



# LOWN INSTITUTE HOSPITALS INDEX

White Paper Series

## Methodology of the Lown Institute Hospitals Index

### OVERVIEW

#### HOSPITAL SET

To establish the set of hospitals included in the ranking, we started with mortality using the Medicare Provider Analysis and Review (MEDPAR) administrative claims dataset. From this list, we obtained information on hospital characteristics from the Fiscal Year Ending (FYE) 2017 American Hospital Association (AHA) annual survey and Medicare Impact File as well as the Center of Medicare and Medicaid Services (CMS) Hospital Compare database. Non-acute care hospitals, federal hospitals (e.g. Veterans Administration) and those outside of the 50 states and Washington, D.C. were excluded, as were hospitals run by Medicare Advantage programs (Kaiser Permanente, for example), and specialty hospitals with more than 20 percent admissions for orthopedic or cardiac procedures. We eliminated hospitals that were closed in 2019 by checking Hospital Compare, a website run by CMS. This left a list of 3,359 hospitals, 542 of which are for-profit, 2,188 private nonprofit, and 629 public nonprofit hospitals.

#### COMPOSITE SCORE

Our rankings are based on three categories of data: quality of care, civic leadership, and value of care. These were weighted at 50, 30, and 20 percent respectively in the final ranking. The three categories comprise seven sub-components, each of which includes several more detailed measurements. The detailed measurements were rolled up into the components, which were rolled up into their respective categories to obtain a final numeric score and final rank for each hospital. Out of the 3,359 hospitals in the Lown Institute Hospitals Index, 3,282 are ranked in the composite. Seventy-seven hospitals were missing values in the category of value of care and therefore could not be ranked in the composite.

Each hospital and hospital system with sufficient data was given a letter grade for their composite and category scores, a percentile out of 100 for their component scores, and a star rating for the details under each component.

Letter grades for the composite score and the three individual categories ranged from A+ through D- and were distributed equally by twelfths, meaning the top 8.33 percent of hospitals or hospital systems were given a grade of A+ and the bottom 8.33 percent of hospitals or hospital systems were given a grade of D-. Thus, a hospital could receive, for example, a category score of D- for quality, and an A for both value and civic leadership, and receive a D- for its composite score, because the composite grade was based on its composite numeric score, which fell into the bottom 12th. For each of the seven components, a percentile was determined to rank the hospital or hospital system relative to all others that had sufficient data to place them in the rankings.

Finally, for each specific detail, a star rating between 1 and 5 was given. More stars mean better performance. If a hospital or system fell in the top quintile (top 20 percent) of hospitals, it received a 5-star rating. Likewise, if a hospital or system fell in the bottom quintile (bottom 20 percent) of hospitals, it received a 1-star rating.

## **CIVIC LEADERSHIP**

Our first category, civic leadership, is comprised of three components: Pay Equity, Community Benefit, and Inclusivity. Pay equity is 20% of the civic leadership score, and community benefit and inclusivity each make up 40% of the civic leadership score.

### **PAY EQUITY**

For pay equity, we obtained data for Chief Executive Officer (CEO) compensation from three sources: information on private nonprofit hospitals was taken from Internal Revenue Service (IRS) 990 forms; information on for-profit, publicly-traded hospital systems was obtained from Securities and Exchange Commission (SEC) filings; and information about public hospital CEO pay was found in publicly available records. In cases in which CEO pay was unavailable for publicly traded private hospitals and public hospitals, values were estimated using a regression model based on CEO pay for nonprofit hospitals, combined with other variables such as bed size and hospital revenue. Pay for 1,756 nonprofit hospitals was used to develop a model which was then applied to impute three populations with unavailable pay: 539 for-profit hospitals, 477 public hospitals, and 506 nonprofit hospitals whose 990 forms did not contain the full executive compensation information.

When the information was available for large hospital systems with complicated management structures, we used the salary information for the person determined to have the most influence over day to day hospital operations such as CEOs, chief operating officers, and hospital administrators. For hospitals within systems (2 or more hospitals), we distributed the system CEO's salary among the constituent hospitals using the percentage of total revenue each hospital generated.

We obtained average worker wages from two sources: the CMS Healthcare Cost Report Information System (HCRIS) and the Bureau of Labor Statistics (BLS). HCRIS wage index information contained hourly wages for all employees. We included lower wage staff, such as janitorial and kitchen staff, and medical records personnel, and excluded professional staff such as physicians and nurse practitioners, whose jobs require specialized degrees. For 704 hospitals that had incomplete wage index information in HCRIS, we used BLS estimates of healthcare industry employment data for metropolitan and non-metropolitan statistical areas. These wage estimates also did not include highly paid workers such as executives and physicians. We then estimated hourly wages for CEOs based on a 60-hour week and calculated a ratio of CEO pay to average worker pay.

#### *Pay Equity Limitations*

CEO salary determination was a difficult process mainly due to differing hospital management structures and lack of transparency around employee responsibilities for hospital operations. The research team in many cases had to exercise their best judgement in determining who was designated the main hospital executive. We would welcome hospitals and hospital systems submitting information to us on their management structures to allow us to provide more accurate reporting for the Low Index.

Other errors may have occurred when different hospitals reported the same person with a nickname instead of their legal name, or added middle initials to name, as well as any number of text-based inconsistencies within tax records or public documentation. We have done our best to minimize these issues using algorithms and manual review.

In regards to salary imputation, the linear regression extrapolations were dependent on the nonprofit hospital salary population for modeling, since nonprofits were the most represented hospital type. With respect to for-profit hospitals, only system-level information was available through SEC filings. For-profit imputations were calculated with the addition of a for-profit to nonprofit system ratio. There are a number of

factors that go into the determination of a hospital CEO's salary and we mainly used revenue as our basis for estimation.

We are aware that the BLS wage estimates do not capture the level of employment detail that HCRIS provides due to the exclusion of non-healthcare industries such as secretarial or janitorial work. This causes the BLS wage estimates to skew lower than the HCRIS wages. We used BLS data only when HCRIS data were unavailable.

## COMMUNITY BENEFIT

To calculate community benefit spending by most nonprofit hospitals, we used the [Community Benefit Insight](#) dataset generated from IRS 990 forms filed in FY 2016, the most recent year for which full data were available. We calculated community benefit spending as a share of total expenses, looking at the subset of community benefit spending that we deemed to be meaningful: charity care (free or discounted care provided on the basis of the patient's financial situation); [subsidized health services](#), such as free clinics, some emergency services, telehealth services, and mammograms; [community health improvement](#) activities such as health fairs, community health education classes, immunizations, interpreter services, providing places to exercise; contributions to community organizations; and community building activities that help increase the capacity of the community to address health needs and often address the "upstream" factors, or social determinants, which impact health, such as education, air quality, and access to nutritious food.

We did not use several other categories of community benefit reported on 990 forms, including: shortfall from Medicaid and other government means-tested insurance programs (shortfall is the difference between the amount Medicaid or other programs pay and the costs to hospitals for caring for such patients); health professional training (which is already largely subsidized by the federal government); and research. These types of community benefit have been criticized in the health policy and health services literature as not directly benefiting community health.

Data on community benefit contributions from public and for-profit hospitals was obtained from HCRIS for FY 2016. While public hospitals and for-profits may have been contributing to their communities in multiple ways, charity care was the only measure included in the HCRIS dataset among our subset of meaningful community benefit activities, and thus was the only one we used from HCRIS. For hospitals with only HCRIS data available, the overall score for the "charity care and other community benefit spending" metric was the ratio of their charity care spending as a share of total expenses. For hospitals with both HCRIS and IRS data, the ratio of charity care as a

share of total expenses and IRS community benefits as a share of total expenses were weighted equally to create their overall score for the “charity care and other community benefit spending” metric.

For the state of Maryland only, we took community benefit spending from the Maryland State Cost Report, to account for the reimbursements Maryland hospitals receive for charity care through their state rates. We excluded data from 14 hospitals which reported spending more on community benefit than total expenses as that was likely a reporting error, and just used their charity care from HCRIS. Finally, we adjusted for the fact that approximately 1500 hospitals were in states that did not expand Medicaid and thus would need to provide a larger percent of charity care compared with hospitals in expansion states. To make this adjustment, we calculated percent Medicaid revenue over total gross patient revenue from HCRIS; this is the “Medicaid revenue as a share of patient revenue” metric. For the overall Community Benefit score, percent Medicaid revenue and charity care and other community benefit spending were combined with a weighting of 1:2 respectively to produce the final community benefit score.

#### *Community Benefit Limitations*

While the measurement of community benefits has improved since the 2010 Affordable Care Act (ACA) clarified reporting requirements for IRS Form 990, there are still several limitations to the data available on hospital community benefit spending. For hospitals that did not file a Form 990, the score was based on charity care as a share of total expenses and share of Medicaid revenue, but we do not have the data available to take into account other types of community benefits on which public and for-profit hospitals are spending. Therefore, community benefit spending of public and for-profit hospitals may be undercounted. Additionally, as previously mentioned, there are many hospital systems that file as a group and their community benefit spending on Form 990 is not broken down by individual hospital.

Second, data for hospital billing and collection practices are not readily available. Hospitals (even nonprofits) commonly overcharge patients who are eligible for charity care and sue eligible patients for unpaid bills. These practices go against the social mission of nonprofit hospitals and threaten the financial health of patients and community health. However, our analysis cannot capture this information because no centralized data exist on hospital billing and collection practices. While IRS Form 990 includes questions regarding hospital billing practices, no hospital reported on the form that they had engaged in aggressive collections.

Third, we can only measure the amount of spending on community benefits, not the impact that spending had on community health. While we have focused on a few categories of community benefits we have deemed most meaningful, we lack data on whether the spending by top-ranking hospitals is directed towards community health priorities identified in the Community Health Needs Assessment, which every nonprofit hospital is required by the ACA to conduct. We hope our research will facilitate efforts to increase transparency around hospital community benefit spending, and permit local citizens, officials, and organizations to hold their hospitals accountable to their social mission to improve community health.

## INCLUSIVITY

Inclusivity is a novel metric we developed to measure the degree to which a hospital's patient population reflects the demographics of its catchment area.

We defined the catchment area by using the zip codes of the hospital's own Medicare patient population, sorted by the number of patients each zip code supplied. We then defined the radius of the catchment area as the distance to zip codes whose contribution to the total patient population became insignificant. The zip code at which the additional number of patients dropped significantly was defined as the radius of the catchment area. Thus all people living within the defined radius were deemed to be potential patients of that hospital, and thus defined the denominator of the inclusivity score.

The median radius was 26.6 miles, with urban settings having far smaller radii than rural hospitals. We calculated the demographics by using census data on income and education as proxies for social class, and self-reported race/ethnicity for race. For each of the three demographics—income, education, and race—the measure reflects the difference between the demographics of a hospital's actual patients' zip codes to the demographics of the population within the catchment radius who could have come to the hospital.

To calculate the denominator, we used the U.S. Census Bureau's American Community Survey data for people over the age of 65 on race, income, and education levels within all zip codes that fell within the defined hospital catchment area. We calculated each score using the total population counts and the levels of income and education and proportions of race for each zip code. We attenuated exponentially the contribution of all zip codes beyond the point at which 50 percent of a hospital's patients had come. We created the hospital score by using the actual beneficiary counts, weighted by contribution to the total, and without a distance attenuation. We then compared the

catchment area score to the hospital score: a ratio for income and education levels, and a score summarizing the differences between the racial group populations. We combined these three values for the overall inclusivity score.

### *Inclusivity limitations*

Our method is based on zip code areas, and assumes that people within a zip code are equally likely to visit one hospital within a catchment area. For example, if a zip code had a 80 percent population of low-income earners and 20 percent of high-income earners, we assume, ceteris paribus, that patients going to the hospital from this zip code should match this ratio. We would not be able to observe if all of the patients going to the hospital from this zip code were actually high-income earners (that is, the 80 percent population of low-income earners was completely excluded), and we would give the hospital a better income score than if we had actual income data for hospital patients. Such data are not available.

Our catchment area is also defined as a circle; if the central point of a zip code falls outside the circumference of the circle, it is considered outside the catchment area. In reality, direct distance may not always reflect the true travel distance or travel time for potential patients. Our method treats all beneficiaries within the catchment area at equal direct distances to the hospital as being equally able to reach the hospital, even though the travel times and therefore likelihood of going to that hospital may be different.

Finally, our inclusivity score, by design, rewards hospitals that effectively “over-serve” communities with lower average income, education attainment and higher minority populations. Hospitals whose catchment area demographics and patient demographics are very similar receive a mid-range score in the percentile ranking of the inclusivity scores. Some of these hospitals, however, may be in a situation where it is difficult to improve on this score. For example, a hospital may be in a catchment area that is all very wealthy and their entire patient demographics will reflect this surrounding demographic fact.

## **VALUE OF CARE**

The Value of Care category was based on a single component measure: Avoidance of overuse. We included the rate of overuse of 13 low-value medical services, including: hysterectomy for benign disease; laminectomy and/or spinal fusion without radicular pain; arthroscopy for knee arthritis; vertebroplasty or kyphoplasty for osteoporotic vertebral fractures; carotid endarterectomy in asymptomatic patients (those with no

history of stroke, transient ischemic attack (TIA), TIA without stroke, or focal neurological symptoms); carotid artery imaging for syncope; EEG for syncope; head imaging for syncope; EEG for headache; inferior vena cava filter; pulmonary artery catheter placement in non-surgical conditions; coronary artery stenting for stable coronary angina; and renal artery stenting. We chose these services based on substantial literature on overuse. Five of these services, including vertebroplasty, arthroscopy, renal stenting, inferior vena cava filter, and pulmonary artery catheterization, have been shown in high-quality clinical trials to be ineffective and are nearly always considered overuse. The remaining eight interventions are considered low value when prescribed to patients with certain diagnoses or conditions. For example, a patient with stable angina is considered an inappropriate candidate for a cardiac stent and use of a stent in this case is considered low value or overuse. Similarly, a patient with syncope does not require an EEG.

We used the methods reported in the literature by reputable researchers to calculate rates of overuse. We used the 100 percent Medicare claims datasets (MEDPAR and outpatient) to search for instances when these 13 services were used. Hospitals without a capacity to perform a service, as reflected in their claims histories, were excluded from the rating for that particular service. Hospitals without capacity to perform any of the 13 services were excluded entirely from the overuse ratings. For the always overuse category (vertebroplasty, arthroscopy, renal stenting, inferior vena cava filter, and non-surgical pulmonary artery monitoring), we counted the number of instances the service was delivered. For the services that were inappropriate depending on the condition, we used additional diagnosis and procedure codes to identify appropriateness of use.

We used two different methods to calculate a denominator for services that are sometimes inappropriate and thus considered overuse. For EEG for syncope, EEG for headache, carotid artery screening for syncope, and head imaging for syncope, we used all instances of that diagnosis for the denominator. For hysterectomy, spinal fusion, coronary artery stenting, and carotid endarterectomy, the denominator was all instances of the procedure, both inappropriate and appropriate use. We then conducted volume or sub-population (service or diagnosis volume) adjustment of observed overuse rates and applied a weighted average of the adjusted overuse rate and the population overuse rate, to balance the reliability of the hospital estimate when volumes are small with the population mean. We then segmented our observed overuse rates into six groups using an RFM scale. RFM is a financial/retail methodology that assigns a simple numeric scale to a measured variable in order to quickly assess and bin users by their shopping habits. In our case, we applied it to the observed overuse ratios for each hospital to bin them by their overuse behaviors for each low-value service.

The six groups had a value of 0 for no overuse, and quintile groups among those with overuse were assigned a value from 1 to 5. We then used those scores within a principal component analysis, which reduced the data to one statistically important variable that became our overuse score.

### *Avoiding Overuse Limitations*

We used low value services well-established in the literature, but the true definition of overuse almost always depends on the clinical circumstances, which are not always captured in claims data. Furthermore, errors of coding and reporting by providers could have resulted in errors in our estimates. Particularly for low-volume hospitals, these estimates may be subject to sampling error resulting in changing rates from year to year. Our goal was to estimate rates at the level of the hospital, not of an individual practitioner.

The definition of the observed overuse rate and of the volume adjustment approach varies. Depending on the service, the volume adjustment is based on number of service instances (inappropriate and appropriate), number of instances which meet the diagnosis criteria, or, for those without a denominator, overall patient volume. This variation is addressed somewhat by standardizing with the RFM technique.

We have endeavored to avoid rewarding hospitals for avoiding overuse when they do not in fact have the capacity to perform such a service. The capacity assessment we developed as an indicator is defined using lists of procedure codes that are much broader than the inappropriate ones. However, as with the measurement of overuse itself, our capacity assessment is claims-based and subject to errors at very low volumes. It is possible that some hospitals have been included and rewarded when they do not, in fact, have true capacity.

## **QUALITY OF CARE**

Our Quality of Care category was created from three components: clinical outcomes, patient safety, and patient satisfaction, which were weighted in a ratio of 5:2:1 respectively in calculating the final quality score. This weighting ensured that clinical outcomes had the greatest impact on the final score and no hospital with comparatively poor clinical outcomes appeared near the top of the list, regardless of their performance on other metrics.

## **CLINICAL OUTCOMES**

Clinical outcomes were composed of risk-standardized rates of mortality and readmission, weighted 80:20 respectively. Mortality included rates of in-hospital, 30-day, 90-day, and 1-year mortality, which were weighted in a ratio of 4:4:2:1 respectively. We chose these mortality endpoints to cover measurements in CMS' inpatient quality reporting programs as well as more extended periods, when mortality is a function of both hospital and community. Similarly for readmission, we wanted both a shorter interval that would better reflect inpatient care, and longer follow-up that would reflect post-hospital community support. Readmission was calculated from equally weighted risk-standardized rates of 7- and 30-day readmission.

Hospitalizations and readmissions were identified from the 100 percent Medicare inpatient file for years 2015 through 2017. Beneficiary characteristics and death date were obtained from the Medicare Beneficiary Summary file. Mortality and readmission rates were risk adjusted using the Risk Stratification Index (RSI), a machine-learning algorithm in the public domain that the Lown Institute trained using more than 24 million patient stays from MEDPAR data along with billions of carrier and outpatient claims with prior diagnoses. RSI has been tested on several different national and hospital-based datasets and has been shown to predict outcomes with greater discrimination compared with other publicly available risk adjustment tools. (Sources: [Validation and Calibration of the Risk Stratification Index](#) ; [Broadly Applicable Risk Stratification System for Predicting Duration of Hospitalization and Mortality](#); [Comparison of an Updated Risk Stratification Index to Hierarchical Condition Categories](#))

### *Clinical Outcomes Limitations*

While our clinical outcomes metrics adjust for underlying patient risk, it is likely that some environmental and social factors that impact patient outcomes, such as the availability of healthy food, access to preventive care, pollution, and others, may not be accounted for in our risk adjustment. Patients living in neighborhoods with poor environmental and social conditions often come to the hospital with more advanced cases of a given disease, and these patients are often discharged from the hospital into situations where they are less able to get the continuing care they need. For example, a patient who leaves the hospital for an apartment on the fifth floor of a walk up with no grocery store nearby might not do as well as a patient who can hire an aide to help them recover at home. That means hospitals caring for the poorest and sickest patients may appear to do worse on patient outcomes than is actually the case.

## PATIENT SAFETY

For patient safety we used well established indicators, such as rates of pressure ulcers, accidental punctures, and central intravenous line infections, provided by CMS on its Hospital Compare website for hospitalizations in 2017. This included the CMS composite measure (PSI-90), which comprises 11 different measures of patient safety as well as their hospital acquired infection (HAI) measure. We used these with a few adjustments. Like CMS, we excluded critical access hospitals, only six of which met our criterion for having a value for supplying data on more than three measures out of the PSI-90 and HAI values. (For more detail and a listing of the 11 measures used, please see <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/HAC-Reduction-Program-Fact-Sheet.pdf>.)

## PATIENT SATISFACTION

CMS was also the source for our patient satisfaction ranking. CMS relies on the annual Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey to give a rating of patient experience across 11 factors. We took the average of the 11 linear mean scores of these factors published on 2017 Hospital Compare, which also reports a percentage of patients with each summary response. The linear mean scores for each component are patient-mix and survey-mode adjusted by CMS. Hospital Compare did not report scores for 469 hospitals that had less than 100 responses. We chose to include 314 hospitals with between 50 and 100 responses after data analysis indicated that imputation of these scores would be reasonable to account for CMS's mean calculations and adjustment. We calculated scores for these hospitals by extrapolating to the nearest median score of hospitals with similar survey responses. (For more detail and a listing of the 11 measures used, please see <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS>.)

## HOSPITAL SYSTEMS

We looked at hospital systems as a secondary unit of analysis. We classified systems under the American Hospital Association definition as a group of hospitals “belonging to a corporate body that owns/manages health provider facilities or health-related subsidiaries.” We wished to see how these systems compared against each other within the components and the subsequent higher tiers of the Lown Index. We only classified hospitals that were selected for our ranking into systems. A system may have additional hospitals outside of our Lown Index that were not included as part of the calculations.

Due to the differing construction of each of the components, we employed different methodologies to “roll up” hospital component scores to the system level. Clinical outcomes metrics, patient safety metrics, and the race component of the inclusivity metric were generated by calculating an average of the hospital metric across the system of hospitals weighted by hospital patient volume. A weighted average was also utilized for system-wide patient experience metrics with the weight for each hospital set at the number of completed surveys recorded within the 2017 Hospital Compare dataset. In the case of overuse, weighted averages were applied to move from hospital to system for both overuse rates and segmented RFM scores, using each hospital’s population as the contributing factor.

The remaining component metrics were generated by calculating a hospital system score using full summations across all system hospitals for the populations of interest. For example, the community benefits component was calculated on a system level by summing total amounts of charity care, revenues, and operating expenses for all system hospitals. This framework was used for the education and income metrics within the inclusivity component (which are ratio scores, unlike the race metric) and for pay equity. In the case of race inclusivity, because it is not a ratio but rather a difference score, the size of the score would have been dependent on the number of hospitals in a system. Accordingly, we calculated it as a weighted-average across hospitals. Pay equity included the salaries of all designated local hospital leaders as well as the system CEO.

### *Hospital Systems Limitations*

We used a weighted average to calculate systems scores rather than a pure summation across all hospitals within a system. This means that larger hospitals with higher patient volume are weighted higher within our systems rollup. We did this because doing a pure summation across all hospitals could hide worse performing hospitals in systems with very high patient volumes. Because we are combining results from many individual hospitals in a system, these results may not be a reflection of the culture of a system as much as the summation of the varying cultures of individual hospitals. Finally, when ranking systems by state, even if a system includes only one hospital in that state, the system will appear in that state’s rankings.

## **ABOUT THIS WHITE PAPER**

This white paper is part of a series analyzing specific metrics in the Lown Institute Hospitals Index. This paper was written by Vikas Saini, Shannon Brownlee, Valérie Gopinath, Paula Smith, Kelsey Chalmers, and Judith Garber.

Suggested Citation: Saini, V., Brownlee, S., Gopinath, V., Smith, P., Chalmers, K., Garber, J. Methodology of the Lown Institute Hospitals Index. Brookline, MA: The Lown Institute. 2020.