AHRQ Quality Indicators

Quality Indicator Empirical Methods, v2025

Prepared for:

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

Contract No. 47QRAA18D00BQ

August 2025

Table of Contents

Index of Tables

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) and State Emergency Department Databases (SEDD) for the development of the AHRQ QIs, using the databases 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 SEDD capture emergency visits at hospital-owned emergency departments (EDs) that do not result in hospitalization.

The AHRQ QI program would like to acknowledge the HCUP Partner organizations that participated in the HCUP SID and SEDD:[1]

Alaska Department of Health
Arizona Department of Health Services
Arkansas Department of Health
California Department of Health Care Access and Information
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 Laulima Data Alliance
Hawaii University of Hawai‹i at Hilo
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
Mississippi State Department of Health
Missouri Hospital Industry Data Institute
Montana Hospital Association
Nebraska Hospital Association

[1] HCUP partner list, available at: https://hcup-us.ahrq.gov/partners.jsp?

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 Apprise Health Insights
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 and Human Services
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

Abbreviations and Terms Used in this Document

Table 1. Quality Indicator Empirical Methods Abbreviations

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
HGLR	Hierarchical Group Lasso Regularisation
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

Abbreviation/Term	Descriptions
MS-DRG	Medicare Severity Diagnostic Related Group
MDRG	Modified Medicare Severity Diagnosis Related Group
NCHS	National Center for Health Statistics
CBE	Consensus-Based Entity
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
PQE	Prevention Quality Indicator in Emergency Department Settings
PQI	Prevention Quality Indicator in Inpatient Settings
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) v2025 (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 hospital use. For example, area-level indicators may answer the question: Did a patient receive inpatient care for a condition that might have been avoided, if the patient's area of the country had more or better preventive or outpatient care? 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 in Inpatient Settings (PQIs)** identify hospital admissions that might have been avoided through access to high-quality health care, preventive care, and health promoting resources within a community (first released November 2000, last updated August 2025).
- **Prevention Quality Indicators in Emergency Department Settings (PQEs)** identify emergency department visits that might have been avoided through access to high-quality health care, preventive care, and health promoting resources within a community (first released as beta version September 2023, last updated August 2025).

[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.

- **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 August 2025).
- **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 August 2025).
- **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 August 2025).
- **Maternal Health Indicators (MHIs)** [Beta] broadly address healthcare quality in the domain of maternal health and identify opportunities to reduce complications around childbirth. MHIs are released as a beta module in v2025. More information about the methods used to calculate the MHIs is available in Appendix E. Maternal Health Indicators.

Table 2 summarizes the quality domains addressed by each module.

Table 2. 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 – Inpatient Settings	X	
Prevention Quality – Emergency Department Settings	X	
Pediatric Quality – Inpatient Quality		X
Pediatric Quality – Patient Safety		X
Pediatric Quality – Prevention Quality	X	
Maternal Health	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.[4]

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 may provide exemplary care, while others provide sub-standard care. Some areas may serve people that are at higher risk for complications or

[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.
[4] For v2025 software, the reference population for hospital level indicators uses three years of data. The reference population for area level indicators (non-MHI) uses one year of data.

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.** The indicator reports the number of times that a hospital reported an event, for example a retained surgical item or unretrieved device fragment (not present on admission). This volume, or count, indicator does not have denominators.
- **Observed rate.** Area-level rates are the number of hospitalizations or emergency department visits 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 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 to the 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 events for this hospital (or area) compared 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 is not numerically stable. As a weighted average of the risk-adjusted rate from the user's input dataset and the rate observed

in the reference population, the smoothed rate is calculated with a shrinkage estimator to result in a rate (1) 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) 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), especially 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 differ 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 these 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 weighting 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 that PSI 90 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 and PQEs), 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. The beta version of the PQEs was released in 2023, and the beta version of the MHIs was first released in September, 2024. Additional details about the development of each module are 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.

Another set of indicators, measured as visit rates from emergency department encounters, was developed between 2011 and 2015. The Prevention Quality Indicators in Emergency Department Settings (PQE) module was intended as a set of indicators complementary to the PQIs that would also reflect the capacity of the health care system to limit the use of hospital care for conditions better treated in primary care facilities. These measures were originally specified using ICD-9-CM diagnosis codes and population-based risk adjustment similar to current area indicators but adopting a Poisson regression methodology. With specifications revised to ICD-10-CM codes, the measures are estimated using the same risk and reliability adjustment approaches as other area indicators. Starting with v2024, the PQE module is integrated into the AHRQ QI software alongside other area-level modules.

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 are complementary AHRQ QI Software Instruction documents on the AHRQ QI website (https://qualityindicators.ahrq.gov) that provide an overview of the software and details about data elements and programs used to calculate the AHRQ QIs (see Appendix A).

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 in Inpatient Settings (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 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. The indicators identify hospital admissions that 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 3 and Appendix B.1).

Table 3. List of AHRQ Prevention Quality Indicators in Inpatient Settings (PQIs)

Abbrev	Indicator Name (v2025)	Area or Hospital Level
PQI 01	Diabetes Short-Term Complications Admission Rate	Area

Abbrev	Indicator Name (v2025)	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 Prevention Quality Indicators in Emergency Department Settings (PQEs)

The PQEs are a set of measures designed to capture access to quality care and the wellness (community health) of a population in a given region, by using hospital administrative data to identify rates of emergency department visits for "ambulatory care sensitive conditions." These are conditions for which short and long-term access to quality care can reduce the likelihood of emergency department use or for which early intervention can prevent complications or more severe disease. These measures are influenced by disease prevalence, environmental factors (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 and emergency department data, they provide insight into the health of the community and the community-based health care system. For example, patients with diabetes may be admitted to the emergency department 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. Analogous to the inpatient PQIs, these indicators identify emergency department visits that would be less likely with access to high-quality outpatient or preventive care. The numerator is a count of visits to the ED for the condition of interest, and the denominator is an estimate of the number of persons at risk for such an ED encounter.

The PQEs 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 to identify and investigate communities potentially in need of interventions.

Because the PQEs are calculated using readily available hospital administrative data, they are an easy-to-use and inexpensive screening tool. Like the inpatient PQIs, 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 PQE module contains a total of 5 indicators (Table 4 and Appendix B.1).

Table 4. List of AHRQ Prevention Quality Indicators in Emergency Department Settings (PQEs)

Abbrev	Indicator Name (v2025)	Area or Hospital Level
PQE 01	Visits for Non-Traumatic Dental Conditions in ED Settings	Area
PQE 02	Visits for Chronic Ambulatory Care Sensitive Conditions in ED Settings	Area
PQE 03	Visits for Acute Ambulatory Care Sensitive Conditions in ED Settings	Area
PQE 04	Visits for Asthma in ED Settings	Area
PQE 05	Visits for Back Pain in ED Settings	Area

A.3 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 5 and Appendix B.2).[5] Most of the IQIs are based on surgical procedures and are reported at the hospital-level, although some are based on medical conditions.[6] The IQIs are grouped into two categories, in-hospital mortality indicators and utilization indicators:

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

[5] 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.
[6] Area-level IQIs were retired in v7.0 ICD-10-CM/PCS specifications and software.

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 5. List of AHRQ Inpatient Quality Indicators (IQIs)

Abbrev	Indicator Name (v2025)	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
IQI 22	Vaginal Birth After Cesarean (VBAC) Delivery Rate, Uncomplicated	Procedure	Hospital

Abbrev	Indicator Name (v2025)	Procedure or Condition	Area or Hospital Level
IQI 33	Primary Cesarean Delivery Rate, Uncomplicated	Procedure	Hospital

[a] Includes stratum-specific indicators.

A.4 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.[7]

The PSI module contains a total of 17 indicators—15 primary indicators expressed as rates, one count measure, and one composite indicator (Table 6 and Appendix B.3).[8]

There are 17 hospital-level PSIs for medical conditions and surgical procedures that have been shown to have complication/adverse event rates/counts that vary substantially across hospitals and for which evidence suggests that high complication/adverse event rates/counts may be associated with deficiencies in the quality of care. One of these 16 indicators is measured as the number of complication/adverse events and the other 15 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, five 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 6. List of AHRQ Patient Safety Indicators (PSIs)

Abbrev	Indicator Name (v2025)	Area or Hospital Level
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
PSI 08	In-Hospital Fall-Associated Fracture Rate[b]	Hospital

[7] 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).
[8] PSI 02, Death Rate in Low-Mortality Diagnosis Related Groups (DRGs), was removed from the software in v2025.

Abbrev	Indicator Name (v2025)	Area or Hospital Level
PSI 09	Postoperative Hemorrhage or Hematoma Rate[c]	Hospital
PSI 10	Postoperative Acute Kidney Injury Requiring Dialysis Rate[d]	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[a]	Hospital
PSI 15	Abdominopelvic Accidental Puncture or Laceration Rate[e,f]	Hospital
PSI 17	Birth Trauma Rate – Injury to Neonate[g]	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[h]	Hospital

[a] Includes stratum-specific indicators
[b] Previously entitled "Postoperative Hip Fracture" prior to v6.0; previously entitled "In Hospital Fall with Hip Fracture Rate" prior to v2023; includes component-specific indicators beginning in v2023. For more details about the expanded PSI 08 specifications, rationale, and testing in v2023, see the PSI 08 Expansion Announcement.
[c] Previously entitled "Perioperative Hemorrhage or Hematoma Rate" prior to v2021
[d] Previously entitled "Postoperative Physiologic and Metabolic Derangement" prior to v5.0
[e] Previously entitled "Accidental Puncture or Laceration Rate" prior to v6.0.
[f] Previously entitled "Unrecognized Abdominopelvic Accidental Puncture or Laceration Rate"
[g] Calculated in the PDI software module
[h] Previously entitled "Patient Safety of Selected Indicators" prior to v6.0

A.5 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 7 (and Appendix Tables B.1 and B.2) list all of the PDIs and indicate 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, seven are hospital-level indicators and four are area-level indicators of quality of pediatric care. The three composite indicators are area-level indicators.

Table 7. List of AHRQ Pediatric Quality Indicators (PDIs)

Abbrev	Indicator Name (v2025)	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 claims data 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 different software platforms. These software products and associated documentation are available on the AHRQ QI website:

- SAS: https://qualityindicators.ahrq.gov/software/sas_qi
- Windows Products:
    a. WinQI: https://qualityindicators.ahrq.gov/software/win_qi
    b. CloudQI: https://qualityindicators.ahrq.gov/software/cloudqi

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).[9]

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 8. 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. Note that starting with v2023, MDC 14 and MDC 15 are identified in the software via individual code lists (MDC14PRINDX and MDC15PRINDX).

Table 8. 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). PQEs include visits of both adult and pediatric populations.

Table 9 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.

[9] For v2024 and v2025, the hospital-level AHRQ QIs are created using three calendar years of data and the area-level AHRQ QIs are created using one calendar year of data.

Table 9. 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 PQE 01, PQE 05
	18+ Years or Obstetric	IQI 21–22, IQI 33 PSI 05, PSI 07
	18 to 39 Years	PQI 15, PQE 04
	18 to 64 Years	PQE 03
	18 to 89 Years or Obstetric	PSI 04
	40+ Years	PQI 05 IQI 12, IQI 30 PQE 02
	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, PQE 03
	2 to 17 Years	PDI 14
	5 to 17 Years	PQE 01, PQE 04
	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, PQEs, 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 within

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. AHRQ v2025 QI software uses the FY 2025 POA Exempt List.[10]

Table 10 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. If the secondary diagnosis is not exempt, then it is 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.[11]

[10] POA Exempt List, available at: https://www.cms.gov/medicare/payment/fee-for-service- providers/hospital-aquired-conditions-hac/coding

[11] 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.

Table 10. 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.[12] Other 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 MDC categories. For records assigned to these MS-DRGs, the official CMS MS-DRG grouper software retains 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, PQEs and area-level PDIs for areas. Area-level rates are constructed using denominators that capture the size of the area's population from census (or user supplied) data.[13]

[12] Centers for Medicare & Medicaid Services. https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/ms-drg-classifications-and-software.

[13] 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

Area-level indicators contained in the PQI module identify inpatient stays that might have been less likely with access to high-quality community care and resources. Indicators contained in the PQE module identify emergency department visits similarly sensitive to access to quality care in other settings. Like PQIs, the area-level indicators contained in the PDI module identify inpatient stays for children that are less likely with access to high quality community care.[14]

Area-level indicators have numerators, denominators and observed rates. In addition, 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.

Cases are excluded from the reference population numerator of area-level indicators if the patient resides in a state that did not contribute to the HCUP State Inpatient Databases (SID). Cases are excluded from the PQE reference population if the patient resides in a state that did not contribute to both the State Emergency Department Databases (SEDD) and SID. In addition, PQE cases are excluded if the patient is treated at an emergency department in a state that differs from the patient's state of residence. PQE 05, Visits for Back Pain in ED Settings, requires linkage between visits and cases are excluded for states that do not permit linkage.

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.

(but not 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.

[14] 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.

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/V2025/AHRQ_QI_v2025_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 need 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, and poverty level in that area. For most QIs, risk-adjusted rates calculated by indirect standardization are used. In statistical language, the risk adjustment controls 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 (v2025)

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 area level modules contain 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 are obtained from the U.S. Census Small Area Income and Poverty Estimates (SAIPE).[15] AHRQ v2025 software uses 2022 SAIPE data. All U.S. counties are assigned to a poverty decile (POVCAT). The poverty deciles then are used as risk factors in the risk adjustment model.[16] Indicators can be adjusted for age and sex, or for age, sex and poverty decile. For all area-level indicators, the risk factors used in risk adjustment are age, sex, and poverty deciles. See Appendix C. List of Risk Factors for Area-Level Quality Indicator Modules Appendix 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-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,

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

[16] In v2024, poverty data were not included for all Connecticut FIPS codes. Therefore, Connecticut was not included in the reference population for calculating socioeconomic status (SES) risk-adjusted rates but was included for the calculation of non-socioeconomic status (non-SES) risk-adjusted rates. Observed rates for Connecticut were still included in benchmark tables. In v2025, poverty data were included for Connecticut. Therefore, Connecticut was included in the reference population for calculating SES risk-adjusted rates.

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

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

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 use 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 the same as the average of predicted event rates. In the modeling process, we assess model calibration properties, but the O-E ratio (observed rate to expected rate ratio) may not be exactly equal to one. In software development (not part of the publicly released software), we multiply 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 include in the AHRQ software the national O-E ratio calculated using our reference population . We also provide users the options of calibrating to the reference population or to users' populations.

1. The reference population-based O-E ratio is recommended in most situations, and it is also the default choice in the software.
2. The 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 not 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[17] and described by Hosmer and Lemeshow[18] 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[19] estimated from the reference (general) population and the pre-determined area-specific noise variance and risk-adjusted rate.[20] 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 is 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 is similar to the area's risk-adjusted rate, whereas the smoothed rate for an area with unstable estimates is 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:

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

[17] Iezzoni L, Ed. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago: Health Administration Press; 2013.
[18] 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.
[19] The pre-determined values are embedded in the software.
[20] The smoothing factors are included in the software.

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[21] 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.[22]

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:

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

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

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

[22] 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 the 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.

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 but also depend on other factors, such as the risk-adjusted rates, noise variance, and 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 hospitalizations 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 (CCSR) for Diagnoses used to risk adjust the hospital-level IQI and PDI rates. The COVID-19 risk factor also used POA starting with v2023. 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 Modified MS-DRG (MDRG) covariates.[23]

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),

[23] The most recent version of ICD-10-CM/PCS MS-DRG software and list of MS-DRGs are available at: https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/ms-drg-classifications-and-software**.** The modified DRG (MDRG) pools individual CMS-DRGs and MS-DRGs into a larger category. See the parameter estimates documentation for details on the MS-DRG to MDRG crosswalk.

being at high risk for a non-preventable event or having an event or underlying clinical precedents present on admission.

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 by important clinical risk factors or procedure types (IQI 09, IQI 11, IQI 17, PSI 04, and PSI 14).
3. Risk adjustment of rates to account for case mix. More details 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.[24] AHRQ v2025 software uses the most recent version of MS-DRGs: v42.1. 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[25]

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 is 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).[26]
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).

[24] ICD-10-CM/PCS MS-DRG, list of MS-DRGs, available at: https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/ms-drg-classifications-and-software.
[25]Numerator exclusions are defined based on similar reasons listed here for PSI 05, an indicator of volume/counts.
[26] For most hospital-level indicators, 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 may provide exemplary care while others provide sub-standard care. This is an important dimension of difference. 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 provide information 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 is not numerically stable.

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 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 replace 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 replace the APR-DRGs for procedure-based IQIs.

For PSIs, AHRQ Elixhauser Comorbidity codes[27] and comorbidity count categories are used in the software to cover comorbidity conditions that can be either present on admission or after admission. AHRQ v2025 software uses the most recent version of the Elixhauser software: v2025.1. For PDIs, AHRQ CCSR software is used because it covers pediatric conditions, whereas Elixhauser Comorbidity categories 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 11 provides an overview of the four classes of risk factors.

Table 11. 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]	MDCs[b]		MDCs[b]

[27] Elixhauser Comorbidity Software Refined for ICD-10-CM: https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp

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

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.
[e] Starting with v2024, only PSI 04 uses the stratified risk adjustment models (the other stratified measures use the standard risk adjustment method, see section *F.3 Feature Selection*).

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.[28] 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 binary 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 use 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 the same as the average of predicted event rates. In the modeling process, we assess model calibration properties, but the O-E ratio (observed rate to expected rate ratio) may not be exactly equal to one. In software development (not part of the publicly released software), we multiply the predicted rate

[28] 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 that the new predicted rates are perfectly calibrated to the observed rates. To be consistent, we include in the AHRQ software the national O-E ratio calculated using our reference population . We also provide users the options of calibrating to the reference population or to users' populations.

1. The reference population-based O-E ratio is recommended in most situations, and it is also the default choice in the software.
2. The 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[29] and described by Hosmer and Lemeshow[30] 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[31] 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 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.

---

[29] Iezzoni, Lisa, Ed. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago: Health Administration Press; 2013.
[30] 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.
[31] 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.

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[32] 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.[33]

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:

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

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

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

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 weight depends on the specific composite of interest. Table 12 shows each of the composite indicators and the type of weight (default) used to derive the indicator.

Table 12. 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 relative frequency of the outcome of interest, in this case, a potentially preventable adverse event.
- *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 in the population that experiences component events. Component indicators that lead to significant excess mortality and morbidity will have the highest harm weights, whereas those that have lower mortality and morbidity associated with them will have lower harm weights. For additional information, see the "Quality Indicator User Guide: Patient Safety Indicators (PSI) Composite Measures, July 2025" at:

  https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2025/PSI_Composite_Measures.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:

$$\mathbf{M} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$$

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

$$\hat{\boldsymbol{\mu}} = \boldsymbol{\mu} + \mathbf{v}$$

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:

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

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:

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

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

where:

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

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 $j \neq k$),

$$Cov(\mathbf{v}^j, \mathbf{v}^k) = E\left[\left(\boldsymbol{\mu}^j - \widehat{\boldsymbol{\mu}}^j\right)\left(\boldsymbol{\mu}^k - \widehat{\boldsymbol{\mu}}^k\right)\right] \qquad (2.1)$$

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

$$Cov(\mathbf{v}^j, \mathbf{v}^k) = \boldsymbol{\Omega}_{\boldsymbol{\mu}}^{jk}\left[\left(1 - \widehat{\boldsymbol{\beta}}^j\right)\left(1 - \widehat{\boldsymbol{\beta}}^k\right)\right] \qquad (2.2)$$

Note that this is just the signal covariance times one minus the signal ratio for each of the measures. Thus, if the signal ratio is zero 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 zero.

Also note that if one measure is missing, then the signal ratio is simply set to zero. 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 (2.2), we substitute

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

into (2.1) 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

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

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

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

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

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, it is desirable to use smoothed rates because 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 one 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 from expert panels such as the National Quality Forum (NQF) Patient Safety Committee[34] and experts from the AHRQ QI Workgroups.[35, 36] Additional detail on the AHRQ QI measure development, implementation, maintenance, and retirement process is posted on the AHRQ QI website at: https://qualityindicators.ahrq.gov/measures/qi_resources.

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 sometimes quarterly) coding updates (e.g., ICD-10-CM/PCS, MS-DRGs, MDCs, POA coding guidelines), reference population changes, census population updates, periodic clinical panel review, the Consensus-Based Entity's (CBE) 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 13 provides a list of all versions of the AHRQ QI specifications, the date of release, and the year(s) upon which the specifications for the reference population are built.

---

[34] National Quality Forum Patient Safety Final Report 2015, https://www.qualityforum.org/Publications/2016/02/Patient_Safety_2015_Final_Report.aspx.
[35] AHRQ QI Composite Workgroups, https://www.qualityindicators.ahrq.gov/Modules/composite_workgroup.aspx.
[36] 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 13. AHRQ QI Specification Releases

AHRQ QI Version	Coding Scheme	Release Date	Modules	Year of Reference Population
2025	ICD-10-CM/PCS	August 2025	All	2020,2021,2022[a]
2024.0.2	ICD-10-CM/PCS	June 2025	PQE	2021
2024.0.1	ICD-10-CM/PCS	September 2024	MHI Beta	2021
2024	ICD-10-CM/PCS	July 2024	All	2019,2020,2021[b]
2023	ICD-10-CM/PCS	September 2023	PQE	2019
2023	ICD-10-CM/PCS	August 2023	All	2019,2020,2021[b]
2022	ICD-10-CM/PCS	July 2022	All	2019
2021.0.2	ICD-10-CM/PCS	March 2022	IQI	2018
2021.0.1	ICD-10-CM/PCS	September 2021	PSI, PDI, IQI	2018
2021	ICD-10-CM/PCS	July 2021	All	2018
2020	ICD-10-CM/PCS	July 2020	All	2017
2019	ICD-10-CM/PCS	Summer 2019	All	2016
2018	ICD-10-CM/PCS	Summer 2018	All	---
7.0	ICD-10-CM/PCS	Spring 2017	All	---
6.0	ICD-10-CM/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, PQE Prevention Quality Indicators in Emergency Department Settings.

[a] For v2025, the reference population for the hospital quality indicators includes 2020, 2021, and 2022. The reference population for the area quality indicators (PQIs, area-level PDIs, and PQEs) includes 2022 only for v2025.

[b] For v2024 and v2023, the reference population for the hospital quality indicators included 2019, 2020, and 2021. The reference population for the area quality indicators (PQIs, area-level PDIs, and PQEs) included 2021 only for v2024 and included both 2019 and 2020 for v2023.

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.[37] 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 and SEDD[38] contain all-payer, encounter-level information on inpatient discharges and emergency department visits that do not result in hospitalization, respectively, from the universe of community hospitals in participating states. The SID and SEDD include 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 State Inpatient Databases (SID) are State-specific files that contain all inpatient care records in forty-seven participating states plus District of Columbia. Together, the SID encompass about 97 percent of all U.S. community hospital discharges.[39] Some states include discharges from specialty facilities, such as acute psychiatric hospitals. The HCUP SID and SEDD data serve as the reference (or general) population for the AHRQ QIs (PQI, IQI, PSI, and PDI use SID only; PQE uses SID and SEDD), 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://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2025/Version_2025_Benchmark_Tables_PQI.pdf

[37] For a complete list of HCUP Partner organizations that participated in the HCUP SID and SEDD, please see the Acknowledgements section of this document.
[38] HCUP State Inpatient Databases (SID). HCUP State Emergency Department Databases (SEDD). Healthcare Cost and Utilization Project (HCUP). Agency for Healthcare Research and Quality (AHRQ), Rockville, MD. www.hcup-us.ahrq.gov/sidoverview.jsp.
[39] HCUP Databases. Healthcare Cost and Utilization Project (HCUP). January 2025. Agency for Healthcare Research and Quality, Rockville, MD. https://hcup-us.ahrq.gov/sidoverview.jsp.

- PQE Benchmark: https://qualityindicators.ahrq.gov/Downloads/Modules/PQE/V2025/Version_2025_Benchmark_Tables_PQE.pdf
- IQI Benchmark: https://www.qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2025/Version_2025_Benchmark_Tables_IQI.pdf
- PSI Benchmark: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2025/Version_2025_Benchmark_Tables_PSI.pdf
- PDI Benchmark: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2025/Version_2025_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://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2025/Parameter_Estimates_PQI_v2025.pdf
   - PQE Parameter Estimates: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PQE/V2025/Parameter_Estimates_PQE_v2025.pdf
   - IQI Parameter Estimates: https://www.qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2025/Parameter_Estimates_IQI_v2025.pdf
   - PSI Parameter Estimates: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2025/Parameter_Estimates_PSI_v2025.pdf
   - PDI Parameter Estimates: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2025/Parameter_Estimates_PDI_v2025.pdf

Table 14 provides details on the availability of HCUP SID and SEDD, 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 14. AHRQ QI Reference Population

Data Year	Number of SID States[a]	Number of SID Hospitals[b]	Total Discharges included in SID	Percentage of SID Discharges[c]	Number of SEDD States[d]	Number of SEDD Hospitals[b]	Total Discharges included in SEDD	Percentage of SEDD Discharges[c]
2022	48	4,215	32,929,549	96	40	3,537	99,867,429	99
2021	48[e]	4,261	33,700,641	96	40	3,553	93,525,595	99
2020	49	4,262	32,681,925	96	n.a.	n.a.	n.a.	n.a.
2019	49	4,252	35,612,594	96	n.a.	n.a.	n.a.	n.a.
2018	48	4,290	35,549,549	97	n.a.	n.a.	n.a.	n.a.
2017	48	4,326	35,747,363	98	n.a.	n.a.	n.a.	n.a.
2016	48	4,039	35,612,904	98	n.a.	n.a.	n.a.	n.a.
2014	45	4,430	33,645,600	94	n.a.	n.a.	n.a.	n.a.
2013	44	4,398	33,670,781	94	n.a.	n.a.	n.a.	n.a.
2012	44	4,440	34,440,381	94	n.a.	n.a.	n.a.	n.a.
2011	46	4,575	35,504,333	90	n.a.	n.a.	n.a.	n.a.
2010	45	4,550	35,722,417	89	n.a.	n.a.	n.a.	n.a.

Abbreviations: SID, State Inpatient Database; SEDD, State Emergency Department Database; n.a, not applicable
[a] Potentially includes 48 states, plus the District of Columbia. The number of states included in the reference population for specific indicators may be smaller due to POA and data limitations (Nevada is excluded from v2024 and v2025). The PQE reference population only includes states that provide both SID and SEDD, which reduces the number of states with SID data included in the PQE reference population.
[b] Number of hospitals include community, non-rehabilitation, non-long-term acute care hospitals.
[c] Represents the percentage of SID discharges included in the reference population. Hospital-level QIs that rely on POA indicator information may have fewer discharges in the reference population.
[d] Potentially includes 39 states and the District of Columbia. The number of states included in the reference population may be smaller due to data limitations (Nevada and New Mexico are excluded from v2024 and v2025).
[e] Number of states available in v2024. For v2023, in which 2021 data were also used, 27 states plus the District of Columbia were available at the time of reference population development.
n.a = PQE module first official release in v2024, using 2021 SID and SEDD data.

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 and excluding rehabilitation and long-term acute care (LTAC) hospitals. AHRQ QI Software v2025 uses the 2022 HCUP SID (PQIs, area-level PDIs, and PQEs) and 2022 SEDD (PQEs only) as the reference population for risk-adjustment models for area measures. Discharges from 47 states

and District of Columbia in 2022 were used to develop v2025 area-level indicators in inpatient settings.[40] Emergency department visits from 39 states and the District of Columbia were used for PQEs.

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 discharges are excluded from the numerator. Similarly, discharges for some residents of counties that are included are missing because these residents travel to nonparticipating states. National rates are slightly underestimated because discharges 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 15 below. The reference population includes patients residing in HCUP states and admitted to hospitals in HCUP states.

Table 15. Treatment of State Border Crossing Discharges in the Inpatient Area-Level Indicator (PQI and Area PDI) 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

Note: PQE reference population requires both SID and SEDD participation and that ED visit is in the patient's state of residence.

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

The PQEs are subject to related restrictions. In 2022, data from 39 states and District of Columbia, are included from both SID and SEDD: AK, AR, AZ, CA, CO, CT, DC, FL, GA, HI, IA, IL, IN, KS, KY, MA, MD, ME, MI, MN, MO, MS, MT, NC, ND, NE, NH, NJ, NY, OH, OR, RI, SC, SD, TN, TX, UT, VT, WI, and WY. These states are included in the reference populations for PQE 01 – PQE 04. A subset of 24 states provide the data elements needed to link records to determine when an individual incurred more than one encounter in a year: AK, AR, CA, CO, FL, GA, HI, IA, IN, MA, MD, ME, MO, MS, NE, NY, OR, SC, SD, TN, UT, VT, WI, and WY. These states are included in the reference population for PQE 05. An additional requirement for PQEs is that the individual reside in the state where ED treatment was received. By contrast, PQI records are included if the individual is discharged from a hospital in any state contributing to the SID.

[40] States with data from 2022 SID included in the area-level reference populations 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, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, WV, and WY. For area-level quality indicators, data from Colorado and Hawaii are also not used in the development of PDI 16, and PDI 18 due to missing age in days (AGEDAY data element).

In the software, users can utilize the POPYEAR macro to select the census year used as the denominator for generating observed rates.

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[41] and also excludes rehabilitation and LTAC hospitals.

The AHRQ QI Software v2025 used the 2020–2022 HCUP SID as the reference population for hospital measures (PSIs, IQIs, hospital-level PDIs).

- For **2022 and 2021**, the reference population included SID data from 44 states and District of Columbia: 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, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, and WV.
- For **2020**, the reference population included SID data from 45 states and District of Columbia, with Wyoming as the additional state: 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, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, UT, VA, VT, WA, WI, WV, and WY.

Using three years of data allows for the inclusion of quarterly COVID-19 effects in the risk adjustment models and provides more robust estimation for indicators with small observed rates.[42] There are indicators of the diagnoses being present on admission (POA) in the SID for 45 states and District of Columbia in 2020, 44 states and District of Columbia in 2021 and 2022, while 2 states are excluded due to incomplete information in the POA data element in 2020 and 3 states are excluded in 2021 and 2022. Discharges from these participating states are used to develop hospital-level indicators. States not reporting age in days were not used in the development of PSI 17, and states not reporting dates associated with all surgical procedures were not used to develop PSI 04, PSI 09, PSI 10, PSI 11, PSI 12, PSI 14, and PSI 15 (Table 16). 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.[43] 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

---

[41] Community hospital is based on the AHA definition and refers to "all nonfederal, short-term general and special hospitals whose facilities and services are available to the public." See https://www.cdc.gov/nchs/hus/sources-definitions/hospital.htm for more information.

[42] Refer to the Log of Coding Updates and Revisions listed in Appendix A for the rationale behind using a reference population with multiple years for hospital-level indicators.

[43] 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 at: https://www.hcup-us.ahrq.gov/reports/methods/2015-06.pdf.

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 two times the standard deviation plus the mean of the percentage for hospitals that are required to report POA to CMS.

Table 16. AHRQ Hospital-Level Indicator Reference Population

Data Year	Number of States in Hospital-Level Indicator Reference Population	States not reporting indicators of the diagnoses being POA	States not reporting procedure day (Excluded from PSI 04, PSI 09, PSI 10, PSI 11, PSI 12, PSI 14, and PSI 15)	States not reporting day in age (Excluded from PSI 17)
2020	46	CT, DE	WI, OK, NH	CO, HI
2021	45	CT, DE, WY	WI, OK, NH	CO, HI, and a few hundred cases in TX
2022	45	CT, DE, WY	WI, OK, NH	CO, HI and a few hundred cases in TX

Abbreviations: POA, present on admission

C. Other Data Used for Area-Level Indicator Development

The v2025 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[44] (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, the Census Bureau revises the estimates from all years following 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. Intercensal estimates in the form used for the QIs were produced following the 2000 census but are not yet available for the years following the 2010 census. Census estimates are used to derive the denominators for area-level indicators. The v2025 2000-2024 AHRQ QI Population File is available at:

https://qualityindicators.ahrq.gov/Downloads/Software/SAS/V2025/2000-2024_Population_Files_V2025.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 v2025 AHRQ area-level QIs use SAIPE estimates from 2022.[45]

D. Coding Updates

[44] "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.

[45] Available at https://www.census.gov/programs-surveys/saipe.html

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 Coordination and Maintenance Committee. The v2025 AHRQ QI software updates all measure specifications to reflect coding updates for ICD-10-CM/PCS codes effective as of October 1, 2024.[46]

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[47] and POA exempt codes;[48] however, the QIs use a modified version of the CMS OR procedure list to better capture procedures occurring in an OR setting, which is also aligned fully across AHRQ programs and tools such as PClassR.[49]

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 v2025.1 (PSI and hospital-level PDI risk model),[50] AHRQ's Clinical Classification System Refined for ICD-10-CM Diagnoses v2025.1 (IQI risk model and hospital-level PDI risk model),[51] AHRQ's Clinical Classification System Refined for ICD-10-PCS Procedures v2025.1 (IQI risk model model),[52] and AHRQ Procedure Classes v2025.1 (PClassR, hospital-level PDI risk model).[53] Updates to these systems are incorporated in the risk models annually.

---

[46] Information on ICD-10-CM/PCS coding updates is available at: https://www.cms.gov/medicare/coding-billing/icd-10-codes.

[47] Most recent ICD-10-CM/PCS MS-DRG operating room procedures and procedure codes available at: https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/ms-drg-classifications-and-software.

[48] Centers for Medicare & Medicaid Services. https://www.cms.gov/medicare/payment/fee-for-service-providers/hospital-aquired-conditions-hac/coding.

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

[50] Elixhauser Comorbidity Software Refined for ICD-10-CM Healthcare Cost and Utilization Project (HCUP). November 2024. Agency for Healthcare Research and Quality, Rockville, MD. https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR-Reference-File-v2024-1.xlsx.

[51] Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. Healthcare Cost and Utilization Project (HCUP). November 2024. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp.

[52] Clinical Classifications Software Refined (CCSR) for ICD-10-PCS Procedures. Healthcare Cost and Utilization Project (HCUP). November 2024. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/prccsr.jsp.

[53] Procedure Classes Refined for ICD-10-PCS. Healthcare Cost and Utilization Project (HCUP). November 2024. 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) updates the Uniform Bill and includes 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.[54] 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 guidelines may necessitate 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 and SEDD (PQE only) state files. The methods used for these calculations are described in Chapter III.E.2 and Chapter III.F.5. These calculations are updated annually.[55] National benchmark rates are currently provided by AHRQ.[56]

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 (v2025)

[54] ICD-10-CM Official Guidelines for Coding and Reporting, FY 2025 (October 1, 2024 – September 30, 2025). Centers for Medicare & Medicaid Services (CMS), and National Center for Health Statistics (NCHS), available at https://www.cms.gov/medicare/coding-billing/icd-10-codes

[55] These calculations were not updated in years when the reference population was unavailable. See Table 14. AHRQ QI Reference Population for more details.

[56] 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.[57] 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 to 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 are 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 key population characteristics that affect the need for, and access to, health care services. 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 QIs measure population health, the 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.

F.2. Construction of Candidate Covariates for Risk Adjustment

To prepare the data for feature selection and risk adjustment, we need to create candidate risk factors, some of which are QI specific. The following is a high-level summary of all candidate risk factors used for each QI module.

[57] The AHRQ QI software v6.0 (ICD-9-CM) included risk adjustment, while the 7.0 and v2018 (ICD-10-CM/PCS) 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 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.

For the PSIs, candidate variables considered for risk adjustment for v2025 include demographic factors (age, sex, and an age-by-sex interaction[58]), COVID-19 indicators,[59] quarterly indicators[60] and their interactions with COVID-19, MDCs, MDRGs, transfer status (Transfer), a ventilator dependence flag (ODC903), and AHRQ/Elixhauser comorbidity variables.[61] Additional PSI-specific variables include complication severity indicators (RegVarPS04_k_ANY and RegVarPS04_k_SEVERE) and Do Not Resuscitate (DNR) status for PSI 04 strata (k = "DVT_PE", "PNEUMONIA", "SEPSIS", "SHOCK", "GIHEMORRHAGE"); Chronic VTE (CHVTEPEDVTD and CHVTEOTHD) for PSI 12; the presence of a solitary kidney (SOLKIDD) for PSI 10; and cardiogenic shock, cardiac arrest, and anoxic brain damage (CSHOCKANOXBDD) for PSI 03; a variable indicating medical/surgical discharges (MEDICDR) for PSI 03, 06, 07, 08, and 15; procedure type or immune risk categories for specific PSIs (HPPS13 for PSI 13 and HPPS15 for PSI 15); and Elixhauser comorbidity count categories for all PSIs except PSI 14[62].

For the IQIs, candidate variables considered for risk adjustment for v2025 include demographic factors (age, sex, and an age-by-sex interaction[63]), COVID-19 indicators, quarterly indicators and their interactions with COVID-19, MDCs, transfer status (Transfer), and CCSR for Diagnoses and Procedures dummy variables.[64] Additional IQI-specific variables include an acute myocardial infarction (MRTAMISTD) for IQI 15, a "do not resuscitate" flag (DNR) for IQI 15, 16, 17, 18, 19, and 20; and cardiogenic shock, cardiac arrest, and anoxic brain damage (CSHOCKANOXBDD) and systolic heart failure (SYSHFD) for IQI 12, 15, 16, and 30.

For the hospital-level PDIs, candidate variables considered for risk adjustment for v2025 include demographic factors (age, sex, and an age-by-sex interaction), COVID-19 indicators, quarterly indicators and their interactions with COVID-19, MDRGs, transfer from acute care - outborn (transfer_alt), a ventilator dependence flag (ODC903), and CCSR for Diagnoses dummy variables. Additional PDI-specific variables include a variable indicating medical/surgical discharges (MEDICDR) for PDI 01, 05, and 12; birth weight category dummy variables (BWHTCAT) for IQI 01, 05, 08, 09, 12, and NQI 03; and procedure type or immune risk categories for specific PDIs: GPPD08 for PDI 08, GPPD10 for PDI 10, GPPD12 for PDI 12, HPPD01 for PDI 01, and HPPD10 for PDI 10. PNL 002 CCSR (Short gestation; low birth weight; and fetal growth retardation) is excluded since this CCSR overlaps with the denominator inclusion/exclusion criteria. Count categories of Elixhauser Comorbidity codes are also included as candidate risk factors to account for multimorbidity except NQI 03. For NQI 03, age in days instead of age, and interaction between sex and age in days are included as covariates.

---

[58] 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."
[59] The COVID-19 diagnosis indicator is subject to POA.
[60] For v2025, three years of SID data are used as reference population data for the hospital QIs. As a result, twelve quarter indicator variables are created, with the last quarter serving as the reference category in the model.
[61] Elixhauser Comorbidity Software Refined for ICD-10-CM Healthcare Cost and Utilization Project (HCUP). February 2024. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/comorbidity_icd10.jsp.
[62] PSI 14 used a different feature selection method (HGLR) compared to other PSIs (LASSO). The inclusion of comorbidity counts in the risk adjustment is not needed because HGLR has the potential to identify comorbidity counts as a feature through the selection.
[63] The reference (omitted) category for the age-sex interaction categories for the IQI is "65-69 year-old women." The oldest and youngest age categories may be insufficiently populated to produce stable results.
[64] 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.

For the PQIs, PQEs and area-level PDIs, risk adjustment covariates include sex and age categorized in five-year intervals. Adjustment for poverty category deciles, as a measure of socioeconomic risk, is also available as an option in the software. The v2025 software uses 2022 Small Area Income and Poverty Estimates (SAIPE) to create the POVCAT variable for the socioeconomic status (SES) risk adjustment models. Specifically, we calculate county population size weighted deciles based on the poverty percentages from the SAIPE data and assign each county to one of ten deciles. We then create ten binary POVCAT variables for those deciles, using POVCAT1 (least impoverished) as the reference category in the logistic regression models. These POVCAT variables are included only in the SES models.

An important update in the risk adjustment modeling for the v2025 software is inclusion of planning region identifiers for Connecticut[65] in the SAIPE poverty data. This enhancement allows Connecticut data to be included in the risk adjustment models that include adjustment for poverty. In contrast, the v2024 SES model, which used the same measure of poverty, excluded Connecticut from the reference population due to the absence of planning region poverty estimates.

The rationale for including new code set based candidate variables is included in Table 17. To help users understand how POA information is applied to each code set, we include all the code sets that are used to create risk factors and how the POA information is used in Table 18.

Table 17. A list of new code set based candidate risk factors used in v2025 and the rationale for including them

Code Set (Risk Factor)	Description	Affected QIs	Clinical Rationale
CHVTEPEDVTD	Chronic VTE PSI 12 trigger diagnosis codes	PSI 12	A chronic VTE involving the pulmonary circulation, lower extremity veins, or the inferior vena cava can directly extend during a period of immobilization during or after surgery, increasing the risk of an acute VTE.
CHVTEOTHD	Chronic VTE unrelated site diagnosis codes	PSI 12	A chronic VTE at an unrelated site, such as the superior vena cava or upper extremities, suggests a hypercoagulable condition that may increase the risk of an acute VTE, but would not directly extend to cause a PSI 12 event.
SOLKIDD	Solitary kidney diagnosis codes	PSI 10	The presence of a solitary kidney, whether due to a previous nephrectomy or a congenital anomaly, increases the risk of acute kidney injury requiring dialysis after any physiologic insult.

[65] CT state replaced county codes with planning regions with different boundaries, names and FIPS codes after 2022.

Code Set (Risk Factor)	Description	Affected QIs	Clinical Rationale
CSHOCKANOXBDD[66]	Cardiogenic shock, cardiac arrest, and anoxic brain damage diagnosis codes	PSI 03, IQI 12, 15, 16, and 30	Cardiogenic shock, cardiac arrest, and anoxic brain injury may develop prior to admission, and greatly increase the risk of death (and certain hospital-associated complications) after admission, despite optimal intensive care.
SYSHFD	Systolic heart failure diagnosis codes	IQI 12, 15, 16, and 30	Systolic heart failure, also known as heart failure with reduced ejection fraction (HRrEF), is associated with increased risk of death after an acute event such as a myocardial infarction or a procedure to restore coronary blood flow.

Table 18. Code sets used for risk adjustment risk factors and related Present on Admission (POA) information

Code Set	Description for Code Set	Affected QIs	POA[67]	Diagnosis Codes Locations for the Code Set
SYSHFD	Systolic heart failure diagnosis codes	IQI 12, 15, 16 and 30	Yes	Any
ODC_VEN	Ventilator Dependence Diagnosis Codes	All PDI and PSI	No	Any
SEVDPEDX	PSI04 DVT PE Regression variable: Any Triggering Complication POA	PSI04 deep vein thrombosis/pulmonary embolism stratum	Yes	Any
SEVPNEUDX	Severe Pneumonia Diagnosis Codes	PSI04 pneumonia stratum	Yes	Any
SEVSEPDX	Severe Sepsis Diagnosis Codes	PSI04 sepsis stratum	Yes	Any
SEVSCKDX	Severe Shock or Cardiac Arrest Diagnosis Codes	PSI04 shock/cardiac arrest stratum	Yes	Any
SEVGIHDX	Severe Gastrointestinal Hemorrhage or Acute Ulcer Diagnosis Codes	PSI04 gastrointestinal hemorrhage stratum	Yes	Any

[66] CSHOCKANOXBDD was used as a candidate risk factor for IQI 12, 15, 16, and 30 in v2025 and added to PSI 03 starting v2025.
[67] Please refer to the SAS software for details about how each code set is used for creating risk factors and how POA is defined and used.

Code Set	Description for Code Set	Affected QIs	POA[67]	Diagnosis Codes Locations for the Code Set
DNR	Do Not Resuscitate Diagnosis Codes	All PSI 04 strata, IQI 15, 16, 17, 18, 19, and 20	Yes	Any
CSHOCKANOXBDD	Cardiogenic shock, cardiac arrest, and anoxic brain damage diagnosis codes	PSI 03, IQI 12, 15, 16, and 30	Yes	Any
CHVTEPEDVTD	Chronic VTE PSI 12 trigger diagnosis codes	PSI 12	Yes	Any
CHVTEOTHD	Chronic VTE unrelated site diagnosis codes	PSI 12	Yes	Any
SOLKIDD	Solitary kidney diagnosis codes	PSI 10	Yes	Any

The comorbidity counts categories are created through the following steps:

- **Step 1.** *Create comorbidity counts per discharge:*
  Generate three types of comorbidity counts for each discharge: an unweighted sum, a weighted sum using mortality weights, and a weighted sum using readmission weights
- **Step 2.** *Sort and bucket discharges:*
  For each type of comorbidity count, sort all discharges and divide them into 10 equal-sized buckets. This results in three sets of buckets—one for each count type (unweighted, mortality-weighted, and readmission-weighted).
- **Step 3.** *Cluster the buckets:*
  Apply a k-means clustering algorithm to the 10 buckets for each count type to form three clusters (low, medium, and high), based on observed event rates.
- **Step 4.** *Select the optimal count type:*
  Identify the count type whose clusters show the largest between-group difference in event rates. This indicates the strongest association with the outcome.
- **Step 5.** *Define the comorbidity count categories:*
  Use the selected count type and its corresponding low-, medium-, and high-comorbidity groups as categorical risk factors in the final model.

The CCSR tool for ICD-10-CM diagnoses was developed as part of HCUP. There are two types of CCSR: (1) CCSR for ICD-10-CM Diagnoses (DXCCSR) and (2) CCSR for ICD-10-PCS Procedures (PRCCSR). The IQIs and PDIs consider DXCCSR as candidate risk factors. Only the procedure-based IQIs consider PRCCSR as candidate risk factors. Both of these variable sets are generated from the most

recently available CCSR HCUP data tools. Given the similarities in CCSR's position codes 1, 2, and 3[68], we create the CCSR binary indicator by combining them together to create a binary indicator that is equal to 1 for position codes equal to 1, 2, or 3 and equal to 0 otherwise. For example, we create a binary indicator d_DXCCSR_DIG020, in which d_DXCCSR_DIG020 = 1 if DXCCSR DIG020 is coded as 1, 2, or 3 and d_DXCCSR_DIG020 = 0 if DXCCSR DIG020 is coded as 0.

- **DXCCSRs** are binary variables (e.g., d_DXCCSR_DIG025) generated at the discharge level, independent of QIs. DXCCSRs are used in RA for PDIs and IQIs. Only the binary variables are used in the RA models.
- **PRCCSRs** are binary variables (e.g., d_IQ08_PRCCSR_CAR010) for procedure-based IQIs 08, 09, 11, 12, 30, and 31. Similar to DXCCSR, PRCCSR's creation is at the discharge level but has nuanced specifications based on the QI. For instance, for IQI 08, specific prefixes like IQI08 are used in conjunction with PRESOPP and PRESO2P to create PRCCSR for risk adjustment. This results in a variable "IQ08_PRCCSR_CAR010" with four levels. It also means the same PRCCSR might have a different value if it was created for another IQI and therefore we need to add a prefix "IQ08" to distinguish them. For v2023, this variable was created at a finer level such as "ENDO" vs "OPEN" for IQI 11 (e.g., d_IQ11_ENDO_PRCCSR_CAR010). Starting with v2024, we create this variable at the QI level to facilitate the HGLR feature selection for IQI 11.

The following CCSR categories for ICD-10-CM diagnoses in the "Factors influencing health status and contact with health services" category are removed from the candidate variable list for the IQI and PDI risk adjustment for v2025 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, FAC018, FAC019, FAC020, FAC021, FAC024, FAC025, FAC028, FAC029, and FAC030. The COVID-19 category (INF012) is also removed.

The following CCSR categories for ICD-10-PCS procedures are removed from IQI candidate risk factor list because of the high-likelihood that these factors are endogenous or represent complications resulting from the IQI denominator procedure: ADM001, ADM002, ADM004, ADM005, ADM006, ADM014, ADM017, ADM018, ADM019, ADM021, CAR024, ENT001, ESA001, ESA002, ESA003, ESA004, ESA005, ESA006, ESA007, ESA008, ESA009, ESA010, EST004, GIS021, GNR005, IMG001, IMG002, IMG003, IMG004, IMG008, IMG009, LYM005, MAM002, MAM005, MAM008, MAM010, MAM011, MAM013, MAM015, OTR001, OTR004, RES001, RES005, RES007, RES008, RES010, RES014, and URN006.

Several MDRG variables (MDRG 7799, 8898, 8899, 8897) are excluded from consideration as candidate variables because assignment to these MDRGs could be due to an in-hospital complication requiring a procedure such as ECMO or tracheostomy, or another operating room procedure not related to the principal diagnosis. Similarly, the MDRG 1501, which includes MS-DRGs for neonates died or transferred to another acute care facility and ungroupable MS-DRGs, is not considered as a candidate variable for risk-adjustment. The following QI-specific MDRGs are removed based on clinical review: MDRG 1801 and 1802 for PSI 13 and PDI 10; MDRG 416 for PSI 06 and PDI 05; MDRG 805, 901, 902, 1004, 913 for PSI 03; MDRG 806, 807, 811, and 2407 for PSI 08; MDRG 542 for PSI 12; and MDRG

[68] 0 – The CCSR was not triggered by any ICD-10-CM diagnosis code on the input record; 1 – The CCSR was triggered by only the principal (or first-listed diagnosis) on the input record; 2 – The CCSR was triggered by both the principal (or first-listed diagnosis) and any secondary diagnosis on the input record; 3 – The CCSR was triggered by only secondary diagnosis code(s) on the input record.

610 for PSI 14. In most cases, these MDRGs were excluded because they represent potential consequences of the PSI event itself (e.g., wound debridement and grafting for PSI 03, prolonged mechanical ventilation for PSI 06 and PDI 05, fracture repair for PSI 08, pulmonary embolectomy for PSI 12, abscess drainage for PSI 13, and repair of a herniated dehiscence for PSI 14).

F.3 Feature Selection

The primary goal of feature selection is to reduce the number of risk factors included in the final risk adjustment models but maintain the model performance using certain evaluation criteria. A study by Osborne et al. on registry-based quality measurement examined whether risk adjustment models with fewer variables could perform comparably to more complex models for indirect adjustment.[69] Their motivation stemmed from the high costs associated with collecting additional data elements in hospital settings. The aim was to streamline data collection without significantly compromising model accuracy.

In contrast, the AHRQ Quality Indicators (QIs) do not depend on costly data collection methods. Instead, they leverage routinely available administrative data such as age, sex, transfer status, and diagnostic codes. While some QI models incorporate over 100 variables, these are derived from a limited set of core elements, with many variables generated as categorical representations (e.g., multiple categories within MDRGs), which may lead to more than 300 binary indicators. Notably, each patient record is assigned exactly one MDRG, which allows the model to assign risk levels based on specific diagnoses and procedures. As such, although the models may appear complex, they rely on easily accessible data. Still, a parsimonious model—one that includes only the most essential predictors—can offer clear advantages: improved interpretability, reduced risk of overfitting, lower computational demands, and greater usability. These streamlined models are also well-suited for generating predicted probabilities at the individual discharge level.

For area-level QIs, the models use the complete set of covariates for sex, age in 5-year age groups, and an interaction between sex and age. There is also an optional set of covariates for poverty category deciles based on the county of patient residence. The poverty category control 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 level of the local population. For other applications, adjusting for poverty could mask important disparities in population health. Feature selection is not used for area-level indicators.

For hospital-level QIs, feature selection is applied to reduce the number of risk factors for the final risk adjustment model. The resulting parsimonious models improve interpretability and help mitigate the risk of overfitting. In the following sections, we provide a detailed, step-by-step explanation of the process used to identify the final set of risk factors for each risk adjustment model.

F.3.1 Risk factors filtering

After creating the candidate binary risk factor variables, we conduct preliminary screening of the variables to reduce the initial candidate risk factor set to k risk factors: ($X_1, \ldots, X_k$). This process is tailored for each QI and is designed to improve computational efficiency during feature selection and reduce the risk of modeling issues such as memory overflow or non-convergence. The key considerations and rationales for this screening process are as follows:

---

[69] Osborne NH, Ko CY, Upchurch GR, Dimick JB. (2010). Evaluating parsimonious risk-adjustment models for comparing hospital outcomes with vascular surgery. *Journal of vascular surgery*, *52*(2), 400–405. https://doi.org/10.1016/j.jvs.2010.02.293.

- **Low Event Counts:** Based on cross-tabulations between each covariate and the outcome of interest, a variable must have at least 30 denominator cases (for example, more than thirty discharges for an MDRG category) and a minimum cell size[70] of 2 to be retained. This requirement helps ensure that variables included in the model have meaningful information while reducing issues related to sparse data. The minimum cell size threshold is determined empirically to balance model performance and computational efficiency. Setting the threshold at 2 allows for the inclusion of as many risk factors as possible, relying on the feature selection algorithm to identify the most relevant variables. However, for some quality indicators, such as those IQIs using a large number of CCSR variables, this approach may result in a large pool of candidate risk factors. Including too many variables with low event counts for some QIs can significantly increase processing time and may cause the feature selection process or the final logistic models to fail to converge. In such cases, we gradually increased the minimum event count threshold from 2, in some cases up to 200. While this adjustment may exclude some candidate risk factors, it retains a sufficient number of variables with adequate event sizes. This ensures that the final models converge without errors or warnings and that overall model performance, as measured by C statistics and Hosmer Lemeshow plots, remains satisfactory.
- **Reference Group Variables:** We remove the binary variable representing the reference category within mutually exclusive groups. The reference category is typically (1) the most common, (2) the lowest risk, or (3) a median category. Although the choice of reference group influences the interpretation of coefficients, it does not impact predicted probabilities or overall model performance.
- **Collinearity:** Collinearity occurs when two or more variables carry highly redundant information. While collinearity does not affect a model's predictive performance, it can inflate standard errors and destabilize coefficient estimates. Our software implementation calculates the Variance Inflation Factor (VIF) for each covariate. Variables with VIFs exceeding 1,000—indicative of near-perfect collinearity—are excluded. Variables with VIFs greater than 10 but below 1,000 are retained and the feature selection algorithm is used to determine if they should be kept or dropped.[71] This approach strikes a balance between retaining a maximum number of informative variables and mitigating risks associated with excessive collinearity and computational burden. All models successfully converged after applying LASSO or HGLR-based feature selection.
- **Complete or Quasi-Complete Separation:** Complete separation occurs when a linear combination of predictor variables perfectly classifies the outcome variable, meaning the model can predict the outcome without error. Quasi-complete separation arises when the outcome variable can be separated by one or more independent variables to a substantial degree, though not perfectly. After filtering out risk factors using previous steps, we mitigate the risk of both complete and quasi-complete separation. Additionally, the feature selection process is designed to further address and reduce any remaining separation issues. All AHRQ QI regression models are monitored for convergence during both the variable selection phase and the final model estimation stage to ensure the estimates and standard errors do not have any extreme values.

[70] The min cell size is calculated as the minimum value of the following four frequency numbers for each covariate i: X_i and QI Y: number of discharges in ($X_i = 0$, $Y = 0$), ($X_i = 1$, $Y=1$), ($X_i = 1$, $Y = 0$), and ($X_i = 0$, $Y = 1$). Where $X_i = 1$ when the covariate is present on the discharge and 0 otherwise, and Y=1 if the QI numerator is 1 and 0 otherwise.

[71] The LASSO or HGLR model selection procedure is also able to drop variables that are highly correlated given its heavy penalty on the variable coefficients.

It is also worth noting that the structure of QI models inherently limits the potential for collinearity. Any collinearity that arises typically occurs between, rather than within, domains such as age-sex categories, transfer status, Elixhauser Comorbidity Software Refined, Clinical Classifications Software Refined (CCSR), and MDRGs. In some cases, overlap between similar classifications across these systems is the source of collinearity.

F.3.2 Risk factors selection

For area-level QIs, no feature selections are needed. Therefore, in the following, we focus on the hospital-level QIs.

After creating the candidate risk factors, we proceed with feature selection and final model estimation. This process is iterative, as model performance must be assessed in the model tuning process to determine whether the feature selection yields a parsimonious model that avoids underfitting or overfitting while maintaining strong model performance. The detailed model evaluation criteria are discussed in section F.4.2. To effectively implement the feature selection approach and evaluate model performance, we begin by randomly selecting 80% of discharges from the reference population to serve as the training dataset, while reserving the remaining 20% as the test dataset. The training dataset is used to run feature selection algorithms, apply cross-validation for parameter tuning, and assess model performance. The test dataset is then used to evaluate the predictive performance of the selected models. For hospital-level QIs, C statistics and Hosmer-Lemeshow plots[72] are generated for the training dataset, test dataset, and the full dataset based on the following considerations:

- A large discrepancy between statistics from the training and test datasets may indicate potential overfitting issues that should be addressed.
- Statistics from the test dataset reflect the model's predictive performance on data not used for training the models.
- Statistics from the full dataset provide a summary of overall model performance on the entire reference population. Parameter estimates from the full dataset are used in the final AHRQ software.

For PSI 04 and other hospital-level QIs without strata, variables for inclusion in the final risk adjustment models are selected by the least absolute shrinkage and selection operator (LASSO) selection method.[73] Due to computation resource limitations, one million discharges are randomly selected if the reference population on which the model is run 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 parameter estimates and failure to select features that improve prediction of the outcome.

LASSO feature selection is implemented using the glmnet package in R, with ten-fold cross-validation to identify the optimal model parameters. Given that discrimination performance—measured by the cross validation (CV) based C statistic—is the primary evaluation metric for model performance, we use it as the objective in the loss function to guide selection of the optimal lambda value, which is the key parameter in the LASSO algorithm. We then plot the cross-validated C statistic against lambda indices, along with the corresponding number of selected risk factors, to visualize the relationship between model

[72] After feature selection, certain risk factors are added back to the model based on their clinical importance or face validity. As a result, model evaluation statistics are generated only after these variables have been reintroduced.
[73] Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (*Methodological*)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

performance, the number of risk factors selected, and the lambda values. The final decision is informed by both clinical review of the relevance of selected risk factors and model performance, taking into account the model's discrimination power (measured by the C statistic) and calibration property (evaluated using the Hosmer-Lemeshow plot). For v2025, we selected lambda.1se for most QIs. However, for PDI 12, a lambda value between lambda.1se and lambda.min was chosen based on clinical review and model performance evaluation.

For most formerly stratified hospital-level QIs (PSI 14, IQI 09, IQI 11, and IQI 17), starting with v2024, the Hierarchical Group Lasso Regularisation (HGLR) feature selection method[74] (Ray, 2023) is applied. In previous versions, those QIs utilized stratified risk adjustment (RA) models due to the diverse nature of their denominator populations. Although stratification is a recognized method for assessing effects across varied groups, the performance of these risk adjustment models can vary significantly. This variability is often the result of very low event rates, and the challenges associated with selecting features specific to each stratum. A more effective approach to addressing population diversity in risk adjustment models would be to incorporate linear interactions into the models. Additionally, models predicting health outcomes frequently suffer from poor calibration, which may result from overlooking interactions. Therefore, starting with v2024, the HGLR method is implemented for these indicators.

HGLR feature selection is implemented through the glinternet package (version 1.0.12) in R. In addition to the filtered risk factors, we also include a stratum indicator (Z), which may have up to four levels. Using the option "interactionCandidates" in the glinternet.cv function, we restrict the interaction selection to a pool of candidate interactions with Z. Practically speaking, this means that only interactions such as Z * X_i will be selected, where X_i is the risk factor i in the candidate covariate pool. If that interaction is selected, HGLR will retain the main effects Z and X_i in the model. We use five-fold CV to reduce running time. The logistic loss metric (negative Bernoulli log-likelihood) is used to quantify CV-based errors. We also use up to 200 lambda indexes (nLambda) and use the lambdaMinRatio parameter to tune the model performance in the glinternet.cv function. We then plot the logistic loss metric against the lambda indices, alongside the number of selected risk factors, to visually assess whether a minimum loss value is achieved within the selected range of lambda indices. If the minimum loss is not observed within this range, we reduce the lambdaMinRatio value and rerun the HGLR algorithm until the optimal model is reached. Similar to the lambda tuning strategy used in LASSO selection, the final choice of lambda is guided by both clinical review of the relevance of selected risk factors and model performance, considering discrimination (measured by the C statistic) and calibration (assessed using the Hosmer-Lemeshow plot). For v2025 final models, we use lambda.1std for all four QIs that used the HGLR feature selection strategy.

HGLR requires more computational power than LASSO for feature selection. To address this, we employ a sampling strategy to reduce the sample size without compromising model performance. Given the rarity of QI events, all events are retained in the training set, while only a subset of non-events is sampled for training the HGLR model.[75,76] During the modeling process, we test a range of sampling rates to ensure

---

[74] Ray, M., Zhao, S., Wang, S. *et al.* Improving hospital quality risk-adjustment models using interactions identified by hierarchical group lasso regularisation. *BMC Health Serv Res* **23**, 1419 (2023).

[75] This strategy is similar to the oversampling non-event approach used in the paper (Ray, 2023).

[76] For IQI 17 (v2025), running the model after subsampling the non-events (~300,000 non-events and ~100,000 events) took over 30 hours for a single run, making model tuning extremely challenging. Given the substantial increase in the event rate, we chose to take a simple random sample of both events and non-events, which reduced the runtime to approximately 15 hours and facilitated more extensive model tuning.

that all HGLR models converge successfully, and that the runtime remains acceptable without compromising the model performance.

F.3.3 Risk factors adjustment after feature selection

After feature selection, certain dropped risk factors may be added back into the model based on their face validity and clinical relevance in influencing the QIs. Note that variables filtered out during the preliminary screening in section F.3.1 are not reconsidered for inclusion and are therefore not subject to the following adjustments.

For the PSIs, we add back the full set of main effects for age and sex, as the risk for nearly all PSI events tends to increase with age and may be influenced by hormonal or reproductive factors. This step ensures that the main effects corresponding to selected age and sex interaction terms remain in the model, and that age effects are interpretable across the entire distribution of age in the reference population. Including these main effects facilitates interpretation and avoids biased coefficient estimates, incorrect standard errors, and misleading p-values that can result from omitting main effects. For PSI 04 specifically, the transfer variable is added back due to user inquiries regarding its relevance and face validity in risk adjustment for mortality measures.[77] Additionally, if any complication-triggering indicators for "any" or "severe" complications (e.g., RegVarPS04_k_ANY, where *k* refers to the five PSI 04 strata) are dropped during feature selection, they are reintroduced for the corresponding stratum. For example, if RegVarPS04_PNEUMONIA_ANY and RegVarPS04_PNEUMONIA_SEVERE are not selected for the PSI 04 PNEUMONIA model, they are added back to that model alone.

For the IQIs, we similarly add back an indicator-appropriate set of main effects for age and sex because only a subset of age terms was selected by LASSO for some models. Starting with v2025, the transfer variable is subject to data-driven selection. This decision is motivated by our preference for a data-driven approach for feature selection, in the absence of strong clinical rationale for forcing features into models.[78]

For the hospital-level PDIs, age main effects are also added back. Starting in v2025, the transfer_alt variable follows the same selection logic as transfer in the IQIs. Starting in v2025, while the sex (MALE) variable is used as a candidate feature in all PDIs, it is not reintroduced if dropped during selection. The rationale is that sex differences in physiology and postoperative risk are virtually absent before puberty, which was confirmed empirically through feature selection. Specifically, only NQI 03 consistently selected sex as a significant main effect, both in the v2024 and v2025 LASSO results. Forcing sex into models when there is no clinical rationale, and the effect estimates are close to null, has little benefit and may contribute to overfitting and misinterpretation. Additionally, given their relevance to procedure type and immune-related risk, GPPD and HPPD risk factors are added back if they were excluded during feature selection. For NQI 03, given its importance in the neonatal nature of the measure, birth weight risk factors (BWHTCAT) are always forced into the model.

Beyond these QI-specific considerations, the following general rules are also applied:

---

[77] Given the expectation that transfer status is associated with a higher risk of PSI 04, the regression coefficients for transfer variables in the PSI 04 stratum models are required to be positive. For PSI 04, any transfer variable with a negative coefficient was excluded in the final model.

[78] In the PSI module, we still used it as a forced variable due to user inquiries regarding its inclusion and face validity.

- For any interaction terms selected in the final model, both of their main effects are forced back in if they were previously dropped given similar rationales described for the PSIs.
- Starting with v2025, data-driven methods are applied to select COVID main effects, quarter dummies, and their interactions, without forcing them into the model. The final models maintain the structure in which, if an interaction between COVID-19 and a quarter dummy is selected, both the COVID-19 main effect and the corresponding quarter dummy are included in the model. In addition, we identify the earliest and latest quarters included in the model and add back any missing intermediate quarters to ensure the contiguity of quarter dummies in the final risk models. This addresses cases where, for instance, the model selected Q2 to Q10 but omitted Q5—suggesting it detected spikes in quarter effects while skipping some periods. Including the missing quarters improves interpretability by making the reference group more coherent, such as defining it as those cases occurring at the start or end of the COVID pandemic.

F.4 Estimate and Evaluate the Models

F.4.1 Model estimation

After the previous steps, the final set of risk factors for each QI has been identified. The next step is to estimate the final regression models using appropriate statistical techniques and SAS procedures.

For area-level QIs, we use U.S. Census population counts by county, age, and sex, as the denominator, and event counts derived from the discharge-level file as the numerator. A county's poverty rate decile may be used as an additional risk factor. These data are used to fit logistic regression models—both without socioeconomic status covariates (non-SES models) and with socioeconomic status covariates (SES models). The non-SES models include age, sex, and age-by-sex interaction terms, while the SES models also include poverty category variables for the county's resident population. Model estimation is implemented in SAS using the PROC LOGISTIC procedure. The resulting parameter estimates are used in the AHRQ area QI software to generate risk-adjusted and smoothed rates at the county level.

For hospital-level QIs, we estimate risk-adjustment models using Generalized Estimating Equations (GEE) to account for within-hospital correlation and to properly adjust for patient case mix. The QI event indicator is used as the dependent variable, and the selected risk factors from previous steps are used as independent variables. The parameter estimates from the regression model are used in the AHRQ hospital QI software to generate risk-adjusted and smoothed rates at the hospital level.

The models are estimated in SAS using PROC GENMOD, which allows specification of different working correlation structures to account for clustering within hospitals:

- **TYPE = IND (independent):** Produces the same coefficient estimates as logistic regression but adjusts the standard errors for within-hospital clustering.
- **TYPE = CS (compound symmetry):** Produces different coefficients and standard errors compared to logistic regression.

Empirical testing revealed two main issues with the **TYPE = CS** option:

1. Significantly longer computation times: we need to use PROC LOGISTIC when PROC GENMOD could not converge.
2. Poor model calibration for many QIs: we need to use additional ratio adjustment to account for the poor national level poor calibration.

Based on these considerations, we choose the **TYPE = IND** option because it adjusts standard errors and p-values to account for within-hospital correlation, maintains good calibration properties, and ensures that the observed rate equals the expected rate for each quality indicator.

For hospital-level QIs, the v2025 software uses three years of HCUP data (2020-2022) as the reference population for risk-adjustment. By utilizing data from three years, the software can factor in the impact of COVID-19 on a quarterly basis and offer more precise estimates for indicators with lower observed rates. In v2025, we use the following risk adjustment models to accommodate the COVID time trend effect for risk adjustment.

$$Y \sim \mu + \boldsymbol{X} + COVID + \sum_{i \in A} Q_i + \sum_{j \in B} Q_j * COVID$$

Here, Y represents each QI, $\boldsymbol{X}$ represents the risk factors other than the COVID and quarter dummies in the final risk adjustment model, $COVID$ is a binary indicator where $COVID$ = 1 if a discharge has COVID present on admission and 0 otherwise, Q_i is a binary quarter indicator where $i = 1,2,\ldots,11$ represents eleven quarters from Q1-2020 to Q3-2022, set A is the set of quarters included in the model, and set B is the set of quarters used for interaction terms. Note that for v2025, $COVID$, Q_i, and their interaction terms $Q_j * COVID$ are subject to selection. Because the main effects for the interaction terms are kept in the model, set B is a subset of set A. Both set A and B are QI specific.

For hospital-level QI models, statistics are calculated on the training dataset, test dataset, and full dataset to assess model performance. For area-level QI model, statistics are calculated on the full dataset only. The C-statistics from the test dataset for hospital-level QIs and the full dataset for area-level QIs are reported in the parameter estimate files available on the AHRQ website. The final multivariable model parameters are published on the AHRQ website (see Chapter III.B).

F.4.2 Model evaluation

In the AHRQ software, 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, predicting that the outcome will occur in approximately the right proportions, over a wide range of predicted probability. We apply those two statistics to evaluate the risk adjustment model performance for both the hospital-level and area-level QIs.

Discrimination

The first measure for the logistic regression discrimination power is the C-statistic. The statistic is 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[79] have coined three widely adopted labels for discrimination performance based on the C-statistic:

[79] 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.

- $0.7 \leq$ C-statistic < 0.8 indicates acceptable discrimination
- $0.8 \leq$ C-statistic < 0.9 indicates excellent discrimination
- C-statistic ≥ 0.9 indicates outstanding discrimination

The C-statistics for the AHRQ QI risk adjustment models are published on the AHRQ QI website in the Parameter Estimates Document: (see Chapter III.B). The C-statistics alone should not be used to indicate if a risk adjustment model is valid or not. Instead, C-statistics are used to compare different models or compare models from different years to make sure the model selection performs as expected to minimize the potential overfitting and underfitting. AHRQ does not recommend using a threshold to disqualify the risk adjustment model just based on C-statistics.

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 risk 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[80] logistic regression goodness-of-fit statistic is based on a chi-square test statistic calculated using the observed and expected counts across the 10 risk deciles. To assess whether there is over or under-prediction in high or low-risk discharges, it is suggested to create a plot of the observed rates and expected rates by risk deciles or a calibration belt.[81] As the sample size in HCUP data is typically very large, using the p-value based on Hosmer and Lemeshow chi-square test statistics is not advised, as it often results in significant results even for very well calibrated models.[82]

In addition to the two key statistics, we also use the Area Under the Precision-Recall Curve (AUPRC) to evaluate model performance, particularly for QIs with imbalanced datasets for hospital-level QIs. AUPRC assesses the quality of event classification based on precision and recall, providing valuable insights when event rates are low.

F.5 Calculate Rates

F.5.1 General Description

In order to make appropriate 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:

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

[80] 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.
[81] Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. PLoS One. 2011 Feb 23;6(2):e16110. doi: 10.1371/journal.pone.0016110. PMID: 21373178; PMCID: PMC3043050.
[82] Kramer, A. A., & Zimmerman, J. E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited*. Critical Care Medicine, 35(9), 2052. https://doi.org/10.1097/01.CCM.0000275267.64078.B0.

Because models for binary outcome events are used, there can be small differences between the observed rate and the expected and risk-adjusted rates in the reference population. For PROC GENMOD with independent correlation structure and PROC LOGISTIC in SAS, the ratio of the observed rate to the expected rate in the reference population is very close to one if not identical.

After the new risk adjustment models are fitted, 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 PSI 04, 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 QIs, 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. In this case, it is recommended that users should not consider a single stratum in isolation (due to the instability of stratum-specific results), but rather the overall rate across all strata should be the focus of attention.

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 use Morris' method. Morris' method[83] 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 assumption 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. It is possible that the iterative method may lead to a negative signal variance. So, when the second issue occurs, a full Bayes-based method is used 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 denominators of varying sizes (i.e., eligible discharges) in the QI calculations. Statistically, this variation means that each hospital contributes a different amount of information ; large hospitals with thousands of discharges contribute more information than small hospitals with, say, fewer

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

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.

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 (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 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.[84]

The weights are defined and calculated as follows:

[84] Zrelak PA, Utter GH, McDonald KM, Houchens RL, Davies SM, Skinner HG, Owens PL, Romano PS. Incorporating harms into the weighting of the revised Agency for Healthcare Research and Quality Patient Safety for Selected Indicators Composite (Patient Safety Indicator 90). Health Serv Res. 2022 Jun;57(3):654-667. doi: 10.1111/1475-6773.13918.

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.

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 fitted 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 one 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 are 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 v2025 reference population. The v2025 AHRQ QI software contains weights for PSI 90 based on 10 component PSI indicators (Table 19). The component weights are re-scaled to sum to one.

Table 19. Weights of PSI 90 Component Indicators, v2025

Indicator abbreviation	Indicator name	Harm weight[a]	Volume weight	Component weight
PSI 03	Pressure Ulcer Rate	0.3080	0.1429	0.2186
PSI 06	Iatrogenic Pneumothorax Rate	0.1381	0.0513	0.0352
PSI 08	In-Hospital Fall-Associated Fracture Rate	0.1440	0.0708	0.0506
PSI 09	Postoperative Hemorrhage or Hematoma Rate	0.0570	0.1192	0.0338
PSI 10	Postoperative Acute Kidney Injury Requiring Dialysis Rate	0.3584	0.0285	0.0507
PSI 11	Postoperative Respiratory Failure Rate	0.2219	0.1952	0.2152
PSI 12	Perioperative Pulmonary Embolism or Deep Vein Thrombosis (PE/DVT) Rate	0.1557	0.2083	0.1611
PSI 13	Postoperative Sepsis Rate	0.3102	0.1243	0.1915
PSI 14	Postoperative Wound Dehiscence Rate	0.1441	0.0237	0.0169
PSI 15	Abdominopelvic Accidental Puncture or Laceration Rate	0.1474	0.0359	0.0236

[a] These harm weights are based on versions of PSI 08 from v2022 and earlier, in that they are specific to hip fractures. The corresponding harm weights for other types of hospital-associated fractures are likely to differ, because other types of fractures require different treatment from hip fractures. Revised harm weights will be estimated in the future.

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 been 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 QIs 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 and SEDD 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 of 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 that 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.[85] 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. See section F.4.2 Model evaluation for more detailed information.

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).[86] These panels recommend refinements to indicator specifications and rate the overall usefulness of the indicators.[87]

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), calibration, and other supporting statistics such as AUPRC. For more details, please see section F.4.2 Model evaluation.

---

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

[86] Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader J, and Kahan JP, The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND Corporation, 2001. https://www.rand.org/pubs/monograph_reports/MR1269.html.

[87] 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.

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.

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

Clinical Classifications Software Refined (CCSR) for ICD-10-CM Diagnoses. Healthcare Cost and Utilization Project (HCUP). February 2024. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/dxccsr.jsp.

Clinical Classifications Software Refined (CCSR) for ICD-10-PCS Procedures. Healthcare Cost and Utilization Project (HCUP). February 2024. Agency for Healthcare Research and Quality, Rockville, MD. www.hcup-us.ahrq.gov/toolssoftware/ccsr/prccsr.jsp.

Elixhauser Comorbidity Software Refined for ICD-10-CM Healthcare Cost and Utilization Project (HCUP). February 2024. Agency for Healthcare Research and Quality, Rockville, MD. https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR-Reference-File-v2024-1.xlsx.

Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G. Calibration belt for quality-of-care assessment based on dichotomous outcomes. PLoS One. 2011 Feb 23;6(2):e16110. doi: 10.1371/journal.pone.0016110. PMID: 21373178; PMCID: PMC3043050.

Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader J, and Kahan JP, The RAND/UCLA Appropriateness Method User's Manual. Santa Monica, CA: RAND Corporation, 2001. https://www.rand.org/pubs/monograph_reports/MR1269.html.

HCUP State Inpatient Databases (SID). Healthcare Cost and Utilization Project (HCUP). October 2022. 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, Lemeshow S. Goodness of fit tests for the multiple logistic regression model. Communications in statistics-Theory and Methods. 1980;9(10), 1043-1069.

ICD-10-CM Official Guidelines for Coding and Reporting, FY 2024 (April 1, 2024 – September 30, 2024). Centers for Medicare & Medicaid Services (CMS), and National Center for Health Statistics (NCHS). https://www.cms.gov/files/document/fy-2024-icd-10-cm-coding-guidelines-updated-02/01/2024.pdf.

ICD-10-CM/PCS MS-DRG v42.1, list of MS-DRGs, available at: https://www.cms.gov/files/zip/icd-10-ms-drgs-v421-effective-april-1-2025.pdf

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

Kramer AA, & Zimmerman JE. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited*. Critical Care Medicine, 35(9), 2052. https://doi.org/10.1097/01.CCM.0000275267.64078.B0

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.

National Quality Forum Patient Safety Final Report 2015, https://www.qualityforum.org/Publications/2016/02/Patient_Safety_2015_Final_Report.aspx

Measure Evaluation Criteria. National Quality Forum (NQF). https://www.qualityforum.org/measuring_performance/submitting_standards/measure_evaluation_criteria.aspx.

Osborne NH, Ko CY, Upchurch GR, Jr, & Dimick, JB. (2010). Evaluating parsimonious risk-adjustment models for comparing hospital outcomes with vascular surgery. *Journal of vascular surgery*, *52*(2), 400–405. https://doi.org/10.1016/j.jvs.2010.02.293

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

Ray, M., Zhao, S., Wang, S. *et al.* Improving hospital quality risk-adjustment models using interactions identified by hierarchical group lasso regularisation. *BMC Health Serv Res* **23**, 1419 (2023)

Schone E, Hubbard M, and Jones D. 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.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (*Methodological*)*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

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

Zrelak PA, Utter GH, McDonald KM, Houchens RL, Davies SM, Skinner HG, Owens PL, Romano PS. Incorporating harms into the weighting of the revised Agency for Healthcare Research and Quality Patient Safety for Selected Indicators Composite (Patient Safety Indicator 90). Health Serv Res. 2022 Jun;57(3):654-667. doi: 10.1111/1475-6773.13918.

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/measures/PQI_TechSpec
- PQE: https://qualityindicators.ahrq.gov/measures/PQE_TechSpec
- IQI: https://qualityindicators.ahrq.gov/measures/IQI_TechSpec
- PSI: https://qualityindicators.ahrq.gov/measures/PSI_TechSpec
- PDI: https://qualityindicators.ahrq.gov/measures/PDI_TechSpec
- MHI: https://qualityindicators.ahrq.gov/measures/MHI_TechSpec

AHRQ QI Benchmark Data Tables

- PQI: https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2025/Version_2025_Benchmark_Tables_PQI.pdf
- PQE: https://qualityindicators.ahrq.gov/Downloads/Modules/PQE/V2025/Version_2025_Benchmark_Tables_PQE.pdf
- IQI: https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2025/Version_2025_Benchmark_Tables_IQI.pdf
- PSI: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2025/Version_2025_Benchmark_Tables_PSI.pdf
- PDI: https://qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2025/Version_2025_Benchmark_Tables_PDI.pdf
- MHI: https://qualityindicators.ahrq.gov/Downloads/Modules/MHI/V2025/Version_2025_Benchmark_Tables_MHI.pdf

AHRQ QI Parameter Estimates Tables

- PQI: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2025/Parameter_Estimates_PQI_v2025.pdf

- PQE: https://qualityindicators.ahrq.gov/Downloads/Modules/PQE/V2025/Parameter_Estimates_PQE_v2025.pdf
- IQI: https://www.qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2025/Parameter_Estimates_IQI_v2025.pdf
- PSI: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2025/Parameter_Estimates_PSI_v2025.pdf
- PDI: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2025/Parameter_Estimates_PDI_v2025.pdf

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

- https://www.qualityindicators.ahrq.gov/Downloads/Software/SAS/V2025/AHRQ_QI_v2025_ICD10_Population_File.pdf

AHRQ QI Software Instructions

- SAS: https://www.qualityindicators.ahrq.gov/Downloads/Software/SAS/V2025/Software_Inst_SASQI_v2025_August_2025.pdf
- WinQI: https://www.qualityindicators.ahrq.gov/Downloads/Software/WinQI/V2025/Software_Inst_WINQI_V2025_August_2025.pdf
- CloudQI: https://qualityindicators.ahrq.gov/Downloads/Software/CloudQI/V2025/Software_Inst_CloudQI_v2025_August_2025.pdf

AHRQ QI Software Release Notes

- SAS: https://qualityindicators.ahrq.gov/Downloads/Software/SAS/V2025/AHRQ_SASQI_v2025_Software_Rel_Notes.pdf
- WinQI/CloudQI: https://qualityindicators.ahrq.gov/Downloads/Software/WinQI/V2025/AHRQ_Windows_v2025_Software_Rel_Notes.pdf

AHRQ QI Log of Coding Updates and Revisions

- PQI: https://qualityindicators.ahrq.gov/Downloads/Modules/PQI/V2025/ChangeLog_PQI_v2025.pdf
- PQE: https://qualityindicators.ahrq.gov/Downloads/Modules/PQE/V2025/ChangeLog_PQE_v2025.pdf

- IQI: https://qualityindicators.ahrq.gov/Downloads/Modules/IQI/V2025/ChangeLog_IQI_v2025.pdf
- PSI: https://qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2025/ChangeLog_PSI_v2025.pdf
- PDI: https://qualityindicators.ahrq.gov/Downloads/Modules/PDI/V2025/ChangeLog_PDI_v2025.pdf
- MHI: https://qualityindicators.ahrq.gov/Downloads/Modules/MHI/V2025/ChangeLog_MHI_v2025.pdf

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

- SID: http://www.hcup-us.ahrq.gov/db/state/siddbdocumentation.jsp
- SEDD: https://hcup-us.ahrq.gov/db/state/sedddbdocumentation.jsp

Appendix B. Comprehensive List of Quality Indicators

Appendix Table 1. Area-Level Quality Indicators

Preventive Quality Indicators in Inpatient Settings	
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
Preventive Quality Indicators in Emergency Department Settings	
PQE 01	Visits for Non-Traumatic Dental Conditions in ED
PQE 02	Visits for Chronic Ambulatory Care Sensitive Conditions in ED
PQE 03	Visits for Acute Ambulatory Care Sensitive Conditions in ED
PQE 04	Visits for Asthma in ED
PQE 05	Visits for Back Pain in ED
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

Pediatric Quality Indicators	
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 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

Patient Safety Indicators	
PSI 06	Iatrogenic Pneumothorax Rate
PSI 07	Central Venous Catheter-Related Blood Stream Infection Rate
PSI 08[b]	In-Hospital Fall-Associated 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[a]	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

Pediatric Quality Indicators	
PDI 92	Pediatric Quality Chronic Composite

[a]Includes stratum-specific indicators.
[b]Includes component-specific indicators.

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

Indicator	POA-Dependent Quality Indicator
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	X
NQI 03	X
PDI 01	X
PDI 05	X
PDI 08	X
PDI 09	X
PDI 10	X

Indicator	POA-Dependent Quality Indicator
PDI 12	X
PDI 14	
PDI 15	
PDI 16	
PDI 18	
PDI 90	
PDI 91	
PDI 92	

Appendix C. List of Risk Factors for Area-Level Quality Indicator Modules Appendix

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

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

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

Appendix Table 5. 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[a], 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[a]
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[a]
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[a]
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

[a] PSI 08 for v2022 and earlier was specific to hip fractures. Starting in v2023, PSI 08 was expanded to include other types of hospital-associated fractures. Because the corresponding harm weights for other types of hospital-associated fractures are likely to differ from the harm weights based on hip fractures alone, revised harm weights will be estimated in the future.

Appendix E. Maternal Health Indicators

Background and Overview

A. Maternal Health Indicators Module

This appendix describes the empirical methods used to develop and calculate the Agency for Healthcare Research and Quality (AHRQ) Maternal Health Indicators (MHI) v2025. The intent of the module is to allow users to leverage claims data to generate area-level measures of maternal health. A beta version of the MHIs was originally released in v2024 and included measures that address severe maternal morbidity (SMM) and death associated with delivery hospitalization. In v2025, eight new measures (beta) were added to the module to address SMM and in-hospital death that occur during the postpartum period and behavioral health conditions that occur during the delivery hospitalization and postpartum period. For more information on the development of the MHIs included in v2025, please see the Scientific Rationale and Empirical Testing: Refinements to the Severe Maternal Morbidity Measure.

The MHI module contains a total of 11 indicators (Appendix Table 6).

Appendix Table 6. List of AHRQ Maternal Health Indicators (MHIs)

Abbrev	Indicator Name (v2025)	Area or Hospital Level
MHI 01	Inpatient Severe Maternal Morbidity Rate, at Delivery (20 indicators)	Area
MHI 02	Inpatient Severe Maternal Morbidity Plus In-Hospital Mortality Rate, at Delivery (20 indicators plus in-hospital mortality)	Area
MHI 03	Refined Inpatient Severe Maternal Morbidity Plus In-Hospital Mortality Rate, at Delivery, Beta (20 indicators plus in-hospital mortality)	Area
MHI 04	Inpatient Mental Health and Substance Use Disorders Rate, at Delivery, Beta	Area
MHI 05	Inpatient Severe Maternal Morbidity Rate, Delivery through 42 Days Postpartum, Beta (20 indicators)	Area
MHI 06	Inpatient Severe Maternal Morbidity Plus In-Hospital Mortality Rate, Delivery through 42 Days Postpartum, Beta (20 indicators plus in-hospital mortality)	Area
MHI 07	Refined Inpatient Severe Maternal Morbidity Plus In-Hospital Mortality Rate, Delivery through 42 Days Postpartum, Beta (20 indicators plus in-hospital mortality)	Area
MHI 08	Emergency Department and Inpatient Mental Health and Substance Use Disorders Rate, Days 1 to 42 Postpartum, Beta	Area
MHI 09	Emergency Department and Inpatient Encounters for Intentional Self-Harm Rate, Days 1 to 42 Postpartum, Beta	Area
MHI 10[a]	Emergency Department and Inpatient Substance Use Disorders and Accidental Overdose Rate, Days 1 to 42 Postpartum, Beta	Area
MHI 11	Emergency Department and Inpatient Perinatal Mood and Anxiety Disorders Rate, Days 1 to 42 Postpartum, Beta	Area

[a] Includes stratum-specific indicators.

B. Patient Population

The MHIs are reported as rate per 10,000 delivery-associated inpatient stays as is common among several SMM measures. Therefore, the denominator for the MHIs includes all deliveries within a one-year period, identified as inpatient discharges with an ICD-10-CM diagnosis code or ICD-10-PCS procedure code for delivery. The denominator is limited to those with female sex and age between 12 and 55 years, inclusive, and who reside in the same state as the hospital associated with the delivery. Discharges with abortions, either spontaneous or elective, based on diagnosis or procedure codes, are also excluded from the denominator.

C. Empirical Development of AHRQ MHIs

C.1 Numerator, Denominator, and Observed Rates for MHIs

Numerator

Numerators are based on the condition or procedure of interest and the time period specified for the measure. For indicators that measure events during the delivery hospitalization (MHI 01 – MHI 04), specified diagnoses or procedures included in the delivery discharge are counted in the numerator.[88] For indicators that measure events through 42 days after delivery (MHI 05 – MHI 11), patients are linked across time and setting. If a patient has an inpatient admission (MHI 05 – MHI 11) or an emergency department visit (MHI 08 – MHI 11) with one of the specified diagnoses or procedures during the postpartum period (0-42 days from delivery for MHI 05 – MHI 07 [includes delivery discharge]; 1-42 after the delivery for MHI 08 – MHI 11 [excludes delivery discharge]), the patient is included in the numerator. Only the first discharge with a numerator event is included and multiple admissions/ED visits are not counted more than once. For example, if a patient has one inpatient admission for mental health and substance use disorders one week after delivery discharge and one emergency department visit for mental health and substance use disorders one month after delivery discharge, the patient is only included once in the numerator for MHI 08.

Patients are excluded from the reference population numerator of MHIs if the patient resides in a state that did not contribute to the HCUP SID (MHI 01 – MHI 07) or did not contribute to both the HCUP SID and SEDD (MHI 08 – MHI 11). MHI 05 – MHI 07 require linkage between inpatient visits, and cases are excluded from states that do not permit linkage within SID.[89] MHI 08 – MHI 11 require linkage between inpatient and emergency department visits, and cases are excluded from states that do not permit linkage within and between SID and SEDD.[90]

Denominator

The denominator is based on the number of deliveries that occurred during a one-year period. MHI 01 – MHI 04, which count events that occur during the delivery hospitalization, use one year of SID data; therefore, the denominator is the total number of deliveries that occur within one year (for v2025, 2022 SID data are used). MHI 05 – MHI 07, which count events that occur through 42 days postpartum, use five quarters of data to allow for a 42-day follow-up period. The denominator includes deliveries that occur in the first four quarters of the five-quarter period (for v2025, deliveries are counted from Quarter 1 [Q1] 2021 – Quarter 4 [Q4] 2021) with numerator events being observed through Q1 2022.[91] MHI 08 –

[88] For more information on the specified diagnoses or procedure included in the numerator, see the MHI Technical Specifications, available at https://qualityindicators.ahrq.gov/measures/MHI_TechSpec.
[89] 21 states with linked SID were available in 2021 and 2022: AK, AR, CA, CO, DE, FL, GA, HI, IN, IA, MA, MD, ME, MS, NE, NM, NY, VT, WA, and WI.
[90] 15 states with linked SID and SEDD were available in 2021 and 2022: AK, AR, CA, CO, FL, GA, HI, IA, IN, MD, NE, NY, UT, VT, and WI.
[91] HCUP data only allow for quarter as the most granular date.

MHI 11, which count events that occur 1-42 days postpartum, require five quarters of data to allow for a 42-day follow-up period. The denominator includes deliveries that occur in the first year of the five-quarter period (for v2025, deliveries are counted from Q1 2021 – Q4 2021) with numerator events being observed through Q1 2022.

Deliveries that meet the following conditions are excluded from the denominator:

- The discharge is missing sex, age, quarter, year, or principal diagnosis.
- The patient is male.
- The patient is younger than 12 years old or older than 55 years old.
- The discharge includes a diagnosis or procedure for abortion, either spontaneous or elective.

Observed Rate

The observed rate of an MHI module indicator is:

- MHI 01 – MHI 04: The number of persons who had the condition or procedure of interest during the delivery discharge divided by the number of index deliveries.
- MHI 05 – MHI 07: The number of persons who had the condition or procedure of interest during the delivery discharge or during an inpatient readmission in the 42-day postpartum period divided by the number of index deliveries.
- MHI 08 – MHI 11: The number of persons who had the condition of interest during an inpatient readmission or emergency department visit in the 1 – 42 days postpartum period divided by the number of index deliveries.[92]

The observed rates are scaled to the rate per 10,000 deliveries. Observed rates are calculated by race/ethnicity, poverty level, state, and expected payer, as well as any user-defined stratum.

C.2 Reference Population

The MHIs are developed using hospital discharge abstracts and billing data from HCUP SID and SEDD. The SID and SEDD include 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 MHIs use three different reference populations, depending on indicator. Similar to the PQIs, the MHIs are developed using a reference population limited to community hospitals and excluding rehabilitation and long-term acute care (LTAC) hospitals. The reference populations for MHI 05 – MHI 07 and MHI 08 – MHI 11 are further limited to states that provided the necessary information to allow for linkage of patients across time and setting. Appendix Table 7 shows information on each of the three reference populations.

Appendix Table 7. Maternal Health Indicators Reference Populations

Indicators	Follow-up Period	Reference population	Denominator (v2025)	Numerator (v2025)	States
MHI 01 – MHI 04	Not applicable	One year of SID	2022 SID	2022 SID	47 states plus District of Columbia[1]

[92] For MHI 08 – MHI 11, the delivery discharge is excluded from the numerator.

Indicators	Follow-up Period	Reference population	Denominator (v2025)	Numerator (v2025)	States
MHI 05 – MHI 07	42 days postpartum	Five quarters of linked SID	Q1 2021 – Q4 2021 linked SID	Q1 2021 – Q1 2022 linked SID	21 states[2]
MHI 08 – MHI 11	1-42 days postpartum	Five quarters of linked SID and SEDD	Q1 2021 – Q4 2021 linked SID and SEDD	Q1 2021 – Q1 2022 linked SEDD	15 states[3]

[1] The 47 states plus DC included in the MHI 01 – MHI 04 reference population are: 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, NY, OH, OK, OR, PA, RI, SC, SC, TN, TX, UT, VA, VT, WA, WI, WC, and WY.
[2] The 21 states included in the MHI 05 – MHI 07 reference population are: AK, AR, CA, CO, DE, FL, GA, HI, IA, IN, MA, MD, ME, MS, NE, NM, NY, UT, VT, WA, and WI.
[3] The 15 states included in the MHI 08 – MHI 11 reference population are: AK, AR, CA, CO, FL, GA, HI, IA, IN, MD, NE, NY, UT, VT, and WI.

The reference population observed rates are published on the AHRQ QI website in MHI-specific Benchmark tables (https://qualityindicators.ahrq.gov/Downloads/Modules/MHI/V2025/Version_2025_Benchmark_Tables_MHI.pdf). The Benchmark tables include reference population observed rates by race/ethnicity, poverty level, state, and expected payer.