Methodology
U.S. News & World Report 2022-2023 Best Hospitals Procedures & Conditions Ratings

Zach Adams, MS
Tavia Binger, MSPH
Harold Chen
Avonelle Davis, MSPPM
Katherine Hilton
Ruoyu Ji, MS
Neal Kar, MPH
Kelsey Lu, MS
Min Hee Seo, PhD
Chelsey Wen, MPH
Ben Harder

July 26, 2022

To Whom It May Concern:

U.S. News & World Report's "Best Hospitals: Procedures & Conditions Ratings" study is the sole and exclusive property of U.S. News & World Report, L.P., which owns all rights, including but not limited to copyright, in and to the attached data and material. Any party wishing to cite, reference, publish or otherwise disclose the information contained herein may do so only with the prior written consent of U.S. News. Any U.S. News-approved reference or citation must identify the source as "U.S. News & World Report's Best Hospitals" and must include the following credit line: "Copyright © 2022 U.S. News & World Report, L.P. Data reprinted with permission from U.S. News." For permission to cite or use, contact permissions@usnews.com.

Executive Summary

This report describes the methodology underlying U.S. News & World Report's 2022-2023 Best Hospitals: Procedures & Conditions ratings of U.S. hospitals' performance in 20 benchmark procedures and conditions. Hospitals ratings, for each procedure and condition we have sufficient data to produce one for, are displayed on scorecards on usnews.com.

The procedures and conditions ratings significantly extend the mission of Best Hospitals: to provide a decision tool that helps the public identify hospitals that best meet their needs. Since 1990, the Best Hospitals rankings have focused on hospitals that excel in treating especially challenging inpatient diagnoses. However, a comparatively small number of patients need such hospitals compared with those who need relatively routine inpatient care. The procedures and conditions in which U.S. News began to rate hospitals in 2015 are much more typical of those needs and represent an integral part of the standard repertoire for most community hospitals. The ratings provide the public with information, using the best data sources we could locate, for consumers choosing, in consultation with their physicians, a local source of competent care.

U.S. News is committed to transparency and therefore publishes detailed descriptions of the methodologies used to rank and rate hospitals. Questions and constructive suggestions can be submitted to bhmethodology@usnews.com. The 2022-2023 ratings evaluate hospitals in the following procedures and conditions:

Abdominal aortic aneurysm repair (AAA)
Aortic valve surgery (AVR)
Back surgery (spinal fusion)
Chronic obstructive pulmonary disease (COPD)
Colon cancer surgery
Heart failure (CHF)
Diabetes
Heart attack
Heart bypass surgery (CABG)
Hip fracture
Hip replacement
Kidney failure
Knee replacement
Lung cancer surgery
Ovarian cancer surgery
Pneumonia
Prostate cancer surgery
Stroke
Transcatheter aortic valve replacement (TAVR)
Uterine cancer surgery

Unless otherwise noted, the metrics discussed on this document refer only to the ratings cohorts cited above. Ratings in other procedures and conditions may be added over time.

More than 4,400 hospitals are evaluated in at least one of the ratings cohorts, using methods developed by health researchers at U.S. News & World Report. Each hospital meeting the rating criteria is assigned to one of three overall performance bands – high performing, average and below average – so patients and families can quickly identify hospitals whose performance meets or exceeds the national norm. In the 2022-2023 ratings, 2,148 hospitals receive a high performing rating in one or more procedures and conditions, and 4 hospitals receive a high performing rating in all 20 procedures and conditions.

Sources of data include Medicare administrative claims, Medicare Care Compare, the American Hospital Association annual survey, publicly available data from clinical registries, and external designations.

These ratings reflect care received by patients age 65 and older. Older patients are at greater risk – they tend to have higher incidence and severity of comorbidities upon admission and illnesses that are more advanced than those of younger patients. While the quality of care of over-65 patients is generally regarded as indicative of a hospital's capabilities, U.S. News' assessments are not necessarily applicable to younger patients.

Changes Implemented in 2022-2023

- We introduced ratings for three additional cohorts: ovarian cancer surgery, prostate cancer surgery, and uterine cancer surgery.
- We refined our heart attack cohort to exclude cases of cardiac arrest and cardiogenic shock.
- We refined our kidney failure cohort to exclude cases of end-stage renal disease, flagged either by diagnosis or by Medicare status.
- We included outpatient volume in addition to inpatient volume for our hip replacement and prostate cancer surgery cohort models.
- We excluded visits in which a patient had a diagnosis of COVID-19. For each hospital's outcome measures, we also excluded visits from March 2020 and from other months in 2020 in which the hospital's COVID-19 rate among Medicare inpatients exceeded the national average or 15%, whichever is less. In addition, we calculated volume measures for each cohort using the higher of two observed visit counts: the count of visits from the most recent time period (2016-2020, for most cohorts), or the corresponding count from the time period used in last year's Procedures & Conditions analysis (i.e., 2015-2019).
- We introduced a new "giving patients time at home" outcome measure which indicated whether the amount of time a patient spent at home within the 30 days after discharge is above or below average for the cohort. Visits identified as transfers

from skilled nursing facilities (SNF) are excluded from the measure definition.

- We refined the denominator for our "discharging patients directly home" measure: in addition to excluding visits where the admission code indicated admission from a SNF, visits identified as transfers from SNF are excluded from the measure definition.
- We revised our reperfusion therapy measure in the stroke cohort to include cases in which a patient received reperfusion therapy in another hospital's emergency department before being transferred.
- We introduced a risk-adjustment in the uterine and ovarian cancer cohorts to account for whether a patient also had a secondary diagnosis of the other cancer.
- Rules were revised for cases in which patients met criteria for multiple cohorts during the same inpatient visit. In general, such patients were excluded from a condition cohort if, during the same visit, they met the inclusion criteria for a procedure cohort. For example, a visit in which a patient underwent TAVR and had a principal diagnosis of stroke is included only in the TAVR cohort.
- We introduced a new public transparency measure in the TAVR cohort. Hospitals received credit if they publicly reported quality metrics through the STS/ACC TVT registry, maintained through a collaboration between Society of Thoracic Surgeons (STS) and the American College of Cardiology (ACC).
- We introduced overuse restrictions on two procedure cohorts. A hospital classified as "high performing" by confirmatory factor analysis (CFA) alone is assigned a rating of "average" instead of "high performing" if it received a 1-star rating from researchers at the Lown Institute for avoiding overuse of arthroscopic knee surgery or carotid endarterectomy. This rule applies to the knee replacement cohort and the stroke cohort, respectively, and supplements the spinal fusion/laminectomy overuse rule implemented starting last year. As a result of this rule, 15 hospitals in the knee replacement cohort and 58 hospitals in the stroke cohort were downgraded from high performing to average.
- We revised the criteria defining Best Regional Hospitals. In order for a hospital to qualify as a Best Regional Hospital, it must be high performing in at least seven procedures/conditions, and have at least three more high performing procedures/conditions than below average procedures/conditions.

Table of Contents

Introduction

First published in 2015, Best Hospitals: Procedures & Conditions (formerly Best Hospitals for Common Care) is a key component of the U.S. News & World Report suite of healthcare consumer decision-support tools. For 2022-23, hospitals are rated in 20 common inpatient procedures and conditions:

Abdominal aortic aneurysm repair (AAA)
Aortic valve surgery (AVR)
Chronic obstructive pulmonary disease (COPD)
Colon cancer surgery
Heart failure (CHF)
Diabetes
Back surgery (Spinal fusion)
Heart attack
Heart bypass surgery (CABG)
Hip fracture
Hip replacement
Kidney failure
Knee replacement
Lung cancer surgery
Ovarian cancer surgery
Pneumonia
Prostate cancer surgery
Stroke
Transcatheter aortic valve replacement (TAVR)
Uterine cancer surgery

Although these procedures and conditions are services common to community hospitals, many studies demonstrate wide variability between hospitals in the quality of the care provided. Access to information about the performance of local hospitals enables patients, in consultation with their physicians, to better select hospitals that are the most likely to offer better, safer care.

In focusing on large numbers of patients with relatively straightforward needs, these ratings complement the Best Hospitals specialty rankings published annually by U.S. News since 1990. Those rankings identify facilities with demonstrable ability to handle a much smaller but far more challenging patient population of difficult and high-risk cases.

Quality of care has no ready definition or definitive metric, and there is no consensus on the best way to measure it. Some of its aspects are readily quantifiable while others are more challenging to measure. Moreover, what matters to one patient, such as reported levels of patient satisfaction, may be of little concern to another patient, who might prioritize rates of survival or complications. Offering not only an overall rating, but a window into the individual elements that make up the rating, recognizes the need for both.

Domains of Quality

Like the Best Hospitals specialty rankings, the procedure and condition ratings use the Donabedian paradigm, which reflects a relationship between structure, process, and outcomes. Avedis Donabedian described this now widely accepted dynamic in 1966[1], which is applied to hospital care as follows:

- *Structure* refers to hospital resources connected with patient care, such as the number of nurses, availability of certain specialists, and accreditations and certifications by outside organizations.
- *Process* refers to the way in which diagnoses, treatments, and practices to avoid harm to patients are rendered – whether steps known to be effective in preventing infections and medical errors, for example, are built into hospital routine.
- *Outcomes* refers to the results of care, including death, harm to patients, preventable readmissions, unusually long hospitalizations, and other consequences.

Failing to acknowledge the influence of random variation in quality metrics can produce results that misleadingly identify one hospital as superior or inferior to another. The methodology for the procedures and conditions ratings takes into account not only how each hospital performs on different measures but also the level of statistical certainty of those performance metrics. Larger sample sizes produce higher statistical confidence, which can result in a high-volume hospital with modestly above average results being rated more highly than a low-volume hospital with comparatively better observed results. This is because the second hospital's performance is more likely due to chance.

An important goal of the methodology is to give patients a clear bottom line. Despite the complexity of the measurement issues and the usefulness of particular types of information such as death and readmission rates, patients deserve an overall conclusion: How well does a hospital perform in a specific procedure or condition, like heart bypass surgery, compared to other hospitals? These ratings aggregate the measures in each cohort of care into an overall assessment by placing a hospital into one of three composite bands: high performing, average, and below average.

Data Sources

1. **Publicly available indicators**. Measures of performance are obtained from the public websites of Care Compare maintained by the federal Centers for Medicare & Medicaid Services (CMS), the Society of Thoracic Surgeons (STS), the American Heart Association (notably abbreviated GWTG in this document, to refer to their Get With The Guidelines program), the American Nurses Credentialing Center (ANCC), the American College of Cardiology (ANCC), and the National Cancer Institute (NCI).

[1] Donabedian, A. 1966. Milbank Memorial Fund Quarterly. Evaluating the Quality of Medical Care. 44(3), Part 2, 166-206. doi: 10.2307/3348969. https://www.jstor.org/stable/3348969?seq=1

2. **Medicare Beneficiary Summary Files (MBSF).** Administered by CMS, the Medicare beneficiary summary files contain demographic and coverage information pertaining to Medicare beneficiaries. All data are de-identified prior to being provided to U.S. News.

3. **Medicare Inpatient Limited Data Set Standard Analytical Files (LDS SAF).** Administered by CMS, the Inpatient LDS SAF contain inpatient hospitalization claims filed on behalf of patients enrolled in traditional Medicare. The LDS SAF provides a thorough administrative record for each patient across all inpatient encounters related to an episode of care. All data are de-identified prior to being provided to U.S. News.

4. **Medicare Outpatient Limited Data Set Standard Analytical Files.** The Outpatient LDS SAF contain final action claims filed by institutional providers for outpatient services covered by the Medicare Part B benefit. As with the other LDS SAF, all data are de-identified prior to being provided to U.S. News. Data from these files are used in order to attribute the volume of procedures performed in the outpatient setting by each hospital in the knee replacement, hip replacement, and prostate cancer cohorts, as well as to identify cases in which patients in the stroke cohort were initially seen in an emergency department before being transferred and admitted to another hospital, and whether or not they received reperfusion therapy. The analysis uses both the Base Files, which contain the base claim record and header information, as well as the Revenue Center Files, which contain line level HCPCS codes for each procedure.

5. **Medicare Skilled Nursing Facility (SNF) Limited Data Set Standard Analytical Files.** The SNF LDS SAF contain final action claims filed by institutional providers for skilled nursing facility services covered by the Medicare Part A benefit. As with the other LDS SAF, all data are de-identified prior to being provided to U.S. News. Data from these files are used to augment information on discharge, admission, and time at home.

6. A**merican Hospital Association (AHA) Annual Survey.** Through its Health Forum arm, the AHA surveys all U.S. hospitals annually (including AHA nonmembers) to obtain operational and clinically relevant information, such as types and levels of staffing. The collected data is the most complete of its kind available on U.S. hospitals.

7. **Hospital Consumer Assessment of Healthcare Providers and Systems Survey (HCAHPS)**. The federal government releases quarterly results of ongoing surveys of recently discharged inpatients conducted by more than 4,000 hospitals. The results comprise a rolling 12-month assessment of inpatients' opinions about their stay in various respects such as staff communication, treatment of pain and overall opinion of the hospital. The procedure and condition ratings incorporate overall patient opinions into the methodology. Other HCAHPS survey results are displayed but not integrated

into the ratings. Because the government aggregates HCAHPS data across each hospital, patients' opinions about their care in specific departments cannot be determined.

8. **Orthopedic Board Certification Data.** Information on physicians' board certification status in orthopedic surgery is obtained from the American Board of Orthopaedic Surgery (ABOS), the National Board of Physicians and Surgeons (NBPS), and the American Osteopathic Association (AOA), via Doximity.
(Disclosure: U.S. News & World Report holds an equity interest in Doximity.)

9. **Total Volume Data from the American Hospital Directory (AHD).** Data from AHD contain hospital-level total volume and Medicare Advantage (MA) volume by year for approximately 4,400 hospitals. AHD calculates this information using the CMS MEDPAR dataset. Because the Inpatient SAF files contain information for only traditional fee-for-service (FFS) Medicare beneficiaries, this data set is used to estimate the proportion of a hospital's inpatient services provided to MA patients in order to adjust the volume measure and account for all visits, not just FFS visits.

Selection of Procedures and Conditions

Procedures and conditions are selected based on the frequency of admission in the Medicare population, the ability to make hospital-to-hospital comparisons, and the presence of a sufficient degree of risk or complexity that the quality of a hospital's performance could be important.

Table 1: *Procedures and Conditions and Number of Patient Visits by Cohort, 2016-2020**

	Estimated Medicare Volume	
	Rated Hospitals	*All Hospitals*
Abdominal aortic aneurysm repair	111,405	113,749
Aortic valve surgery	374,672	377,316
Back surgery (Spinal fusion)	451,912	453,924
Chronic obstructive pulmonary disease (COPD)	1,617,079	1,618,422
Colon cancer surgery	197,429	203,436
Diabetes	321,655	329,877
Heart attack	1,221,505	1,228,619
Heart bypass surgery	487,278	487,361
Heart failure	3,258,835	3,260,005
Hip fracture	1,211,888	1,213,548
Hip replacement	1,089,777	1,092,976
Kidney failure	1,377,201	1,382,132
Knee replacement	2,031,854	2,033,647
Lung cancer surgery	117,453	120,123
Ovarian cancer surgery	24,370	26,699
Pneumonia	3,535,201	3,535,632
Prostate cancer surgery	149,936	156,397
Stroke	1,106,163	1,112,218
Transcatheter aortic valve replacement (TAVR)	279,350	279,664
Uterine cancer surgery	32,347	35,679

*Estimates include fee-for-service and estimated Medicare Advantage visits.

Table 1 above lists the procedures and conditions evaluated for publication, along with Medicare inpatient visit volume at both rated hospitals (those with 15 or more adjusted patient visits over the evaluation window) and at all hospitals, rated and unrated. For the stroke cohort, this reflects the number of patient visits during the period from October 2016 through December 2020, as this is when the NIH Stroke Scale (NIHSS) became available; for other cohorts, this reflects the number of patient visits during the five-year period from January 2016 through December 2020 (referred to as 2016-2020 throughout this document). As the table shows, most Medicare patient visits in these cohorts received care at hospitals that received a rating.

The cohorts created in this project are not identical to those created by CMS or other organizations in their performance indicators. In defining inclusion and exclusion criteria, three aims are paramount for maximizing statistical and clinical accuracy:

1. **Maximal homogeneity:** patients are as alike as possible other than with regard to factors that could be adequately managed through risk-adjustment.
2. **Maximal sample size:** selection of procedure and condition cohorts is limited to those with sufficiently large volume of care for statistical robustness and meaningfulness.
3. **Minimal coding variation:** coding definitions are relatively immune to large variations due to differences in coding practices. In considering this issue, it is particularly important to try to avoid systematic biases that might benefit particular organizations

and encourage gaming, as opposed to random coding variations that would simply add noise and reduce precision.

These three goals are not in harmony. While (1) argues for narrowly defined patient cohorts, (2) and (3) argue for broader inclusion criteria. This dynamic factors into determining which procedures and conditions we rate.

When we rate procedures or conditions for which CMS has also developed quality measures, our cohort and outcome inclusion criteria may differ from CMS. We aim to not distort outcome measures with decisions about the way in which patients are treated or procedures are coded. Using procedure codes to exclude patients from a cohort or to risk-adjust may be inappropriate if the choice of code and/or procedure is within a doctor's or hospital's discretion. In such cases, exclusion or risk-adjustment by procedure code could encourage upcoding, or perversely reward a hospital for performing a higher-risk procedure when a lower-risk alternative may be indicated, such as selection of open surgery over a minimally invasive procedure.

To the extent that a hospital's use of different interventions and associated procedure codes is a reliable indicator of a patient's risk, the desire for homogeneity suggests using procedure codes for risk-adjustment or to define exclusion criteria. However, to the extent that the use of different procedures represents a hospital's decisions in treating an otherwise homogeneous group of patients, procedure codes should not be used in this way. This last issue is of particular concern, since using procedure codes in this way could encourage manipulation of data. With these considerations in mind, we define our cohorts as follows:

Procedures

Abdominal aortic aneurysm repair. This cohort includes predominantly endovascular (closed) repair of abdominal aortic aneurysm, with the exception of risk-adjusted survival, where we included open repair and adjusted for approach. This cohort excludes repairs in other locations, as well as ruptured aneurysms and those with a claim admission type code of "1", indicating an emergent procedure. Patients undergoing emergent surgery typically are unable to choose which hospital they visit.

Aortic valve surgery. The cohort includes isolated open surgical aortic valve replacement and excludes concurrent coronary artery bypass. Transcatheter aortic valve therapies, which have become increasingly common since the time period covered by this analysis, are analyzed separately in the TAVR cohort, described in further detail below.

Colon cancer surgery. This cohort includes colon resection for colon cancer.

Back surgery (spinal fusion). This cohort includes thoracolumbar, lumbar, and lumbosacral spinal fusions, performed on patients with degenerative spinal conditions, and excludes spinal fractures or dislocations, spinal cord injuries, congenital or other anomalies, inflammatory spondylopathy, osteoporosis, and traumas, which may indicate non-elective spine surgery.

Heart bypass surgery. This cohort includes isolated open coronary artery bypass graft (CABG) and excludes concurrent valve replacement, repair, and other significant cardiac procedures.

Hip fracture. This cohort includes surgical repairs for pathologic or traumatic fractures of the hip, femoral head, or upper femur, and excludes fractures which received only medical care, minor procedures, or percutaneous interventions.

Hip replacement. This cohort includes primary arthroplasty of the hip for osteoarthritis and excludes partial joint replacement, revision, concurrent fracture, and concurrent hip and knee replacement. The volume measure includes procedures performed on an outpatient basis from January 2020 through the end of the analytic period.

Knee replacement. This cohort includes primary arthroplasty of the knee for osteoarthritis and excludes partial joint replacement, revisiont, and concurrent hip and knee replacement. The volume measure includes procedures performed on an outpatient basis from January 2018 through the end of the analytic period.

Lung cancer surgery. This cohort includes lobectomy, pneumonectomy, and sublobar resection, for lung cancer.

Ovarian cancer surgery. The cohort includes primary oophorectomy, hysterectomy, salpingectomy, and trachelectomy, for ovarian cancer.

Prostate cancer surgery. The cohort includes prostatectomy, as well as resection or excision of related structures often removed during the process of prostatectomy, including bilateral seminal vesicles, vas deferens, and pelvic lymph nodes, for prostate cancer. The volume measure includes procedures performed on an outpatient basis during the analytic period.

TAVR. This cohort includes all approaches (e.g. transfemoral and transapical) of isolated transcatheter aortic valve replacement. This procedure has emerged in recent years as a feasible, safe, and less invasive alternative to surgical aortic valve replacement (AVR). In 2011, TAVR was approved as an alternative to AVR for high risk patients. Since then approval has expanded, and volume of TAVR in the Medicare SAF database now eclipses that of surgical AVR.

Uterine cancer surgery. The cohort includes primary hysterectomy, oophorectomy, salpingectomy, or trachelectomy, for uterine cancer.

Conditions

CHF. This cohort includes principal nonhypertensive congestive heart failure, congestive heart failure, and certain other heart failure subgroups.

COPD. This cohort includes principal chronic obstructive pulmonary disease and bronchiectasis.

Diabetes. This cohort includes principal Type I and Type II diabetes mellitus, as well as certain "other specified" diabetes mellitus.

Heart attack. This cohort includes principal acute myocardial infarction and excludes cardiac arrest and cardiogenic shock, as defined by CMS.

Kidney failure. This cohort includes principal acute kidney failure and excludes diagnosis of or Medicare status code indicating end stage renal disease, as well as concurrent kidney transplant.

Pneumonia. This cohort includes both isolated principal pneumonia and principal sepsis with secondary pneumonia and without a secondary severe sepsis, as defined by CMS.

Stroke. The stroke cohort includes principal ischemic stroke, as defined by CMS.

Visits that meet criteria for both a procedure and a condition cohort during the same inpatient visit are usually limited to inclusion in the procedure cohort. However, if a visit is associated with either the TAVR or AVR cohort and the CHF cohort, or a visit is associated with the CABG cohort and either the heart attack or CHF cohort, the visit is included in both the procedure cohort and the condition cohort.

Inclusion of Providers and Cases

All hospitals represented in the 2020 AHA survey were initially considered for inclusion in the ratings analysis, unless categorized on the survey by a control (CNTRL) code (40-48) indicating federal government ownership.

Hospitals were also excluded if they lacked a valid six-digit Medicare provider number (MPN) to attribute to their AHA entity. In some cases, we attributed visits from multiple MPNs to a single AHA entity. This occurred when, in the judgment of U.S. News, the AHA entity encompassed the operations of two or more clinically integrated facilities or campuses that maintained separate MPNs during any portion of the analytic period.

In the condition cohorts only, we excluded hospitals with primary service (SERV) codes indicating service types other than general acute care, tuberculosis and other respiratory diseases, and heart, from rating eligibility, except in relevant specialties. Cancer hospitals were included only for colon cancer surgery, lung cancer surgery, ovarian cancer surgery, prostate cancer surgery, and uterine cancer surgery cohorts; respiratory hospitals were included only for lung cancer surgery, COPD, and pneumonia cohorts; and heart hospitals were excluded for diabetes, pneumonia, and kidney failure cohorts.

Table 2: *Number of Hospitals, by Cohort*

	Rated	**All**
Abdominal aortic aneurysm repair	1,262	1,592
Aortic valve surgery	912	1,068
Back surgery (Spinal fusion)	1,771	2,083
Chronic obstructive pulmonary disease (COPD)	4,124	4,286
Colon cancer surgery	2,369	3,283
Diabetes	2,892	4,138
Heart attack	2,737	3,907
Heart bypass surgery	1,088	1,119
Heart failure	4,137	4,286
Hip fracture	3,029	3,326
Hip replacement	2,813	3,217
Kidney failure	3,505	4,202
Knee replacement	3,157	3,331
Lung cancer surgery	1,152	1,574
Ovarian cancer surgery	457	1,002
Pneumonia	4,231	4,285
Prostate cancer surgery	1,180	2,095
Stroke	3,135	4,081
Transcatheter aortic valve replacement (TAVR)	688	736
Uterine cancer surgery	534	1,305

A small number of additional hospitals are excluded from ratings in individual cohorts where their volume is not large enough to allow estimation for at least one outcome used in that cohort. This occurred, for instance, with hospitals that began offering knee replacement near the end of our analytic period, but performed no surgeries during the surveillance period for postoperative infection in that cohort.

Cohort ratings are displayed on usnews.com for all other hospitals with adjusted cohort volume of at least fifteen visits (fee-for-service cases plus estimated managed care cases). For hospitals with adjusted cohort volumes of fewer than fifteen visits, we display information on selected metrics, but not overall composite ratings or claims-based outcome measures. The number of hospitals rated in each of the cohorts is shown in Table 2 Note that the numbers in Table 2 are inclusive of hospitals that have closed, or closed their evaluated service line(s), in order to include the fullest volume in our analysis.

Inpatient visits are aggregated from multiple claims (if needed), then excluded from cohort eligibility if they are missing key information for modeling purposes, contain data that were logically inconsistent, or otherwise indicate data entry errors, i.e.:

- The patient did not appear in the MBSF
- The patient sex was not identified

- The length of stay was greater than 365 days
- The patient date of death was prior to the admission date or relevant procedure date
- The patient had multiple dates of death
- The discharge was against medical advice

Visits from patients less than 65 years old are also excluded, because they represent a distinct population with a different medical profile than other Medicare patients.

Finally, visits in which a patient had a diagnosis of COVID-19, using the definition derived from the Centers for Disease Control and Prevention (CDC)[2], were excluded. For outcome measures, visits from March 2020 and for other months in which a hospital's COVID-19 rate among Medicare inpatients exceeded the national average inpatient prevalence or 15%, whichever was lower, were also excluded. Volume measures were calculated using either the count of visits from the current analytic time period (2016-2020, for most cohorts), or the corresponding count from last year's analytic time period (i.e., 2015-2019, for most cohorts), whichever was greater.

OUTCOMES

Outcomes are primarily derived from 2015-2020 LDS SAF dataset, which enables us to capture and attribute them to the index hospital, even if a patient experienced that outcome outside of that hospital or at a different facility. The surveillance periods from which index visits are drawn vary, depending on the pre- and post-admission or surgery surveillance requirements specific to each measure, in order to capture the most recent data available that meet those requirements.

All claims-based outcomes are risk-adjusted using a multi-level (hierarchical) logistic regression model that controls for potential confounders, with a random intercept for hospital identity. Details on the results and performance of risk-adjustment models for each cohort are listed under "Evaluation of Risk-Adjustment Models". In all instances, hospital-specific random intercepts are treated as such in composite modeling in order to make maximum use of the information contained in the variable, and to minimize the risk of measurement error due to categorization. Categorical groupings and descriptions of hospital-specific random intercepts are displayed on scorecards. The details are listed under "Categorical Display".

The following claims-based, risk-adjusted outcomes are used in the final composite models to evaluate each hospital's performance relative to others in the cohort. The relative contributions of each outcome to the overall cohort ratings are depicted under "Indicators and Correlations With Scores." Surveillance windows for index cases are provided in parentheses after each description.

1. **Mortality within 30 days (labeled "Survival" on scorecards)**. Reflects death within 30 days of surgery for procedure cohorts, or 30 days of admission for condition cohorts.

[2] Centers for Disease Control and Prevention (2020). ICD-10-CM Official Coding and Reporting Guidelines. https://www.cdc.gov/nchs/data/icd/COVID-19-guidelines-final.pdf and https://www.cdc.gov/nchs/data/icd/Announcement-New-ICD-code-for-coronavirus-19-508.pdf.

(9/1/2016 - 12/1/2020 for stroke, 12/1/2015 - 12/1/2020 for other cohorts)

2. **Unplanned readmission within 30 days ("Readmission prevention").** Unless otherwise noted, reflects unplanned inpatient readmission within 30 days of discharge, similar to the CMS hospital-wide 30-day unplanned readmission measure definition[3]. For some cohorts, this measure may reflect additional cohort-specific criteria[4,5,6]. (12/1/2015 - 12/1/2020)
3. **Surgical site infection ("Infection prevention"), hip replacement, knee replacement, AAA, CABG, and AVR cohorts**. Reflects development of a surgical site infection following the index procedure. Published literature[7,8,9,10,11] indicates that a careful approach to constructing claims-based infection measures can accurately identify hospitals with unusually low or high infection rates. (12/31/2014 - 12/31/2019 for hip replacement and knee replacement, 11/1/2015 - 11/1/2020 for AAA, CABG, and AVR)
4. **Revision within 1 year ("Prevention of revision surgery"), hip replacement and knee replacement cohorts**. Reflects subsequent procedure to address problems with a joint replacement within 1 year of the original surgery. (12/31/2014 - 12/31/2019)
5. **Prolonged hospitalizations ("Prevention of prolonged hospitalization")**. Reflects length of stay duration in the highest quartile. (12/1/2015 - 12/1/2020)
6. **Discharge to a location other than the patient's home ("Discharging patients**

---

[3] Horwitz, L. I., Partovian, C., Lin, Z., Grady, J. N., Herrin, J., Conover, M., Drye, E. E. (2014). Development and use of an administrative claims measure for profiling hospital-wide performance on 30-day unplanned readmission. *Annals of Internal Medicine, 161*(0), S66–S75. http://doi.org/10.7326/M13-3000

[4] Center for Medicare & Medicaid Services. (n.d.). *Coronary Artery Bypass Graft (CABG) Surgery 30-day readmission rate*. Measure Details. https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=1426.

[5] Center for Medicare & Medicaid Services. (n.d.). *30-Day Unplanned Readmissions for Cancer Patients*. Measure Details. https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=6030.

[6] Center for Medicare & Medicaid Services. (n.d.). *Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) 30-day readmission rate*. Measure Details. https://cmit.cms.gov/CMIT_public/ViewMeasure?MeasureId=899.

[7] Calderwood, M. S., A. Ma, Y. M. Khan, M. A. Olsen, D. W. Bratzler, D. S. Yokoe, D. C. Hooper*, et al.* "Use of Medicare Diagnosis and Procedure Codes to Improve Detection of Surgical Site Infections Following Hip Arthroplasty, Knee Arthroplasty, and Vascular Surgery." *Infect Control Hosp Epidemiol* 33, no. 1 (Jan 2012): 40-9.

[8] Letourneau, A. R., M. S. Calderwood, S. S. Huang, D. W. Bratzler, A. Ma, and D. S. Yokoe. "Harnessing Claims to Improve Detection of Surgical Site Infections Following Hysterectomy and Colorectal Surgery." *Infect Control Hosp Epidemiol* 34, no. 12 (Dec 2013): 1321-3.

[9] Calderwood, M. S., K. Kleinman, D. W. Bratzler, A. Ma, R. E. Kaganov, C. B. Bruce, E. C. Balaconis*, et al.* "Medicare Claims Can Be Used to Identify Us Hospitals with Higher Rates of Surgical Site Infection Following Vascular Surgery." *Med Care* 52, no. 10 (Oct 2014): 918-25.

[10] Calderwood, M. S., K. Kleinman, D. W. Bratzler, A. Ma, C. B. Bruce, R. E. Kaganov, C. Canning*, et al.* "Use of Medicare Claims to Identify Us Hospitals with a High Rate of Surgical Site Infection after Hip Arthroplasty."*Infect Control Hosp Epidemiol* 34, no. 1 (Jan 2013): 31-9.

[11] Calderwood, M. S., Kleinman, K., Murphy, M. V., Platt, R., Huang, S. S. "Improving Public Reporting and Data Validation for Complex Surgical Site Infections After Coronary Artery Bypass Graft Surgery and Hip Arthroplasty." Open Forum Infectious Diseases 1, no. 3 (Dec 2014).

directly home"). Reflects discharge to a location other than home, such as a long-term acute care facility or a different hospital. More details are provided in Appendix A. (10/1/2016 - 12/31/2020 for stroke, 12/31/2015 - 12/31/2020 for other cohorts)

7. **Stroke on procedure date ("Prevention of stroke"), CABG, AVR, and TAVR cohorts.** Reflects stroke on the index procedure date.
(12/31/2015 - 12/31/2020)
8. **Time spent at home within 30 days of discharge ("Giving patients time at home").** Reflects whether the amount of time spent at home within the 30 days after a hospital visit was above or below average.
(12/1/2015 - 12/1/2020)

Other claims-based, risk-adjusted outcome measures are included in the model selection process, but not in the final composite measure, including time to joint revision within five years, complications of total joint replacement (NQF #1550[12]), and a readmission measure for cancer cohorts (closely following specifications developed by the Alliance of Dedicated Cancer Centers).

Process Measures

We evaluate a variety of process measures, obtained primarily from the CMS Care Compare website as well as the inpatient claims datasets. Most are excluded prior to modeling, due to missing data or other data validity concerns, while others did not demonstrate good empirical fit. The following measures are included in the composite model for one or more cohorts:

- **Worker flu immunization.** Percentage of healthcare personnel at the hospital who received a timely vaccination during flu season. Derived from the CMS Care Compare Database.
- **Noninvasive ventilation.** Hospitals receive credit in the CHF or COPD cohort if, in each cohort respectively, as identified on their Medicare claims, at least 20% of cohort visits over the analytic period requiring ventilation were performed noninvasively.
- **Patient experience.** Overall hospital linear mean score of recently discharged patient experience from the HCAHPS survey from the 7/1/2020-3/31/2021 data[13]. We use this score over the star rating because it is a continuous measure that provides more information.
- **Board certification.** Percentage of hip replacement, knee replacement, and hip fracture repair visits, respectively, performed by board-certified orthopedic surgeons. The measure accounts for both MDs and DOs. Board-certified orthopedic surgeons are identified in data (a) provided by ABOS and AOA to Doximity prior to April 12, 2022 and (b) self-reported by

[12] Centers for Medicare & Medicaid Services. (2019, September). *Hospital-Level Risk-Standardized Complication Rate (RSCR) Following Elective Primary Total Hip Arthroplasty (THA) and/or Total Knee Arthroplasty (TKA) (NQF #1550).* https://innovation.cms.gov/files/fact-sheet/bpciadvanced-fs-nqf1550.pdf.

[13] The current version of the survey is available at http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS.html.

NBPS-certified orthopedic surgeons to Doximity prior to April 12, 2022. Surgeons are linked to hospitals where they operate using National Provider Identifier information on Medicare claims rather than hospital affiliations reflected in doctors' Doximity profiles.

- **Reperfusion therapy.** Percentage of ischemic stroke patients treated with reperfusion therapy, either with intravenous tissue plasminogen activator or mechanical thrombectomy. When both treatments are used in accordance with guidelines, studies demonstrate that they improve short-term and long-term outcomes, such as reduced mortality and the incidence of hemorrhage, and an increased likelihood of patients being discharged to home.[14,15,16,17] ICD codes of alteplase administration or mechanical thrombectomy among all ischemic stroke patients quantify how often the hospital provided time-sensitive medical interventions during the period from October 2016 through December 2020. HCPCS codes denoting reperfusion therapy identify cases in which therapy was provided in an outpatient setting (i.e., emergency department) prior to inpatient admission.

Structural Measures

Structural measures of health care evaluate staff, services, equipment and other resources used to deliver care. Structural indicators that have been associated with good outcomes for patients were included. In addition to volume, seven structural indicators are employed.

- **Volume.** There is widespread evidence that hospitals performing a procedure more frequently get better outcomes. Volume derived from Medicare claims is therefore included as an indicator. In the AVR rating, the volume of TAVR procedures is combined with the volume of AVR since the two procedures are used to treat the same medical conditions. In the CABG rating, the volume of CABG procedures is combined with the volume of AVR procedures.

---

[14] Fonarow, G. C., Zhao, X., Smith, E. E., Saver, J. L., Reeves, M. J., Bhatt, D. L., Xian, Y., Hernandez, A. F., Peterson, E. D., & Schwamm, L. H. (2014). Door-to-needle times for tissue plasminogen activator administration and clinical outcomes in acute ischemic stroke before and after a quality improvement initiative. *JAMA*, *311*(16), 1632–1640. https://doi.org/10.1001/jama.2014.3203

[15] Saver, J. L., Goyal, M., van der Lugt, A., Menon, B. K., Majoie, C. B., Dippel, D. W., Campbell, B. C., Nogueira, R. G., Demchuk, A. M., Tomasello, A., Cardona, P., Devlin, T. G., Frei, D. F., du Mesnil de Rochemont, R., Berkhemer, O. A., Jovin, T. G., Siddiqui, A. H., van Zwam, W. H., Davis, S. M., Castaño, C., … HERMES Collaborators (2016). Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-analysis. *JAMA*, *316*(12), 1279–1288. https://doi.org/10.1001/jama.2016.13647

[16] Powers, W. J., Rabinstein, A. A., Ackerson, T., Adeoye, O. M., Bambakidis, N. C., Becker, K., Biller, J., Brown, M., Demaerschalk, B. M., Hoh, B., Jauch, E. C., Kidwell, C. S., Leslie-Mazwi, T. M., Ovbiagele, B., Scott, P. A., Sheth, K. N., Southerland, A. M., Summers, D. V., & Tirschwell, D. L. (2019). Guidelines for the Early Management of Patients With Acute Ischemic Stroke: 2019 Update to the 2018 Guidelines for the Early Management of Acute Ischemic Stroke: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke*, *50*(12), e344–e418. https://doi.org/10.1161/STR.0000000000000211

[17] Saver, J. L., Fonarow, G. C., Smith, E. E., Reeves, M. J., Grau-Sepulveda, M. V., Pan, W., Olson, D. M., Hernandez, A. F., Peterson, E. D., & Schwamm, L. H. (2013). Time to treatment with intravenous tissue plasminogen activator and outcome from acute ischemic stroke. *JAMA*, *309*(23), 2480–2488. https://doi.org/10.1001/jama.2013.6959

In order to account for total knee replacement transitioning to the outpatient setting[18], we combine knee volume from the Outpatient LDS SAF with inpatient knee volume. We apply the same strategy for hip replacement and prostate cancer surgery.

Volume totals in each procedure or condition cohort were adjusted to account for MA cases. Hospitals with very low volumes – defined as fewer than 15 cases over five years – were not rated because their numbers were too low to establish whether the quality of care was different from average.

- **Nurse staffing**. The number of nurses involved in direct patient care at a hospital is known to play a major role in the quality of care[19,20,21,22,23,24]. For this project, we conceptualize a nurse staffing index as a ratio reflecting inpatient and outpatient nursing. The numerator is the total number of staff registered nurses (RNs), converted to full-time equivalents (FTEs). For example, two half-time nurses add up to one FTE. Only non-supervisory nurses with an RN degree from an approved nursing school and a current state registration are included.

  Making sense of nurse staffing requires comparing the number of staff to the total workload. The two most commonly used approaches are total patient days and adjusted average daily census of patients, and we use the latter for the composite models for the procedures and conditions ratings, as it better conceptualizes the total workload of nursing, taking into account both inpatient and outpatient revenue adjusted by the number of days that the facility is open during the reporting period. The adjusted average daily census of patients obtained from the AHA survey reflects the number of days of inpatient care plus an estimate of the volume of outpatient services, expressed in units equivalent to an inpatient day in terms of level of effort. The latter is derived by first multiplying the number of outpatient visits by the ratio of outpatient revenue per outpatient visit to inpatient revenue per inpatient day (to get the number of patient days attributable to outpatient services), then adding that to the number of inpatient days. The nurse staffing ratio for each year is mathematically expressed as the following: (inpatient days + (inpatient days * (outpatient revenue/inpatient revenue))/number of days in the reporting period).[25]

  The nurse staffing index is then a ratio of FTE registered nurses divided by adjusted patient days. We use the three-year average of this index, calculated for each year, in order to reduce the impact of year-to-year variation in reporting, so the average of the 2018, 2019,

[18] Total Knee Arthroplasty (TKA) Removal from the Medicare Inpatient-Only (IPO) List and Application of the 2-Midnight Rule. (2019, January 24). MLN Matters, SE19002.

[19] Unruh, L. (2003) Licensed Nurse Staffing and Adverse Events in Hospitals. Medical Care. 41(1) (pp142-152)

[20] Stanton MW, Rutherford MK. (2004) Hospital nurse staffing and quality of care. Agency for Healthcare Research and Quality. Research in Action Issue 14. AHRQ Pub. No. 04-0029.

[21] Spetz J, Donaldson N, Aydin C, Brown DS. (2008) How Many Nurses per Patient? Measurements of Nurse Staffing in Health Services Research. Health Services Research. 43(5) (pp1674-1692)

[22] Lankshear AJ, Sheldon TA, Maynard A. (2005) Nurse Staffing and Healthcare Outcomes. Advances in Nursing Science. 28(2) (pp163-174)

[23] Hickham DH, Severance S, Feldstein A. (2003) The Effect of Health Care Working Conditions on Patient Safety. AHRQ Evidence Report/Technology Assessment (74)

[24] Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR, Harris M. (2011) Nurse Staffing and Inpatient Hospital Mortality. New England Journal of Medicine 364(11) (pp1037-1045)

[25] This can be found in the survey code book for the AHA annual survey.

and 2020 nurse staffing ratios derived from the AHA survey for this year's ratings.

- **ICU specialists**. Intensivists are board-certified physicians with subspecialty or fellowship training in critical-care medicine. They specialize in managing critically ill patients in hospital intensive care units (ICUs). A hospital receives credit if it reported having at least one full-time equivalent intensivist available, on staff or from another source, in any ICU other than neonatal or pediatric. Research indicates that better outcomes are associated with the presence of intensivists.[26, 27] This measure is derived from the 2020 AHA survey. Aortic valve surgery, heart attack, and TAVR cohorts received a credit if they had either one specialized intensive-care unit physician or a cardiac intensive-care unit.
- **Cardiac intensive care unit.** Cardiac Intensive Care Units (CICU) are specialized units that are designed to manage patients who are critically ill with serious heart conditions or who are recovering from heart surgery. Hospitals receive credit if they reported having a cardiac intensive care unit. This measure is derived from the 2020 AHA survey. In the aortic valve surgery, heart attack, and TAVR cohorts, hospitals receive credit if they had either one specialized intensive-care unit physician or a cardiac intensive-care unit.
- **Advanced heart program.** Indicates whether a hospital provided CHF patients with either left ventricular assist devices (LVAD) implantation or heart transplantation according to Medicare inpatient claims data within the analysis time frame.
- **NCI-designated Cancer Center and/or American College of Surgeons (ACS) Commission on Cancer.** In cancer surgery cohorts, this measure identifies whether a hospital is recognized as a designated cancer center by the NCI, a member of the ACS Commission on Cancer, or both. Hospitals with both received the maximum score on this measure, hospitals with neither received the minimum score, and hospitals with one organization but not the other received an intermediate score. The NCI funds clinical trials and other advances in care, and the ACS provides tools and resources to help hospitals deliver high quality, patient-centered care.
- **Public transparency.** Public transparency measures were incorporated into six ratings, based on hospitals' public reporting status in relevant clinical registries. This is done in part to encourage all hospitals, regardless of performance, to release their data and by doing so expand the data universe. As a result, it has the advantages of allowing researchers to evaluate the results of hospital ratings, facilitating informed decision making by patients, and demonstrating a public commitment of pursuing quality improvement. Table 3 shows which registries correspond to each cohort.
- **GWTG recognized hospital.** Hospitals receive credit in this measure by voluntarily reporting quality metrics to the public through websites maintained by the American Heart Association under its GWTG quality improvement programs. In order to receive a credit, hospitals must have opted into the public reporting program and been appearing on their

---

[26] Pronovost PJ, Holzmueller CG, Clattenburg L, Berenholtz S, Martinez EA, Paz JR, Needham DM. "Team care: beyond open and closed intensive care units." *Current Opinion in Critical Care*. 2006; 12(6):604-8.

[27] Sapirstein A, Needham DM, Pronovost PJ. "24-hour intensivist staffing: balancing benefits and costs." *Critical Care Medicine*. 2008; 36(1):367-8.

public reporting site by 12/31/2021.

- **ACC recognized hospital.** Hospitals receive credit for participating in the ACC National Cardiovascular Disease Registry data-reporting initiatives if they also agreed to allow their ACC-calculated results to be publicly reported on the ACC's website. To receive credit for ACC public reporting, hospitals must have voluntarily agreed to allow data from at least one of these registries to be posted on the ACC registry website, www.CardioSmart.org.
- **STS recognized hospital.** Hospitals receive credit in heart bypass surgery and aortic valve surgery cohorts if they permit STS to publicly report their performance data. Published research by STS-affiliated researchers[28] and independent analysis by U.S. News found that hospitals that do not publicly report via STS performed worse than STS reporters on quality measures such as risk-adjusted mortality, morbidity and readmissions following heart surgery. While not establishing the direction of causality, these observed correlations between STS-mediated transparency and better outcomes support the use of transparency as an indicator of higher quality of care.[29]
- **STS/ACC TVT registry recognized hospital.** Hospitals receive credit for participating in the STS/ACC TVT Registry, created and maintained through a collaboration between STS and the ACC, if they also allowed their results to be publicly reported on the registry website.

Table 3: *Public Transparency*

Cohort	Public Reporting Program	Opt-in Date	Source
AVR	STS	12/31/2021	Society of Thoracic Surgeons
CABG	STS	12/31/2021	Society of Thoracic Surgeons
TAVR	STS/ACC	12/31/2021	American College of Cardiology
Heart attack	ACC†	12/31/2021	American College of Cardiology
(either registry)	GWTG-Heart*	12/31/2021	American Heart Association
Heart failure	ACC†	12/31/2021	American College of Cardiology
(either registry)	GWTG-Heart*	12/31/2021	American Heart Association
Stroke	GWTG-Stroke	12/31/2021	American Heart Association

* *For GWTG-Heart credit, hospital has to publicly report data in at least one of the following registries: AFib, Coronary Artery Disease, Heart Failure, or Resuscitation.*
† *For ACC credit, hospital has to publicly report data in at least one of the following registries: EP Device Implant (formerly ICD), Chest Pain - MI, or CathPCI.*

Risk-Adjustment for Medicare Claims-Based Outcomes

When comparing outcomes between hospitals, adjusting for differences in the patients treated at each hospital is critical. A hospital with a 50% mortality rate might be superior to a

[28] Shahian, David M., et al. "The Society of Thoracic Surgeons voluntary public reporting initiative: the first 4 years." Annals of surgery 262.3 (2015): 526-535.
[29] Data was extracted from the STS website (https://publicreporting.sts.org/) on January 5, 2022 and contains information up until December 2019.

hospital with a 10% mortality rate if most of the patients at the first hospital are expected to die and most of the patients at the second hospital are low risk.

We use multilevel logistic regression models to adjust for differences in case mix between hospitals. Multilevel models are a form of regression that allocates variance between variables on two or more levels. We use the empirical Bayes estimate of the hospital intercept as an estimate of each hospital's value for a given outcome. Multilevel modeling accounts for clustering of patient observations within hospitals and allows for more precise rating of hospitals with lower patient volume and fewer outcomes.

We select covariates for inclusion in risk-adjustment models based on the literature, discussions with clinicians in relevant specialties and a causal-inference model aimed at achieving unbiased estimation of the effect of treatment at a particular hospital on a given outcome.

The causal model (Appendix B) indicates that an unbiased estimate of the effect of treatment at a given hospital as compared to a hospital selected at random from among those eligible for rating in a cohort, requires adjustment for age, sex, comorbidities, severity of index condition, socioeconomic status (SES), admission urgency, inbound transfer status, and year of admission. In certain instances, we control for the severity of the index condition. Because severity is correlated with many of the other covariates for which we adjusted, we suspect residual confounding is negligible. "Strengths and Limitations" contains further discussion of this issue.

Risk-Adjustment Variables

- **Age at admission**. Age in years as a continuous variable, obtained from the denominator or Master Beneficiary Summary file.
- **Sex**. Male or female.
- **Inbound transfer status.** Transfer from the initial receiving hospital may indicate a complex case. Visits are classified as an inbound transfer if the patient was treated at another acute care hospital on the day of admission, if the claim admission source variable indicates inbound transfer, or if the preceding visit indicates outbound transfer.
- **Year of hospital admission** Quality of care tends to improve over time, so year-over-year risk of adverse outcomes should decrease.
- **Elixhauser comorbidities.** Comorbidities identified by Elixhauser et al[30] are highly predictive of mortality.[31] All 29 comorbidities identified with AHRQ's Elixhauser comorbidity software version 2020.1, released in Oct 2019, which overlaps with the study period of 2016-2020, are individually adjusted for.
- **Medicare status code**. The reason(s) why the patient is eligible for Medicare: age, disability, or end-stage renal failure. Medicare status code is conceptualized as a proxy for comorbidities.

---

[30] Elixhauser, Anne, et al. Comorbidity measures for use with administrative data. *Medical care* 36.1 (1998): 8-27.

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

- **Socioeconomic status**. Patients with lower incomes are typically sicker when they arrive at the hospital, and may face more challenges in obtaining or managing their care after they are discharged. This can affect their risk of death, readmission and complications. When hospitals differ by the socioeconomic status of their patients, this can create bias in comparing outcomes. Our risk-adjustment models include "dual-eligibility" as a measure of socioeconomic background, and patients who are eligible for both Medicare and Medicaid are treated as a separate risk group.
- **ICD version.** The ICD version each visit was coded under is controlled for in models for two cohorts: hip replacement and knee replacement. Cohort visits dated September 30, 2015 or earlier use ICD-9, while all other visits in these and all other cohorts use ICD-10.
- **Condition cohort-specific covariates.** Binary variables indicating whether a patient had ever left against medical advice, been admitted for the same condition, or had a history of mechanical ventilation are included in the CHF and COPD models. A binary measure indicating whether a patient had a diagnosis of sepsis is included for the pneumonia cohort. Binary variables indicating whether a patient had a diagnosis of ST elevation myocardial infarction (STEMI) of anterior wall, STEMI of inferior wall, or non-ST elevation myocardial infarction (NSTEMI) are included in the heart attack models. Binary variables indicating whether a patient was diagnosed with diabetes ketoacidosis (DKA) and hypoglycemia are included in the diabetes models. A continuous NIHSS score, and binary variables indicating whether a patient was transferred from an emergency department or had previously been diagnosed with a transient ischemic attack (TIA), are included in the stroke models. For stroke visits in which NIHSS was not recorded, it is imputed using multiple imputation, generating imputed values by fitting a linear mixed effects model with patient-level and hospital-level attributes in order to incorporate the hierarchical structure of patient-visit data.
- **Surgical cohort-specific covariates.** A binary variable indicating whether the operation was performed on both joints simultaneously (bilaterally) is included in the hip replacement and knee replacement models. A binary variable indicating approach (open or endoscopic) is included in the AAA mortality model. A binary variable indicating diagnosis of CHF or heart attack is included in the CABG models. An ordinal variable indicating the type of degenerative condition (e.g., scoliosis) is included in the back surgery models. A binary variable indicating whether a patient also had a secondary diagnosis of the other cancer is included in the ovarian and uterine models.
- **History of stroke.** A variable indicating history of stroke in the year prior to surgery is included in the stroke model for the TAVR, AVR, and CABG cohorts.

Evaluation of Risk-Adjustment Models

The accuracy of risk-adjustment models is measured by two statistics, the C-statistic and the Hosmer-Lemeshow goodness of fit statistic. The C-statistic estimates the probability that if one subject who experienced an outcome (death, for example) and another who did not are drawn randomly from the data, the model will assign a higher probability of death to the person who died.

A C-statistic of .5 indicates the model has no better than random chance at predicting the outcome. A C-statistic in the .60-.69 range indicates limited discrimination, .70-.79 indicates acceptable discrimination and above .8 indicates good discrimination.

Typically, the C-statistic for mortality models implemented using clinical data range from approximately .75-.85[32]. Our models for outcomes are generally of similar predictive quality as those based on clinical data. Our models for readmission and others have lower predictive power, with C-statistics similar to those in the published literature drawing on claims data. The Hosmer-Lemeshow goodness of fit statistic looks at whether the observed number of outcomes matches the expected number predicted by the model in samples of the population. As this test is not informative for samples over 25,000, we use a procedure designed to evaluate Hosmer-Lemeshow fit in large samples, in which multiple Hosmer-Lemeshow tests are conducted on small samples of the data. A Hosmer-Lemeshow test results in a p-value, which conventionally indicates likely bad fit when below 0.05 unlikely bad fit when closer to 1. For the stroke cohort, 10 sets of fit indices are combined together using Rubin's rule after imposing multiple imputation.[33]

[32] e.g.: Kozower, Benjamin D., et al. "STS database risk models: predictors of mortality and major morbidity for lung cancer resection." *The Annals of Thoracic Surgery*90.3 (2010): 875-883; Hamel, Mary Beth, et al. "Surgical outcomes for patients aged 80 and older: morbidity and mortality from major noncardiac surgery." *Journal of the AmericanGeriatrics Society* 53.3 (2005): 424-429.

[33] Rubin, D. B. (2004). Multiple imputation for nonresponse in surveys (Vol. 81). John Wiley & Sons

Table 4: *Predictive Accuracy of Risk-adjustment Models*

Cohort	Outcome	C-statistic	Mean (min, max) of Large-sample Hosmer-Lemeshow Tests
Abdominal aortic aneurysm repair	Survival	0.865	0.56 (0.01,0.96)
	Discharging patients directly to home	0.851	0.45 (0.10,0.88)
	Readmission prevention	0.689	0.54 (0.03,1.00)
	Infection prevention	0.897	0.74 (0.00,1.00)
Aortic valve surgery	Survival	0.802	0.46 (0.03,0.97)
	Discharging patients directly to home	0.809	0.50 (0.03,0.84)
	Prevention of prolonged hospitalizations	0.823	0.57 (0.03,0.95)
	Readmission prevention	0.699	0.28 (0.06,0.72)
Back surgery (Spinal fusion)	Survival	0.798	0.62 (0.00,1.00)
	Discharging patients directly to home	0.772	0.38 (0.03,0.82)
Chronic obstructive pulmonary disease (COPD)	Survival	0.749	0.51 (0.24,0.94)
	Discharging patients directly to home	0.741	0.43 (0.01,0.88)
Colon cancer surgery	Survival	0.842	0.45 (0.00,0.96)
	Giving patients time at home	0.780	0.23 (0.00,0.88)
	Prevention of prolonged hospitalizations	0.844	0.36 (0.01,0.64)
Diabetes	Survival	0.768	0.57 (0.10,0.95)
	Discharging patients directly to home	0.708	0.37 (0.06,0.72)
Heart attack	Survival	0.804	0.56 (0.06,0.90)
	Discharging patients directly to home	0.809	0.43 (0.11,0.90)
Heart bypass surgery	Survival	0.803	0.53 (0.00,0.95)
	Discharging patients directly to home	0.800	0.54 (0.02,1.00)
	Prevention of prolonged hospitalizations	0.802	0.43 (0.06,0.91)
	Prevention of stroke	0.836	0.72 (0.31,0.95)
Heart failure	Survival	0.708	0.48 (0.02,0.98)
	Discharging patients directly to home	0.716	0.39 (0.05,0.99)
Hip fracture	Survival	0.761	0.47 (0.00,0.95)
	Readmission prevention	0.708	0.51 (0.06,0.81)
Hip replacement	Survival	0.840	0.81 (0.06,1.00)
	Prevention of prolonged hospitalizations	0.826	0.69 (0.34,0.95)
	Prevention of revision surgery	0.666	0.62 (0.32,0.92)
	Infection prevention	0.762	0.41 (0.01,0.99)
Kidney failure	Survival	0.758	0.42 (0.06,0.96)
	Giving patients time at home	0.631	0.44 (0.07,0.75)
Knee replacement	Survival	0.791	0.75 (0.07,1.00)
	Prevention of prolonged hospitalizations	0.808	0.32 (0.00,0.63)
	Readmission prevention	0.733	0.49 (0.06,0.96)
	Prevention of revision surgery	0.681	0.44 (0.00,0.92)
Lung cancer surgery	Survival	0.864	0.52 (0.01,0.97)
	Discharging patients directly to home	0.822	0.35 (0.00,0.89)
	Prevention of prolonged hospitalizations	0.795	0.45 (0.05,0.93)
Ovarian cancer surgery	Survival	0.885	0.48 (0.02,0.96)
	Discharging patients directly to home	0.829	0.51 (0.07,0.92)
	Prevention of prolonged hospitalizations	0.804	0.32 (0.01,0.76)
	Readmission prevention	0.649	0.30 (0.00,0.79)
Pneumonia	Survival	0.745	0.22 (0.00,0.69)
	Discharging patients directly to home	0.753	0.33 (0.01,0.64)
Prostate cancer surgery	Survival	0.907	0.66 (0.01,0.97)
	Discharging patients directly to home	0.936	0.73 (0.28,0.93)
	Prevention of prolonged hospitalizations	0.835	0.38 (0.02,0.94)
Stroke	Survival	0.891	0.45 (0.01,0.97)
	Discharging patients directly to home	0.747	0.33 (0.00,0.90)
Transcatheter aortic valve replacement (TAVR)	Survival	0.789	0.54 (0.11,0.93)
	Discharging patients directly to home	0.813	0.39 (0.14,0.86)
	Readmission prevention	0.662	0.46 (0.05,0.71)
	Prevention of stroke	0.828	0.47 (0.01,0.95)
Uterine cancer surgery	Survival	0.864	0.50 (0.08,0.96)
	Discharging patients directly to home	0.850	0.28 (0.00,0.73)
	Prevention of prolonged hospitalizations	0.815	0.15 (0.00,0.55)
	Readmission prevention	0.677	0.37 (0.02,0.81)

Construction of Composite Ratings

There are two major issues in constructing a composite rating of quality of surgical or medical care: determining how much weight each indicator should receive and accounting for measurement error. Some approaches, such as averaging a set of indicators with equal weight on each, do not address measurement error. More sophisticated statistical procedures can determine empirically how much weight each indicator should be assigned. They can also account for the degree to which an indicator is measured inaccurately due to incomplete risk-adjustment, random variation due to low sample size, and other factors.

Best Hospitals: Procedures & Conditions relies on a statistical method known as confirmatory factor analysis, which assigns empirical weights to the indicators. This approach has been previously used to evaluate provider quality of care.[34] Confirmatory factor analysis is based on the statistical principle that variables sharing a common cause will be correlated. Here, we hypothesize that the various candidate indicators for a given condition or procedure are caused by an underlying, or latent, variable that represents quality of surgical or medical care rendered by a hospital. Thus, for each indicator, the model can estimate the extent to which the values are the result of a relationship with quality of care. The remaining variance in the indicator is attributed to measurement error. The degree to which an indicator is correlated with other indicators helps to determine its weight in the equation for the composite scores.

We develop models by evaluating model statistics for all possible combinations of a field of structure, process, and outcome indicators. From the resulting list of candidate models exhibiting acceptable fit statistics, we select final models offering an optimal combination of number of indicators (models with more indicators produce more accurate factor scores), number of outcomes, model fit, and consistency with models in related cohorts. The selected models show acceptable fit statistics in the majority of the bootstrapped samples in all cohorts.

We evaluate our confirmatory factor analysis models using three measures: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). The literature provides a variety of standards for acceptable model fit using these statistics. We seek final models with a CFI and TLI of .9 or greater, and RMSEA of .1 or lower, while also considering our theoretical understanding of the factors that are most relevant for quality of care. Most models display fit characteristics better than the cutoff value.

We estimate model fit statistics with the robust weighted least squares multivariate (WLSMV) estimator after imputing missing data with relevant hospital-level characteristics. We do not assign quality scores to hospitals based on imputed data. To avoid using this imputed data for that purpose, we estimate hospital factor scores separately with the robust maximum likelihood (MLR) model using a full information maximum likelihood with empirical Bayes (FIML) estimator. This latter

[34] e.g. Keller, S., A. J. O'Malley, R. D. Hays, R. A. Matthew, A. M. Zaslavsky, K. A. Hepner, and P. D. Cleary. "Methods Used to Streamline the CAHPS Hospital Survey." *Health Serv Res* 40, no. 6 Pt 2 (Dec 2005): 2057-77.

model is appropriate for use with missing data, but does not provide the fit statistics necessary to guide model development when categorical indicators are included. Fit statistics can change depending on the estimator used, so there is no assurance that fit estimated with WLSMV is the same as fit that would be estimated with MLR. However, we find key model results, including factor loadings, fit statistics, and factor scores, to be robust across these two estimators.[35]

We assign each rated hospital in a cohort to one of three bands: below average, average, or high performing. Inference that a hospital was below average or high performing was made at the 75% confidence level. Health researchers more commonly use a 95% confidence level, an approach that is geared toward minimizing the number of false positive results (in this context, incorrectly identifying average hospitals as better or worse than the mean). However, because false negatives (identifying poor-performing hospitals as average) can have serious consequences for patients, we seek to strike a balance between minimizing false positive and false negative results.

Two conditions further modify the hospital ratings. First, *if all the outcome measures in a cohort's CFA are categorized with "poor" ratings, its overall rating in that cohort is downgraded to "average" if it would otherwise be categorized as "high performing" by the CFA results alone.* For example, the stroke cohort contains two outcome measures, survival and discharging patients directly to home. Any hospital that received "poor" ratings for both of these outcomes and "high performing" overall would have its stroke rating downgraded to "average". As a result, 130 ratings across 8 cohorts are downgraded from "high performing" to "average".

Second, using ratings computed by data scientists at the Lown Institute using 2018-2020 Medicare fee-for-service claims data, we apply downgrades in three cohorts for hospitals that receive the lowest rating in avoiding overuse in relevant procedures. Details of these measures can be found on the Lown Hospitals Index website.[36]

- For back surgery (spinal fusion), *if a hospital receiveda 1-star rating from the Lown Institute for avoiding overuse of spinal fusion/laminectomy, its U.S. News overall rating is downgraded to "average" if it would otherwise be categorized as "high performing" by the CFA results alone* As a result, 6 hospitals are downgraded from "high performing" to "average".
- For knee replacement, *if a hospital received a 1-starrating from the Lown Institute for avoiding overuse of arthroscopic knee surgery, its U.S. News overall rating is downgraded to "average" if it would otherwise be categorized as "high performing" by the CFA results alone.* As a result, 15 hospitals are downgraded from "high performing" to "average".
- For stroke, *if a hospital received a 1-star ratingfrom the Lown Institute for avoiding overuse of carotid endarterectomy, its U.S. News overall rating is downgraded to "average" if it would otherwise be categorized as "high performing" by the CFA results alone.* As a result, 58 hospitals are downgraded from "high performing" to "average".

[35] When all indicators are continuous measures, the CFA with a MLR estimator yields fit statistics. Hence, for cohorts that incorporate indicators that are all continuous measures, their factor scores, fit statistics, and factor loadings are all generated using a MLR estimator.

[36] For more details, visit https://lownhospitalsindex.org/2022-winning-hospitals-avoiding-overuse. Last accessed: May 19, 2022.

Indicators and Correlations With Scores

The following tables list the indicators included in each cohort's final composite model. The quality score correlation, or standardized factor loading, indicates the relative strength of the relationship in a cohort between a given indicator and hospitals' quality scores. The quality score correlation is determined by the statistical model; it is not a weight and is not applied as a factor of a summative formula. Instead, it is applied to a maximum likelihood estimation algorithm that produces the overall quality score for each hospital. The greater the value of the correlation, the stronger the relationship to the quality score. It may be noted that outcome measures in some cohorts are relatively weakly correlated with quality scores. That is to be expected if the incidence of negative outcomes is very low, as it is, for example, for mortality in the hip replacement and knee replacement cohorts, or if there is little variation in the measure from one hospital to another.

Table 5: *Confirmatory Factor Analysis Fit Statistics, by Cohort*

	CFI	**TLI**	**RMSEA**
Abdominal aortic aneurysm repair	0.987	0.982	0.076
Aortic valve surgery	0.928	0.904	0.065
Back surgery (Spinal fusion)	0.968	0.935	0.037
Chronic obstructive pulmonary disease (COPD)	0.969	0.948	0.099
Colon cancer surgery	0.974	0.964	0.093
Diabetes	0.964	0.929	0.082
Heart attack	0.952	0.929	0.083
Heart bypass surgery	0.927	0.903	0.051
Heart failure	0.981	0.974	0.082
Hip fracture	0.999	0.998	0.005
Hip replacement	0.925	0.900	0.078
Kidney failure	0.996	0.993	0.039
Knee replacement	0.935	0.909	0.065
Lung cancer surgery	0.942	0.922	0.080
Ovarian cancer surgery	0.937	0.911	0.085
Pneumonia	0.966	0.931	0.082
Prostate cancer surgery	0.975	0.962	0.100
Stroke	0.982	0.974	0.045
Transcatheter aortic valve replacement (TAVR)	0.998	0.998	0.079
Uterine cancer surgery	0.983	0.967	0.026

Table 6: *Indicator Correlations, Abdominal Aortic Aneurysm Repair*

Indicator	**Quality Correlation**
Survival	0.192
Discharging patients directly to home	0.195
Readmission prevention	0.158
Infection prevention	0.110
ICU specialists	0.489
Number of patients	0.414
Patient experience	0.543
Worker influenza immunization	0.401

Table 7: *Hospital Distribution by Performance Band, Abdominal Aortic Aneurysm Repair*

Band	**Description**	**Number of Hospitals**
1	Below average	158
2	Average	963
3	High performing	141

Table 8: *Indicator Correlations, Aortic Valve Surgery*

Indicator	Quality Correlation
Survival	0.376
Discharging patients directly to home	0.464
Prevention of prolonged hospitalizations	0.362
Readmission prevention	0.352
ICU specialists	0.486
Number of patients	0.590
Nurse staffing	0.447
Patient experience	0.488
Public transparency	0.460

Table 9: *Hospital Distribution by Performance Band, Aortic Valve Surgery*

Band	Description	Number of Hospitals
1	Below average	223
2	Average	490
3	High performing	199

Table 10: *Indicator Correlations, Back Surgery (Spinal Fusion)*

Indicator	**Quality Correlation**
Survival	0.167
Discharging patients directly to home	0.235
Number of patients	0.309
Nurse staffing	0.506
Patient experience	0.735

Table 11: *Hospital Distribution by Performance Band, Back Surgery (Spinal Fusion)*

Band	**Description**	**Number of Hospitals**
1	Below average	308
2	Average	1,191
3	High performing	272

Table 12: *Indicator Correlations, Chronic Obstructive Pulmonary Disease (COPD)*

Indicator	Quality Correlation
Survival	0.255
Discharging patients directly to home	0.415
ICU specialists	0.754
Noninvasive breathing aid (>=20%)	0.698
Number of patients	0.818
Nurse staffing	0.282

Table 13: *Hospital Distribution by Performance Band, Chronic Obstructive Pulmonary Disease (COPD)*

Band	Description	Number of Hospitals
1	Below average	1,207
2	Average	1,616
3	High performing	1,301

Table 14: *Indicator Correlations, Colon Cancer Surgery*

Indicator	**Quality Correlation**
Survival	0.363
Giving patients time at home	0.196
Prevention of prolonged hospitalizations	0.116
ICU specialists	0.677
NCI and/or ACS cancer center	0.674
Number of patients	0.792
Nurse staffing	0.339
Worker influenza immunization	0.132

Table 15: *Hospital Distribution by Performance Band, Colon Cancer Surgery*

Band	**Description**	**Number of Hospitals**
1	Below average	637
2	Average	1,282
3	High performing	450

Table 16: *Indicator Correlations, Diabetes*

Indicator	Quality Correlation
Survival	0.200
Discharging patients directly to home	0.135
ICU specialists	0.855
Number of patients	0.704
Nurse staffing	0.325

Table 17: *Hospital Distribution by Performance Band, Diabetes*

Band	Description	Number of Hospitals
1	Below average	892
2	Average	1,561
3	High performing	439

Table 18: *Indicator Correlations, Heart Attack*

Indicator	Quality Correlation
Survival	0.547
Discharging patients directly to home	0.390
ACC or GWTG Heart registry participation	0.644
Cardiac ICU or ICU specialists	0.766
Number of patients	0.769
Nurse staffing	0.390
Patient experience	0.128

Table 19: *Hospital Distribution by Performance Band, Heart Attack*

Band	Description	Number of Hospitals
1	Below average	874
2	Average	1,171
3	High performing	692

Table 20: *Indicator Correlations, Heart Bypass Surgery*

Indicator	**Quality Correlation**
Survival	0.560
Discharging patients directly to home	0.356
Prevention of prolonged hospitalizations	0.275
Prevention of stroke	0.154
ICU specialists	0.522
Number of heart surgery patients	0.621
Nurse staffing	0.448
Patient experience	0.475
STS score	0.443

Table 21: *Hospital Distribution by Performance Band, Heart Bypass Surgery*

Band	**Description**	**Number of Hospitals**
1	Below average	260
2	Average	606
3	High performing	222

Table 22: *Indicator Correlations, Heart Failure*

Indicator	Quality Correlation
Survival	0.405
Discharging patients directly to home	0.445
ACC or GWTG Heart registry participation	0.797
Advanced heart program	0.831
ICU specialists	0.836
Noninvasive breathing aid (>=20%)	0.670
Number of patients	0.774
Nurse staffing	0.349

Table 23: *Hospital Distribution by Performance Band, Heart Failure*

Band	Description	Number of Hospitals
1	Below average	1,353
2	Average	1,638
3	High performing	1,146

Table 24: *Indicator Correlations, Hip Fracture*

Indicator	Quality Correlation
Survival	0.104
Readmission prevention	0.208
Board certified physicians	0.121
Patient experience	0.854
Worker influenza immunization	0.325

Table 25: *Hospital Distribution by Performance Band, Hip Fracture*

Band	Description	Number of Hospitals
1	Below average	580
2	Average	1,752
3	High performing	697

Table 26: *Indicator Correlations, Hip Replacement*

Indicator	**Quality Correlation**
Survival	0.252
Prevention of prolonged hospitalizations	0.467
Prevention of revision surgery	0.252
Infection prevention	0.289
Board certified physicians	0.100
ICU specialists	0.466
Number of patients (includes outpatient setting)	0.746
Nurse staffing	0.319
Patient experience	0.285

Table 27: *Hospital Distribution by Performance Band, Hip Replacement*

Band	**Description**	**Number of Hospitals**
1	Below average	511
2	Average	1,864
3	High performing	438

Table 28: *Indicator Correlations, Kidney Failure*

Indicator	Quality Correlation
Survival	0.321
Giving patients time at home	0.096
ICU specialists	0.834
Number of patients	0.781
Nurse staffing	0.346

Table 29: *Hospital Distribution by Performance Band, Kidney Failure*

Band	Description	Number of Hospitals
1	Below average	1,004
2	Average	1,412
3	High performing	1,089

Table 30: *Indicator Correlations, Knee Replacement*

Indicator	**Quality Correlation**
Survival	0.211
Prevention of prolonged hospitalizations	0.437
Readmission prevention	0.098
Prevention of revision surgery	0.190
Board certified physicians	0.161
ICU specialists	0.595
Number of patients (includes outpatient setting)	0.719
Worker influenza immunization	0.099

Table 31: *Hospital Distribution by Performance Band, Knee Replacement*

Band	**Description**	**Number of Hospitals**
1	Below average	538
2	Average	2,176
3	High performing	443

Table 32: *Indicator Correlations, Lung Cancer Surgery*

Indicator	Quality Correlation
Survival	0.388
Discharging patients directly to home	0.488
Prevention of prolonged hospitalizations	0.526
ICU specialists	0.574
NCI cancer center	0.812
Number of patients	0.685
Nurse staffing	0.443
Patient experience	0.464
Worker influenza immunization	0.239

Table 33: *Hospital Distribution by Performance Band, Lung Cancer Surgery*

Band	Description	Number of Hospitals
1	Below average	298
2	Average	601
3	High performing	253

Table 34: *Indicator Correlations, Ovarian Cancer Surgery*

Indicator	**Quality Correlation**
Survival	0.395
Discharging patients directly to home	0.319
Prevention of prolonged hospitalizations	0.197
Readmission prevention	0.115
NCI cancer center	0.748
Number of patients	0.600
Nurse staffing	0.387
Patient experience	0.496

Table 35: *Hospital Distribution by Performance Band, Ovarian Cancer Surgery*

Band	**Description**	**Number of Hospitals**
1	Below average	69
2	Average	313
3	High performing	75

Table 36: *Indicator Correlations, Pneumonia*

Indicator	**Quality Correlation**
Survival	0.335
Discharging patients directly to home	0.485
ICU specialists	0.641
Nurse staffing	0.394
Worker influenza immunization	0.096

Table 37: *Hospital Distribution by Performance Band, Pneumonia*

Band	**Description**	**Number of Hospitals**
1	Below average	510
2	Average	3,187
3	High performing	534

Table 38: *Indicator Correlations, Prostate Cancer Surgery*

Indicator	**Quality Correlation**
Survival	0.233
Discharging patients directly to home	0.241
Prevention of prolonged hospitalizations	0.341
NCI and/or ACS cancer center	0.517
Number of patients (includes outpatient setting)	0.722
Nurse staffing	0.392
Patient experience	0.404

Table 39: *Hospital Distribution by Performance Band, Prostate Cancer Surgery*

Band	**Description**	**Number of Hospitals**
1	Below average	181
2	Average	852
3	High performing	147

Table 40: *Indicator Correlations, Stroke*

Indicator	**Quality Correlation**
Survival	0.148
Discharging patients directly to home	0.184
GWTG Stroke registry participation	0.668
ICU specialists	0.681
Number of patients	0.890
Nurse staffing	0.358
Reperfusion therapy	0.796

Table 41: *Hospital Distribution by Performance Band, Stroke*

Band	**Description**	**Number of Hospitals**
1	Below average	1,162
2	Average	919
3	High performing	1,054

Table 42: *Indicator Correlations, Transcatheter Aortic Valve Surgery*

Indicator	**Quality Correlation**
Survival	0.467
Discharging patients directly to home	0.366
Readmission prevention	0.373
Prevention of stroke	0.136
ICU specialists	0.666
Number of patients	0.556
Nurse staffing	0.416
Public transparency	0.324
Worker influenza immunization	0.217

Table 43: *Hospital Distribution by Performance Band, Transcatheter Aortic Valve Surgery*

Band	**Description**	**Number of Hospitals**
1	Below average	117
2	Average	465
3	High performing	106

Table 44: *Indicator Correlations, Uterine Cancer Surgery*

Indicator	**Quality Correlation**
Survival	0.249
Discharging patients directly to home	0.611
Prevention of prolonged hospitalizations	0.563
Readmission prevention	0.331
Number of patients	0.230

Table 45: *Hospital Distribution by Performance Band, Uterine Cancer Surgery*

Band	**Description**	**Number of Hospitals**
1	Below average	80
2	Average	373
3	High performing	81

Validation of Factor Scores

The primary means of evaluating construct validity of our measurement models and resulting factor scores is a multi-trait matrix, by which we compare the relative correlations of hospital ratings across cohorts. Specifically, we hypothesize that hospital factor scores for heart bypass surgery, aortic valve surgery, and TAVR would be more closely correlated with each other than with the other procedure cohorts, and that the two cardiac surgeries would be least correlated with the condition cohorts. Similarly, we hypothesize that hip replacement and knee replacement ratings would be highly intercorrelated, and less well correlated with other procedures, and that they, like the cardiac surgeries, would be least correlated with condition cohorts. Finally we hypothesize that factor scores among condition cohorts would be strongly intercorrelated, and less well correlated with procedure ratings. The correlations, shown in Appendix C, provide strong evidence of construct validity. We also hypothesize that hospitals who are ranked (i.e. perform extremely well) in specialty care would more often be rated high performing in related P&C cohorts.

We further investigate validity by examining concordance of the CABG and AVR ratings with ratings published by STS. The U.S. News and STS ratings cover different time periods and patient populations. The U.S. News ratings are based on three domains of quality, while the STS ratings do not use structural indicators. The U.S. News ratings employ statistical testing at the $p<.25$ level, while STS ratings employ a standard of $p<.05$, and because of this difference, one would expect that the U.S. News ratings would identify more hospitals as performing above or below average. We hypothesize modest agreement between the two sets of ratings, with very few instances of marked disagreement, in which a hospital received the lowest rating from one organization and the highest from the other. Tables 46 and 47 show findings consistent with this hypothesis.

Table 46: *Concordance with STS Aortic Valve Surgery Star Rating*

	STS Star Rating		
	1	2	3
Below Average	2	83	0
Average	8	302	11
High Performing	0	146	28

Table 47: *Concordance with STS Heart Bypass Surgery Star Rating*

	STS Star Rating		
	1	2	3
Below Average	1	64	1
Average	15	365	23
High Performing	2	155	53

VALIDATION OF STROKE SEVERITY RISK-ADJUSTMENT

The NIHSS score, which quantifies stroke severity on a numeric scale from 0 to 42, is widely used for risk-adjustment. Studies demonstrate that the NIHSS score has a strong association with patient condition, and potentially could improve model discrimination and performance.[37] Indeed, Table 48 shows that the inclusion of a NIHSS score covariate improved model fit appreciably.

One challenge is the scarcity of NIHSS score in claims data. Among all ischemic stroke visits between October 2016 to 2020, only about 50% document this score. Our examination of claims data suggests that the availability of the NIHSS score is highly associated with hospital-, patient-, and visit-level attributes; in other words, the pattern of the missing values is systematic and cannot be ignored. Consequently, a case-wise deletion may also cause substantial bias. We instead choose to multiply impute the missing NIHSS score, generating imputed values by fitting a hierarchical linear regression with hospital-level random effects. This modeling strategy is considered a superior strategy for handling missing data.[38]

Table 48: *C-statistics of Stroke Risk Models*

Outcome	w/o NIHSS risk adjustment	w/ NIHSS risk adjustment
Survival	0.76	0.89
Discharging patients directly to home	0.72	0.75

We assess the reliability of the stroke ratings by examining the concordance of stroke ratings when NIHSS score is present and absent from the risk-adjusted patient outcome measures. Table 49 shows both scenarios yield similar results, with few discrepancies. For example, 36 of the hospitals we rated as High Performing would have received a rating of Average if we had not used NIHSS as a covariate in the risk-adjustment models.

Table 49: *Concordance of Stroke Ratings (w/ and w/o NIHSS)*

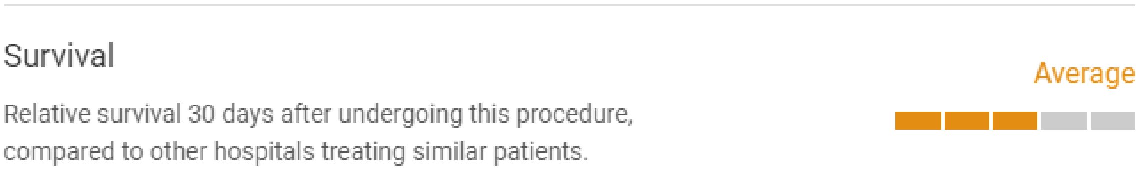
		Overall Band w/o NIHSS		
		Below Average	Average	High Performing
	Below Average	1143	19	0
Overall Band w/ NIHSS	Average	2	767	150
	High Performing	0	36	1018

[37] Fonarow, G. C., Pan, W., Saver, J. L., Smith, E. E., Reeves, M. J., Broderick, J. P., Kleindorfer, D. O., Sacco, R. L., Olson, D. M., Hernandez, A. F., Peterson, E. D., & Schwamm, L. H. (2012). Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. JAMA, 308(3), 257–264. https://doi.org/10.1001/jama.2012.7870

[38] Lall, R. (2017). How Multiple Imputation Makes a Difference. Political Analysis, 24(4), 414-433. doi:10.1093/pan/mpw020

Categorical Display

In our confirmatory factor analysis, we use the continuous form of each measure when possible. For the purpose of making information more accessible for patients, we display categorical groupings (bands) and descriptions of each continuous outcome or process measure on scorecards. See example of the survival rating below.



Our approach to estimating each hospital's outcome band falls under the general rubric of statistical significance testing. The band cutoffs are ***different for each hospital and each measure***. This band is reflective of a hospital's estimated risk-adjusted value on the outcome compared to other hospitals, as well as its Medicare claims volume and the incidence of that outcome. We compare each hospital's risk-adjusted outcome value to a normal distribution, taking into account both the point estimate and the precision—the greater a hospital's volume, the more certain we are of its estimated outcome value. For rare outcomes, such as death after knee replacement, relatively few hospitals have a rate designating it as above average. The bands displayed provide a heuristic for each underlying continuous metric we use to evaluate a hospital's performance.

Strengths and Limitations

Our study makes use of many datasets, which allow us to consider indicators from most, if not all, domains relevant to hospital quality measurement. We employ statistical procedures that simultaneously minimize measurement error and empirically combine indicators to maximize quality measurement validity. We conduct extensive research on the validity of our results, including using multitrait matrices and comparing with external datasets, and we benefit from input from diverse stakeholders, including patients, health service researchers, clinicians, and hospital leaders.

Quality measurement derived from the Medicare population is generally believed to be representative of what would emerge from the overall population, and affords sufficient statistical power to distinguish between providers, even when procedures may be relatively rare. The LDS SAF data imperfectly mirror the overall hospital inpatient population because other than those with disabilities or end-stage kidney disease, and Medicare members in the analysis are age 65 and older. However, these data are widely used in academic literature to permit meaningful comparisons of rates of death, complications, readmission, infection and other outcomes on a like-to-like basis across most hospitals. How these older patients fare represents a test of hospital performance that is more revealing than results would be from a population that includes younger and healthier patients. Broad "all-payer" data that would permit such an evaluation for all hospitals, moreover, is unavailable, and the population tracked is large and clearly defined.

A noteworthy limitation of the ratings is that the outcome indicators rely on administrative data, which could lead to bias in several ways. As previously discussed, controlling for severity of the index condition is required to achieve adequate case-mix adjustment. We believe we have largely mitigated this problem by adjusting for a number of variables that are correlated with severity of the index condition, such as transfer status and urgency of admission, and by using other statistical procedures that account for measurement error. It is possible, however, that our results are biased by residual confounding. Similarly, ascertainment of some outcomes, e.g. stroke or surgical-site infection, requires accurate coding across hospitals.[39] Prior studies have demonstrated, for example, that capturing stroke with different coding algorithms in administrative data results in a tradeoff between sensitivity and specificity.[40] Differences in claims coding practices could result in bias.

Another issue is our use of datasets with incomplete hospital-level data. Some of the reported datasets may have robust data for hospitals participating in the related programs, but only have a limited set of hospitals participating (or with data made available). Not all hospitals, for example, report process-of-care measures via Care Compare. We use two methods to deal with incomplete data. To build and evaluate composite models, we imputed data for missing indicators. To calculate factor scores, we relied on a FIML estimator. Both of these approaches assume that the data are missing at random. If the data are missing dependent on values of the process measures themselves, or on other unmeasured variables, the missing data could result in biased estimates. There is no way to guarantee that this assumption has not been violated. However, we determined that missing Care Compare process measures are primarily associated with hospital size, so we do not suspect that the data are missing conditional on levels of the process variables. As discussed earlier, the use of different estimators in our CFA may each result in different estimation of factor scores and fit statistics, but we found loading coefficients to be similar for the two estimators.

The statistical procedures used to estimate composite scores cannot assure that the label a researcher applies to the composite score (quality of care, in this case), is in fact germane to the content of the score itself. The factor scores we estimated might measure a latent variable different from the one we sought to measure. We addressed this possibility through extensive evaluation of construct validity. As illustrated above, those efforts were strongly supportive of our conceptualization of the factor scores as a measure of hospital quality.

Future Opportunities

Like healthcare delivery itself, quality measurement warrants continuous improvement. Among the opportunities we recognize to improve this methodology, those that stand out include:

---

[39] Calderwood, M. S., A. Ma, Y. M. Khan, M. A. Olsen, D. W. Bratzler, D. S. Yokoe, D. C. Hooper, et al. "Use of Medicare Diagnosis and Procedure Codes to Improve Detection of Surgical Site Infections Following Hip Arthroplasty, Knee Arthroplasty, and Vascular Surgery." Infect Control Hosp Epidemiol 33, no. 1 (Jan 2012): 40-9.

[40] Tirschwell DL, Longstreth WT Jr. Validating administrative data in stroke research. Stroke. 2002; 33(10): 2465-2470. doi:10.1161/01.str.0000032240.28636.bd

further incorporation of outpatient claims data, particularly for patient populations who may be treated in either inpatient or outpatient settings; analysis of additional procedures and conditions, to provide decision support to more patients; and the development of additional candidate measures, including a larger portfolio of risk-adjusted outcome measures and additional measures of process, appropriateness and value. In addition, U.S. News recognizes that racial and socioeconomic disparities plague the healthcare system in this country, and acknowledges the importance of addressing the role these disparities play in outcomes of care. We have begun to measure and publicly report on health equity at the hospital level.

Best Regional Hospitals

U.S. News first published Best Regional Hospitals in 2011 to offer patients a heuristic comparison of community hospitals located in or near the community where they reside. A goal of Best Regional Hospitals is to help healthcare consumers identify suitable hospitals without necessitating travel. A Best Regional Hospital is a hospital that offers a full range of services (as opposed to a specialty hospital) and that either is nationally ranked in one of the eleven data-driven Best Hospitals specialties (excluding rehabilitation) or has seven or more ratings of high performing in the Best Hospitals procedures and conditions. Note that high performing recognitions in the specialties are not counted toward the required minimum; in our view, a hospital must perform at a high level in a variety of common procedures and conditions in order to warrant recognition as one of the best hospitals in its state or metro area. In addition to the aforementioned eligibility criteria, a hospital must also have at least three more high performing procedures or conditions than below average procedures or conditions in order to be a Best Regional Hospital.

In a given state or metro area, a hospital on the Best Hospitals Honor Roll outranks all other hospitals not on the Honor Roll, regardless of point totals. Other hospitals located in each region are ranked according to the number of points they earn: Hospitals earn two points for each of the eleven data-driven Best Hospitals specialties (excluding rehabilitation) in which they are nationally ranked and one point for each specialty and each of the twenty procedures and conditions in which they are rated high performing. In addition, hospitals lose one point for each procedure or condition in which they were rated below average. A combined score for the AVR and TAVR procedures is used rather than assigning points for each individually, because these procedures are different approaches to treating the same disease. Similarly, a combined score for the ovarian and uterine cancer surgery procedures is used rather than assigning points for each individually. In order to be considered high performing for the combined score, a hospital has to be high performing in at least one of the procedures and at least average in the other. If a hospital only provides one of the two procedures, its score for that procedure is used.

Best Regional Hospitals eligibility details are outlined in Appendix D, while scoring details are outlined in Appendix E.

Geographical Definitions

Regional rankings are displayed for every state and for the 100 metro areas with the largest populations in the 2020 census, provided there is at least one Best Regional Hospital located in the state or metro area. In 2022-2023, 493 hospitals are recognized as Best Regional Hospitals. Two states, Alaska and Wyoming, have no Best Regional Hospitals. In all, hospitals are ranked in 91 metro areas.

U.S. News generally uses the U.S. Census Bureau list of Metropolitan Statistical Areas (MSAs) to define metro areas, but we depart from MSAs in cases where we use larger Combined Statistical Areas (CSAs) or combined adjacent MSAs to include nearby smaller cities with nationally ranked hospitals. For example, we use the Detroit CSA instead of the Detroit MSA; we use the

Denver CSA instead of the Denver MSA; we combine the Durham-Chapel Hill and Raleigh-Cary MSAs to define the Raleigh-Durham metro area; we combine the Ogden-Clearfield and Salt Lake City MSAs to define the Salt Lake City metro area; and we combine the Winston-Salem and Greensboro-High Point MSAs to define the Greensboro/Winston-Salem metro area.

Some metropolitan areas, such as Cincinnati and New York City, cross state lines. That is also true for Washington, D.C., which is included in Best Regional Hospitals as a metro area (encompassing parts of Maryland and Virginia) but not a state. Rankings are not published for U.S. territories.

U.S. News groups counties and county equivalents, like parishes, into approximately 200 regions that reflect geography, local customs, and regional health care markets. Best Regional Hospitals are recognized but not numerically ranked in regions that are not major metro areas.

Appendix A: Discharge to a location other than home

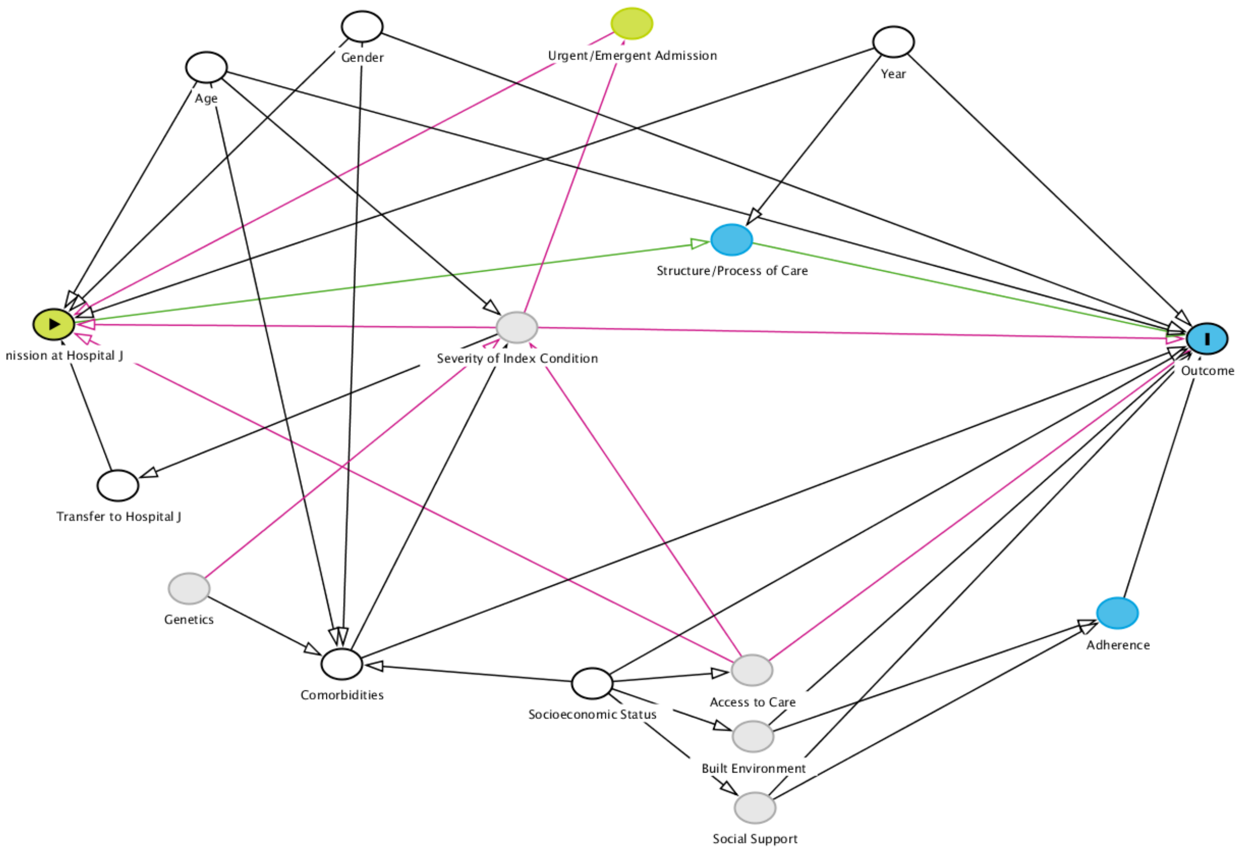
The denominator for this measure includes only patients who were discharged following a visit qualifying as an index visit in one of the 20 Procedures and Conditions cohorts. Discharge status codes of 07 (left against medical advice or discontinued care), 20 (expired, did not recover - Christian Science), 21 (court/law enforcement), 30 (still a patient), 40 (expired at home, hospice claim), 41 (expired in facility, hospice claim), 42 (expired place unknown, hospice claim), 50 (hospice – home), or 87 (to court/law enforcement with a planned acute care hospital inpatient readmission) are excluded, as are visits with a missing or invalid discharge status code. Similarly, visits with an inpatient source admission code of 05 (transfer from a SNF or ICF) and 08 (court/law enforcement) are also excluded. Similarly, visits that are determined to have been admissions from a SNF, because in Medicare SNF claims data, the patient was observed in a SNF immediately prior to being admitted to a hospital, are excluded.

Discharge to a location other than home is indicated by one of the following patient discharge status codes: 0, 02, 03, 04, 05, 08, 09, 43, 51, 61, 62, 63, 64, 65, 66, 69, 70, 71, 72, 82, 83, 84, 85, 88, 89, 90, 91, 92, 93, 94, 95.

Discharge codes 01 (home/self care), 06 (home with care of organized home health service organization), 81 (home/self care with planned readmission), and 86 (home with care of organized home health service organization with planned readmission) are included as a successful discharge directly to home.

Appendix B: Causal model for risk-adjustment

The following directed acyclic graph[41] shows the hypothesized relationship between covariates, hospital selection and outcomes.



[41] Johannes Textor, Juliane Hardt, and Sven Knuppel. Dagitty: A graphical tool for analyzing causal diagrams. Epidemiology, 22(5):745, 2011.

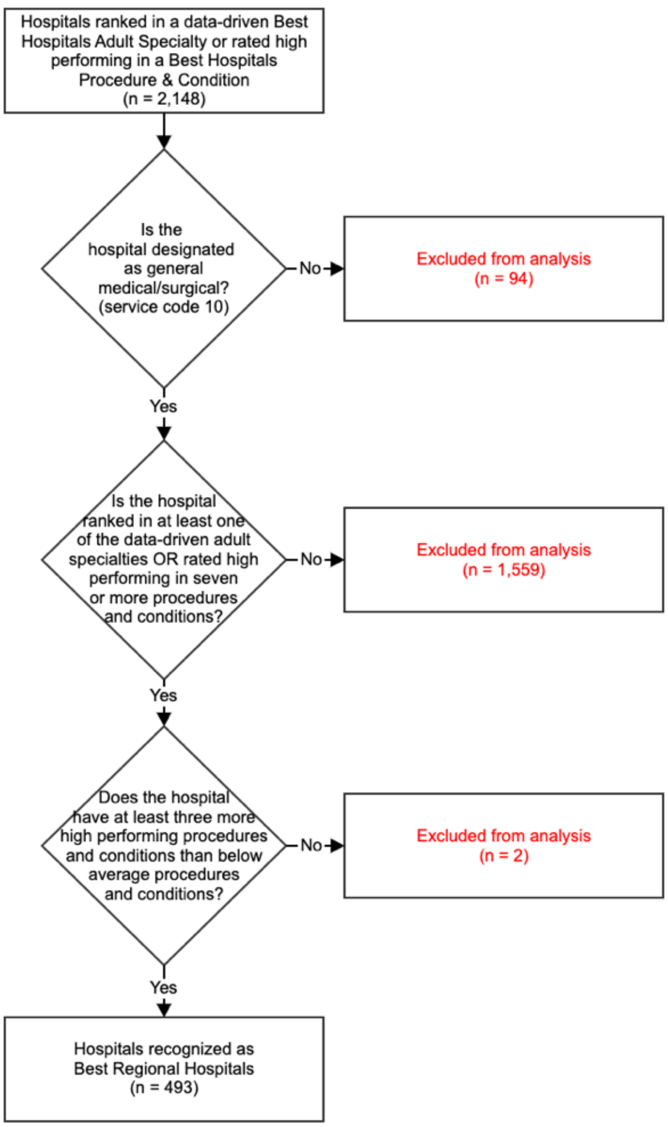
Appendix C: Multi-trait correlation table

Table C1: *Multi-Trait Correlation Table, Hospital Cohort Scores*

	AAA	AVR	TAVR	CABG	HIP	KNEE	FUSION	FRACTURE	COLON	LUNG	PROSTATE	UTERINE	OVARIAN	CHF	COPD	AMI	DIABETES	AKF	PNEUMONIA	STROKE
AAA	1.000	0.701	0.614	0.705	0.564	0.534	0.742	0.786	0.635	0.696	0.574	0.123	0.583	0.473	0.254	0.529	0.375	0.504	0.589	0.326
AVR	0.701	1.000	0.746	0.872	0.599	0.501	0.718	0.609	0.656	0.774	0.633	0.096	0.599	0.573	0.265	0.642	0.408	0.545	0.640	0.448
TAVR	0.614	0.746	1.000	0.755	0.488	0.393	0.531	0.410	0.612	0.663	0.515	0.059	0.525	0.610	0.315	0.608	0.427	0.577	0.669	0.382
CABG	0.705	0.872	0.755	1.000	0.636	0.568	0.685	0.566	0.730	0.769	0.650	0.087	0.571	0.678	0.373	0.753	0.506	0.606	0.605	0.517
HIP	0.564	0.599	0.488	0.636	1.000	0.895	0.512	0.341	0.653	0.587	0.520	0.148	0.361	0.605	0.480	0.606	0.534	0.573	0.535	0.547
KNEE	0.534	0.501	0.393	0.568	0.895	1.000	0.353	0.246	0.677	0.504	0.430	0.124	0.235	0.688	0.606	0.637	0.632	0.679	0.568	0.625
FUSION	0.742	0.718	0.531	0.685	0.512	0.353	1.000	0.903	0.416	0.667	0.587	0.169	0.700	0.258	0.060	0.361	0.199	0.235	0.461	0.197
FRACTURE	0.786	0.609	0.410	0.566	0.341	0.246	0.903	1.000	0.258	0.550	0.488	0.147	0.604	0.017	−0.118	0.132	0.037	−0.033	0.185	−0.010
COLON	0.635	0.656	0.612	0.730	0.653	0.677	0.416	0.258	1.000	0.729	0.710	0.158	0.651	0.823	0.642	0.784	0.749	0.758	0.567	0.721
LUNG	0.696	0.774	0.663	0.769	0.587	0.504	0.667	0.550	0.729	1.000	0.694	0.152	0.778	0.610	0.374	0.574	0.470	0.584	0.642	0.506
PROSTATE	0.574	0.633	0.515	0.650	0.520	0.430	0.587	0.488	0.710	0.694	1.000	0.200	0.691	0.559	0.262	0.505	0.451	0.444	0.484	0.442
UTERINE	0.123	0.096	0.059	0.087	0.148	0.124	0.169	0.147	0.158	0.152	0.200	1.000	0.225	0.076	0.005	0.083	0.051	0.038	0.167	0.069
OVARIAN	0.583	0.599	0.525	0.571	0.361	0.235	0.700	0.604	0.651	0.778	0.691	0.225	1.000	0.427	0.204	0.335	0.238	0.367	0.559	0.307
CHF	0.473	0.573	0.610	0.678	0.605	0.688	0.258	0.017	0.823	0.610	0.559	0.076	0.427	1.000	0.856	0.876	0.875	0.884	0.740	0.821
COPD	0.254	0.265	0.315	0.373	0.480	0.606	0.060	−0.118	0.642	0.374	0.262	0.005	0.204	0.856	1.000	0.688	0.781	0.891	0.674	0.788
AMI	0.529	0.642	0.608	0.753	0.606	0.637	0.361	0.132	0.784	0.574	0.505	0.083	0.335	0.876	0.688	1.000	0.750	0.793	0.617	0.778
DIABETES	0.375	0.408	0.427	0.506	0.534	0.632	0.199	0.037	0.749	0.470	0.451	0.051	0.238	0.875	0.781	0.750	1.000	0.857	0.662	0.729
AKF	0.504	0.545	0.577	0.606	0.573	0.679	0.235	−0.033	0.758	0.584	0.444	0.038	0.367	0.884	0.891	0.793	0.857	1.000	0.739	0.848
PNEUMONIA	0.589	0.640	0.669	0.605	0.535	0.568	0.461	0.185	0.567	0.642	0.484	0.167	0.559	0.740	0.674	0.617	0.662	0.739	1.000	0.607
STROKE	0.326	0.448	0.382	0.517	0.547	0.625	0.197	−0.010	0.721	0.506	0.442	0.069	0.307	0.821	0.788	0.778	0.729	0.848	0.607	1.000

Appendix D: Best Regional Hospitals

The following diagram outlines the guidelines that determine whether or not a hospital is recognized as a Best Regional Hospital.



Note:

For eligibility and scoring purposes here, the TAVR and AVR cohorts in Best Hospitals Procedures & Conditions were combined into a single cohort. Similarly, the uterine and ovarian cancer surgery cohorts were combined. Therefore, despite there being 20 procedures & conditions cohorts, a hospital was effectively scored in 18.

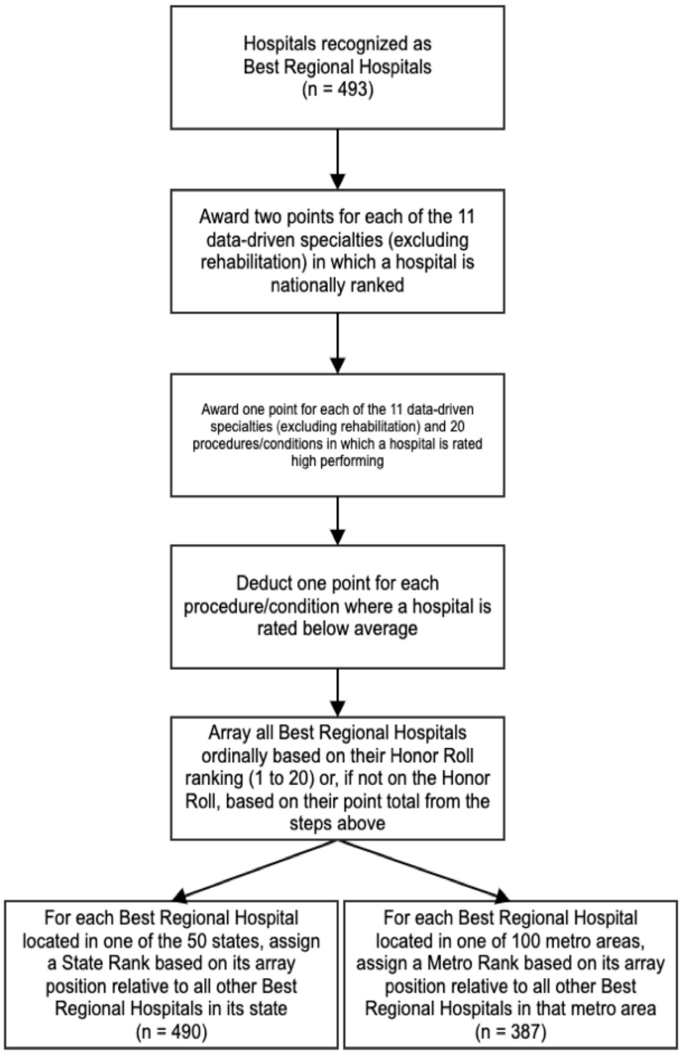
A hospital rated high performing in both individual cohorts or rated high performing in one and average in the other received a high performing rating in the combined cohort.

A hospital rated below average in both individual cohorts or below average in one and average in the other received a below average rating in the combined cohort.

A hospital receiving a rating in only one of the individual cohorts received the single cohort rating for the combined cohort.

Appendix E: Calculation of State and Metro Rankings

The following diagram outlines the scoring methodology that determines the state and metro area rankings of Best Regional Hospitals.



Notes:

1. Point counts here were not used to determine which hospitals are Best Regional Hospitals, only to determine rank ordering of Best Regional Hospitals within a state or metro area.
2. For eligibility and scoring purposes here, the TAVR and AVR cohorts in Best Hospitals Procedures & Conditions were combined into a single cohort. Similarly, the uterine and ovarian cancer surgery cohorts were combined. Therefore, despite there being 20 procedures & conditions cohorts, a hospital was effectively scored in 18.
A hospital rated high performing in both individual cohorts or rated high performing in one and average in the other received a high performing rating in the combined cohort.
A hospital rated below average in both individual cohorts or below average in one and average in the other received a below average rating in the combined cohort.
A hospital receiving a rating in only one of the individual cohorts received the single cohort rating for the combined cohort.

Appendix F: Study Periods for Key Indicators and Cohorts

Outcome Measures		
Indicator	**Source File**	**Time Period**
Survival	Inpatient LDS SAF	9/1/2016 - 12/1/2020 for stroke; 12/1/2015 - 12/1/2020 for other cohorts
Readmission prevention	Inpatient LDS SAF	12/1/2015 - 12/1/2020 for procedure cohorts (AAA, AVR, back surgery (spinal fusion), colon cancer surgery, CABG, hip fracture, hip replacement, knee replacement, ovarian cancer surgery, lung cancer surgery, prostate cancer surgery, TAVR, and uterine cancer surgery)
Infection prevention	Inpatient LDS SAF	12/31/2014 - 12/31/2019 for hip replacement and knee replacement; 11/1/2015 - 11/1/2020 for AAA, CABG, and AVR
Prevention of revision surgery	Inpatient LDS SAF	12/31/2014 - 12/31/2019 for hip replacement and knee replacement

Prevention of prolonged hospitalization	Inpatient LDS SAF	12/1/2015 - 12/1/2020 for procedure cohorts
Discharging patients directly home	Inpatient LDS SAF; Skilled Nursing LDS SAF	10/1/2016 - 12/31/2020 for stroke; 12/31/2015 - 12/31/2020 for other cohorts
Prevention of stroke	Inpatient LDS SAF	12/31/2015 - 12/31/2020 for CABG, AVR, and TAVR
Giving patients time at home	Inpatient LDS SAF; Outpatient LDS SAF; Skilled Nursing LDS SAF	9/1/2016 - 12/1/2019 for stroke; 12/1/2015-12/1/2020 for other cohorts
Process Measures		
Worker flu immunization	CMS Care Compare Database	10/1/2020 - 3/31/2021
Noninvasive ventilation	Inpatient LDS SAF	1/1/2016 - 12/31/2020
Patient experience	CMS Hospital Consumer Assessment of Healthcare Providers and Systems Survey (HCAHPS)	7/1/2020 - 3/31/2021

Board certification	American Board of Orthopaedic Surgery (ABOS); National Board of Physicians and Surgeons (NBPS); American Osteopathic Association (AOA)	(a) provided by ABOS and AOA to Doximity prior to 4/12/2022 and (b) self-reported by NBPS-certified orthopedic surgeons to Doximity prior to 4/12/2022
Reperfusion therapy	Inpatient LDS SAF; Outpatient LDS SAF	10/1/2016 - 12/31/2020
Structure Measures		
Volume	Inpatient LDS SAF	10/1/2016 - 12/31/2020 for stroke; 1/1/2016 - 12/31/2020 for ovarian cancer surgery, uterine cancer surgery, and prostate cancer surgery 1/1/2016 - 12/31/2020* for all other cohorts[42]
	Outpatient LDS SAF	1/1/2020 - 12/31/2020 for hip replacement; 1/1/2018 - 12/31/2020 for knee replacement; 1/1/2016 - 12/31/2020 for prostate cancer surgery
Nurse Staffing	2020 AHA Annual Survey	1/1/2018 - 12/31/2020

[42] *Or 1/1/2015 - 12/31/2019, if greater than 1/1/2016 - 12/31/2020

ICU specialists	2020 AHA Annual Survey	1/1/2020 - 12/31/2020
Cardiac intensive care unit	2020 AHA Annual Survey	1/1/2020 - 12/31/2020
Advanced heart program	Inpatient LDS SAF	1/1/2016 - 12/31/2020
NCI cancer center	National Cancer Institute (NCI)	Hospitals must be designated as NCI clinical or comprehensive cancer centers as of 3/1/2022
ACS cancer center	2020 AHA Annual Survey	1/1/2020 - 12/31/2020
GWTG recognized hospital	American Heart Association Get With The Guidelines (GWTG)	Hospitals must have opted into the public reporting program and been appearing on their public site by 12/31/2021
ACC recognized hospital	American College of Cardiology (ACC)	Hospitals must have opted into the public reporting program by 12/31/2021
STS recognized hospital	Society of Thoracic Surgeons (STS)	Hospitals must have opted into the public reporting program by 12/31/2021
STS/ACC TVT registry recognized hospital	Collaboration between Society of Thoracic Surgeons (STS) and the American College of Cardiology (ACC)	Hospitals must have opted into the public reporting program by 12/31/2021