AHRQ Quality Indicators™

Quality Indicator Empirical Methods, v2022

Prepared for:

Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
5600 Fishers Lane
Rockville, MD 20857
https://qualityindicators.ahrq.gov/

Contract No. HHSA290201800003G

July 2022

Table of Contents

Acknowledgments **1**

Abbreviations and Terms Used in this Document **3**

Chapter I. Background and Overview **5**

A. Background on Agency for Healthcare Research and Quality (AHRQ) Quality Indicators (QIs) ..5

B. AHRQ QI Results: Counts, Rates, and Scores 6

C. Brief History of the AHRQ QIs 8

D. Overview of the Empirical Methods Document 9

Chapter II. AHRQ QIs Modules and Methods **10**

A. AHRQ QI Modules 10

B. Specifications 15

C. Data 16

D. Patient Population 16

E. Area-Level Quality Indicators 19

F. Hospital-Level Quality Indicators 25

G. Recommendations on How to Report Trends 38

Chapter III. Empirical Development of the AHRQ QIs **40**

A. Overview of the Development Process 40

B. Discharge Reference Population 41

C. Other Data Used for Area-Level Indicator Development 45

D. Coding Updates 46

E. Reference Population: Numerators, Denominators, and Observed Rates 47

F. Reference Population: Risk Model Development and Parameter Estimates (v2022) 47

G. Composite Development 55

H. Empirical Testing – Evaluating AHRQ QI Specifications and Risk Models 59

Chapter IV. References **62**

Chapter V. Appendices **63**

Appendix A. Other Helpful Documents 63

Appendix B. Comprehensive List of Quality Indicators 64

Appendix C. Comprehensive Lists of Risk Factors for Quality Indicator Modules Appendix 69

Appendix D. Patient Harms Captured in the AHRQ Patient Safety and Adverse Events Composite 71

Index of Tables

Table 1. Quality Domains Addressed by Area-Level and Hospital-Level Modules 6
Table 2. List of AHRQ Prevention Quality Indicators (PQIs) 10
Table 3. List of AHRQ Inpatient Quality Indicators (IQIs) 12
Table 4. List of AHRQ Patient Safety Indicators (PSIs) 13
Table 5. List of AHRQ Pediatric Quality Indicators (PDIs) 15
Table 6. Analysis Data Inclusion Rule 16
Table 7. Age Groups and Indicators 16
Table 8. Values for the Present-on-Admission Data Element 18
Table 9. AHRQ QI Risk Adjustment Covariates for Hospital-Level Indicators 29
Table 10. AHRQ QI Composite and Weight 35
Table 11. AHRQ QI Specification Releases 41
Table 12. AHRQ QI Reference Population 43
Table 13. Treatment of State Border Crossing Discharges in the Reference Population 44
Table 14. Weights of PSI 90 Component Indicators, v2022 58

Acknowledgments

The Agency for Healthcare Research and Quality (AHRQ) Quality Indicators (QI) program uses the Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID)[1] (available at https://www.hcup-us.ahrq.gov/sidoverview.jsp) for the development of the AHRQ QIs, using HCUP SID as the reference (general or standard) population. HCUP is a family of healthcare databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by AHRQ. HCUP databases bring together the data collection efforts of State data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of encounter-level health care data. HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs, and outcomes of treatments at the national, State, and local market levels. The HCUP SID encompass about 97 percent of all annual inpatient discharges in the United States.

The AHRQ QI program would like to acknowledge the HCUP Partner organizations that participated in the HCUP SID:

Alaska Department of Health and Social Services
Alaska State Hospital and Nursing Home Association
Arizona Department of Health Services
Arkansas Department of Health
California Department of Health Care Access and Information (HCAI)
Colorado Hospital Association
Connecticut Hospital Association
Delaware Division of Public Health
District of Columbia Hospital Association
Florida Agency for Health Care Administration
Georgia Hospital Association
Hawaii University of Hawaii at Hilo
Hawaii Laulima Data Alliance, a non-profit subsidiary of the Healthcare Association of Hawaii
Illinois Department of Public Health
Indiana Hospital Association
Iowa Hospital Association
Kansas Hospital Association
Kentucky Cabinet for Health and Family Services
Louisiana Department of Health
Maine Health Data Organization
Maryland Health Services Cost Review Commission
Massachusetts Center for Health Information and Analysis
Michigan Health & Hospital Association
Minnesota Hospital Association (provides data for Minnesota and North Dakota)
Mississippi State Department of Health
Missouri Hospital Industry Data Institute
Montana Hospital Association

[1] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp.

Nebraska Hospital Association
Nevada Department of Health and Human Services
New Hampshire Department of Health & Human Services
New Jersey Department of Health
New Mexico Department of Health
New York State Department of Health
North Carolina Department of Health and Human Services
North Dakota (data provided by the Minnesota Hospital Association)
Ohio Hospital Association
Oklahoma State Department of Health
Oregon Association of Hospitals and Health Systems
Oregon Health Authority
Pennsylvania Health Care Cost Containment Council
Rhode Island Department of Health
South Carolina Revenue and Fiscal Affairs Office
South Dakota Association of Healthcare Organizations
Tennessee Hospital Association
Texas Department of State Health Services
Utah Department of Health
Vermont Association of Hospitals and Health Systems
Virginia Health Information
Washington State Department of Health
West Virginia Department of Health and Human Resources
Wisconsin Department of Health Services
Wyoming Hospital Association

For more information on HCUP, visit https://www.hcup-us.ahrq.gov

Abbreviations and Terms Used in this Document

Abbreviation/Term	Descriptions
AAA	Abdominal Aortic Aneurysm
AHA	American Hospital Association
AHRQ	Agency for Healthcare Research and Quality
AMI	Acute Myocardial Infarction
APR-DRG	All Patient Refined Diagnosis Related Groups
ATT	Average Treatment effect on the Treated
CABG	Coronary Artery Bypass Graft
CCSR	Clinical Classifications Software Refined
CMS	Centers for Medicare and Medicaid Services
COPD	Chronic Obstructive Pulmonary Disease
DNR	Do Not Resuscitate
DRGs	Diagnosis Related Groups
DVT	Deep Vein Thrombosis
ECMO	Extracorporeal Membrane Oxygenation
FIPS	Federal Information Processing Standards
FSCPE	Federal State Cooperative Program for Population Estimates
FY	Fiscal Year
HCUP	Healthcare Cost and Utilization Project
ICD-9-CM	International Classification of Diseases Volume 9 Clinical Modification
ICD-10-CM/PCS	International Classification of Diseases Volume 10 Clinical Modification or Procedure Code System
IQI	Inpatient Quality Indicator
LASSO	Least Absolute Shrinkage and Selection Operator
LTAC	Long-Term Acute Care
MDC	Major Diagnostic Category
MS-DRG	Medicare Severity Diagnostic Related Group
NCHS	National Center for Health Statistics

Abbreviation/Term	Descriptions
NQF	National Quality Forum
NQI	Neonatal Quality Indicator
NUBC	National Uniform Bill Committee
O-E	Observed-to-Expected
OR	Operating Room
PCI	Percutaneous Coronary Intervention
PDI	Pediatric Quality Indicator
PE	Pulmonary Embolism
POA	Present on Admission
POVCAT	Poverty Decile
PQI	Prevention Quality Indicator
PSI	Patient Safety Indicator
PS	Propensity Score
QI	Quality Indicator
ROC	Receiver Operating Characteristic
ROM	Risk of Mortality
SAF	Standard Analytical Files
SAIPE	Census Bureau Small Area Income and Poverty Estimates
SID	State Inpatient Databases
U.S.	United States
UB	Uniform Bill
VBAC	Vaginal Birth after Cesarean
VIF	Variance Inflation Factor

Chapter I. Background and Overview

A. Background on Agency for Healthcare Research and Quality (AHRQ) Quality Indicators (QIs)

This document describes the empirical methods used to develop and calculate the Agency for Healthcare Research and Quality (AHRQ) Quality Indicators™ (QIs) v2022 (including risk adjustment and smoothing). Using administrative data (e.g., hospital discharge abstracts, billing records, or claims data), the AHRQ QIs measure health care quality and can be used to highlight potential quality concerns, identify areas that need further study and investigation, and track changes over time.

The AHRQ QIs measure quality and utilization at two different levels of analysis, the area level and the hospital level.[2]

- **Area-level indicators** capture all cases of potentially preventable complications that occur in a given population resulting in hospitalization. For example, area-level indicators may answer the question: Was the inpatient admission for a condition that might have been avoided, if the patient's area of the country had more or better preventive or outpatient care? As a practical matter, the default unit of analysis for the area-level AHRQ QIs is the county.
- **Hospital-level indicators** capture potentially preventable complications or adverse events following a medical condition or procedure or mortality following a medical condition or surgical procedure in which evidence suggests that high mortality may be associated with deficiencies in care. For example, hospital-level indicators may answer the question: Did the patient experience an adverse quality-related event while in the hospital? As a practical matter, the default unit of analysis for hospital-level AHRQ QIs is the hospital.

Moreover, the AHRQ QI modules capture various aspects of quality:

- **Prevention Quality Indicators (PQIs)** identify hospital admissions that might have been avoided given access to high-quality health care, preventive care, and health promoting resources within a community (first released November 2000, last updated July 2022).
- **Inpatient Quality Indicators (IQIs)** reflect quality of care inside hospitals,[3] including inpatient mortality for medical conditions and surgical procedures (first released May 2002, last updated July 2022).

[2] The hospital entity as defined by the data source may differ from the hospital entity as defined by the American Hospital Association (AHA). For example, a case where the data source treats two separate facilities as two hospitals, while the AHA Annual Survey treats the two facilities as a single hospital, or vice versa. For consistency across states, the Healthcare Cost and Utilization Project (HCUP) defines hospitals in accordance with AHA Annual Survey of Hospitals. During HCUP data processing, the data source's identification of the hospital is reconciled with the identification of the hospital in the AHA Annual Survey of Hospitals. For detailed information about this linking process, see the special report on HCUP Hospital Identifiers.

[3] Area-level IQIs and PSIs were retired in v7.0, ICD-10-CM/PCS specifications and software. As of v7.0, the IQIs and PSIs do not reflect quality of care across geographic areas.

- **Patient Safety Indicators (PSIs)** reflect quality of care inside hospitals, to focus on potentially avoidable complications and iatrogenic events (first released March 2003, last updated July 2022).
- **Pediatric Quality Indicators (PDIs)** and **Neonatal Quality Indicators (NQIs)** use indicators from the other three modules with adaptations to measure the access and quality of care for children and at-risk neonates (first released April 2006, last updated July 2022).

Table 1. Quality Domains Addressed by Area-Level and Hospital-Level Modules summarizes the quality domains addressed by each module.

Table 1. Quality Domains Addressed by Area-Level and Hospital-Level Modules

Domain	Area-Level Modules	Hospital-Level Modules
Inpatient Quality		X
Patient Safety		X
Prevention Quality	X	
Pediatric Quality – Inpatient Quality		X
Pediatric Quality – Patient Safety		X
Pediatric Quality – Prevention Quality	X	

B. AHRQ QI Results: Counts, Rates, and Scores

Most of the AHRQ QIs are ratios or rates in which the numerator is a count of hospitalizations with the condition or outcome of interest and the denominator is an estimate of the number of people (or hospitalizations) at risk for that outcome over a period of time (generally, over one year).

AHRQ QI observed rates are derived for the entire United States (U.S.) (called the reference population) and for individual areas of the country or hospitals. The observed rates may vary between areas or hospitals due to a number of factors. Some areas and hospitals provide exemplary care, while others provide sub-standard care. Some areas may serve people that are at higher risk for complications or exacerbations of their conditions, while others serve people that are at lower risk. Some hospitals may have sicker patients with more complex conditions, while others may have a lower-risk case mix.

In order to make meaningful comparisons about quality of care, the AHRQ QIs take into account underlying differences across areas or across hospitals that are unrelated to quality. The AHRQ QI technical specifications and methodology provide five different kinds of results, depending on whether comparisons are of interest for that particular indicator:

- **Volume/counts.** One indicator reports the number of times that a hospital performed a medical procedure of interest. This volume, or count, indicator does not have denominators.
- **Observed rate.** Area-level rates are the number of hospitalizations for the condition of interest divided by the number of individuals who live in that area who are at risk for the condition. In contrast, hospital-level rates are the number of hospitals stays in which the patient experienced the QI adverse event divided by the number of hospital stays for patients at risk for the event.
- **Expected rate.** A comparative rate that incorporates information about an external reference population that is not part of the user's input dataset—that is, the rate that would be ***predicted*** if

the expected level of care observed in the reference population and estimated with risk adjustment regression models were applied to the mix of patients with demographic and comorbidity distributions observed in the user's dataset. The expected rate answers the question, "What rate of adverse events would we expect to see if this area or hospital provided the average level of care observed in the reference population, but provided it to the patients with the locally observed distribution of characteristics?" (i.e., average performance from the reference population of the universe of patients applied to locally observed mix of patients with their local risk profiles). When the observed rate is smaller than the expected rate (or the observed/expected ratio is < 1), then there is reason to think that the hospital (or area) is performing better than average on this indicator given the local patient case mix. The expected rate is calculated only for risk-adjusted indicators.

- **Risk-adjusted rate.** A comparative rate that incorporates information about the observed rate, expected rate, and a reference population that is not part of the input dataset. The risk-adjusted rate is the ratio of the observed rate and expected rate multiplied by the reference population observed rate. Therefore, it answers the same question as the ratio of the observed and expected: "How does the rate of adverse events for this hospital (or area) compare to the rate we would expect to see if it provided the average level of care observed in the reference population, but provided it to the patients with the locally observed distribution of characteristics?" If the risk-adjusted rate is higher than the reference rate, the hospital (or area) is performing worse than an average hospital or area in the reference population in providing care to patients with the locally observed distribution of characteristics.

- **Smoothed rate.** The smoothed rate is a weighted average of the reference population rate and the local (hospital or area) risk-adjusted rate. If the data from the individual hospital or area include many observations and provide a numerically stable estimate of the rate, then the smoothed rate will be very close to the risk-adjusted rate, and it will not be heavily influenced by the reference population rate. Conversely, the smoothed rate will be closer to the reference population rate if the hospital or area rate is based on a small number of observations and may not be numerically stable, especially from year to year. As a weighted average of the risk-adjusted rate from the user's input dataset and the rate observed in the reference population discharges, the smoothed rate is calculated with a shrinkage estimator (1) to result in a rate near that from the user's dataset if the hospital's (or area's) rate is estimated in a stable fashion with minimal noise or (2) to result in a rate near that of the reference population if the rate from the input dataset is unstable and based on noisy data. In practice, the smoothed rate brings rates toward the reference population mean (i.e., the rate among all discharges in the reference population) and does this more so for hospitals with lower volume (smaller denominators) and outliers (such as rural hospitals). Rates for larger, high volume, hospitals will tend not to move much with smoothing, even if their risk-adjusted rates differs from the reference population rate.

- **Composite scores.** The composite QI scores combine information from multiple component QIs into a single summary index. There are two different methods used to construct composites in the AHRQ QI software. Area-level QI composites include PQIs 90, 91, 92, and 93 and PDIs 90, 91, and 92. The numerator of the composites is the sum (unweighted) of all hospital stays for the composite conditions of interest. A consistent denominator is used (e.g., population of adults age 18 years and older). In contrast, hospital-level composites (e.g., IQI 90 and 91, PSI 90) rely on a weighing scheme. They are calculated by first computing the smoothed rate for each component indicator and then computing a weighted average of the smoothed rates, where the weights are

determined empirically using methods that differ by QI composite. All weighted composites use weights based on volume (either the numerator volume or denominator volume), except PSI 90 which uses weights based on volume and harm.

C. Brief History of the AHRQ QIs

The AHRQ QIs are measures of health care quality designed for use by program managers, researchers, and others at the Federal, State, and local levels interested in health care quality measurement. The AHRQ QIs provide health care decisionmakers with tools to assess their data, highlight potential quality concerns, identify areas that need further study and investigation, and track changes over time. The modules represent various aspects of quality: prevention (PQIs), inpatient care (IQIs), patient safety (PSIs), and pediatric care (PDIs). The AHRQ QIs are used in free software distributed by AHRQ; the software programs can be applied to hospital inpatient administrative data, which is readily available and relatively inexpensive to use.

The AHRQ QIs were originally developed at the request of Healthcare Cost and Utilization Project (HCUP) Partners in 1999 using evaluation methodologies developed in the AHRQ Evidence-based Practice Centers (EPC). Over the years several refinements have been made to the original indicators by incorporating risk adjustment and a reference population to improve the reliability and validity of the indicators. The PQIs were developed in 2002 the IQIs in 2002, the PSIs in 2003, and finally the PDIs in 2006 using ICD-9 CM codes. In 2012, several other enhancements were added such as present on admission (POA) criteria, laboratory values, and other key clinical values as well as to account for the conversion of AHRQ QIs to ICD-10 CM/PCS. Additional detail about the development of each module is included below.

The AHRQ PQIs were developed in 2002 as measures of access to quality care within a community. They were based on constructs of "ambulatory care sensitive conditions" and "potentially preventable hospitalizations" that were empirically related to access measures or poverty. Between 2005 and the present day, the PQIs have been re-evaluated and refined by expert clinical panels, stakeholder and topic expert panels and through empirical analyses. As additional research informed the PQIs, the purpose of the module was expanded in collaboration with an expert panel in 2015 to include community-based factors that influence health along with access to quality care.

The AHRQ IQIs and PSIs were originally developed in 2002 and 2003, respectively, as measures of quality of clinical care at both the hospital level and across geographic areas. The indicators were developed with input from an expert panel which assessed each indicator for: face validity, precision, minimum bias (i.e., ability to risk adjust), construct validity, opportunity for quality improvement, and fit for the indicator set. Like AHRQ's other quality indicator modules, the IQIs and PSIs were originally intended for surveillance and quality improvement uses. Since their development, both IQIs and PSIs have been adopted into national reporting and payment programs. As such, both sets of measures have increasingly been used for the comparative assessment of hospital performance rather than internal quality improvement alone. To allow for fair comparisons, most measures are risk adjusted for case mix differences across hospitals and are reliability adjusted to account for differential signal strengths.

For accountability measures, the goal of risk adjustment in comparative outcome measures is to account for differences in patients across measured entities (e.g., hospitals) that affect outcome rates and that are unrelated to the quality of care. When such differences are not addressed, the observed rates will reflect both case mix and quality, and will be biased against hospitals who have patients at higher risk for the measured adverse outcome when compared to a national average hospital.

D. Overview of the Empirical Methods Document

In the remainder of this document, we describe the methods for calculation of AHRQ QI results from a user perspective (Chapter II), describe the underlying empirical development of the AHRQ QIs (Chapter III), and provide a list of the references used in the document (Chapter IV), as well as tables of the indicators (Chapter V). Please note that this document is intended to provide information on the methodology of the AHRQ QIs. There is a complementary document on the AHRQ QI website (https://qualityindicators.ahrq.gov) entitled *AHRQ QI Software Instructions* that provides an overview of the SAS software and details about data elements and SAS programs used to calculate the AHRQ QIs.

Chapter II. AHRQ QIs Modules and Methods

In this chapter, we provide a general description of each QI module and a list of indicators included in the module. We then describe the technical specifications that provide detailed information about each indicator, and the types of data and populations used to calculate QI rates. Finally, we describe the methods used to calculate the numerators, denominators, and observed, expected, risk-adjusted, and smoothed rates for the area-level and hospital-level QIs.

A. AHRQ QI Modules

A.1 Prevention Quality Indicators (PQIs)

The PQIs are a set of measures designed to capture access to quality of care and wellness (community health) of a population in a given region, by using hospital administrative data to identify rates of hospitalization for "ambulatory care sensitive conditions." These are conditions for which short and long-term access to quality care can potentially prevent hospitalization or for which early intervention can prevent complications or more severe disease. These measures are influenced by disease prevalence, environmental factors influencing physical health (poverty, housing, pollution, and food access) and health behaviors, and reflect access to care, including affordability, availability, timeliness, accessibility and understanding.

Even though these indicators are based on hospital inpatient data, they provide insight into the health of the community and the community-based health care system. For example, patients with diabetes may be hospitalized for diabetic complications if their conditions are not adequately monitored, if they do not receive the patient education needed for appropriate self-management, or if they do not have access to community resources that help promote self-management. These indicators identify hospital admissions that evidence suggests might have been avoided through access to high-quality outpatient or preventive care. The numerator is a count of admissions for the condition of interest, and the denominator is an estimate of the number of persons at risk for such a hospitalization.

The PQIs can be used as a "screening tool" to help flag potential health care access problems or concerns about population health and help public health agencies, State data organizations, health care systems, and others interested in improving health care quality in their communities to identify and investigate communities potentially in need of interventions.

Because the PQIs are calculated using readily available hospital administrative data, they are an easy-to-use and inexpensive screening tool. They can be used to provide a window into the community — to identify unmet community health care needs, to monitor the extent that complications from a number of common conditions are avoided in the community outpatient setting, and to compare performance of local health care systems across communities.

The PQI module contains a total of 14 indicators—10 primary indicators and four composite indicators (Table 2 and Appendix B.1).

Table 2. List of AHRQ Prevention Quality Indicators (PQIs)

Abbrev	Indicator Name (v2022)	Area or Hospital Level
PQI 01	Diabetes Short-Term Complications Admission Rate	Area

Abbrev	Indicator Name (v2022)	Area or Hospital Level
PQI 03	Diabetes Long-Term Complications Admission Rate	Area
PQI 05	Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate	Area
PQI 07	Hypertension Admission Rate	Area
PQI 08	Heart Failure Admission Rate	Area
PQI 11	Community-Acquired Pneumonia Admission Rate	Area
PQI 12	Urinary Tract Infection Admission Rate	Area
PQI 14	Uncontrolled Diabetes Admission Rate	Area
PQI 15	Asthma in Younger Adults Admission Rate	Area
PQI 16	Lower-Extremity Amputation among Patients with Diabetes Rate	Area
PQI 90	Prevention Quality Overall Composite	Area
PQI 91	Prevention Quality Acute Composite	Area
PQI 92	Prevention Quality Chronic Composite	Area
PQI 93	Prevention Quality Diabetes Composite	Area

A.2 Inpatient Quality Indicators (IQIs)

The IQIs are a set of measures that provide a perspective on hospital quality of care using hospital administrative data. These indicators reflect quality of care inside hospitals and include inpatient mortality for certain procedures and medical conditions and utilization of procedures for which there are questions of overuse, underuse, and misuse.

The IQIs can be used to help hospitals identify potential problem areas that may need further study. The IQIs provide the opportunity to assess quality of care inside the hospital using administrative data found in the typical discharge record, and include two primary types of indicators: (1) mortality indicators for conditions or procedures – for which mortality can vary from hospital to hospital, and (2) utilization indicators for procedures – for which utilization varies across hospitals.

The IQI module contains a total of 17 indicators—15 primary indicators and two composite indicators (Table 3 and Appendix B.2).[4] Most of the IQIs are based on surgical procedures and are reported at the hospital-level, although some are based on medical conditions.[5] The IQIs are grouped into two categories, in-hospital mortality indicators and utilization indicators:

[4] IQI 32 (Acute Myocardial Infarction [AMI] Mortality Rate, Without Transfer Cases) and IQI 34 (Vaginal Birth After Cesarean [VBAC] Rate) were retired in v2021 ICD-10-CM/PCS specifications and software.
[5] Area-level IQIs were retired in v7.0 ICD-10-CM/PCS specifications and software.

1. **In-hospital mortality indicators**. There are 12 in-hospital mortality indicators (three of which have stratum-specific specifications) and two composite indicators for *surgical procedures and medical conditions* that have been shown to have in-hospital mortality rates that vary substantially across hospitals and for which evidence suggests that high in-hospital mortality may be associated with deficiencies in the quality of care. These indicators are measured at the hospital-level. Six of these mortality indicators are for procedures. The other six mortality indicators are associated with medical conditions.

2. **Utilization indicators**. There are three utilization indicators for *surgical procedures* for which there are questions of overuse, underuse, or misuse. The usage of the procedures being examined varies significantly across hospitals, and high or low rates by themselves do not represent poor quality of care; rather, the information is intended to inform consumers about local practice patterns.

Table 3. List of AHRQ Inpatient Quality Indicators (IQIs)

Abbrev	Indicator Name (v2022)	Procedure or Condition	Area or Hospital Level
Mortality Indicators			
IQI 08	Esophageal Resection Mortality Rate	Procedure	Hospital
IQI 09[a]	Pancreatic Resection Mortality Rate	Procedure	Hospital
IQI 11[a]	Abdominal Aortic Aneurysm (AAA) Repair Mortality Rate	Procedure	Hospital
IQI 12	Coronary Artery Bypass Graft (CABG) Mortality Rate	Procedure	Hospital
IQI 15	Acute Myocardial Infarction (AMI) Mortality Rate	Condition	Hospital
IQI 16	Heart Failure Mortality Rate	Condition	Hospital
IQI 17[a]	Acute Stroke Mortality Rate	Condition	Hospital
IQI 18	Gastrointestinal Hemorrhage Mortality Rate	Condition	Hospital
IQI 19	Hip Fracture Mortality Rate	Condition	Hospital
IQI 20	Pneumonia Mortality Rate	Condition	Hospital
IQI 30	Percutaneous Coronary Intervention (PCI) Mortality Rate	Procedure	Hospital
IQI 31	Carotid Endarterectomy Mortality Rate	Procedure	Hospital
IQI 90	Mortality for Selected Inpatient Procedures	Procedure	Hospital
IQI 91	Mortality for Selected Inpatient Conditions	Condition	Hospital
Utilization Indicators			
IQI 21	Cesarean Delivery Rate, Uncomplicated	Procedure	Hospital

Abbrev	Indicator Name (v2022)	Procedure or Condition	Area or Hospital Level
IQI 22	Vaginal Birth After Cesarean (VBAC) Delivery Rate, Uncomplicated	Procedure	Hospital
IQI 33	Primary Cesarean Delivery Rate, Uncomplicated	Procedure	Hospital

[a] Includes stratum-specific indicators.

A.3 Patient Safety Indicators (PSIs)

The PSIs are a set of indicators providing information on safety-related adverse events occurring in hospitals following operations, procedures, and childbirth. The PSIs use administrative data in the typical hospitalization discharge record to identify potential in-hospital complications. They can be used to help hospitals identify adverse events worthy of further study and to assess the incidence of such events for comparative purposes.[6]

The PSI module contains a total of 18 indicators—17 primary indicators and one composite indicator (Table 4 and Appendix B.3).

There are 17 hospital-level PSIs for medical conditions and surgical procedures that have been shown to have complication/adverse event rates that vary substantially across hospitals and for which evidence suggests that high complication/adverse event rates may be associated with deficiencies in the quality of care. These indicators are measured as rates: the number of complications/adverse events divided by the number of discharges with the associated procedure or condition. The hospital-level indicators include only those cases where a secondary diagnosis code flags a potentially preventable complication. Eight of these indicators are for surgical discharges, six are for either medical or surgical discharges, and three are for obstetric discharges. In addition, there is one hospital-level composite that summarizes 10 different patient safety events.

Table 4. List of AHRQ Patient Safety Indicators (PSIs)

Abbrev	Indicator Name (v2022)	Area or Hospital Level
PSI 02	Death Rate in Low-Mortality Diagnosis Related Groups (DRGs)	Hospital
PSI 03	Pressure Ulcer Rate	Hospital
PSI 04[a]	Death Rate among Surgical Inpatients with Serious Treatable Complications	Hospital
PSI 05	Retained Surgical Item or Unretrieved Device Fragment Count	Hospital
PSI 06	Iatrogenic Pneumothorax Rate	Hospital
PSI 07	Central Venous Catheter-Related Blood Stream Infection Rate	Hospital

[6] Area-level PSIs were retired in v7.0 ICD-10-CM/PCS specifications and software (https://qualityindicators.ahrq.gov/News/Retirement%20Notice_v2019_Indicators.pdf).

Abbrev	Indicator Name (v2022)	Area or Hospital Level
PSI 08	In Hospital Fall with Hip Fracture Rate[b]	Hospital
PSI 09	Postoperative Hemorrhage or Hematoma Rate	Hospital
PSI 10	Postoperative Acute Kidney Injury Requiring Dialysis Rate[c]	Hospital
PSI 11	Postoperative Respiratory Failure Rate	Hospital
PSI 12	Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate	Hospital
PSI 13	Postoperative Sepsis Rate	Hospital
PSI 14	Postoperative Wound Dehiscence Rate	Hospital
PSI 15	Abdominopelvic Accidental Puncture or Laceration Rate[d,e]	Hospital
PSI 17	Birth Trauma Rate – Injury to Neonate[f]	Hospital
PSI 18	Obstetric Trauma Rate – Vaginal Delivery with Instrument	Hospital
PSI 19	Obstetric Trauma Rate – Vaginal Delivery without Instrument	Hospital
PSI 90	Patient Safety and Adverse Events Composite[g]	Hospital

[a] Includes stratum-specific indicators
[b] Previously entitled "Postoperative Hip Fracture" prior to v6.0
[c] Previously entitled "Postoperative Physiologic and Metabolic Derangement" prior to v5.0
[d] Previously entitled "Accidental Puncture or Laceration Rate" prior to v6.0.
[e] Previously entitled "Unrecognized Abdominopelvic Accidental Puncture or Laceration Rate"
[f] Calculated in the PDI software module
[g] Previously entitled "Patient Safety of Selected Indicators" prior to v6.0

A.4 Pediatric Quality Indicators (PDIs)

The PDIs are a set of measures that can be used with hospital inpatient discharge data to provide a perspective on the quality of pediatric healthcare and the health of the pediatric population. There are two types of PDIs. The seven area-level PDIs (four primary indicators and three composites) use hospital administrative data to identify rates of hospitalization for "ambulatory care sensitive conditions" within a given region. They are designed to capture a population's overall wellness (community health) and access to quality health care. The seven hospital-level PDIs screen for problems that occur while a patient is hospitalized, and that patients experience as a result of exposure to the healthcare system. These events may be preventable by changes in the system or hospital.

The PDIs are expressly for children under the age of 18. These indicators take into account four factors—differential epidemiology of child healthcare relative to adult healthcare, dependency, demographics, and development—that relate to all aspects of children's healthcare. Neonatal Quality Indicator (NQI) 03 is a PDI calculated for neonates.

Table 5 (and Appendix Tables B.1 and B.2) list all of the PDIs and indicates whether they are measured at the area or the hospital level. The PDI module contains a total of 14 indicators- 11 primary indicators and 3 composite indicators. Of the primary indicators, 7 are hospital-level indicators and 4 are area-level indicators of quality of pediatric care. The 3 composite indicators are area-level indicators.

Table 5. List of AHRQ Pediatric Quality Indicators (PDIs)

Abbrev	Indicator Name (v2022)	Area or Hospital Level
NQI 03	Neonatal Blood Stream Infection Rate	Hospital
PDI 01	Accidental Puncture or Laceration Rate	Hospital
PDI 05	Iatrogenic Pneumothorax Rate	Hospital
PDI 08	Postoperative Hemorrhage or Hematoma Rate	Hospital
PDI 09	Postoperative Respiratory Failure Rate	Hospital
PDI 10	Postoperative Sepsis Rate	Hospital
PDI 12	Central Venous Catheter-Related Blood Stream Infection Rate	Hospital
PDI 14	Asthma Admission Rate	Area
PDI 15	Diabetes Short-Term Complications Admission Rate	Area
PDI 16	Gastroenteritis Admission Rate	Area
PDI 18	Urinary Tract Infection Admission Rate	Area
PDI 90	Pediatric Quality Overall Composite	Area
PDI 91	Pediatric Quality Acute Composite	Area
PDI 92	Pediatric Quality Chronic Composite	Area

B. Specifications

Technical specifications for each of the indicators are posted on the AHRQ QI website (https:// qualityindicators.ahrq.gov). The specifications provide a written description of the measure, numerator, numerator exclusions, denominator, and denominator exclusions. Specifications are based on information found in a typical discharge abstract, billing record or inpatient claim, including age, sex, ICD-10-CM/PCS diagnosis and procedure codes, the Medicare-Severity-Diagnostic Related Group (MS-DRG) and Major Diagnostic Category (MDC) appropriate for the date of discharge, day of procedures, length of stay, source of admission / point of origin, type of admission, and discharge disposition.

Given that not all data claims include MS-DRGs and MDCs, users must derive these from information on the billing record (see section D.4 for more details). Expected values generally align with the Uniform Bill (UB-04) classification scheme. In addition to the written description of the measure, the technical specification documents provide the specific ICD-10-CM/PCS for each clinical construct. The specifications are operationalized in two different software platforms. Both software platforms and associated documentation are freely available on the AHRQ QI website:

- SAS: https:// qualityindicators.ahrq.gov/Software/SAS.aspx
- WinQI: https:// qualityindicators.ahrq.gov/Software/winQI.aspx

C. Data

The AHRQ QIs are specified for use with hospital discharge abstracts, billing records or claims data (administrative data consistent with the UB–04 format). The AHRQ QIs are intended to be calculated on an entire patient population (e.g., all discharges from a hospital in a given time period).[7]

User data must contain information about basic patient demographics (e.g., age, sex), ICD-10-CM/PCS coded clinical diagnoses and procedures, and information about the hospital stay (e.g., length of stay, type of admission, where the stay originated, discharge disposition, discharge quarter). See the Software Instructions Guide for a detailed list of each of the data elements, including the name, a complete description, format, and values, used in the AHRQ QI specifications.

D. Patient Population

D.1 Identification of Adult and Pediatric Discharges

Discharge records in the dataset are analyzed as either adult or pediatric on the basis of age and MDC (Table 6. Analysis Data Inclusion Rule). Discharges in MDC 14 (Pregnancy, Childbirth & the Puerperium) are analyzed as adult regardless of age. Despite pediatric age exclusions built into PSI and IQI indicators, some observations with ages greater than 17 have MDC 15 (Newborns & Other Neonates with Conditions Originating in Perinatal Period). Because these observations could be newborn patients, starting with v2021, the software excludes observations with MDC = 15 from adult IQI and PSI modules.

Table 6. Analysis Data Inclusion Rule

Analysis Data	Inclusion Rule
Adult	(AGE≥18 years or MDC=14) and MDC ≠ 15
Pediatric	AGE<18 years and MDC≠14

With a few exceptions, discharges for adults are used to calculate PQIs, IQIs, and PSIs. Discharges for children and adolescents are used to calculate PDIs, and discharges for neonates are used to calculate Neonatal Blood Stream Infection Rate (NQI 03) and Birth Trauma Rate – Injury to Neonate (PSI 17).

Table 7 shows a summary of the indicators by age group. See Appendix B for a detailed list of all indicators and the patient population of interest.

Table 7. Age Groups and Indicators

Population	Age / Major Diagnostic Category (MDC)	Indicators
Adult	18+ Years	PQI 01, PQI 03, PQI 07, PQI 11–12, PQI 14, PQI 16, PQI 90–93, IQI 08–09, IQI 11–12, IQI 15–18, IQI 20, IQI 31, IQI 90–91 PSI 06, PSI 08–15, PSI 90

[7] The AHRQ QIs are created using one calendar year of data.

Population	Age / Major Diagnostic Category (MDC)	Indicators
	18+ Years or Obstetric	IQI 21–22, IQI 33 PSI 02, PSI 05, PSI 07
	18 to 39 Years	PQI 15
	18 to 89 Years or Obstetric	PSI 04
	40+ Years	PQI 05 IQI 12, IQI 30
	65+ Years	IQI 19
	Vaginal delivery (no age parameters)	PSI 18, PSI 19
Pediatric	Neonates / Newborns	PSI 17 NQI 03
	0 to 17 Years	PDI 01, PDI 05, PDI 10, PDI 12
	3 months to 17 Years	PDI 16, PDI 18
	2 to 17 Years	PDI 14
	6 to 17 Years	PDI 15, PDI 90–92

D.2 Identification of Patient Residing in Area of Interest

A fundamental component of the AHRQ QI area-level indicators (PQIs and some PDIs) is the area of residence of the patient, usually specified by the Federal Information Processing Standards (FIPS) county and state codes (but that could also be determined by ZIP Codes). The area of patient residence determines the catchment area of the numerator (the number of all indicator-specific hospital stays defined by that area) and the denominator (the corresponding U.S. Census population estimate for the area). Patients who do not reside in the area of interest are not included in the calculation of that area's rates.

D.3 Identification of Present on Admission (POA)

A fundamental component of the AHRQ IQI, PSI, and PDI specifications v5.0 and beyond is whether a patient has a clinical condition or complication which is present upon their admission to the hospital. The presence of a clinical condition or complication is used to determine if a discharge should be included as a numerator event or to ensure the accurate identification of comorbidities. If POA information is not available, all clinical conditions on a discharge record, except the principal diagnosis, are considered to have occurred in the hospital, and not present at the time of admission to the hospital.

POA was added to the UB-04 effective October 1, 2007, and hospitals incurred a payment penalty for not including POA on Centers for Medicare and Medicaid Services (CMS) Medicare FFS records beginning October 1, 2008. Each diagnosis on a discharge record must indicate whether the condition was "present at

the time the order for inpatient admission occurs" according to the ICD-10-CM/PCS Coding Guidelines. Additional information about the coding guidelines for POA can be found at:

https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf.

Table 8 lists the possible values of the POA data elements (Y, N, U, W, 1, or missing) along with whether the AHRQ QIs treat the clinical condition or complication as present at the time of admission. The principal diagnosis is always assumed to be POA by definition, regardless of the coding of the POA data element in the principal field. Secondary diagnosis codes first are checked to see whether the diagnosis is exempt from reporting POA. If the secondary diagnosis is exempt, it is considered POA.[8] If the secondary diagnosis is not exempt, then it considered POA if the POA data element is coded with a Y or W. Secondary diagnosis codes are considered not POA if the POA data element is coded with an N, a U, a blank, or a 1.[9]

Table 8. Values for the Present-on-Admission Data Element

ICD-10-CM/PCS Guidelines	Description	Present at Time of Admission
Y – Yes	Diagnosis is present at the time of inpatient admission	Yes
N – No	Diagnosis is not present at the time of inpatient admission	No
U – Unknown	Documentation is insufficient to determine whether condition is present on admission	No
W – Clinically undetermined	Hospital is unable to clinically determine whether condition is present on admission	Yes
1 – Unreported/not used; also includes UB-04 values previously coded as 1	Reported as exempt from reporting on a nonexempt diagnosis	No

Source: Centers for Medicare & Medicaid Services. https://www.cms.gov/medicare/medicare-fee-for-service-payment/hospitalacqcond/coding.html.

D.4 Identification of Major Diagnostic Category (MDC)

Another fundamental component of the AHRQ QI specifications is the MS-DRG and MDC to which a discharge is assigned.

MS-DRGs and MDC are derived from the CMS MS-DRG grouper algorithm, which assigns the MDC based on the principal diagnosis.[10] Different versions of the MS-DRG grouper produce slightly different results with respect to certain high resource intensity MS-DRGs. Specifically, MS-DRGs 001-019 and 981-989 are classified as "pre-MDC" MS-DRGs, which means that they are associated with such high length of stay and/or cost that they supersede the usual assignment of MS-DRGs within body system or

[8] Centers for Medicare & Medicaid Services. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Downloads/2020-POA-Exempt-Codes.zip.
[9] Data before January 1, 2011 do not include POA indicator information, and contain an 'X' in this field, indicating that POA indicator information is not available.
[10] Centers for Medicare & Medicaid Services. https://www.cms.gov/files/zip/icd-10-ms-drg-definitions-manual-files-v381-text.zip

MDC categories. For records assigned to these MS-DRGs, some versions of the grouper software retain the MDC that would be assigned based on the principal diagnosis and procedure codes, whereas other versions of the grouper software overwrite the MDC assignment with a blank, missing, or nonnumeric value such as "PRE." Pre-MDC assignments together with "invalid" or "ungroupable" MS-DRGs, MS_DRGs 998 and 999 are not considered in the AHRQ QI specifications.

E. Area-Level Quality Indicators

E.1 Overview of Area-Level Indicators

Area-level indicators capture cases of potentially preventable hospital stays or complications that occur in the population in a given geographic area. The AHRQ QI software and reference population calculate the PQIs and area-level PDIs for areas. Area-level rates are constructed using denominators that capture the size of the area's population using census (or user supplied) data.[11]

Area-level indicators contained in the PQI module identify hospital admissions that evidence suggests might have been avoided through access to high-quality community care and resources. The area-level indicators contained in the PDI module are adapted from indicators from the other modules.[12]

Area-level indicators have numerators, denominators and observed rates. In addition, some area-level indicators have expected rates, risk-adjusted rates and smoothed rates.

E.2. Numerator, Denominator, and Observed Rates for Area-Level Indicators

E.2.1 Numerator and Numerator Exclusions

Numerators are based on the condition or procedure of interest. The specifications often stipulate that cases should be excluded from the numerator for one of the following reasons:

1. The outcome of interest is very difficult to prevent or has an unclear conceptual relationship to access to quality care or community resources.
2. The patient was transferred from another health care facility (to avoid double counting a single encounter).
3. Encounters are missing data elements that are required for indicator construction.
4. Obstetric cases are excluded from some measures by default because discharges with a principal diagnosis relevant to those measures exclude obstetric discharges.

---

[11] Previous versions of area-level indicators included two types of condition-specific denominators. First, some indicators allowed the denominator to be specified with the diabetic population only and calculated with the SAS QI (but not in the WinQI) software through the condition-specific denominator at the state-level feature. However, the disease-specific denominator file has been temporarily removed from the software beginning with v2021 for further review and refinement. Second, three area-level indicators (Perforated Appendix Admission Rate [PQI 02 and PDI 17] and Low Birth Weight [PQI 09]) had discharge-based condition-specific denominators, meaning that the denominator was the count of discharges for a specific condition among patients residing in an area. These three measures were retired in v2019 specifications and software.

[12] Area-level IQIs and PSIs were retired in v7.0 ICD-10-CM/PCS specifications and software. As of v7.0 ICD-10-CM/PCS, none of the IQIs or PSIs reflect quality of care across geographic areas.

In addition, cases are excluded from the numerator if the patient resides in a state that did not contribute to the HCUP State Inpatient Databases (SID).

E.2.2 Denominator

The denominator is based on the census population estimate for the patient's geographic area of residence. Note that the age- and sex-specific population denominator estimates correspond to the age and sex criteria of the numerator (e.g., adult population for adult indicators, adult female population for female-specific indicators, pediatric population for pediatric indicators). Geographic area is defined at the county level, specifically the FIPS county codes.

For information about how the denominators are calculated from census data, see Chapter III.C and the QI Population Documentation File at:
https://qualityindicators.ahrq.gov/Downloads/Software/SAS/V2022/AHRQ_QI_v2022_ICD10_Population_File.pdf

E.2.3 Observed Rate

The observed rate of an area-level indicator is the number of persons with the condition or procedure of interest divided by the number of persons in the geographic area of interest. Note that the age and sex-specific population denominator estimates correspond to the age and sex criteria of the numerator. As noted above, the denominator is a population estimate from a U.S. Census Bureau dataset.

Older versions of the AHRQ QI software allowed users to calculate quarterly observed rates. However, quarterly rates needed to be interpreted with caution, given seasonal variation for many conditions and the potential decrease in reliability associated with reduced numerator counts. Since v2019, the AHRQ QI software does not include quarterly calculations.

E.3. Comparing Indicators Across Geographic Areas

E.3.1 Overview of Expected, Risk-Adjusted, and Smoothed Rates for Area-Level Indicators

In order to make meaningful comparisons of the area-level rate for one area with a national average area, it is helpful to account statistically for population characteristics such as age, sex, poverty level in that area. For most QIs, risk-adjusted rates calculated by indirect standardization are used. In statistical language, the risk adjustment control for demographic differences via regression analyses (area-level indicators use logistic regression). This chapter discusses the risk factors that are used with the area-level indicators. All area-level indicators are risk adjusted for demographics. None of the area-level indicators are risk adjusted for clinical factors.

Three sets of QI rates are calculated for risk-adjusted area-level indicators: expected or predicted rates, risk-adjusted rates, and smoothed rates.

Expected and risk-adjusted rates both acknowledge that geographic areas are unique and differ in two important ways from the representative profile observed in the reference population. First, there is heterogeneity in the care that is available, in the community resources, or in exposures from the environment. Second, most areas differ in the demographic composition of their residents. The expected rate is that which would prevail if heterogeneity from sources other than demographics were removed, but local demographic characteristics were allowed to vary. The risk-adjusted rate then uses the difference between the rate observed in a given area and that expected rate to project the rate that would result in the reference population if local differences other than demographic prevailed.

The *expected rate* answers the question, "What rate of admissions would we expect to see if this geographic area provided the average access to care observed in the reference population, but provided it to patients with the locally observed distribution of characteristics?" (i.e., average performance from the reference population of the universe of patients applied to locally observed mix of residents). When the observed rate is smaller than the expected rate (or the observed / expected ratio is < 1), then there is reason to think that the geographic area is performing better than average on this indicator.

The *risk-adjusted rate* is the product of the ratio of the observed and expected rate and the reference population rate. Risk adjustment permits the rate for a given geographic area to be compared with the rate for the reference population. The risk-adjusted rate answers the question, "What rate of admissions is expected if the standard of care applied to local residents were applied to the reference population?" (i.e., locally observed performance on a representative mix of patients from the reference population). If the risk-adjusted rate is higher than the reference rate (or if observed rates are higher than expected rates), it means that the admission rate for a given geographical area is worse than expected based on the experience of patients in the reference population with a similar distribution of characteristics.

The *smoothed rate* is a weighted average of the reference population rate and the locally observed geographic area rate. If the data from the individual geographic area include many observations and provide a numerically stable estimate of the rate, then the smoothed rate will be very close to the risk-adjusted rate, and it will not be heavily influenced by the reference population rate. Conversely, the smoothed rate will be closer to the reference population rate if the geographic area rate is based on a small number of observations and may not be numerically stable, especially from year to year.

E.3.2 Risk Factors for Risk Adjustment for Area-Level Indicators (v2022)

For area rates, the risk adjustment models adjust for age-group proportions by sex. The models include age groups (in 5-year increments) for each sex. The PQI module contains an option to incorporate a poverty variable, defined as the percent of the population under the federal poverty line for each area. County-level poverty data is obtained from the U.S. Census Small Area Income and Poverty Estimates.[13] In v2022 only coefficients based on 2019 poverty data are included and are applied to all years of user data. All U.S. counties are assigned to a poverty decile (POVCAT) based on these data. The poverty deciles then are used as risk factors in the risk adjustment model. For all area-level indicators, the risk factors used in risk adjustment are age, sex, and poverty deciles. See Appendix C for a list of risk factors by module.

E.3.3 Expected or Predicted Rate for Area-Level Indicators

The *expected or predicted rate* for an area-level QI is the rate that would be observed if the amount and quality of outpatient and preventive care available across the general population were available to individuals living in specific geographic areas. Expected rates are predicted for each area using risk adjustment model coefficients that summarize the age and sex distribution of the area's population and optionally, the poverty decile within which the area's poverty rate falls.

An expected (or predicted) rate for each QI is derived for each area of interest in the dataset. The risk adjustment for an area's expected rate is calculated using parameter estimates that were previously estimated using the entire reference (general) population for each QI (see Appendix A for addition QI-

[13] 2019 U.S. Census Bureau Small Area Income and Poverty Estimates, https://www.census.gov/programs-surveys/saipe/data/datasets.html.

related documentation, including parameter estimates tables). Because each area in the user's sample has a distinct sex and age distribution, the expected rates at the area level may vary from the reference (general or standard) population's expected rate for each QI. We define the observed (O_m) and expected rates (E_m) of area m by, respectively,

$$O_m = \frac{1}{n_m} \sum_{i \in A_m} Y_i$$

$$E_m = \frac{1}{n_m} \sum_{i \in A_m} \hat{Y}_i$$

Here, A_m is the collection of persons in the population at risk; n_m is the population size; Y_i is the outcome for person i; and ; $\hat{Y}_i$ is the person-level expected or predicted probability for person i.

E.3.4 Risk-Adjusted Rate for Area-Level Indicators

A risk-adjusted rate is derived for each QI for each area of interest. The risk adjustment for each area is calculated using the embedded reference (general or standard) population risk-adjusted rate and the area-specific observed rate and expected rate for each QI. The risk-adjusted rate, using an indirect standardization approach, equals the reference (general or standard) population rate (α) multiplied by the ratio of observed rate in the user's sample to expected rate in the user's sample:

$$RAR_m = \alpha \cdot \frac{O_m}{E_m}$$

Because each area in the user's sample has a distinct observed rate and a distinct expected rate for each QI, each area will have a distinct risk-adjusted rate that may vary from the reference (general or standard) population rate for each QI.

We used logistic regression models to build risk adjustment models for QIs that need risk adjustment. For complicated risk adjustment models, the national HCUP reference population observed rate may not be exactly same as the average of predicted event rates. In the modeling process, we assessed model calibration properties, but the O-E ratio (observed rate to expected rate ratio) may not be exactly equal to 1. In software development (not part of the publicly released software), we multiplied the predicted rate for each person by this constant (O-E ratio) to make sure the new predicted rates are perfectly calibrated to the observed rates. To be consistent, we included the national O-E ratio that was calculated based on our reference population in the AHRQ software v2022. We also provided users the options of calibrating to the reference population or to users' populations.

1. Reference population based O-E ratio is recommended in most situations and it is also the default choice in the software.
2. Users' own population based O-E ratio option is kept in the software for users who want to calibrate the predicted rates to users' population.

When area rates are compared to reference population rates, differences may be observed for several reasons. Some of the most important reasons may be related to the availability of quality preventive and outpatient care, and other reasons may contribute as well, but after risk adjustment, the differences should <u>not</u> be attributable to differences in the age and sex profiles in the areas.

E.3.5 Risk-Adjusted Rate Variance for Area-Level Indicators

The standard error of the risk-adjusted rate for each area is calculated using a method recommended by Iezzoni[14] and described by Hosmer and Lemeshow[15] that represents the amount of within-area variance due to sampling (i.e., as the number of patients per area increases, this variance tends to zero).

Using a Taylor expansion or "delta method" for the variance of the ratio of two stochastic variables, we compute the variance of the risk-adjusted rate:

$$\mathrm{Var}(RAR_m) \cong \alpha^2 \frac{\mathrm{E}(O_m)^2}{E_m^2}\left(\frac{\mathrm{Var}(O_m)}{\mathrm{E}(O_m)^2} - 2\frac{\mathrm{Cov}(O_m, E_m)}{\mathrm{E}(O_m)\cdot E_m} + \frac{\mathrm{Var}(E_m)}{E_m^2}\right)$$

It is common practice in these calculations to neglect the variance of the predictor E_m and to consider a normal distribution for the risk-adjusted rate (only true in the limit $n_m \to \infty$).[16] In this case, the above formula simplifies to:

$$\mathrm{Var}(RAR_m) \cong \alpha^2 \frac{\mathrm{Var}(O_m)}{E_m^2}$$

and the 95% confidence intervals are calculated assuming normality.

E.3.6 Smoothed Rates for Area-Level Indicators

For each area in the dataset, a smoothed rate can be calculated for each QI. The smoothed rate for each area is calculated using the pre-determined signal variance[17] estimated from the reference (general) population and the pre-determined area-specific noise variance and risk-adjusted rate.[18] Because each area in the user's sample has a distinct noise variance and a distinct risk-adjusted rate for each QI, each area will have a distinct smoothed rate that may vary from the reference (general) population smoothed rate for each QI.

Specifically, each area's *smoothed rate* is a weighted average of the risk-adjusted rate and the reference (general) population rate; the smoothed rate is calculated with an empirical Bayes shrinkage estimator (i.e., shrinkage weight) (1) to result in a rate that will be near that from the input dataset if the area's rate is estimated in a stable fashion with minimal noise or (2) to result in a rate near that of the reference (general) population if the rate from the area is unstable and based on noisy data. Thus, the smoothed rate for an area with stable estimates will be similar to the area's risk-adjusted rate, whereas the smoothed rate for an area with unstable estimates will be similar to the reference (general) population rate.

The accent "~" is used to denote the reliability adjustment. The formula for the smoothed rate is as follows:

$$\widetilde{RAR}_m = \lambda_m \cdot RAR_m + (1 - \lambda_m) \cdot \alpha$$

---

[14] Iezzoni L, Ed. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago: Health Administration Press; 2013.
[15] Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance model based on logistic regression. Statistics in Med. 1995;14(19):2161-72.
[16] Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance model based on logistic regression. Statistics in Med. 1995;14(19):2161-72.
[17] The pre-determined values are embedded in the software.
[18] The smoothing factors are included in the software for v2022.

where the reliability weight λ_m for area m is a function of the population signal variance τ^2 and area-level noise variance σ_m^2. Specifically, the reliability weight is the ratio of the signal variance (i.e., true variation in area quality reflected by the risk-adjusted rates) to the total variance, which includes sampling error:

$$\lambda_m = \frac{\tau^2}{\tau^2 + \sigma_m^2}$$

The noise variance is an estimate of variability in the QI outcome within the area (county) of interest, and the signal variance is an estimate of variability in the QI outcome across all areas of interest.

$$Noise\ Variance\ \ \hat{\sigma}_m^2 = \left(\frac{\alpha}{n_m E_m}\right)^2 \sum_{i \epsilon A_m} \hat{Y}_i(1 - \hat{Y}_i)$$

$$Signal\ Variance\ \ \hat{\tau}^2 = \frac{\sum_{m=1}^{M} \frac{1}{(\hat{\tau}^2 + \sigma_m^2)^2} \left\{ \frac{M}{M-1} (RAR_m - \overline{RAR})^2 - \hat{\sigma}_m^2 \right\}}{\sum_{m=1}^{M} \frac{1}{(\hat{\tau}^2 + \sigma_m^2)^2}}$$

Here, M is the number of areas with persons at risk for the measure, α is the observed rate for the reference population; $\hat{Y}_i$ is the person-level expected or predicted probability for person i; and for area m, A_m is the collection of persons in the population at risk, n_m is the population size, E_m is the expected rate, RAR_m is the risk-adjusted rate, and $\overline{RAR}$ is the weighted[19] average of hospital risk-adjusted rates. Note that $\hat{\tau}^2$ appears on both sides of the signal variance equation; it is estimated in an iterative fashion.[20]

E.3.7 Smoothed Rate Variance for Area-Level Indicators

The smoothed rate is an empirical Bayes posterior estimate of the hospital's risk-adjusted rate—that is, it is calculated from the reliability-weighted combination of the risk-adjusted rate and reference population rate. The variance of the smoothed rate is given by:

$$\mathrm{Var}\left(\widetilde{RAR}_m\right) = \tau^2(1 - \lambda_m)$$

E.4. Composite Rates for Area-Level Indicators

The area-level composite QI are unweighted combinations of conceptually related component QIs. The area-level QI composites are created by grouping records together using a logical "OR" operation to assign them to a composite numerator when they appear in any of the relevant component numerators. For example, the numerator for PQI 93 includes all records that qualify for any diabetes-related PQI (PQI 01, PQI 03, PQI 14, or PQI 16). Observed, risk-adjusted, and smoothed rates and their variances for the area-level composites are then computed using the same methods described for the individual component area-level QI.

E.5 Interpretation of Rates for Area-Level Indicators

The area-level QIs reflect the healthcare system, not hospital care, and may be used as "screening tools" to identify problems with ambulatory care access or quality of care provided across the system or community health. These QI serve as a trigger for more in-depth investigation in order to explain

[19] The weights are $\frac{1}{(\hat{\tau}^2+\sigma_m^2)^2}$.

[20] Morris, CN. Parametric empirical Bayes inference: theory and applications. J Am Statistical Assoc. 1983 Mar;78(381):47-55.

disparities in avoidable hospitalization rates for ambulatory care sensitive conditions, patient safety events or procedure utilization. Such information can help public health agencies, State data organizations, health care systems, and others interested in improving health in their communities to target populations for interventions, form policy or evaluate impact of interventions and policy. Although many factors can influence area-level QI rates, the indicators provide a good starting point for assessing access to quality health services or health promoting resources in the community and the health of individuals residing in the community.

The observed, risk-adjusted and smoothed rates for area-level indicators are scaled to the rate per 100,000 population. AHRQ assesses reliability of the area-level QI rates among areas. Rates for areas with very small populations are often less reliable; smoothed rates will account for the low reliability. AHRQ recommends using smoothed rates for all comparisons.

Overall, the signal-to-noise estimates based on a national, all-payer population for the PQI measures are 0.71 to 0.96. For this population, most indicators are stable for all but the smallest areas (under 2,000-3,000 adults). However, reliability estimates are not only a function of size and also depend on other factors such as the risk-adjusted rates, noise variance, prior distribution assumptions. As such, AHRQ does not calculate a "minimum population size" for the area level measures.

F. Hospital-Level Quality Indicators

F.1 Overview of Hospital-Level Indicators

The AHRQ hospital-level indicators include in-hospital mortality indicators, utilization indicators, and adverse-event indicators. These hospital-level indicators are part of the IQI, PSI, and PDI modules.

- **Hospital-level indicators** address questions such as: Did the patient have an inpatient procedure for which there are questions of overuse, underuse, or misuse? Did the patient experience an adverse quality-related event while in the care of a specific healthcare provider?
- **In-hospital mortality indicators** are for medical conditions and surgical procedures that have been shown to have mortality rates that vary substantially across institutions and for which evidence suggests that high mortality may be associated with deficiencies in the quality of care.
- **Utilization indicators** track procedures in which there are questions of overuse, underuse, or misuse. The usage of the procedures being examined varies significantly across hospitals and areas, and high or low rates by themselves do not represent poor quality of care; rather, the information is intended to inform consumers about local practice patterns.
- **Adverse-event indicators** are for medical conditions and procedures that have been shown to have complication/adverse event rates that vary substantially across institutions and for which evidence suggests that high rates may be associated with deficiencies in the quality of care. Adverse-event indicators usually include only those cases in which a secondary diagnosis code flags a potentially preventable complication. A few indicators are based on procedure codes that imply a potential preventable adverse event.

All hospital-level indicators have numerators, denominators, and observed rates. In addition, most hospital-level indicators are measured as rates—the number of hospitalizations with the outcome (mortality, adverse event) of interest divided by the population at risk for the outcome (or procedure). Hospital-level indicators are more complicated than area-level indicators because they have *indicator-specific denominators* to identify only the hospitalizations that were at risk for the outcome of interest,

and use a customized list of regression covariates that are selected when the QI software is updated annually using methods described in Chapter III.

F.2 Special Cases: Operationalizing Hospital-Level Numerators and Denominators

Some of the complexity of the hospital-level indicators is evident in the operationalization of the numerator and denominator specifications, including present-on-admission status, distinction between comorbidities and complications, and indicator-specific comorbid risk factors embedded in the numerator and denominator definitions.

F.2.1 Importance of Present on Admission (POA): Complications vs Comorbidities

As noted in Chapter II.D.3, POA is an important element in the AHRQ QI specifications. POA indicates whether a diagnosis is present at the time of admission (comorbidity) or arose during a hospitalization (complication).

For the hospital-level AHRQ QIs, an indicator-specific complication is counted in the numerator, while the indicator-specific comorbid condition is excluded from the calculation of the hospital-level AHRQ QI. Some of the indicators identify adverse conditions that develop as medical complications during the hospitalization of interest. Evidence suggests that high rates may be associated with lower quality of care. For example, PSI 03 measures the rate of pressure ulcers. However, some of these complications may have been POA, which would not be related to the quality of inpatient care.

The hospital-level PSIs and the hospital-level PDIs use POA to define the numerator event (implemented as denominator exclusion) and identify comorbidities for risk adjustment. POA is also incorporated into the AHRQ Clinical Classifications System Refined (CCSRs) for Diagnoses used to risk adjust the hospital-level IQI rates. See Appendix B for the complete list of POA dependent indicators.

F.2.2 Importance of Major Diagnostic Category (MDC)

The hospital-level AHRQ QI specifications rely heavily on MDC. MDCs are used in two ways: (1) to capture or exclude obstetric cases in the denominator, and (2) to exclude broad categories of clinical conditions which may raise the likelihood that a numerator event is not preventable. The MDC is also used in risk models to adjust for broad categories of clinical conditions in addition to the more focused MS-DRG covariates.[21]

F.3 Numerators, Denominators and Observed Rates for Hospital-Level Indicators

F.3.1 Numerator

Numerators are based on the outcome of interest (mortality or adverse event).

F.3.2 Denominator and Denominator Exclusions

The denominator is defined to include patients at risk for the numerator event. Patients may be excluded from the denominator based on being at very low risk of having numerator event (e.g., normal newborns), being at high risk for a non-preventable event or having an event or underlying clinical precedents present on admission.

[21] ICD-10-CM/PCS MS-DRG v38.1, list of MS-DRGs, available at: https://www.cms.gov/icd10m/version38-1-fullcode-cms/fullcode_cms/P0001.html

Three primary strategies are used to account for variations in case mix between hospitals. More than one approach may be employed for a single indicator. The strategies include:

1. Inclusion and exclusion criteria that limit the denominator to clinically homogeneous populations.
2. Stratification of observed and risk-adjusted rates by important clinical risk factors or procedure types (IQI 09, IQI 11, IQI 17, PSI 04, PSI 14).
3. Risk adjustment of rates to account for case mix. Note that for stratified measures, risk-adjusted rates are available for each stratum and for the overall rate. More detail on risk adjustment can be found later in this chapter in Section F.4.

General Description

The denominator of the hospital-level indicators is typically defined as a medical and/or surgical discharge, or by a specific surgical procedure. Medical and surgical discharge types are defined by lists that group MS-DRGs into medical and surgical groups and generally correspond with the CMS designation as a surgical/medical MS-DRG.[22] A list of operating room (OR) procedures is used to define denominator inclusion and exclusion criteria for some measures where the intended denominator includes only major OR procedures that are not performed as a result of the complication of interest.

Denominator Exclusions[23]

Generally, discharges may be excluded from the denominator for one (or more) reasons:

1. The outcome of interest has been coded as POA.
2. The outcome of interest is very difficult to prevent and therefore not an indication of substandard care.
3. The discharge is at very low risk for the adverse event and are therefore excluded to keep from diluting the QI denominator.
4. The exclusion enhances face validity with clinicians (e.g., exclude patients from being at risk of a pressure ulcer [PSI 03] if they have not been hospitalized for at least three days).
5. The patient was transferred to another health care facility (to avoid either double counting a single encounter or incomplete capture of the event).[24]
6. Encounters missing data elements that are required for indicator construction (e.g. missing principal diagnosis code, disposition of patient or the source of admission).

[22] ICD-10-CM/PCS MS-DRG v38.1, list of MS-DRGs, available at: https://www.cms.gov/icd10m/version38-1-fullcode-cms/fullcode_cms/P0001.html.
[23]Numerator exclusions are defined based on similar reasons listed here for PSI 05, an indictor for count.
[24] In the instance of IQI 32, transfers from another facility are excluded from the denominator. For most other HQIs, transfers to another facility are excluded and transfers from other facilities are accounted for during risk adjustment.

F.3.3 Observed Rate

Observed rates for hospital-level indicators are calculated by dividing the number of discharges with the outcome of interest (mortality, adverse event) by the number of discharges for patients at risk of the outcome (denominator).

F.4 Comparing Indicators across Hospitals, Units, or Time

F.4.1 Overview of Expected, Risk-Adjusted, and Smoothed Rates for Hospital-Level Indicators

In order to make meaningful comparisons of the hospital-level indicators from one hospital to another, one unit or another, and/or from one time period to another, it is helpful to account statistically for differences in demographics and clinical case mix of each of the hospitals, units, or time periods (if there are changes in referral sources).

Expected and risk-adjusted rates both acknowledge that individual hospitals are unique and differ in two important ways from the representative profile observed in the reference population. First, there is heterogeneity in the quality of care that is provided. Some hospitals provide exemplary care. Others provide sub-standard care. This is an important dimension of differences. Second, most individual hospitals serve patients with a distribution of covariates (demographics and comorbidities) that differs from the reference population. Some hospitals serve populations that are at higher risk for adverse events, and some serve populations that are at lower risk. This is a dimension that makes it difficult to make meaningful comparisons of observed rates. The expected and risk-adjusted rates each peg one of these two dimensions (quality of care or patient mix) to that observed in the reference population and then comment on the second dimension, as observed in the local data.

The ***expected rate*** answers the question, "What rate of adverse events would we expect to see if this hospital provided the average level of care observed in the reference population, but provided it to patients with the locally observed distribution of characteristics?" (i.e., average performance from the reference population of the universe of patients applied to locally observed mix of patients with their local risk profiles). When the observed rate is smaller than the expected rate (or the observed / expected ratio is < 1), then there is reason to think that the hospital is performing better than average on this indicator.

The ***risk-adjusted rate*** is calculated by multiplying the ratio of the observed rate and expected rate with the reference population observed rate. The risk-adjusted rate answers the converse question, "What rate of adverse events would we see in this hospital if they provided the locally observed quality of care to patients whose distribution of characteristics matched those in the reference population?" (i.e., locally observed performance on a representative mix of patients from the reference population). If the risk-adjusted rate is higher than the reference rate (or if observed rates are higher than expected rates), it means the performance of the hospital is worse than what would be expected based on the experience of patients in the reference population with a similar distribution of characteristics.

The ***smoothed rate*** is a weighted average of the reference population rate and the local risk-adjusted hospital rate. If the data from the individual hospital include many observations and provide a numerically stable estimate of the rate, then the smoothed rate will be very close to the risk-adjusted rate, and it will not be heavily influenced by the reference population rate. Conversely, the smoothed rate will be closer to the reference population rate if the hospital rate is based on a small number of observations and may not be numerically stable, especially from year to year.

F.4.2 Risk Factors for Hospital-Level Indicators

For accountability measures, the goal of risk adjustment in comparative outcome measures is to account for differences in patients across measured entities (e.g., hospitals) that affect outcome rates and that are unrelated to the quality of care. When such differences are not addressed, differences in the measure score will reflect both case mix and quality, and will be biased against hospitals who have patients at higher risk for the measured adverse outcome.

All hospital-level indicators are risk adjusted with the exception of the volume/count indicators.

Identifying clinical condition categories is challenging for all age groups and outcomes. The IQI module uses AHRQ's Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses and Clinical Classifications Software Refined (CCSR) for ICD-10-PCS Procedures to collapse individual diagnosis and procedure codes into a smaller number of meaningful categories. AHRQ's CCSR categories for diagnoses and procedures are used to capture risk factors that relate to comorbidities, procedure subtype, and procedure complexity. The IQI software creates the procedure based CCSR categories only for procedures that occur on or prior to the day of the IQI denominator procedure. This accounts for procedure subtype and complexity or concurrent procedures but will not adjust for procedures resulting from post-operative complications. Starting in v2021, the AHRQ CCSR for ICD-10-CM Diagnoses replaced the APR-DRGs, based on Refined-DRGs and All-Payer DRGs systems, in the condition-based IQIs. Starting in v2022, the CCSR for ICD-10-PCS Procedures replaced the APR-DRGs for procedure-based IQIs.

For PSIs, AHRQ Elixhauser Comorbidity codes and comorbidity count categories are used in the software to cover comorbidity conditions that can be either present on admission or after admission. For PDIs, AHRQ CCSR software is used because it covers pediatric conditions, whereas the MS-DRGs do not. Comorbidity count categories of AHRQ Comorbidity codes are also included for non-neonatal indicators for PDIs.

Four classes of risk factors are considered for the AHRQ QI hospital-level indicators, including demographics, severity of illness, clinical/comorbidities, and discharge-specific information. Table 9 provides an overview of the four classes of risk factors. Appendix C provides a detailed description of each of the risk factors.

Table 9. AHRQ QI Risk Adjustment Covariates for Hospital-Level Indicators

Category	IQI	PSI	PDI	NQI
Demographics	Sex[a]	Sex[a]	Sex[a]	Sex[a]
	Age[a]	Age[a]	Age in days (90 days–1 year)[a] Age in years (1 year+)[a]	Age in days (0 or 1 day)[a]
Severity of Illness		Modified MS-DRG[b]	Modified MS-DRG[b]	Modified MS-DRG[b]
		MDCs[b]		
Clinical / Comorbidities	AHRQ CCSR for ICD-10-CM Diagnoses[c]		AHRQ CCSR for ICD-10-CM Diagnoses[c]	

Category	IQI	PSI	PDI	NQI
	AHRQ CCSR for ICD-10-PCS Procedures[b,d]			
		AHRQ Comorbidities (using latest guidance on POA requirements)[b]		
		Count categories of AHRQ Comorbidities (using latest guidance on POA requirements[25])[b]	Count categories of AHRQ Comorbidities (using latest guidance on POA requirements)[b]	
		Do Not Resuscitate (POA)		
			Indicator-specific risk stratifiers	
				Birth weight (500g groups)[a]
Discharge-specific information	Transfer-in status[b]	Transfer-in status[b]	Transfer-in status[b]	Transfer-in status[b]
		Surgical/Medical discharge (MS-DRG)[b]	Surgical/Medical discharge (MS-DRG)[b]	
Stratified risk groups	Indicator-specific risk stratifiers	Indicator-specific risk stratifiers		

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; CCSR, Clinical Classification Software Refined; IQI, Inpatient Quality Indicator; MDC, Major Diagnostic Category; MS-DRG, Medicare Severity-Diagnostic Related Group; NQI, Neonatal Quality Indicator; PDI, Pediatric Quality Indicator; PSI, Patient Safety Indicator; QI, Quality Indicator.

[a] Categories are mutually exclusive.
[b] Variable or variable categories are selected into model for some indicators.

[c] AHRQ CCSR for ICD-10-CM Diagnoses are modified and additional comorbidity groups are also included.
[d] AHRQ CCSR for ICD-10-PCS Procedures are modified to only include procedures occurring on or before the day of the denominator procedure.

[25] AHRQ HCUP Elixhauser Comorbidity Software Refined Tool, v2022.1: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR_v2022-1.zip

F.4.3 Expected Rate for Hospital-Level Indicators

Expected rates are predicted for each hospital using risk adjustment model coefficients that summarize the demographic and clinical case mix of the hospital. An expected (or predicted) rate for each QI is derived for each hospital. Using reference population risk adjustment parameters and indirect standardization, each eligible discharge (i.e., one that is included in the denominator of the indicator) is scored for its expected (or predicted) probability for the outcome of interest using PROC SCORE.[26] PROC SCORE produces new predictions from a model. For the QI module implementation, this SAS procedure takes a new set of discharges (i.e., from the user's dataset) and calculates probabilities from the risk adjustment model; these probabilities are the discharge-level expected outcomes, which are then aggregated by hospital to yield the hospital-level expected rate. This output score is simply the sum across all covariates in the risk adjustment model of the scalar multiplication of the presence or absence of a covariate (1 or 0) times the value of the coefficient from the risk adjustment model for that covariate.

Denoted by:

Y_i, the observed (0, 1) outcome for discharge *i*

$\hat{Y}_i$, the expected (predicted) rate for discharge *i*

A_h, the set of discharges in hospital *h*

n_h, the number of discharges at hospital *h*

α, the reference population rate (average outcome in the entire sample)

We define the observed and expected rates of hospital *h* by, respectively,

$$O_h = \frac{1}{n_h} \sum_{i \in A_h} Y_i$$

$$E_h = \frac{1}{n_h} \sum_{i \in A_h} \hat{Y}_i$$

F.4.4 Risk-Adjusted Rate for Hospital-Level Indicators

The AHRQ QIs use indirect standardization to calculate the risk-adjusted rate. The risk-adjusted rate is given by the indirectly standardized ratio multiplied by the reference population rate:

$$RAR_h = \alpha \cdot \frac{O_h}{E_h}$$

We used logistic regression models to build risk adjustment models for QIs that need risk adjustment. For complicated risk adjustment models, the national HCUP reference population observed rate may not be exactly same as the average of predicted event rates. In the modeling process, we assessed model calibration properties, but the O-E ratio (observed rate to expected rate ratio) may not be exactly equal to 1. In software development (not part of the publicly released software), we multiplied the predicted rate

[26] SAS. SAS/STAT 9.2 User's Guide. The SCORE Procedure (Book Excerpt). https://support.sas.com/documentation/cdl/en/statugscore/61828/PDF/default/statugscore.pdf.

for each discharge by this constant (O-E ratio) to make sure the new predicted rates are perfectly calibrated to the observed rates. To be consistent, we included the national O-E ratio that was calculated based on our reference population in the AHRQ software v2022. We also provided users the options of calibrating to the reference population or to users' populations.

1. Reference population based O-E ratio is recommended in most situations and it is also the default choice in the software.
2. Users' own population based O-E ratio option is kept in the software for users who want to calibrate the predicted rates to users' population.

F.4.5 Risk-Adjusted Rate Variance for Hospital-Level Indicators

The standard error of the risk-adjusted rate for each hospital is calculated using a method recommended by Iezzoni[27] and described by Hosmer and Lemeshow[28] that represents the amount of within-hospital or area variance due to sampling (i.e., as the number of patients per hospital or individuals per area increases, this variance tends to zero). This standard error is used to calculate lower and upper bound 95% confidence intervals around the risk-adjusted rate as risk-adjusted rate +/– 1.96 * risk-adjusted rate standard error.

Using a Taylor expansion or "delta method" for the formula for the variance of the ratio of two stochastic variables, we compute the variance on the risk-adjusted rate:

$$\mathrm{Var}(RAR_h) \cong \alpha^2 \frac{\mathrm{E}(O_h)^2}{E_h^2}\left(\frac{\mathrm{Var}(O_h)}{\mathrm{E}(O_h)^2} - 2\frac{\mathrm{Cov}(O_h, E_h)}{\mathrm{E}(O_h)\cdot E_h} + \frac{\mathrm{Var}(E_h)}{E_h^2}\right)$$

It is common practice in these calculations to neglect the variance of the predicted values $\hat{Y}_i$ and to consider a normal distribution for the risk-adjusted rate (as $n_h \to \infty$).[29] In this case, the above formula simplifies to:

$$\mathrm{Var}(RAR_h) \cong \alpha^2 \frac{\mathrm{Var}(O_h)}{E_h^2}$$

and the 95% confidence intervals are calculated assuming normality. However, arguments to support using nonapproximate equations[30] for the RAR confidence intervals (in particular, when n_h is small) may be considered in future releases of the AHRQ QI software.

F.4.6 Smoothed Rate for Hospital-Level Indicators

Each hospital's smoothed rate is a weighted average of the risk-adjusted rate and the reference population rate calculated from discharges in the reference population; the smoothed rate is calculated with an empirical Bayes shrinkage estimator (1) to result in a rate that will be near that calculated from the input

---

[27] Iezzoni, Lisa, Ed. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago: Health Administration Press; 2013.
[28] Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance model based on logistic regression. Statistics in Med. 1995;14(19):2161-72.
[29] Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance model based on logistic regression. Statistics in Med. 1995;14(19):2161-72.
[30] For example, see: Luft HS, Brown BW Jr. Calculating the probability of rare events: why settle for an approximation? Health Serv Res. 1993;28(4):419-39.

dataset if the hospital's rate is estimated in a stable fashion with minimal noise, or (2) to result in a rate near that of the reference population if the rate from the hospital is unstable and based on noisy data. Thus, the smoothed rate for a hospital with stable estimates will be similar to the hospital's risk-adjusted rate, whereas the smoothed rate for a hospital with unstable estimates will be more similar to the rate calculated in the discharges of the reference population.

The accent "~" is used to denote the reliability adjustment. The formula for the smoothed rate is as follows:

$$\widetilde{RAR}_h = \lambda_h \cdot RAR_h + (1 - \lambda_h) \cdot \alpha$$

where the reliability weight λ_h for hospital h is a function of the reference population signal variance τ^2 and hospital's noise variance σ_h^2. Specifically, the reliability weight is the ratio of the signal variance (i.e., true variation in hospital quality reflected by the risk-adjusted rates) to the total variance, which includes sampling error:

$$\lambda_h = \frac{\tau^2}{\tau^2 + \sigma_h^2}$$

The noise variance is calculated for each hospital based on the user's data. The signal variance is a parameter calculated from the reference population. The two variances are estimated as follows:

$$\text{Noise Variance } \hat{\sigma}_h^2 = \left(\frac{\alpha}{n_h E_h}\right)^2 \sum_{i \in A_h} \hat{Y}_i (1 - \hat{Y}_i)$$

$$\text{Signal Variance } \hat{\tau}^2 = \frac{\sum_{h=1}^{H} \frac{1}{(\hat{\tau}^2 + \sigma_h^2)^2} \left\{ \frac{H}{H-1} (RAR_h - \overline{RAR})^2 - \hat{\sigma}_h^2 \right\}}{\sum_{h=1}^{H} \frac{1}{(\hat{\tau}^2 + \sigma_h^2)^2}}$$

where $\overline{RAR}$ is the weighted[31] average of hospital risk-adjusted rates; H is the number of hospitals with patients at risk for the QI, α is the reference population rate; $\hat{Y}_i$ is the patient-level predicted probability; and for hospital h, A_h is the set of patients, n_h is the number of patients, E_h is the expected rate, and RAR_h is the risk-adjusted rate. Note that $\hat{\tau}^2$ appears on both sides of the signal variance equation; it is estimated in an iterative fashion.[32]

For small hospitals, the reliability weight λ_h is usually small since the noise variances are usually larger than large hospitals.

F.4.7 Smoothed Rate Variance for Hospital-Level Indicators

The smoothed rate is an empirical Bayes posterior estimate of the hospital's risk-adjusted rate—that is, it is calculated from the reliability-weighted combination of the risk-adjusted rate and reference population mean. The variance of the smoothed rate is given by:

---

[31] The weights are $\frac{1}{(\hat{\tau}^2 + \sigma_h^2)^2}$.

[32] Morris, CN. Parametric empirical Bayes inference: theory and applications. J Am Statistical Assoc. 1983 Mar;78(381):47-55.

$$\mathrm{Var}\left(\widetilde{RAR}_h\right) = \tau^2(1-\lambda_h)$$

F.5 Weighted Composite Scores for Hospital-Level Indicators

F.5.1 Overview of Composite Methodology

The general method for computing a hospital-level composite measure is to calculate a weighted average of a set of risk and reliability-adjusted (e.g., smoothed) component quality indicators. The individual smoothed quality indicators are referred to as "component" indicators, and the weighted average of the components is the "composite". The composite weights are selected based on the intended interpretation of the composite QI and are determined empirically.

F.5.2 Composite Value

The basic steps for computing the composite are as follows:

Step 1. Compute the risk-adjusted rate and confidence interval.

The AHRQ QI risk-adjusted rate and confidence interval are computed as described above.

Step 2. Scale indicators compute the Observed-to-Expected (O/E) ratio by scaling the risk-adjusted rate using the reference population.

To combine the component indicators across a common scale, each indicator's risk-adjusted rate is divided by the reference population rate to yield the observed to expected ratio (O/E ratio) ratio. The O/E ratio for hospital *h* is 1.0 if the observed QI rate is equal to the expected QI rate determined from the risk adjustment parameters applied to the data. For component indicator *c* of hospital *h*, the O/E ratio is given by:

$$OE_{hc} = \frac{O_{hc}}{E_{hc}} = \frac{RAR_{hc}}{\alpha_c}$$

where subscript *c* indexes the component indicator. For example, α_c is the reference population rate for component indicator *c*, and RAR_{hc} is the analogous risk-adjusted rate for hospital *h*.

Step 3. Compute the reliability-adjusted ratio.

The reliability-adjusted O/E ratio is computed as the weighted average of the risk-adjusted ratio and the reference population ratio, which is defined to be equal to 1, since the observed rate equals the expected rate in the population. The weights are determined by the reliability weight for the hospital (or other unit of analysis). The accent "~" is used to denote the reliability adjustment.

$$\widetilde{OE_{hc}} = \lambda_{hc}OE_{hc} + (1-\lambda_{hc}) = \lambda_{hc}(OE_{hc}-1)+1$$

Note that multiplying the above expression by the reference population rate α, the smoothed rate is recovered.

Step 4. Select the component weights.

The composite measure is the weighted average of the scaled and reliability-adjusted ratios for the component indicators. The default type of weights applied is dependent on the specific composite of interest. Table 10 shows each of the composite indicators and the type of weight (default) used to derive the indicator.

Table 10. AHRQ QI Composite and Weight

Abbrev	Indicator Name	Weight (by default)		
		Numerator	Denominator	Harm
IQI 90	Mortality for Selected Inpatient Procedures		X	
IQI 91	Mortality for Selected Inpatient Conditions		X	
PSI 90	Patient Safety and Adverse Events Composite (beginning in v6.0)	X		X

Alternative options for weights include the following:

- *Numerator weight*. A numerator weight is based on the relative frequency of the numerator for each component indicator in the reference population. In general, a numerator weight reflects the amount of harm in the outcome of interest, in this case, a potentially preventable adverse event. One also might use weights that reflect the amount of excess mortality or complications associated with the adverse event or the amount of confidence that one has in identifying events (i.e., the positive predictive value).

- *Denominator weight*. A denominator weight is based on the relative frequency of the denominator for each component indicator in the reference population. In general, a denominator weight reflects the degree of risk of experiencing the outcome of interest in a given population. For example, the denominator weight might be based on the demographic composition of a health plan, the employees of a purchaser, a State, an individual hospital, or a single patient.

- *Harm weight.* Harm weighting is based on an analysis that assigns each component indicator a weight that reflects the contribution of that indicator to excess harmful outcomes that occur in the population that experience the component events. Component indicators that both are common and lead to significant excess mortality and morbidity will have the highest weights, whereas those that are less common or have lower mortality and morbidity associated with them will have lower weights. For additional information, see the "Quality Indicator User Guide: Patient Safety Indicators (PSI) Composite Measures, July 2022" at:

  https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2022/PSI_Composite_Development.pdf.

Step 5. Construct the composite measure.

The composite measure is the weighted average of the component indicators using the selected weights and the scaled and reliability-adjusted indicators. For hospital *h*, the composite value is calculated by:

$$COMPOSITE_h = \sum_c w_c \, \widetilde{OE}_{hc}$$

where w_c denotes the weight applied to component indicator *c*.

When a hospital's component indicator fails the minimum denominator criterion (i.e., it has fewer than three denominator cases), PSI 90 sets the O/E ratio = 1 for that component indicator. If a hospital fails the

denominator criteria for all component indicators, the hospital's PSI 90 value then equals one. Hospitals that are missing many of the component indicators will have less informative PSI 90 scores (not distinguishable from average performance).

F.5.1 Composite Variance

The probability interval of the composite measure is based on its standard error, which is the square root of the variance. The variance is computed based on the signal variance-covariance matrix and the reliability weights.

Let $\mathbf{M}$ be a $1 \times K$ vector of observed quality measures (for a given hospital, suppress hospital subscript for convenience), noisy measures of the true underlying $1 \times K$ quality vector $\boldsymbol{\mu}$, such that:

$$\mathbf{M} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \qquad (11.1)$$

where $\boldsymbol{\epsilon}$ is a $1 \times K$ noise vector with zero mean and $K \times K$ variance-covariance matrix $\mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}$. Let the $K \times K$ signal variance-covariance be $\mathrm{Var}(\boldsymbol{\mu}) = \boldsymbol{\Omega}_{\boldsymbol{\mu}}$.

Let $\widehat{\boldsymbol{\mu}}$ be a $1 \times K$ vector indicating the posterior (filtered) estimate of $\boldsymbol{\mu}$**,** such that:

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \mathbf{v} \qquad (11.2)$$

where $\mathbf{v}$ is a $1 \times K$ vector with zero mean and $K \times K$ variance-covariance matrix $Var(\mathbf{v})$ representing the prediction error of the posterior estimates.

The goal is to estimate the variance for any weighted average of the posterior estimates. For a given $1 \times K$ weighting vector $\mathbf{w}$, this is given by:

$$Var(\mathbf{vw}) = \mathbf{w}' Var(\mathbf{v})\mathbf{w} \qquad (11.3)$$

where $\mathbf{w}'$ indicates the transpose of $\mathbf{w}$.

Thus, we need an estimate of $Var(\mathbf{v})$. We simplify the calculation by assuming that the filtered estimates are formed in isolation for each measure (univariate) and that the estimation error is assumed not correlated across measures (e.g., each measure is based on a different sample of patients or independent patient outcomes).

Forming each measure in isolation, using superscripts $k = 1, \ldots, K$ to indicate the measure, we have:

$$\widehat{\mathbf{u}}^k = \mathbf{M}^k \widehat{\boldsymbol{\beta}}^k = \mathbf{M}^k \left(\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk} + \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{kk}\right)^{-1} \boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk} \qquad (11.4)$$

$$Var(\mathbf{v}^k) = \boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk}(1 - \widehat{\boldsymbol{\beta}}^k) = \boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk} - \boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk}\left(\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk} + \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{kk}\right)^{-1}\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk},$$

where:

$$\widehat{\boldsymbol{\beta}}^k = (\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk} + \boldsymbol{\Omega}_{\boldsymbol{\epsilon}}^{kk})^{-1}\boldsymbol{\Omega}_{\boldsymbol{\mu}}^{kk} \qquad (11.5)$$

is the signal ratio of measure k, the reliability of the measure, and is the r-squared that measures how much of the variation in the true measure can be explained with the filtered measure. Note that in this simplified case the filtered estimate is a univariate shrinkage estimator. For the non-diagonal elements of the covariance matrix (for $j \neq k$),

$$Cov(\mathbf{v}^j, \mathbf{v}^k) = E\left[\left(\boldsymbol{\mu}^j - \widehat{\boldsymbol{\mu}}^j\right)\left(\boldsymbol{\mu}^k - \widehat{\boldsymbol{\mu}}^k\right)\right] \qquad (11.6)$$

assuming independent estimation error in the two measures, one gets the following simplified expression (see supplemental notes below for the derivation):

$$Cov(\mathbf{v}^j, \mathbf{v}^k) = \boldsymbol{\Omega}_{\boldsymbol{\mu}}^{jk}\left[\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right)\right] \qquad (11.7)$$

Note that this is just the signal covariance times 1 minus the signal ratio for each of the measures. Thus, if the signal ratio is 0 for each measure, the covariance in the estimates is simply the signal covariance. As either measure gets a stronger signal ratio (becomes more precise), the covariance in the estimates shrinks to 0.

Also note that if one measure is missing, then the signal ratio is simply set to 0. The filtered estimate is shrunk all the way back to the (conditional) mean, and the variance and covariance are as defined above.

The standard error on the composite is the square root of the variance, which is then used to compute the 95% probability interval.

Supplemental Notes:

To derive formula (11.6), we substitute

$$\widehat{\boldsymbol{\mu}} = \mathbf{M}\widehat{\boldsymbol{\beta}} = (\boldsymbol{\mu} + \boldsymbol{\epsilon})\widehat{\boldsymbol{\beta}}$$

into (11.5) and obtain (for $j \neq k$)

$$Cov(\mathbf{v}^j, \mathbf{v}^k) = E\left[\left(\boldsymbol{\mu}^j - \left(\boldsymbol{\mu}^j + \boldsymbol{\epsilon}^j\right)\widehat{\boldsymbol{\beta}}^j\right)\left(\boldsymbol{\mu}^k - \left(\boldsymbol{\mu}^k + \boldsymbol{\epsilon}^k\right)\widehat{\boldsymbol{\beta}}^k\right)\right]$$

$$= E\left[\left(\boldsymbol{\mu}^j\left(1 - \widehat{\boldsymbol{\beta}}^j\right) - \boldsymbol{\epsilon}^j\widehat{\boldsymbol{\beta}}^j\right)\left(\boldsymbol{\mu}^k\left(1 - \widehat{\boldsymbol{\beta}}^k\right) - \boldsymbol{\epsilon}^k\widehat{\boldsymbol{\beta}}^k\right)\right]$$

$$= E\left[\boldsymbol{\mu}^j\boldsymbol{\mu}^k\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right) + \boldsymbol{\mu}^k\boldsymbol{\epsilon}^j\left(1 - \widehat{\boldsymbol{\beta}}^k\right)\widehat{\boldsymbol{\beta}}^j + \boldsymbol{\mu}^k\boldsymbol{\epsilon}^j\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\widehat{\boldsymbol{\beta}}^k + \boldsymbol{\epsilon}^j\boldsymbol{\epsilon}^k\widehat{\boldsymbol{\beta}}^j\widehat{\boldsymbol{\beta}}^k\right]$$

$$= E\left[\boldsymbol{\mu}^j\boldsymbol{\mu}^k\right]\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right) + E\left[\boldsymbol{\mu}^k\boldsymbol{\epsilon}^j\right]\left(1 - \widehat{\boldsymbol{\beta}}^k\right)\widehat{\boldsymbol{\beta}}^j + E\left[\boldsymbol{\mu}^j\boldsymbol{\mu}^k\right]\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\widehat{\boldsymbol{\beta}}^k + E\left[\boldsymbol{\epsilon}^j\boldsymbol{\epsilon}^k\right]\widehat{\boldsymbol{\beta}}^j\widehat{\boldsymbol{\beta}}^k.$$

Assuming and $E[\boldsymbol{\mu}] = 0$, we have

$$E\left[\boldsymbol{\mu}^j\boldsymbol{\mu}^k\right] = E\left[\boldsymbol{\epsilon}^j\boldsymbol{\mu}^k\right] = E\left[\boldsymbol{\epsilon}^j\boldsymbol{\epsilon}^k\right] = 0$$

$$Cov(\mathbf{v}^j, \mathbf{v}^k) = E\left[\boldsymbol{\mu}^j\boldsymbol{\mu}^k\right]\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right)$$

$$= Cov\left(\boldsymbol{\mu}^j, \boldsymbol{\mu}^k\right)\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right) - E\left[\boldsymbol{\mu}^j\right]E[\boldsymbol{\mu}^k]\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right)$$

$$= Cov(\boldsymbol{\mu}^j, \boldsymbol{\mu}^k)(1 - \widehat{\boldsymbol{\beta}}^j)(1 - \widehat{\boldsymbol{\beta}}^k).$$

F.6 Interpretation of Counts, Rates, and Scores

Counts are reported for adverse events or indicators where risk adjustment is challenging. As such, risk adjustment is not used for counts. For adverse events, the ideal benchmark is zero. For other counts, national-level benchmarks are provided in the QI benchmark data tables (see Chapter III.B for links to the benchmark data tables).

- **Rates** are reported for non-composite measures. Observed rates are used for non-comparative purposes while risk-adjusted rates and smoothed rates are better used when comparing hospitals or areas to a national average hospitals or area. For all QIs , a rate below reference rate indicates better quality than expected for that hospital's case mix. When comparing hospitals to a benchmark, using smoothed rates are desirable given that they adjust for small sample sizes. Although it is possible to compare risk-adjusted rates to a benchmark, it is advised to incorporate confidence intervals/uncertainty estimates and use the appropriate statistical interpretation of results. National benchmarks are available in the QI benchmark data tables (see Chapter III.B for links to the benchmark data tables).

- **Scores** are reported for hospital-level composite measures (observed to expected ratio). Scores incorporate both risk adjustment and smoothing (i.e., reliability adjustment). A composite below 1 indicates better quality than expected for that hospital's case mix; however, the composite is an estimate, and any comparisons should account for uncertainty.

The reliability of the hospital-level indicators varies by indicator. Often less common events have lower reliability, but reliability is also impacted by the distribution of events in the reference population which is influenced by the characteristics of the total population. Reliability is calculated for each hospital. To account for potential issues with reliability smoothed rates are recommended for most hospital-level measures. Differences between hospitals in both observed and risk-adjusted rates are often more stable using two or more years of data.

G. Recommendations on How to Report Trends

For any comparative analysis (e.g., using pre and post periods), it is important to note the reference population over which the QI models were estimated. For risk and reliability adjustment, the expected QI rate is calibrated to the reference population specific to that QI version.

Calculating and reporting trends in QI rates over time, depends on the research question. For example, are the trends meant to illustrate how hospital quality has changed over time against a contemporaneous benchmark? In this example, the analyst could apply the *recent* version of the QI software to both "pre" and "post" data; in particular, the pre-period QI rate would reflect current hospital quality against the quality that would have been expected had they treated the same type of patients in the post period.

On the other hand, a cross-sectional analysis might apply the QI versions that are concurrent with the observation period of the pre- and post-period discharge populations. In this way, the trends would illustrate how underlying hospital quality changes over time, also taking into account how the reference population had changed over time.

A comparative analysis can also be designed by geographic area or between hospital types. Similarly, the analyst would need to consider whether the underlying risk and reliability adjustment of the QI module is

appropriate for measuring hospital quality. The QI module is calibrated to a specific reference population on which hospital and area comparisons are made using the risk- and reliability-adjusted QI rates.

Chapter III. Empirical Development of the AHRQ QIs

In this chapter, we describe the underlying methods used to develop the QI software. Specifically, we describe the reference population data, the calculations performed to update the reference population, possible risk factors used in the risk models derived during QI development, development of risk (and harm) models that provide the parameter estimate used in the software, and a summary of the testing and evaluation that is performed on each indicator.

A. Overview of the Development Process

One of the hallmarks of the AHRQ QI programs is the continuous enhancement and annual refinement of all indicators based on user feedback, review of clinical practice changes, validation studies, empirical testing for validity and reliability, and input for expert panels such as the National Quality Forum (NQF) Patient Safety Committee[33] and experts from the AHRQ QI Workgroups.[34, 35] Additional detail on the AHRQ QI measure development, implementation, maintenance, and retirement process is posted on the AHRQ QI website at: http:// qualityindicators.ahrq.gov/Modules/.

In order for the QIs to remain scientifically acceptable and useful, they must be maintained and potentially enhanced on a regular cycle. QIs need to be updated based on such factors as: recent evidence published in the literature (particularly as publications are made available using the specific QI) and from user feedback, technical specification updates including annual (and sometime quarterly) coding updates (e.g., ICD-9-CM, ICD-10-CM/PCS, MS-DRGs, MDCs, POA coding guidelines), reference population changes, census population updates, periodic clinical panel review, the NQF endorsement and maintenance process, and newly available data and methodological advances in the industry. Each of the material maintenance steps must be considered within the broader measure life cycle.

Each year, the AHRQ QI project takes into account the aforementioned changes and refines the AHRQ QI technical specifications. Refinements may include but are not necessarily limited to the following: integration of new codes, removal of clinically irrelevant codes, new risk models with updated risk adjustment parameter estimates, updated reference population observed, expected, risk-adjusted, and smoothed rates, updated weights for hospital-level composites based on the frequency of the events, and updated variance estimates based on the most recent reference population information. Annually, the AHRQ QI project releases a list (or log) of changes that have been implemented with each release of the AHRQ QI specifications.

Table 11 provides a list of all versions of the AHRQ QI specifications, the date of release, and the year the reference population upon which the specifications are built.

---

[33] NQF Patient Safety Final Report 2015, https://www.qualityforum.org/Publications/2016/02/Patient_Safety_2015_Final_Report.aspx
[34] AHRQ QI Composite Workgroups, https://qualityindicators.ahrq.gov/Modules/composite_workgroup.aspx
[35] Federal registry notice of the AHRQ QI Workgroups, available at: https://www.federalregister.gov/documents/2006/04/04/06-3207/ahrq-quality-indicators-workgroup-on-inpatient-and-patient-safety-composite-measures

Table 11. AHRQ QI Specification Releases

AHRQ QI Version	Coding Scheme	Release Date	Modules	Year of Reference Population
2022	ICD-10-CM/PCS/PCS	July 2022	All	2019
2021.0.2	ICD-10-CM/PCS/PCS	March 2022	IQI	2018
2021.0.1	ICD-10-CM/PCS/PCS	September 2021	PSI, PDI, IQI	2018
2021	ICD-10-CM/PCS/PCS	July 2021	All	2018
2020	ICD-10-CM/PCS/PCS	July 2020	All	2017
2019	ICD-10-CM/PCS/PCS	Summer 2019	All	2016
2018	ICD-10-CM/PCS/PCS	Summer 2018	All	---
7.0	ICD-10-CM/PCS/PCS	Spring 2017	All	---
6.0	ICD-10-CM/PCS/PCS	Summer 2016	All	---
6.0	ICD-9-CM	Summer 2016 – Spring 2017	All	2013
5.0	ICD-10-CM/PCS	October 2015	All	---
5.0	ICD-9-CM	March 2015	All	2012
4.5a	ICD-9-CM	July 2014	PSI only	
4.5	ICD-9-CM	May 2013	All	2010
4.4	ICD-9-CM	March 2012	All	2009
4.3a	ICD-9-CM	September 2012	All	2008
4.3	ICD-9-CM	August 2011	All	2008
4.2	ICD-9-CM	September 2010	All	2007
4.1	ICD-9-CM	December 2009	All	2006
3.2	ICD-9-CM	February – March 2008	All	2005
3.1	ICD-9-CM	March 2007	PQI, IQI, PSI	2004
3.0a	ICD-9-CM	May 2006	PSI only	2003
3.0	ICD-9-CM	February 2006	PSI only	2003

Abbreviations: ICD-9-CM, International Classification of Diseases Volume 9 Clinical Modification; ICD-10-CM/PCS, International Classification of Diseases Volume 10 Clinical Modification or Procedure Code System; PQI, Prevention Quality Indicators; IQI, Inpatient Quality Indicators, PSI, Patient Safety Indicators

Ellipse (--) indicates that no data was available to derive national rates or risk adjustment models.

B. Discharge Reference Population

The AHRQ QIs are developed using hospital discharge abstracts and billing data from HCUP. HCUP is a family of health care databases and related software tools and products developed through a Federal-

State-industry partnership[36]. HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. The HCUP SID[37] contains all-payer, encounter-level information on inpatient discharges from the universe of community hospitals in participating states. The SID includes clinical and resource information typically found on a billing record (Uniform Bill – 04, UB-04), such as patient demographics, up to 127 ICD-10-CM/PCS diagnoses and procedures (up to 35 diagnosis and 30 procedure codes are retained in the reference population), length of stay, expected payer, admission and discharge dates, and discharge disposition.

The reference population file is limited to community hospitals and beginning with 2012 data also excludes rehabilitation and long-term acute care (LTAC) hospitals. Information on the type of hospital was obtained by the American Hospital Association (AHA) Annual Survey of Hospitals. AHA defines community hospitals as "all non-Federal, short-term, general, and other specialty hospitals, excluding hospital units of institutions." Included among community hospitals are specialty hospitals such as obstetrics-gynecology, ear-nose-throat, orthopedic, and pediatric institutions. Also included are public hospitals and academic medical centers.

The HCUP databases represent more than 97 percent of all annual community hospital discharges in the United States. Some States include discharges from specialty facilities, such as acute psychiatric hospitals. The HCUP SID data serve as the reference (or general) population for the AHRQ QIs, upon which national benchmarks for numerators, denominators, observed rates, risk models, expected rates and risk-adjusted rates, and smoothed rates are derived. Specifically, the reference population plays two important roles:

1. The ***reference population rate*** for each QI is calculated and serves as a comparative standard. One can analyze data to determine which entities have rates that are higher or lower than those of the overall reference population. The reference population rates are published on the AHRQ QI website in module-specific documents named Benchmark Tables (formerly known as Comparative Data Tables):

    - PQI Benchmark: https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2022/Version_2022_Benchmark_Tables_PQI.pdf
    - IQI Benchmark: https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2022/Version_2022_Benchmark_Tables_IQI.pdf
    - PSI Benchmark: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2022/Version_2022_Benchmark_Tables_PSI.pdf

[36] For a complete list of HCUP Partner organizations that participated in the HCUP SID, please see the Acknowledgements section of this document.
[37] HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality (AHRQ), Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp.

- PDI Benchmark: https://qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2022/Version_2022_Benchmark_Tables_PDI.pdf

2. The ***risk adjustment models*** are re-estimated annually using the most recent reference population dataset. This process is described in Chapter III.F of this document. The models are included in the QI software to allow calculation of risk-adjusted rates. The risk adjustment model covariates and regression coefficients are published on the AHRQ website in module-specific documents named Parameter Estimates Tables:

   - PQI Parameter Estimates: https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2022/Parameter_Estimates_PQI_v2022.pdf
   - IQI Parameter Estimates: https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2022/Parameter_Estimates_IQI_v2022.pdf
   - PSI Parameter Estimates: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2022/Parameter_Estimates_PSI_v2022.pdf
   - PDI Parameter Estimates: https://qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2022/Parameter_Estimates_PDI_v2022.pdf

Table 12 provides details on HCUP SID data availability, including the year-specific number of states, number of hospitals and total discharges that potentially could be included in the AHRQ QI reference population universe. However, variations from these estimates exist, as not all data is available at the time needed and states may vary in the availability of data elements (e.g., present on admission information or the number of days between admission and procedure).

Table 12. AHRQ QI Reference Population

Data Year	Number of States[a]	Number of Hospitals[b]	Total Discharges included in SID	Percentage of discharges[c]
2019	49	4,252	35,612,594	96
2018	48	4,290	35,549,549	97
2017	48	4,326	35,747,363	98
2016	48	4,039	35,612,904	98
2014	45	4,430	33,645,600	94
2013	44	4,398	33,670,781	94
2012	44	4,440	34,440,381	94
2011	46	4,575	35,504,333	90

2010	45	4,550	35,722,417	89

Abbreviations: SID, State Inpatient Database
[a] Potentially includes 50 states plus the District of Columbia.
[b] Number of hospitals include community, non-rehabilitation, non –long-term acute care hospitals.
[c] Represents the percent of discharges from U.S. community hospitals included in the reference population. Hospital-level QIs which rely on POA indicator information may have fewer discharges in the reference population.

B.1 Reference Population for Area-Level Indicators

Beginning with v5.0 (2015), all area-level indicators are developed using a reference population limited to community hospitals that excludes rehabilitation and long-term acute care (LTAC) hospitals.

ICD-10-CM/PCS v2022 used the 2019 HCUP SID. In 2019, 49 states, including DC, were available for area-level indicator development. States in the reference population for 2019 represent approximately 97 percent of the United States population, and include: AK, AR, AZ, CA, CO, CT, DC, DE, FL, GA, HI, IA, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, WV, and WY.

Residents of counties in states not contributing to HCUP are excluded from rate calculations. They are excluded because care received in those states, which is most of the care received by their residents, is missing from the reference population. Residents from some of the excluded counties travel to participating states to receive care, but their admissions are excluded from the numerator. Similarly, admissions for some residents of counties that are included are missing because these residents travel to nonparticipating states. National rates are slightly underestimated because admissions for participating states' residents traveling to nonparticipating states are not found in the numerator but the residents are included in the denominator. The information needed to adjust calculations for the undercount is lacking currently, so this method produces the most accurate possible rates.

This methodology can be seen in Table 13 below. The reference population includes patients resident in HCUP states and admitted to hospitals in HCUP states.

Table 13. Treatment of State Border Crossing Discharges in the Reference Population

	Admission in HCUP State	Admission in Non-HCUP State
Patient county in HCUP State	Observed in SID and included reference population	Not observed in SID
Patient county in non-HCUP State	Observed in SID, not included reference population	Not observed in SID

Abbreviations: HCUP, Healthcare Cost and Utilization Project; SID, State Inpatient Database

B.2 Reference Population for Hospital-Level Indicators

Beginning with v5.0 (2015), all hospital-level indicators are developed on a reference population with complete POA information. The reference population file is limited to community hospitals and also excludes rehabilitation and LTAC hospitals.

The v2022 software uses the 2019 HCUP SID. In 2019, 46 states in the SID included indicators of the diagnoses being POA (CT, DE, and WY did not report POA data), and 42 included the days to procedure from admission, and had accurate MDC coding based on principal diagnosis not on pre-MDC classifications. Note that Colorado, Hawaii, and New York were also not used in the development of PSI

17. Wisconsin, Oklahoma, West Virginia, Delaware were not used to develop PSI 04, PSI 09, PSI 10, PSI 11, PSI 12, PSI 14, and PSI 15. Edit checks on POA were developed during an HCUP evaluation of POA coding in the 2011 SID at hospitals that were required to report POA to CMS.[38] The edits identify general patterns of suspect reporting of POA. The edits do not evaluate whether a valid POA value (e.g., Y or N) is appropriate for the specific diagnosis.

There are three hospital-level edit checks:

1. Indication that a hospital has POA reported as Y on all diagnoses on all discharges
2. Indication that a hospital has POA reported as missing on all non-Medicare discharges
3. Indication that a hospital reported POA as missing on all nonexempt diagnoses for 15 percent or more of discharges. The cut-point of 15 percent was determined by 2 times the standard deviation plus the mean of the percentage for hospitals that are required to report POA to CMS.

States in the POA reference population for 2019 represent approximately 96 percent of the United States population, and include: AK, AR, AZ, CA, CO, DC, FL, GA, HI, IA, IL, IN, KS, KY, LA, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NM, NV, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, and WV.

C. Other Data Used for Area-Level Indicator Development

The v2022 AHRQ QI specifications rely on population estimates derived from other data sources, including the U.S. Census Bureau. Every year, the Census Bureau releases postcensal population estimates[39] (as of July 1 of each year) that are generated with the assistance of the Federal State Cooperative Program for Population Estimates (FSCPE) using residence, total births, total deaths, and net migration. With each new issue of July 1 estimates from the Census Bureau, the Census Bureau makes revisions to all years back to the last decennial census. Each decade, after a decennial census, the Census Bureau produces a set of intercensal estimates that provide annual population estimates that are adjusted to smooth the transition from one decennial census to the next. These estimates are used to derive the denominator for area-level indicators. The v2022 2000-2020 AHRQ QI Population File is available at:

https://qualityindicators.ahrq.gov/Downloads/Software/SAS/V2022/2000-2021_Population_Files_V2022.zip

As described in Chapter II.E, the area-level indicators also include an optional poverty variable obtained from Census Bureau Small Area Income and Poverty Estimates (SAIPE). The v2022 AHRQ area-level QIs use SAIPE estimates from 2019, available at:

https://www2.census.gov/programs-surveys/saipe/datasets/2019/2019-state-and-county/est19all.xls

---

[38] Barrett ML, Owens PL, Bolhack J, Sheng M. Examination of the Coding of Present-on-Admission Indicators in Healthcare Cost and Utilization Project (HCUP) State Inpatient Databases (SID). 2015. HCUP Methods Series Report #2015-06 ONLINE. September 1, 2015. U.S. Agency for Healthcare Research and Quality. Available: https://www.hcup-us.ahrq.gov/reports/methods/2015-06.pdf

[39] "Population projections are estimates of the population for future dates. They are typically based on an estimated population consistent with the most recent decennial census and are produced using the cohort-component method." U.S. Census. Population Projections. https://www.census.gov/programs-surveys/popproj.html. Accessed June 23, 2021.

D. Coding Updates

D.1 ICD-10-CM/PCS Coding Updates and Coding Guidelines

On October 1, 2015 (FY 2016), ICD-10-CM/PCS became the CMS standard for administrative data. Beginning in FY 2017 (October 1, 2016), new ICD-10-CM/PCS codes and revisions to existing codes are added annually. The codes are maintained by the ICD-10 Coding and Maintenance Committee. The v2022 AHRQ QI software updates all measure specifications to reflect coding updates for ICD-10-CM/PCS codes effective as of October 1, 2021.[40]

Information on ICD-10-CM/PCS coding updates is located on the NCHS and CMS websites, respectively:

- http://www.cdc.gov/nchs/icd/icd10cm.htm
- https://www.cms.gov/medicare/icd-10/2022-icd-10-cm

D.2 Fiscal Year Coding Updates to Classification Schemes

CMS updates the MS-DRGs, MDCs, OR procedures, valid principal procedures, and POA exempt codes for ICD-10-CM/PCS on an annual basis. Annual updates to these classification schemes may impact the numerators of all indicators and the denominators of all hospital-level indicators. Annually, these changes are reviewed to determine how the changes impact the QIs and their risk models and whether coding changes should result in changes to the QI specifications. In general, the QI specifications align with CMS definitions of OR procedures[41] and POA exempt codes;[42] however, the QIs use a modified version of the CMS OR procedure list to better capture procedures occurring in an OR setting.

In addition, organizations external to the AHRQ QI program update algorithms based on the ICD-10-CM/PCS system that are utilized in the risk models for the PSI, PDI and IQI. These include AHRQ Elixhauser Comorbidity Software Refined (PSI and hospital-level PDI risk model),[43] AHRQ's Clinical Classification System Refined for ICD-10-CM Diagnoses (IQI risk model and hospital-level PDI risk model),[44] AHRQ's Clinical Classification System Refined for ICD-10-PCS Procedures (IQI risk model model),[45] and AHRQ Procedure Classes (hospital-level PDI risk model).[46] Updates to these systems were incorporated in the risk models annually up to at least FY2019.

---

[40] For more information about the ICD-10-CM/PCS codes used in AHRQ QIs, see https://qualityindicators.ahrq.gov/News/ICD10_v2018_FAQ.pdf.

[41] ICD-10-CM/PCS MS-DRG v39.0 operating room procedures and procedure codes available at: https://www.cms.gov/icd10m/version39.0-fullcode-cms/fullcode_cms/P0001.html

[42] Centers for Medicare & Medicaid Services. https://www.cms.gov/files/zip/2022-poa-exempt-codes-updated-02152022.zip

[43] Elixhauser Comorbidity Software Refined for ICD-10-CM Healthcare Cost and Utilization Project (HCUP). October 2021. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp.

[44] Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. Healthcare Cost and Utilization Project (HCUP). February 2022. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp.

[45] Clinical Classifications Software Refined (CCSR) for ICD-10-PCS Procedures. Healthcare Cost and Utilization Project (HCUP). April 2022. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/prccsr.jsp.

[46] Procedure Classes Refined for ICD-10-PCS. Healthcare Cost and Utilization Project (HCUP). April 2022. Agency for Healthcare Research and Quality, Rockville, MD. https://www.hcup-us.ahrq.gov/toolssoftware/procedureicd10/procedure_icd10.jsp.

D.3 Changes to Data Elements on the Uniform Bill

As noted above, the reference population for the AHRQ QIs is based on administrative data with data elements consistent with the UB-04. At times, the National Uniform Bill Committee (NUBC) update the Uniform Bill and include changes to or additions to the data elements available on the UB-04, including but not limited to changes in source of admission and present on admission information.

Guidelines for POA Coding are provided in the ICD-10-CM/PCS Official Guidelines for Coding and updated annually by CMS and NCHS.[47] Changes to the POA guidelines impact the PSI and PDI numerators and denominators. These guidelines are reviewed and if necessary changes are made to QI specifications. In addition, POA coding impacts the reference population for the PSI, PDI, and IQIs. Changes to POA coding guideline have the potential of necessitating a change to the POA hospital- and discharge-level edits for the reference population.

Several other data elements are used in the QI specifications. Point of origin describes the “source of the referral for this admission or visit." Previously the Uniform Bill used the "Source of Admission" data element, which differed in that it described the venue immediately prior to hospitalization. Source of Admission is no longer used in the UB-04 but some states (notably California) use Source of Admission. To account for the transition time, the QIs use both source of admission and point of origin based criteria when feasible. Discharge status is also used in the AHRQ QI specifications. Annual updates to the UB-04 are reviewed and if applicable changes are made to the specifications.

E. Reference Population: Numerators, Denominators, and Observed Rates

E.1 Calculating Numerators, Denominators and Observed Rates

For each QI, numerators, denominators, and observed rates are calculated using hospital discharge data from an aggregation of the HCUP SID State files. The methods used for these calculations are described in Chapter III.E.2 and Chapter III.F.4. These calculations are updated annually.[48] National benchmark rates are currently provided by AHRQ.[49]

E.2 Evaluating the Numerators, Denominators and Observed Rates

Nationwide rates from the reference population for all QIs by module are compared against previous estimates to check for expected (i.e., changes to indicator specifications) and unexpected rate changes.

F. Reference Population: Risk Model Development and Parameter Estimates (v2022)

[47] ICD-10-CM Official Guidelines for Coding and Reporting, FY 2022 (October 1, 2021 – September 30, 2022. Centers for Medicare & Medicaid Services (CMS), and National Center for Health Statistics (NCHS). https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf. Accessed April 14, 2022.

[48] These calculations were not updated in years when the reference population was unavailable. See Table 11 for more details.

[49] Reference population rates are published on the AHRQ QI website in documents named Benchmark Tables (formerly known as Comparative Data Tables; see Chapter III.B).

F.1 Rationale for Risk Adjustment

The AHRQ QIs use empirically derived risk models based on a clinically coherent set of candidate variables.[50] The goal of risk adjustment should be distinguished from the goal of a prediction model. A prediction model uses all available information to maximize the prediction of an event. A risk model aims standardize observed performance as a function of factors independent from quality of care. Risk models may have lower performance than prediction models (e.g., statistics in a logistic regression model). For hospital-level QIs, risk models incorporate only factors that are present on admission and unrelated to quality, such as the clinical characteristics of patients at admission. Including risk adjustment variables that are the potential consequences of care quality, such as complications of care, length of stay, or hospital characteristics, will improve a model's predictive ability but may adjust away the very quality differences we are trying to illuminate.

The AHRQ QI program carefully assesses the need for each individual risk adjuster. First, candidate variables are independent from quality of care. Second, variables must be observable and valid using administrative data across hospitals. Third, the variables should reflect characteristics or factors that are plausibly clinically related to the outcome. Fourth, the candidate variables must be frequent enough to obtain reasonably precise estimates of risk, but adequately homogenous such that risk is not masked. Fifth, the risk factors should vary systematically by hospital, such that inclusion adds information to the model.

With these considerations in mind, the hospital-level QI models were developed to include as large a set of clinically meaningful, reliable, and valid risk factors as were found to influence the outcome. Thus, the model goals are shifted towards including as many covariates as theoretically justified and computationally practical, on an indicator-by-indicator basis.

For area-level QIs, risk adjustment aims to account for differences in demographics that are not mutable. In addition, risk adjustment helps to simplify interpretation by removing aspects that may impact hospital utilization but are of less interest to the user. Because users of the area level measures may have different needs for risk adjustment, observed (non-adjusted), age-sex adjusted and age-sex-poverty adjusted models are available. Area-level risk adjustment is limited by the data that are nationally available at the county level. In general, clinical factors are not available. However, because the QIs measure population health, development of chronic disease or the rapid progression of chronic disease may also reflect poor access to care and community based resources to promote health.

There is wide agreement on most aspects of risk adjustment. The NQF provides one consensus guideline on the formal criteria for the design of valid risk adjustment of outcome measures. The NQF's Measure Evaluation Criterion for scientific acceptability of outcome measures[51] states:

For outcome measures and other measures when indicated (e.g., resource use): an evidence-based risk adjustment strategy (e.g., risk models, risk stratification) is specified. The risk model and stratification

---

[50] The previous ICD-9-CM v6.0 software included risk adjustment, while the ICD-10-CM/PCS v6.0, 7.0, and v2018 software did not. This is because the AHRQ QI program requires one full year of data to improve the integrity of the risk models. At the time of their release, the ICD-10-CM/PCS v6.0, 7.0, and v2018 software did not have access to a full year of ICD-10-CM/PCS coded data, and thus did not allow for the calculation of risk-adjusted rates.

[51] Measure Evaluation Criteria. National Quality Forum (NQF).
https://www.qualityforum.org/measuring_performance/submitting_standards/measure_evaluation_criteria.aspx
Accessed April 14, 2022

approach is based on patient factors that influence the measured outcome (but not factors related to disparities in care or the quality of care) and are present at start of care. The models have also demonstrated adequate discrimination and calibration **OR** rationale/data that supports no risk adjustment/ stratification.

F.2. Construction of Candidate Covariates for Risk Adjustment

For the PQIs, potential risk adjustment covariates indicate whether the discharge record meets the technical specification for sex, age in 5-year groups that are used as covariates in the risk adjustment model. Alternatives to poverty category as measures of social risk (optional) are also considered.

For the IQIs, potential risk adjustment covariates indicate whether the discharge record meets the technical specification for sex, age, MDCs, and CCSR for Diagnoses and Procedures dummy variables that are used as covariates in the risk adjustment model.[52]

Age-sex categories are always included in the final risk model. Age-sex categories span 10-year intervals. The reference (omitted) category for the age-sex interaction categories for the IQI is "65-74 year-old women." The oldest and youngest age categories may be insufficiently populated to produce stable results. As a result, age categories may be collapsed such that there are a minimum of three age categories within each sex and any additional categories have at least 5 numerator events in the reference population.

Transfer-in from another acute care facility is included in final models for IQIs related to medical diagnoses (as opposed to IQI related to surgical procedures). For other measures transfer status is eligible for variable selection, except IQI 11 and IQI 17A and IQI 17B, where the empirical relationship lacks face validity. Do Not Rescusitate (DNR) status present on admission is included in the pool of candidate risk adjustment variables for IQIs related to medical diagnoses. For IQI 12 and IQI 30, an indicator for cardiogenic shock, anoxic brain injury, and cardiac arrest present on admission is included in the pool of candidate risk adjustment variables. For IQI 15, an indicator for non-ST-evelation myocardial infarction is included in the pool of candidate risk adjustment variables.

To be included in the pool of candidate risk adjustment variables, there must be at least 30 denominator records for that covariate (e.g. >30 denominator cases for CCSRs).

The CCSR tool for ICD-10-CM diagnoses was developed as part of HCUP. The CCSR aggregates over 70,000 ICD-10-CM diagnosis codes into a manageable number of clinically meaningful categories. Embedded in the AHRQ QI v2022 software, the CCSR generates codes that are used as covariates for hospital-level IQI and PDI risk adjustment. The CCSR codes take two values in AHRQ QI v2022:

- 0 – The CCSR was not triggered by any ICD-10-CM/PCS diagnosis code POA on the input record.
- 1 – The CCSR was triggered by any ICD-10-CM/PCS diagnosis code POA on the input record.

The followings CCSRs in the "Factors influencing health status and contact with health services" category were removed from the IQI and PDI risk adjustment due to the heterogenous diagnoses and/or the Z codes included within these categories: FAC002, FAC003, FAC005, FAC007, FAC008, FAC010, FAC011, FAC012, FAC013, FAC016, FAC017, FAC019, FAC021, FAC024, and FAC025.

[52] CCSR for Diagnoses are modified to be based on diagnoses that are POA and CCSR for Procedures are modified to only include procedure flags on or prior to the day of the denominator procedure.

The CCSR tool for ICD-10-PCS procedures was developed as part of HCUP. The CCSR aggregates over 80,000 ICD-10-PCS procedure codes into a manageable number of clinically meaningful procedure categories. Embedded in the AHRQ QI v2022 software, the CCSR generates codes that are used as covariates for hospital-level IQI risk adjustment for the procedure-based IQIs. The CCSR codes take two values in AHRQ QI v2022:

- 0 – The CCSR was not triggered by any ICD-10- PCS diagnosis code occurring on or before the day of the IQI denominator procedure-day on the input record.
- 1 – The CCSR was triggered by a ICD-10-PCS diagnosis code occurring on or before the day of the IQI denominator procedure-day on the input record.

For the PSIs, potential risk adjustment covariates indicate whether the discharge record meets the technical specification for sex, age, Modified Medicare Severity-Diagnostic Related Groups (MDRGs); Major Diagnostic Categories (MDCs); a list of 38 Elixhauser comorbidity variables[53]; Elixhauser comorbidity count categories; whether the patients was transferred in to the hospital; for PSI 04, variables indicating the severity of the condition; for PSI 03, 06, 07, 08, 15, variable indicating medical/surgical discharges; for PSI 02, Do Not Resuscitate (DNR) status present on admission; for PSI 13 and PSI 15, variables indicating procedure type or immune risk categories.

Age-sex categories are always included in the final risk model. Age-sex categories span 5-year intervals. The reference (omitted) category for the age-sex interaction categories for the PSI is "65-69 year-old women."

Several MDRG variables were excluded from consideration as candidate variables because assignment to these MDRG related MS-DRGs could be due to an in-hospital complication, pre-MDC MS-DRGs, including MS-DRGs for ECMO and tracheostomy, those operating room procedures that are not related to principal diagnosis and certain high resource intensity MS-DRGs, postoperative or post-traumatic infections, ungroupable MS-DRGs or ones that represent a major coding error.

To be included in the pool of candidate risk adjustment variables, there must be at least 30 denominator records for that covariate (e.g. >30 denominator cases for the MDRG).

For the PDIs, risk adjustment was not performed for v2019 but was performed for v2020,v2021, and v2022. Risk adjustment for the PDIs include age, sex, interaction between age and sex, birth weight, MDRG, and CCSR dummy variables as the covariates. For NQI 03, age in days instead of age, and interaction between sex and age in days are included as covariates. We also constructed procedural and diagnostic risk categories based upon clinical input. For PDI 01 and PDI 10, we included variables indicating procedure type or immune risk categories. As of v2022, count categories of Elixhauser Comorbidity code are also included as risk factors in the PDI risk adjustment models to account for comorbidity status.

Several MDRG variables were excluded from consideration as candidate variables because assignment to these MDRG related MS-DRGs could be due to an in-hospital complication, pre-MDC MS-DRGs, including MS-DRGs for ECMO and tracheostomy, those operating room procedures that are not related to principal diagnosis and certain high resource intensity MS-DRGs, neonates died or transferred to

[53] Elixhauser Comorbidity Software Refined for ICD-10-CM Healthcare Cost and Utilization Project (HCUP). October 2020. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp.

another acute care facility, postoperative or post-traumatic infections, ungroupable MS-DRGs or ones that represent a major coding error.

The area-level PDIs do not undergo variable selection, and always include sex*age (in 5-year groups) and poverty category (optional) as covariates in the risk adjustment model.

F.3 Feature Selection

For the area-level indicators, the models use the complete set of covariates for sex, age in 5-year age groups, an interaction with sex*age. There is also an optional set of covariates for poverty category based on the county of patient residence. Poverty may be useful as a covariate for applications that wish to isolate factors unrelated to poverty, or to identify areas that have better outcomes than would be expected based on the poverty of the population. For other applications, adjusting for poverty may mask important disparities in population health.

For hospital-level indicators, the models use demographic and clinical factors. On the basis of cross tabulations between each covariate and the outcome of interest, only those covariates with at least 30 denominator cases are used in the risk adjustment models. The omitted covariate within mutually exclusive categories is the reference group for those categories. Reference categories are usually (1) the most common and/or (2) the least risk, or (3) the median category. The choice of omitted reference category does affect how one might describe the parameter coefficients in words, but it does not affect predicted probabilities or model performance.

Variables for inclusion in the final risk adjustment models are selected by the least absolute shrinkage and selection operator (LASSO) selection method.[54] Due to the computation resource limitations, one million discharges are randomly selected if the reference population is larger than one million. The LASSO method is used because the traditional p-value or stepwise based selection methods use sequential fitting, which could lead to biased estimates of R-square, coefficients, and local optimal models. The advantage of using LASSO is that LASSO is a global optimization procedure to find the global optimal model satisfying certain restrains on the covariates coefficients. For v2022 software, we changed from SAS LASSO selection to R LASSO selection (glmnet function in glmnet library) due to the following main advantages found in R:

1. R program is able to use cross validation to tune the penalty parameter (λ) which will make the prediction based statistics more accurate to avoid potential overfitting/underfitting of the risk adjustment models.
2. R program output has multiple statistics that can be used to tune λ such as C-statistics, mean squared errors, deviance, etc.

The final multivariable model parameters are published on the AHRQ website (see Chapter III.B).

F.4 Estimate the Models

Area-level indicators use logistic models. Hospital-level models are estimated using GEEs (hierarchical modelling) to account for within-hospital correlation. These models are run with PROC GENMOD and use a logit link with an exchangeable correlation matrix.

---

[54] Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the lasso". Journal of the Royal Statistical Society. Series B.

F.4.1 Model Specification

The final model is determined as follows. First, a maximally inclusive set of candidate variables available from the data are evaluated by the module team with clinical and subject matter expertise. Decisions are made about which variables to include as candidate variables, how to handle age-sex categories, and whether to include any additional administrative variables (e.g., transfer-in status). Variables are excluded based on clinical considerations, known unreliability, potential for reflecting complications versus comorbidities, and face validity. These decisions result in an initial model specification that includes all remaining candidate variables (i.e., a saturated model). From the saturated model, variables are considered for removal by a LASSO selection process. The final subset of variables is included in a logistic regression model estimated by generalized estimating equations, clustered on Hospital ID.

F.4.2 Parsimonious Models

A paper by Osborne et al. about registry-based quality measurement evaluated whether risk adjustment models with fewer variables were as useful for indirect adjustment as models with more variables.[55] The authors' motivation for this work was to reduce the number of variables needed for risk adjustment because the cost of collecting additional variables for hospitals was high. The goal was, therefore, to reduce the number of variables that hospitals needed to measure without sacrificing too much in the way of accuracy.

The AHRQ QIs do not rely on expensive data collection methods for additional information, so from the standpoint of resources, we as a project team are not motivated by the concerns in the Osborne et al. paper. It is important to note that although some QI models have more than 100 variables, these are based on just a handful of administrative data elements (age, sex, transfer status, principal and secondary diagnoses) that are subsequently stratified. These data elements give rise to hundreds of categories within the MS-DRG variables, but each record has exactly one MS-DRG assigned. These additional categories help to more accurately assign patient-level risk based on the principal diagnosis. In other words, it assigns a specific level of risk to each MS-DRG, which reflects the clinical context about variation in risk by diagnoses.

Reducing the number of MS-DRG categories serves only to misclassify records with regard to the principal diagnosis, and should only be done when a stable estimate cannot be computed. In fact, the development data set (based on the HCUP reference population) are sufficiently large so that we can reliably estimate specific levels of risk for each MS-DRG in the risk adjustment model. The current approach may be conservative (tend to select fewer variables) relative to the rich data source available.

F.4.3 Collinearity

Collinearity arises when there is complete, or nearly complete, overlap in the information contained between two variables. Collinearity of covariates is well known to have no impact on predictive ability of a model.[56] However, excessive covariance between predictors can lead to large standard errors and unstable coefficients. The *p*-value based inclusion criterion for the model selection process tends to omit variables with large standard errors, eliminating that concern. In v2022 software development, we calculated variance inflation factor (VIF) for each covariate and dropped any covariates with larger than 1,000 VIF value. VIF is a measure of the extent variance of the estimated regression coefficient is

---

[55] Osborne NH, Ko CY, Upchurch GR, Dimick JB. Evaluating parsimonious risk adjustment models for comparing hospital outcomes with vascular surgery. J Vasc Surg. 2010 Aug.
[56] Berry WD, Feldman S. Multiple Regression in Practice. SAGE; 1985. p. 100.

"inflated" by the existence of correlation among the predictor variables in the model. The LASSO model selection procedure is also able to select variables that are not highly correlated given its heavy penalty on the variable coefficients.

All our models converge after the LASSO model selection procedure. At the same time, it is important to point out that the structure of the QI models inherently limits the possibility of collinearity. Collinearity could occur between, but not within, age-sex categories, transfer status, Elixhauser Comorbidity Software Refined, and MS-DRGs. There is no covariance within the mutually exclusive MS-DRGs.

F.4.4 Over-Parameterization

Over-parameterization is a concern that arises when the number of predictor variables is close to the number of records in the sample. With over-parameterization, the variances can be large and consequently the estimates of the regression coefficients can be unstable. The reference population database consists of many thousands, to millions, of observations, depending on QI in question. None of the models have a number of variables that approach the number of records in the reference population. Moreover, variable selection criteria require that a minimum of 30 records be present for each level of each covariate (e.g., at least 30 records for each MS-DRG). Variables that are under-populated are not included in models. The size of the dataset being used to make predictions is irrelevant to parameterization. The models could be used to compute a predicted probability for a single record.

F.4.5 Complete or Quasi-Complete Separation

Complete separation arises when a linear combination of predictor variables perfectly classifies (separates) the outcome variable. Quasi-complete separation is the analogous situation in which the separation is not quite complete. The AHRQ QI regression models are monitored for convergence criterion during variable selection and in the final model estimating stage. For variables that are forced into the model (e.g., age-sex categories) the solution to separation is to identify the variable(s) causing the separation and collapse the variable with the adjacent category closer to the reference group or drop them.

F.5 Calculate Rates

F.5.1 General Description

In order to make fairer comparisons among hospitals with different types of patients, the AHRQ QIs use indirect standardization to calculate risk-adjusted rates. The risk-adjusted rate using an indirect standardization approach equals the reference (general or standard) population observed rate multiplied by the ratio of observed rate in the user's sample divided by expected rate in the user's sample:

$$RAR_h = \alpha \cdot \frac{O_h}{E_h}$$

When risk adjustment models are estimated using GEE, there can be small differences between the observed rate and the expected and risk-adjusted rates in the reference population.

After the new risk adjustment models are fit, expected values (i.e. record-level predicted probabilities) are output so that they can be used to calculate expected rates and risk-adjusted rates. These values can be output directly from the regression procedures, or can be calculated in a subsequent step by applying PROC SCORE and the regression coefficients to the data, Reference population rates and signal variances are calculated.

F.5.2 Special Case: Calculating Rates with Stratified Indicators

For PSIs and IQIs that have clinical strata, the risk-adjusted rate for the overall indicator is calculated as the observed-to-expected ratio multiplied by the reference population rate, where the record-level observed and expected values are summed across categories of risk strata. This approach differs from other AHRQ PSIs and IQIs without strata, in that each discharge-record's expected value is computed using one of the distinct stratum-specific risk adjustment models that correspond to an assigned stratum.

F.6 Calculate Signal-to-Noise Ratio and Variance Estimates

Reliability is a crucial measure for determining measure quality. Reliability is estimated by the variation of true hospital quality of care, known as the signal variance, and the variation of sampling within each hospital, known as the noise variance (see section E.3.6 for the formula used to calculate reliability of area-level indicators). In general, good reliability means that the sampling errors are very small, the variation of true quality of care across all hospitals is large, and that we can use this measure to distinguish hospitals' performance.

The noise variance can be estimated through the risk adjustment models using the predicted risks of discharges. The signal variance is more difficult to estimate and we have two general methods. Morris' method[57] is calculated through the empirical Bayes model (see Chapter II, section E.3.6). It uses an iterative method to estimate the signal variance under the assumptions that the hospital QIs are normally distributed within each hospital and the true hospital quality of care is also normally distributed among hospitals. There are two main issues with this method. The first issue is that the normal distribution assumption may not be true for certain hospital QIs. The second issue is that the iterative method may lead to a negative signal variance. So, when the second issue occurs, we will use a full Bayes-based method which can be implemented with the "PROC MCMC" procedure in SAS. Under this approach, we assume the prior for the true hospital quality of care follows a Gamma distribution, which gives more flexibility compared to the symmetric normal distribution. We use a non-informative prior for both parameters for the Gamma distribution and let the data estimate all the parameters, including the signal variance, through posterior distributions.

Hospitals present a varying number of denominators (i.e., eligible discharges) in the QI calculations. Statistically, this means that each hospital contributes a different amount of information than the next hospital; large hospitals with thousands of discharges contribute more information than small hospitals with, say, fewer than a hundred discharges. In the empirical Bayes framework, the hospital means (i.e., their "true" QI rates) are distributed around the reference population mean. The extent to which the hospital means are spread about the reference population mean is characterized by the signal variance. To calculate the signal variance, the reference population mean may account for the different amounts of information from large and small hospitals through a weighting scheme that places more weight on large hospitals and less weight on small hospitals. This distinction from the unweighted mean depends on the specific interpretation of QI results—that is, whether or not hospitals should be distinguished by their case sizes (i.e., denominators) in the estimation of the empirical Bayes smoothing model.

F.7 Evaluate Models

Two desirable qualities of risk adjustment models are that they discriminate well between discharge records that experience the outcome of interest and those that do not and that they are well calibrated,

[57] Morris, CN. Parametric empirical Bayes inference: theory and applications. J Am Statistical Assoc. 1983 Mar;78(381):47-55.

predicting that the outcome will occur in approximately the right proportions, over a wide range of predicted probability.

F.7.1 Discrimination

One common scalar measure of logistic regression discrimination is the c-statistic. This may be calculated by computing the area under the Receiver Operating Characteristic (ROC) curve. Alternatively, it may be calculated by forming every possible pair in a dataset in which one member of the pair is a discharge with the outcome of interest and the other member is a discharge without the outcome of interest. The c-statistic is the proportion of such pairs in which the predicted probability for the member with the outcome of interest is higher than the predicted probability for the other record. Pairs with tied probabilities each contribute one-half to the numerator and denominator of the proportion. A c-statistic of 0.5 is the same discrimination performance as flipping a coin. A c-statistic of 1.0 indicates perfect discrimination. Hosmer and Lemeshow[58] have coined three widely adopted labels for discrimination performance based on the c-statistic:

- $0.70 \leq$ c-statistic < 0.80 indicates acceptable discrimination
- $0.80 \leq$ c-statistic < 0.90 indicates excellent discrimination
- $0.90 \leq$ c-statistic indicates outstanding discrimination

The c-statistics for the AHRQ QI risk adjustment models are published in on the AHRQ QI website in the Parameter Estimates Document: (see Chapter III.B)

F.7.2 Calibration

Calibration often is described by sorting the dataset on the basis of predicted probability and dividing it into deciles of risk. It is meaningful to compare the proportion of records in each decile that were observed to have the outcome of interest with the proportion of records that are expected to have that outcome. Hosmer and Lemeshow's[59] logistic regression goodness-of-fit statistic is based on a chi-square test statistic calculated using the observed and expected counts across the 10 deciles.

G. Composite Development

G.1 Area-Level Composites

The area-level composite QI are unweighted combinations of conceptually related component QI. The area-level QI composites are calculated as the count of discharges qualifying for any of the component indicators over the total population for all component measures. For example, the numerator for PQI 93 includes all records that qualify for any diabetes-related PQI (PQI 01, PQI 03, PQI 14, or PQI 16) over all adults 18+ years residing in an area. Observed and risk-adjusted rates for the area-level composites are computed using the same methods described for the individual component area-level QI.

G.2 Hospital-Level Composites

The hospital-level composites are all weighted composites (i.e., IQI 90, 91, PSI 90). They are calculated as the weighted average of the component indicator smoothed rate for each component indicator

[58] Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance model based on logistic regression. Statistics in Med. 1995;14(19):2161-72.
[59] Hosmer, D. W., & Lemesbow, S. Goodness of fit tests for the multiple logistic regression model. Communications in statistics-Theory and Methods. 1980;9(10), 1043-1069.

(composite rate = component weight * hospital smoothed component rate). All weighted composites use weights based on volume and reliability, except PSI 90 which uses weights based on volume and harm. See Section G.3.1 for details on the weight calculation.

G.3 Special Case: Hospital-Level Composite – PSI 90

G.3.1 Calculating Harms Weights for PSI 90 Composite

The PSI composite combines smoothed (empirical Bayes shrunken) standardized morbidity ratios (observed/expected ratios) from selected AHRQ PSIs to provide a composite that gives an overview of hospital-level quality as it relates to a set of hospital-related events that are associated with harmful outcomes for patients. In past versions of the AHRQ QI software PSI 90 (v5.0 and earlier) the weight that each component received was proportional to the volume of the events in the component indicator observed in the HCUP reference population (i.e., numerator weighting). The re-weighting of PSI 90 was undertaken to improve the validity and reliability of the composite by refining the component indicators that are included in the composite and aligning the weights with the burden of harm (risk of harmful outcomes) that each component contributes in a reference population. In other words, the new weights account for both the magnitude of harm associated with a patient safety event as well as the volume (number of cases) of the event, whereas in past iterations only the volume was used for weighting.

The new weights are defined and calculated as follows:

Each component PSI indicator, *q*, which is part of PSI 90 receives a weight defined by:

$$weight_q = \frac{volume_q \sum_{h=1}^{H} harm_{qh}\ disutility_{qh}}{\sum_{q=1}^{Q} volume_q \sum_{h=1}^{H} harm_{qh}\ disutility_{qh}}$$

Where:

Q is the total number of component quality indicators, *q*, in PSI 90.

H is the total number of outcome types (harms), *h*, related to each component indicator.

volume is the numerator count, or the number of total QI events within the component indicator in the reference population.

harm is the excess risk (risk difference) of each type of outcome (i.e. harm) within each component indicator estimated from a model comparing people with PSI events to those without PSI events in an “at risk” cohort.

disutility is the complement of a utility weight (1-utility_wt) assigned to each excess occurrence of each type of outcome within each component indicator.

For each component indicator in the PSI 90 composite, two sets of values need to be computed or estimated. The first is the excess risk of the outcomes (risk difference) that may occur as a consequence of the patient safety event associated with the indicator. The second is the set of numerator weights. Although estimates of disutility are required to incorporate disparate types of harms, the values of disutility are treated as not varying.

G.3.2 Harms Included

Harms weights were developed specifically for the AHRQ QIs. Based on literature review and expert opinion from 13 clinical specialists in surgery, internal medicine, nephrology, trauma and emergency care, critical care, nursing, and home healthcare, 37 downstream harms associated with 10 PSIs were defined (See Appendix D). For some PSIs, harms were included for up to one year after the PSI event (such as mortality, skilled nursing facility days, and outpatient dialysis). An expert panel then ranked the harms. These rankings, along with information from relevant studies in the literature, were then used to assign disutilities, or a measure of the severity of the adverse effects, associated with each of the harms.

G.3.3 Estimating Excess Harms

The estimates of excess harms that go into the harm weighting aim to answer the question, how much more likely is a particular harmful outcome in a population of patients who experience a PSI event than in a population of patients who were at risk for the event, but did not experience the event. In other words, what is the risk difference between PSI events and non-events in an at-risk population? These models require the use of longitudinal data that contain information about morbidity and mortality following a PSI event.

For version 2022 of the software, excess harms were modeled using CMS Inpatient and Outpatient Medicare Fee-For-Service data in the 100% standard analytical files (SAF). A separate cohort sample was defined for each component indicator based on the sample of 2012 patient records who were “at risk” (i.e., in the denominator) for the component QI indicator. Index events were identified as patient discharges in 2012 with an eligible QI PSI component event. The comparison group was composed of at risk patients (as defined by the component PSI specification) who did not experience the PSI event. The 2013 data were used solely to provide follow-up information about harms. The follow-up period was one year from the discharge date of the index hospitalization. For each component indicator, the independent variable was the presence or absence of the component PSI event. Separate models were fit for each harm outcome. Outcomes varied among the component PSIs. Example outcomes included all-cause 30-day and 180-day mortality, hospital readmissions, condition-specific complications, and total length of hospital stay (potentially including the postoperative period during the index admission plus all qualifying readmissions within the ascertainment window). The selection of outcomes relied on the underlying conceptual model for the component indicator, the available data elements in the CMS data, and the availability of a meaningful utility weight.

Confounding may arise if factors associated with the probability of experiencing a QI event are also related to the probability of experiencing a consequence (outcome) from the QI event. To account for potential confounding in these analyses, for each component indicator, we used a propensity score weighting approach. The propensity score (PS) was the predicted value (i.e. expected value) from the QI’s risk adjustment model, which accounted for age and sex as well as pre-existing complications and comorbidities. We used a version of propensity weighting suitable for estimating the average treatment effect on the treated (ATT). In other words, we estimated the effect of the safety event on harms among

patients who suffer the safety event. Patient stays with the safety event (QI=1) received a weight of 1 and at-risk patient stays without a safety event (QI=0) received a weight of PS/(1-PS).

Another potential source of confounding may arise from patients who experience multiple PSI events that share common outcomes (e.g., mortality). In this scenario, it is necessary to estimate independent associations between PSI events and outcomes. When multiple component PSIs are related to the same outcome, we included the other component PSIs in the model as covariates for the excess harm effect we were estimating. For example, if we are estimating the excess risk of renal failure in PSI 13, we would use propensity weights appropriate for PSI 13 and would also include PSI 10 as an indicator covariate in the model.

G.3.4 Harm Utility Values

To combine disparate harms into a single overall weight, we applied disutility values that scale the relative utility of health states from a patient perspective. Utilities were anchored at zero for mortality and one for no harmful health outcome. When available, intermediate utility values were drawn from studies that examine patient preference for various health states (e.g. standard gamble studies). When literature-based utility values were not available for patient preference, we used an expert panel of clinicians (physicians and nurses) to rank a list of health states that they have seen in their patients. We applied a regression process to interpolate utility values based on the consensus ranking of the health states. Disutility was calculated as the complement of utility (i.e., 1-utility).

G.3.5 Final PSI 90 Weight

The final PSI 90 weights were computed using the excess harm and disutility values derived from the procedures above and combined with information about the volume of the PSI 90 components in the 2019 reference population. The v2022 AHRQ QI software contains weights for PSI 90 based on 10 component PSI indicators (Table 14). The component weights are re-scaled to sum to 1.0.

Table 14. Weights of PSI 90 Component Indicators, v2022

Indicator abbreviation	Indicator name	Harm weight	Volume weight	Component weight
PSI 03	Pressure Ulcer Rate	0.3080	0.1068	0.1669
PSI 06	Iatrogenic Pneumothorax Rate	0.1381	0.0435	0.0305
PSI 08	In Hospital Fall with Hip Fracture Rate	0.1440	0.0199	0.0145
PSI 09	Postoperative Hemorrhage or Hematoma Rate	0.0570	0.1499	0.0434
PSI 10	Postoperative Acute Kidney Injury Requiring Dialysis Rate	0.3584	0.0314	0.0572
PSI 11	Postoperative Respiratory Failure Rate	0.2219	0.2129	0.2397
PSI 12	Perioperative Pulmonary Embolism or Deep Vein Thrombosis (PE/DVT) Rate	0.1557	0.2288	0.1808
PSI 13	Postoperative Sepsis Rate	0.3102	0.1367	0.2151

Indicator abbreviation	Indicator name	Harm weight	Volume weight	Component weight
PSI 14	Postoperative Wound Dehiscence Rate	0.1441	0.0238	0.0174
PSI 15	Abdominopelvic Accidental Puncture or Laceration Rate	0.1474	0.0461	0.0345

G.3.6 Estimating PSI 90 Variance

The within-hospital variance for the PSI 90 Composite characterizes the statistical uncertainty around the result that arises from sampling at the discharge level. The hospital's discharges in PSI 90 calculation are assumed to have be drawn from an infinite population of similar, eligible discharges; the random differences between sample and population are what constitutes the sampling error for within-hospital variance. For a component indicator, the within-hospital variance is the *noise variance* associated with that indicator; see section F.4 of Quality Indicator Empirical Methods.

The PSI 90 Composite is a weighted sum of the component indicators. Essentially, the AHRQ QI software computes a within-hospital PSI 90 variance based on this weighted sum; the variance calculation can be derived from the signal variance of the component PSI (in the reference population), final PSI 90 weight (specific to the measure's definition; see section G.3.5), and the hospital's reliability weight. This calculation is based on the assumption of independence among the component PSIs – that is, component PSI rates are uncorrelated within hospitals.

From the statistical perspective, the resulting PSI 90 Composite variance may be sensitive to the assumption of independence across component PSIs. In other words, correlated PSIs would contribute *less* information in the composite value (than if they were independent), which indicates that the variance would be underestimated. To assess the sensitivity of the variance, the analyst could apply bootstrap methods to simulate the within-hospital variance-covariance of component indicators in the PSI 90 Composite. In developing and testing a bootstrapped approach, the size of the reference population in the SID and the requisite number of bootstrap iterations should be taken into account.

H. Empirical Testing – Evaluating AHRQ QI Specifications and Risk Models

The AHRQ QI are routinely evaluated to ensure continued scientific soundness. This section describes selected routine testing. In addition to the routine testing, additional analyses are conducted on an ad hoc basis to assess specific aspects of indicator performance as part of the continuous improvement cycle. Testing is completed using the HCUP SID data reference populations, meaning that all testing reflects indicator performance in an all-payer population.

H.1 Reliability

Broadly defined reliability refers to the consistency of a measure. In the context of quality measures, reliability can encompass multiple aspects of constancy:

1. Is a measure consistent when measured by multiple raters or using differing sets of data within the same time period? (inter-rater reliability)
2. Is a measure consistent when measured multiple times within a time period for which the measure is not expected to change? (test-retest reliability)

3. Is performance consistent when measured using different methods? (inter-method reliability)
4. Are measures within a scale or composite consistent? (internal reliability)
5. Does the measure consistently distinguish one measured entity from another? (signal-to-noise)

These types of reliability may be applied to the performance score itself or the categorization of the measured entity, such as the identification outlier hospitals. Each reliability metric describes a distinct aspect; different measure applications may favor different reliability.

To calculate the reliability weight, the QI modules use the signal and noise variances. These estimates come from the empirical Bayes shrinkage model that characterizes the distribution of QI between and within hospitals. In reliability testing, the overall reliability of the QI to distinguish hospitals on the basis of their underlying quality can be calculated as a weighted sum of the hospital-level reliability weights. This diagnostic would characterize the amount of total variation in QI rates than can be explained by the true quality of hospitals (i.e., the signal-to-noise ratio).

Alternative methods for testing reliability use different statistical frameworks. For example, a reliability analysis can be based on a beta-binomial model that posits an underlying beta distribution for the true QI rates and a binomial for the distribution of discharges within a hospital.[60] Other bootstrap-based methods such as test-retest reliability could be applied, whereby the reference discharge population is resampled in split halves to assess the agreement (or correlation) in QI rates between them; this approach would be computationally intensive.

Standards for reliability can differ by sources and purpose. For example, a reliability analysis for the Centers for Medicare & Medicaid Services (CMS) suggested a lower limit for "moderate" reliability at 0.4.[61] In addition to statistical considerations, reporting programs need to consider implications of minimum case sizes in the calculation of any quality measure, in order to ensure that reliability standards are met.

H.2 Validity

Validity testing is tailored for each measure. For instance, for AMI mortality testing examines the relationship of hospital-level rates with AMI process measures and readmission rates. The PQIs validity testing examines the relationship of county level rates with county-level access to care measures (e.g. insurance coverage, physician density), poverty and community characteristics that contribute to hospital utilization and access to care.

Two other types of validity have been assessed historically but this testing is not conducted routinely.

1. All measures have been assessed for face validity by at least one clinical expert panel using the modified RAND Appropriateness Method (i.e. nominal group method).[62] These panels

[60] Adams JL (2009). The reliability of provider profiling: a tutorial. RAND Technical Report #653. Prepared for the National Committee for Quality Assurance.
[61] Results of Reliability Analysis from Mathematica Policy Research. Centers for Medicare & Medicaid Services. February 13, 2012. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf. Accessed June 23, 2021.
[62] K. Fitch et al. (2001). The RAND/UCLA Appropriateness Method User's Manual.

recommend refinements to indicator specifications and rate the overall usefulness of the indicators.[63]

2. For the patient safety measures (PSI and PDI) chart review has been used to assess criterion validity, namely positive predictive value, negative predictive value, sensitivity and specificity of the coding to detect actual events. These studies were conducted using ICD-9-CM data by both research members of the QI development team and outside researchers. However, these studies should be viewed in the context of changes to the ICD-9-CM coding structure since the studies were conducted. In many cases, these studies informed improvements to the PSI specifications and/or to the ICD-9-CM coding structure or instructions that have improved the validity.

H.3 Risk Model Performance

Risk models are evaluated using tests of discrimination (how well the risk adjustment model distinguishes events from non-events) and calibration. The measure of discrimination is the c-statistic, also known as the area under a receiver operating characteristic curve. The c-statistic is computed by assigning each observation a predicted probability of the outcome from the risk adjustment model, based on the value of the observed covariates and the parameter estimates from the risk adjustment model. Two copies of the dataset are sorted, first from highest to lowest predicted probability and second from lowest to highest predicted probability. Random sampling is used to create a set of paired observations. Pairs that consist of one event and one non-event (discordant pairs) are kept and concordant pairs are discarded. The c-statistic represents the proportion of discordant pairs of observations for which the observation with the event had a higher predicted probability from the risk adjustment model than the observation without the event. Common "goodness of fit" tests are not used because these tests tend to be uninformative with large samples.

The metric for calibration is the evaluation of how closely observed and predicted rates compare across deciles of the predicted rate. This analysis splits the sample into deciles based on predicted rates, and then compares these rates with the observed rates for the population in each decile. A well calibrated model, or one that does not over or under-estimate risk, will have comparable observed and predicted rates across the risk spectrum.

[63] Most recently used by AHRQ QI Expert Panel Workgroup in summer of 2018

Chapter IV. References

Adams JL (2009). The reliability of provider profiling: a tutorial. RAND Technical Report #653. Prepared for the National Committee for Quality Assurance.

AHRQ HCUP tool Elixhauser Comorbidity Software Refined Tool, v2022.1: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR_v2022-1.zip

Berry WD, Feldman S. Multiple Regression in Practice. SAGE; 1985. p. 100.

Center for Health Policy/Center for Primary Care and Outcomes Research & Battelle Memorial Institute. Quality Indicator Measure Development, Implementation, Maintenance, and Retirement (Prepared by Battelle, under Contract No. 290-04-0020). Rockville, MD: Agency for Healthcare Research and Quality. May 2011.

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). October 2020. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp

Hosmer DW, Lemeshow S. Confidence interval estimates of an index of quality performance model based on logistic regression. Statistics in Med. 1995;14(19):2161-72.

Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. Communications in statistics-Theory and Methods. 1980;9(10), 1043-1069.

Iezzoni, Lisa, Ed. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago: Health Administration Press; 2013.

Luft HS, Brown BW Jr. Calculating the probability of rare events: why settle for an approximation? Health Serv Res. 1993;28(4):419-39.

Morris CN. Parametric empirical Bayes inference: theory and applications. J Am Statistical Assoc. 1983 Mar;78(381):47-55.

Osborne NH, Ko CY, Upchurch GR, Dimick JB. Evaluating parsimonious risk-adjustment models for comparing hospital outcomes with vascular surgery. J Vasc Surg. 2010 Aug;

Tibshirani R (1996). "Regression Shrinkage and Selection via the lasso". Journal of the Royal Statistical Society. Series B

https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/hospital-value-based-purchasing/Downloads/HVBP_Measure_Reliability-.pdf

2013 U.S. Census Bureau Small Area Income and Poverty Estimates, downloaded from: https://www.census.gov/programs-surveys/saipe/data/datasets.html

ICD-10-CM/PCS MS-DRG v38.1 Definitions Manual, available at https://www.cms.gov/icd10m/version38-1-fullcode-cms/fullcode_cms/P0001.html.

SAS. SAS/STAT 9.2 User's Guide. The SCORE Procedure (Book Excerpt). https://support.sas.com/documentation/cdl/en/statugscore/61828/PDF/default/statugscore.pdf.

Chapter V. Appendices

Appendix A. Other Helpful Documents

Readers may wish to access additional QI-related documentation. The following are some helpful examples:

AHRQ QI Technical Specifications

- PQI: https://qualityindicators.ahrq.gov/Modules/PQI_TechSpec_ICD10_v2022.aspx
- IQI: https://qualityindicators.ahrq.gov/Modules/IQI_TechSpec_ICD10_v2022.aspx
- PSI: https://qualityindicators.ahrq.gov/Modules/PSI_TechSpec_ICD10_v2022.aspx
- PDI: https://qualityindicators.ahrq.gov/Modules/PDI_TechSpec_ICD10_v2022.aspx

AHRQ QI Parameter Estimates Tables

- PQI: https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2022/Parameter_Estimates_PQI_v2022.pdf
- IQI: https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2022/Parameter_Estimates_IQI_v2022.pdf
- PSI: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2022/Parameter_Estimates_PSI_v2022.pdf
- PDI: https://qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2022/Parameter_Estimates_PDI_v2022.pdf

AHRQ QI Population Documentation File (used with area-level indicators)

https://qualityindicators.ahrq.gov/Downloads/Software/SAS/V2022/AHRQ_QI_v2022_ICD10_Population_File.pdf

AHRQ QI Software Instructions

- SAS: https://qualityindicators.ahrq.gov/Downloads/Software/SAS/V2022/Software_Inst_SASQI_v2022_July_2022.pdf
- WinQI: https://qualityindicators.ahrq.gov/Downloads/Software/WinQI/V2022/Software_Inst_WINQI_V2022_July_2022.pdf

AHRQ HCUP SID Documentation (to better understand the source of the reference population)

http://www.hcup-us.ahrq.gov/db/state/siddbdocumentation.jsp

Appendix B. Comprehensive List of Quality Indicators

Appendix Table 1. Area-Level Quality Indicators

Abbrev	Preventive Quality Indicators
PQI 01	Diabetes Short-Term Complications Admission Rate
PQI 03	Diabetes Long-Term Complications Admission Rate
PQI 05	Chronic Obstructive Pulmonary Disease (COPD) or Asthma in Older Adults Admission Rate
PQI 07	Hypertension Admission Rate
PQI 08	Heart Failure Admission Rate
PQI 11	Community-Acquired Pneumonia Admission Rate
PQI 12	Urinary Tract Infection Admission Rate
PQI 14	Uncontrolled Diabetes Admission Rate
PQI 15	Asthma in Younger Adults Admission Rate
PQI 16	Lower-Extremity Amputation among Patients with Diabetes Rate
PQI 90	Prevention Quality Overall Composite
PQI 91	Prevention Quality Acute Composite
PQI 92	Prevention Quality Chronic Composite
PQI 93	Prevention Quality Diabetes Composite
Pediatric Quality Indicators	
PDI 14	Asthma Admission Rate
PDI 15	Diabetes Short-Term Complications Admission Rate
PDI 16	Gastroenteritis Admission Rate
PDI 18	Urinary Tract Infection Admission Rate
PDI 90	Pediatric Quality Overall Composite
PDI 91	Pediatric Quality Acute Composite
PDI 92	Pediatric Quality Chronic Composite

Appendix Table 2. Hospital-Level Quality Indicators

Mortality Indicators	
IQI 08	Esophageal Resection Mortality Rate
IQI 09[a]	Pancreatic Resection Mortality Rate
IQI 11[a]	Abdominal Aortic Aneurysm (AAA) Repair Mortality Rate
IQI 12	Coronary Artery Bypass Graft (CABG) Mortality Rate
IQI 15	Acute Myocardial Infarction (AMI) Mortality Rate
IQI 16	Heart Failure Mortality Rate
IQI 17[a]	Acute Stroke Mortality Rate
IQI 18	Gastrointestinal Hemorrhage Mortality Rate
IQI 19	Hip Fracture Mortality Rate
IQI 20	Pneumonia Mortality Rate
IQI 30	Percutaneous Coronary Intervention (PCI) Mortality Rate
IQI 31	Carotid Endarterectomy Mortality Rate
IQI 90	Mortality for Selected Inpatient Procedures
IQI 91	Mortality for Selected Inpatient Conditions
Utilization Indicators	
IQI 21	Cesarean Delivery Rate, Uncomplicated
IQI 22	Vaginal Birth After Cesarean (VBAC) Delivery Rate, Uncomplicated
IQI 33	Primary Cesarean Delivery Rate, Uncomplicated
Patient Safety Indicators	
PSI 02	Death Rate in Low-Mortality Diagnosis Related Groups (DRGs)
PSI 03	Pressure Ulcer Rate
PSI 04[a]	Death Rate among Surgical Inpatients with Serious Treatable Complications
PSI 05	Retained Surgical Item or Unretrieved Device Fragment Count
PSI 06	Iatrogenic Pneumothorax Rate
PSI 07	Central Venous Catheter-Related Blood Stream Infection Rate
PSI 08	In Hospital Fall with Hip Fracture Rate
PSI 09	Postoperative Hemorrhage or Hematoma Rate

PSI 10	Postoperative Acute Kidney Injury Requiring Dialysis Rate
PSI 11	Postoperative Respiratory Failure Rate
PSI 12	Perioperative Pulmonary Embolism or Deep Vein Thrombosis Rate
PSI 13	Postoperative Sepsis Rate
PSI 14	Postoperative Wound Dehiscence Rate
PSI 15	Abdominopelvic Accidental Puncture or Laceration Rate
PSI 17	Birth Trauma Rate – Injury to Neonate
PSI 18	Obstetric Trauma Rate – Vaginal Delivery with Instrument
PSI 19	Obstetric Trauma Rate – Vaginal Delivery without Instrument
PSI 90	Patient Safety and Adverse Events Composite
Pediatric Quality Indicators	
NQI 03	Neonatal Blood Stream Infection Rate
PDI 01	Accidental Puncture or Laceration Rate
PDI 05	Iatrogenic Pneumothorax Rate
PDI 08	Postoperative Hemorrhage or Hematoma Rate
PDI 09	Postoperative Respiratory Failure Rate
PDI 10	Postoperative Sepsis Rate
PDI 12	Central Venous Catheter-Related Blood Stream Infection Rate
PDI 14	Asthma Admission Rate
PDI 15	Diabetes Short-Term Complications Admission Rate
PDI 16	Gastroenteritis Admission Rate
PDI 18	Urinary Tract Infection Admission Rate
PDI 90	Pediatric Quality Overall Composite
PDI 91	Pediatric Quality Acute Composite
PDI 92	Pediatric Quality Chronic Composite

[a]Includes stratum-specific indicators.

Appendix Table 3. Quality Indicators Dependent on Present on Admission (POA) Information

Indicator	POA-Dependent Quality Indicator
PSI 02	
PSI 03	X
PSI 04	
PSI 05	X
PSI 06	X
PSI 07	X
PSI 08	X
PSI 09	X
PSI 10	X
PSI 11	X
PSI 12	X
PSI 13	X
PSI 14	X
PSI 15	X
PSI 17	
PSI 18	
PSI 19	
PSI 90	
NQI 03	X
PDI 01	X
PDI 05	X
PDI 08	X
PDI 09	X
PDI 10	X
PDI 12	X
PDI 14	

Indicator	POA-Dependent Quality Indicator
PDI 15	
PDI 16	
PDI 18	
PDI 90	
PDI 91	
PDI 92	

Appendix C. Comprehensive Lists of Risk Factors for Quality Indicator Modules Appendix

Appendix Table 4. Risk Factors by Module at the Area-Level

Data Element	PQI	PDI
AGE	X	X
SEX	X	X
POVERTY	X	X

Appendix Table 5. Risk Factors by Module at the Hospital-Level

Category	IQI	PSI	PDI	NQI
Demographics	Sex[a]	Sex[a]	Sex[a]	Sex[a]
	Age[a]	Age[a]	Age in days (90 days–1 year)[a] Age in years (1 year+)[a]	Age in days (0 or 1 day)[a]
Severity of Illness		Modified MS-DRG[b]	Modified MS-DRG[b]	Modified MS-DRG[b]
		MDCs[b]		
Clinical / Comorbidities	AHRQ CCSR for Diagnoses[c]		AHRQ CCSR for Diagnoses[c]	
	AHRQ CCSR for Procedures[b,d]	AHRQ Comorbidities (using latest guidance on POA requirements[64])[b]		
		Count categories of AHRQ Comorbidities (using latest guidance on POA requirements)[b]	Count categories of AHRQ Comorbidities (using latest guidance on POA requirements)[b]	
		Do Not Resuscitate (POA)		
			Indicator-specific risk stratifiers[b]	

[64] AHRQ HCUP tool Elixhauser Comorbidity Software Refined Tool, v2022.1: https://www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR_v2022-1.zip

Category	IQI	PSI	PDI	NQI
				Birth weight (500g groups)
Discharge specific information	Transfer-in status[b]	Transfer-in status[b]	Transfer-in status[b]	Transfer-in status[b]
		Surgical/Medical discharge (MS-DRG)b	Surgical/Medical discharge (MS-DRG)[b]	
Stratified risk groups	Indicator-specific risk stratifiers	Indicator-specific risk stratifiers		

Abbreviations: AHRQ, Agency for Healthcare Research and Quality; CCSR, Clinical Classification Software Refined; IQI, Inpatient Quality Indicator; MDC, major diagnostic category; MS-DRG, Medicare Severity Diagnostic Related Group; NQI, Neonatal Quality Indicator; PDI, Pediatric Quality Indicator; PSI, Patient Safety Indicator; QI, Quality Indicator

[a] Categories are mutually exclusive.
[b] Variable or variable categories are selected into model for some indicators.

[c] AHRQ CCSR for Diagnoses are modified and additional comorbidity groups are also included.
[d] AHRQ CCSR for Procedures are modified to only include procedures on or before the day of the denominator procedure.

Appendix D. Patient Harms Captured in the AHRQ Patient Safety and Adverse Events Composite

Appendix Table 6. Description of Patient Harms Captured in the AHRQ Patient Safety and Adverse Events Composite

Outcome	Description of Events Captured	Applicable Patient Safety Indicator (PSI)
Pressure ulcer treatment	Debridement of a pressure ulcer and/or surgical skin flap procedure during the hospitalization when the pressure ulcer developed, due to tissue damage.	PSI 03
180-day hospital readmission for a pressure ulcer-related complication	Readmission to an acute care hospital within 30 to 180 days of discharge after a PSI 03 event for any of the following conditions that were present on admission: recurrent pressure ulcer, cellulitis, pyoderma, infection, bacteremia, sepsis, acute or chronic osteomyelitis, septic arthritis, necrotizing fasciitis, gangrene, or flap failure.	PSI 03
30-day all-cause mortality	Death due to any cause within 30-days of the discharge after a PSI triggering event.	PSI 06, PSI 08, PSI 09, PSI 15
30-day all-cause readmission	Readmission to an acute care hospital within 30 days of the discharge after a PSI triggering event (excluding any readmissions categorized separately below).	All
180-day all-cause mortality	Death due to any cause within 30 to 180 days of the discharge after the PSI triggering event.	PSI 03, PSI 10, PSI 11, PSI 12, PSI 13, PSI 14
90-day nonsurgical hip fracture complication	Hospital readmission within 30 to 90 days of the discharge after a PSI 08 event for a mechanical or infectious hip fracture complication not requiring surgery.	PSI 08
Hip reoperation within 90 days	Hospital readmission for reoperation on the hip within 90 days of the discharge after a PSI 08 event.	PSI 08
Avascular necrosis	Admission to the hospital within 30 to 365 days of the discharge after a PSI 08 event with aseptic or avascular necrosis.	PSI 08
Anoxic brain damage or shock	Development of brain (cerebral) anoxia and or shock associated with a hemorrhage or hematoma event.	PSI 09
Acute renal failure requiring dialysis	Development of acute kidney injury/failure (stage V) requiring dialysis while hospitalized after a PSI triggering event.	PSI 09, PSI 13

Outcome	Description of Events Captured	Applicable Patient Safety Indicator (PSI)
Dialysis post discharge for up to 6 months	Ongoing need for dialysis for up to 6 months after discharge following a PSI event.	PSI 10
1-year all-cause hospital readmission	All cause hospital readmission within 30 to 365 days of the discharge after a PSI 10 triggering event.	PSI 10
Tracheostomy	Received a tracheostomy due to extended need for mechanical ventilation and/or a complication from intubation.	PSI 11
6-month hospital readmission for a bleeding complication	Hospital readmission within 30 to 180 days of the discharge due to a bleeding complication related to anticoagulation.	PSI 12
Emergency department visits within 180 days for a thrombotic complication	Emergency department visits related to a thrombotic event such as pulmonary embolus, deep vein thrombosis, or postphlebitic syndrome within 180 days of discharge after a PSI 12 event.	PSI 12
180-day hospital readmission for an enterocutaneous fistula	Readmitted to an acute care hospital for intra-abdominal abscess or enterocutaneous fistula within 30 to 180 days of the discharge after a PSI 14 event.	PSI 14
180-day hospital readmission for an incisional hernia	Readmitted to an acute care hospital (including observational stays) for incisional hernia or reclosure of postoperative disruption of the abdominal wall within 30 to 180 days of the discharge after a PSI 14 event.	PSI 14
180-day hospital readmission for an intra-abdominal abscess or enterocutaneous fistula	Development of an intra-abdominal abscess or enterocutaneous fistula up to 30 to 180 days of discharge after a PSI 15 event.	PSI 15
Excess hospital days	Excess hospital length of stay (in days) associated with a PSI event.	All
Long-term skilled nursing facility stay	Long-term skilled nursing facility stays that are 26 consecutive days or longer in a skilled nursing facility or long-term care facility.	All
Short-term skilled nursing home days	Long-term skilled nursing facility stays that are 26 consecutive days or longer in a skilled nursing facility or long-term care facility.	All