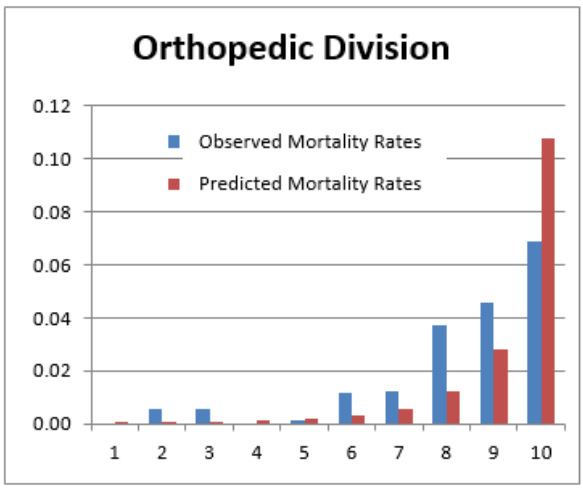


Figure F.6. Risk Decile Plot for the Non-Surgical Pulmonary Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

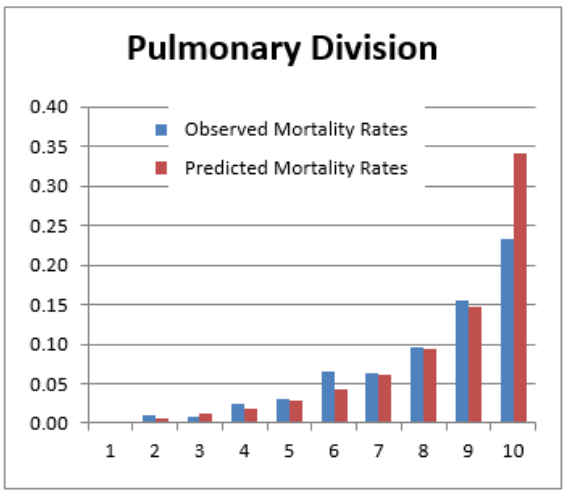


Figure F.7. Risk Decile Plot for the Non-Surgical Renal Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

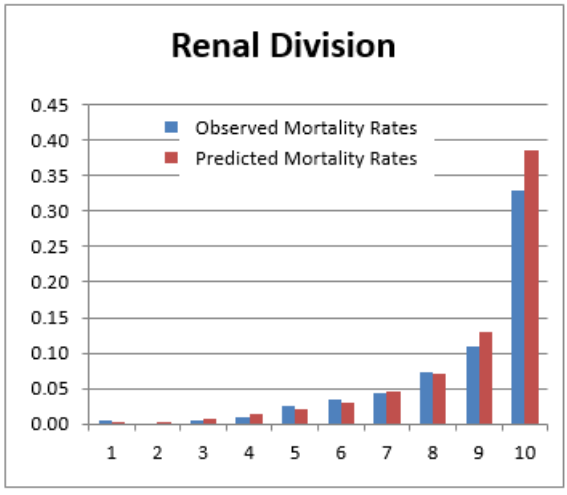


Figure F.8. Risk Decile Plot for the Non-Surgical Other Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

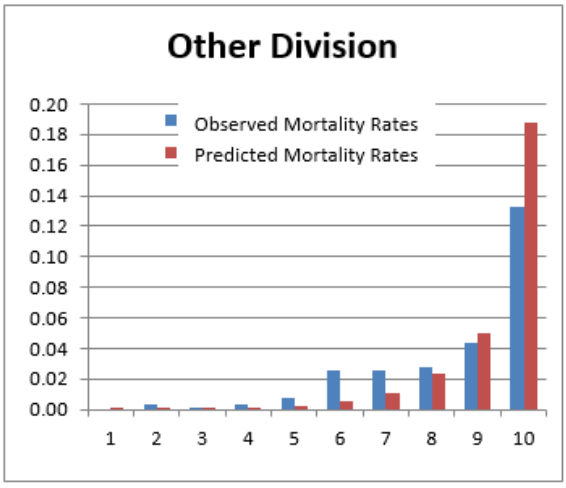


Figure F.9. Risk Decile Plot for the Surgical General Division (Hybrid HWM Measure, using the Clinical Hybrid Re-specification Dataset)

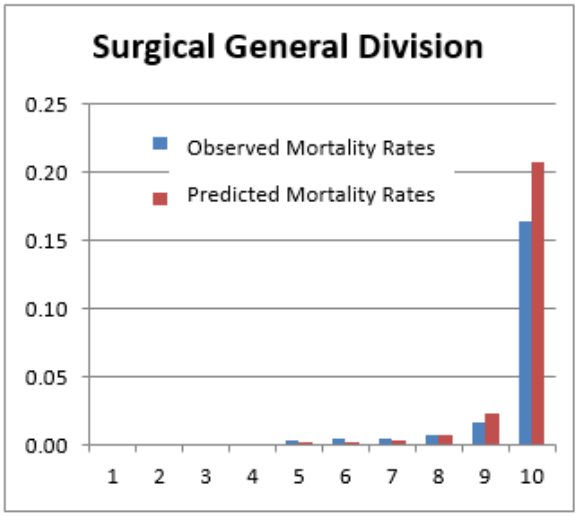


Figure F.10. Risk Decile Plot for the Non-Surgical Cancer Division (Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset)

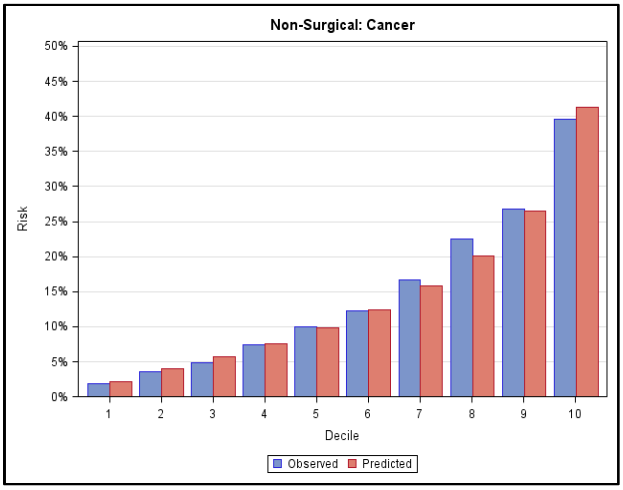


Figure F.11. Risk Decile Plot for the Surgical Cancer Division (Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset)

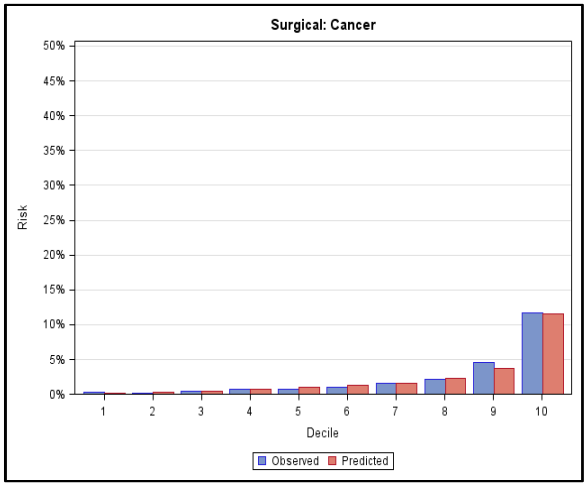


Figure F.12. Risk Decile Plot for the Surgical Cardiothoracic Division (Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset)

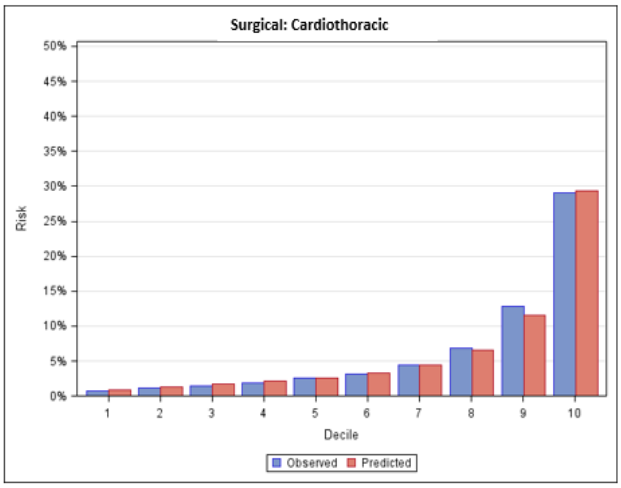


Figure F.13. Risk Decile Plot for the Surgical Neurosurgery Division (Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset)

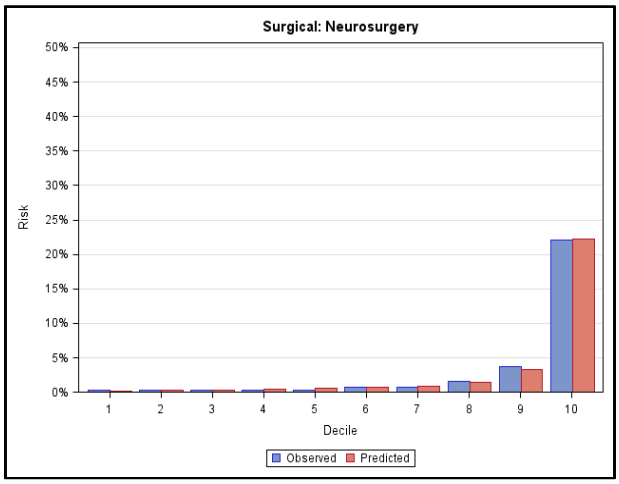


Figure F.14. Risk Decile Plot for the Surgical Orthopedic Division (Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset)

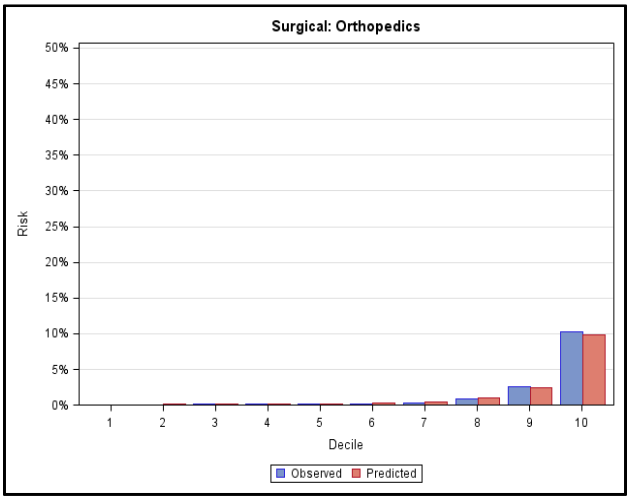


Figure F.15. Risk Decile Plot for the Surgical Other Division (Claims-only HWM Measure, using the Medicare Claims Re-specification Dataset)