LOWN INSTITUTE
HOSPITALS INDEX

White Paper Series

2021 Methodology Lown Institute Hospitals Index for Social Responsibility

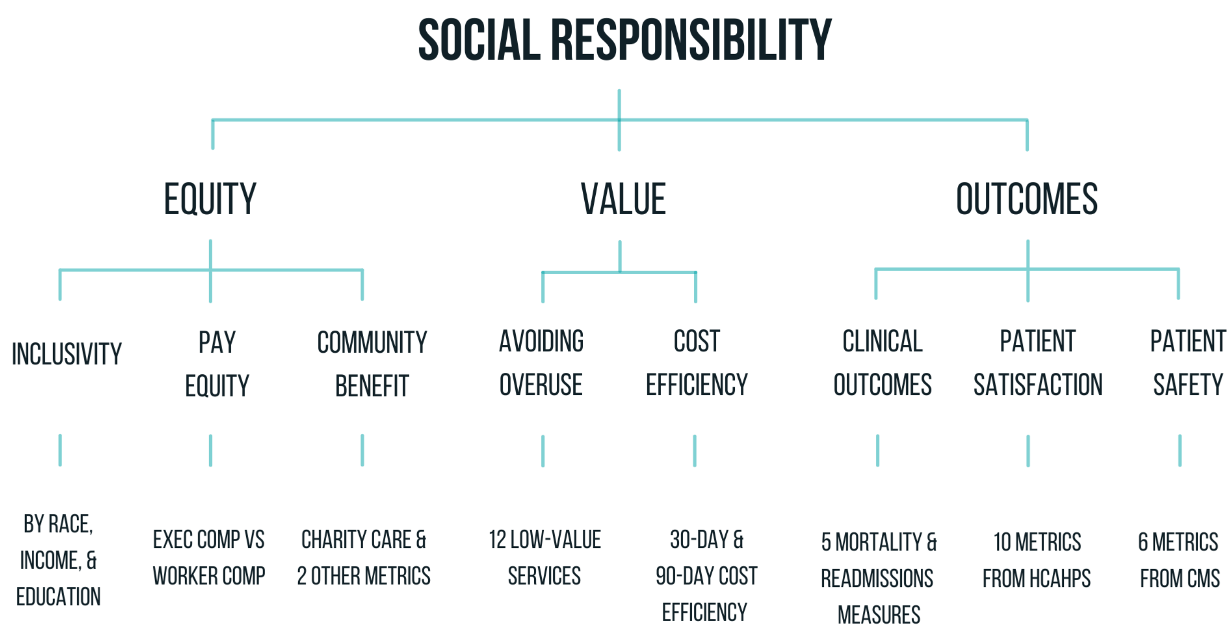
CONTENTS

OVERVIEW

ABOUT THE INDEX

The Lown Institute Hospitals Index is the first ranking of hospital social responsibility, evaluating more than 3,000 hospitals on their performance across health outcomes, value, and equity.

The Lown Institute Hospitals Index has 54 metrics distributed across four tiers (see Figure 1 below).



NEW ON THE INDEX

Here is a brief description of major changes made to the Lown Hospitals Index methodology from 2020 to 2021:

- The “Overall” score is now called “Social Responsibility” and the Tier 2 category formerly called “Civic Leadership” is now “Equity.”
- A new component of cost efficiency has been added to the value category. Due to this addition, the Value category is weighed slightly higher than last year in the overall Social Responsibility ranking (see *cost efficiency* for more).
- This year, hospitals’ Social Responsibility scores are calculated based on their grades in each category, rather than their raw scores. This means that hospitals ranked at the very top will have A grades in each category (see *Grades, stars, and rankings* for more).
- This year, each hospital receives a grade and a ranking for each Tier 3 component. The Tier 2 and Tier 3 grades and Tier 4 stars are now based on underlying hospital performance rather than relative performance (see *Grades, stars, and rankings* for more).
- About 350 more hospitals are included in this year’s rankings compared to last year, for a total of 3,709 hospitals.
- The Tier 4 metric previously called “charity care and other community benefits” has been separated into the two metrics: “charity care” and “community investment.”
- Changes have been made to the clinical outcomes algorithm to improve the risk adjustment (see *clinical outcomes* for more).
- Changes have been made to the overuse algorithm to improve accuracy (see *overuse* for more).
- Changes have been made to the inclusivity metric to improve accuracy (see *inclusivity* for more)

CREATING THE HOSPITAL SET

The Lown Hospitals Index for Social Responsibility includes 3,709 general acute care hospitals in the U.S.

Non-acute care hospitals, federal hospitals (e.g. Veterans Health Administration) and those outside of the 50 states and Washington, D.C. were excluded, as were hospitals run by Medicare Advantage programs (e.g. Kaiser Permanente), and specialty hospitals with more than 45% admissions for orthopedic, cardiac, or surgical procedures. We eliminated hospitals that were closed as of October 2020 by checking against Hospital Compare, a website run by the Centers for Medicare and Medicaid Services (CMS). Hospitals with patient volume below 50 annual patient stays were also eliminated as well as hospitals that did not perform any surgery in each of the three years spanning 2016 to 2018. This left a list of 3,709 hospitals: 568 for-profits, 2,405 private nonprofits, and 736 public nonprofits.

We defined Safety Net hospitals as the top 20% of hospitals based on the proportion of patients eligible for both Medicare and Medicaid. The dual-eligibility ratio was measured as the number of dual-eligible patient days out of all Medicare patient days in the Medicare Provider Analysis and Review (MEDPAR).

Information on hospital characteristics was taken from the Fiscal Year 2018 American Hospital Association (AHA) annual survey and Medicare Impact File as well as the CMS Hospital Compare database.

EQUITY

The equity tier 2 category, (previously known as civic leadership), comprises three components: community benefit, inclusivity, and pay equity weighted in a ratio of 2:2:1, respectively.

NEW THIS YEAR

- The category previously called "Civic Leadership" is now "Equity."
- In the community benefit component, the metric previously called "charity care and other community benefits" has been separated into the two metrics "charity care" and "community investment." This change was implemented to make it easier to compare hospitals on these metrics separately.
- In the inclusivity component, instead of using the center of the hospital's zip code, we now use the hospital's address as the center of the hospital's community area radius.
- In the inclusivity component, for the community area zip code weighting we now use a hospital-specific rate rather than a constant rate across all hospitals .

PAY EQUITY

For pay equity, we obtained data for Chief Executive Officer (CEO) compensation from three different sources corresponding to the tax status of the hospital. Compensation data on for-profit, publicly-traded hospital systems was obtained from Securities and Exchange Commission's (SEC) Edgar database. Public, non-federal hospital CEO salaries were gleaned from available payroll data and other public records. For nonprofit hospitals required to file with the IRS, we accessed the IRS 990 filings on the Amazon Web Services (AWS) Registry of Open Data.

We generated our own comprehensive dataset that linked CMS hospital data to IRS tax filings. To do this, we first created a crosswalk between the two datasets. After isolating tax entities that filed a Schedule H, we matched addresses automatically for 92% of hospitals and manually for the remaining 8%. Using the IRS dataset and the listed hospital administrator in the AHA dataset, we used text matching algorithms to identify CEO names and then manually verified the result. We were able to find 1,992 hospital CEO salaries of the original 2,398 (83.1%) hospitals using this strategy.

In cases in which CEO pay was unavailable for publicly traded (for-profit) private hospitals and public hospitals, values were imputed using predictive mean matching (PMM) based on CEO pay for nonprofit hospitals, combined with other variables such as bed size and hospital revenue. Pay for 1,992 nonprofit hospitals was used to impute values for three populations with unavailable pay: 561 for-profit hospitals, 507 public hospitals, and 406 nonprofit hospitals whose 990 forms did not contain the full executive compensation information. For public hospitals, imputed values were multiplied by average public to nonprofit compensation ratio. For for-profit hospitals, imputed values were multiplied by average for profit to nonprofit system compensation ratio. For hospitals within systems (two 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 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 the work hours listed in their IRS forms, defaulting to 40 when the hours were not listed, and calculated a ratio of CEO pay to average worker pay.

Pay equity limitations

Data anomalies may exist if different hospitals reported the same person with a name other than 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

The community benefit metric measures hospital spending on charity care and community health initiatives, as well as their service of Medicaid patients. Community benefit is a composite of three details: charity care, Medicaid revenue, and community investment, which included several categories of community benefit spending that we deemed to be meaningful. 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 hospitals claim for caring for such patients); health professionals training (which is already largely subsidized by the federal government); and research. For this metric, our goal was to focus on spending that directly benefits community health and the upstream factors that affect it.

We ranked 3,641 hospitals on community benefit. For 2,300 hospitals with data available for all three metrics, each metric was weighed equally in the composite at one-third of the total community benefit score. For 1,341 hospitals with data for two of the metrics available, each metric was weighed equally in the composite as half of the total score.

Charity care

Charity care is free or discounted care provided on the basis of the patient's financial situation. We measured charity care as a share of total hospital expenses as reported in the Centers for Medicare and Medicaid's Hospital Cost Reports (HCRIS). We ranked 3,619 hospitals on charity care; data for 22 hospitals were unavailable. For most hospitals we used 2018 HCRIS data. For 184 hospitals, there was no 2018 HCRIS data available so we substituted 2017 data.

Community investment

We measured hospital spending on community investment, as a share of total hospital expenses, using Fiscal Year Ending 2018 Internal Revenue Service (IRS) tax filings. Nonprofit hospitals are required to report community benefit spending to the IRS to maintain nonprofit status. IRS data on these community benefits were available for 2,330 hospitals. For 895 hospital systems that filed with multiple hospitals as one tax entity, we estimated each individual hospital's community benefit spending by prorating each hospital's share of system revenue.

Community investment comprises a subset of hospital spending including: *Subsidized health services*, such as free clinics, some emergency services, telehealth services, and other services provided at a loss to the hospital; *community health improvement activities* such as health fairs, community health education classes, immunizations, interpreter services; *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 that impact health, such as education, air quality, and access to nutritious food.

Medicaid revenue

In general, hospitals in states that expanded Medicaid spend less on charity care because fewer patients need financial assistance. To account for hospitals' service of Medicaid patients and differences in state policy, we included a metric to estimate the proportion of the hospital's patients that are covered by Medicaid. We measured Medicaid patient revenue as a proportion of total patient revenue using HCRIS data. We ranked 3,633 hospitals on Medicaid patient revenue; eight hospitals did not have data available. For 184 hospitals, 2018 data were not available so we used 2017 data.

Community benefit limitations

The measurement of community benefits spending by hospitals has improved since the 2010 Affordable Care Act (ACA) clarified reporting requirements for IRS Form 990, which all hospitals must use for reporting their spending. However, there are still several limitations to the data that are available. 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 could not take into account other types of community benefits. Therefore, community benefit spending by public and for-profit hospitals may be undercounted.

For private nonprofit hospital systems that filed as a single tax entity, we estimated the community benefit spending for individual hospitals based on their share of system revenue. However, we did not have revenue data on all hospitals within these systems, so a hospital's share of system revenue within the LIHI dataset may not be the same as their share of system revenue among all hospitals in the system. Additionally, certain hospitals may spend more on community benefits than their share of system revenue would indicate. Our calculation does not capture this.

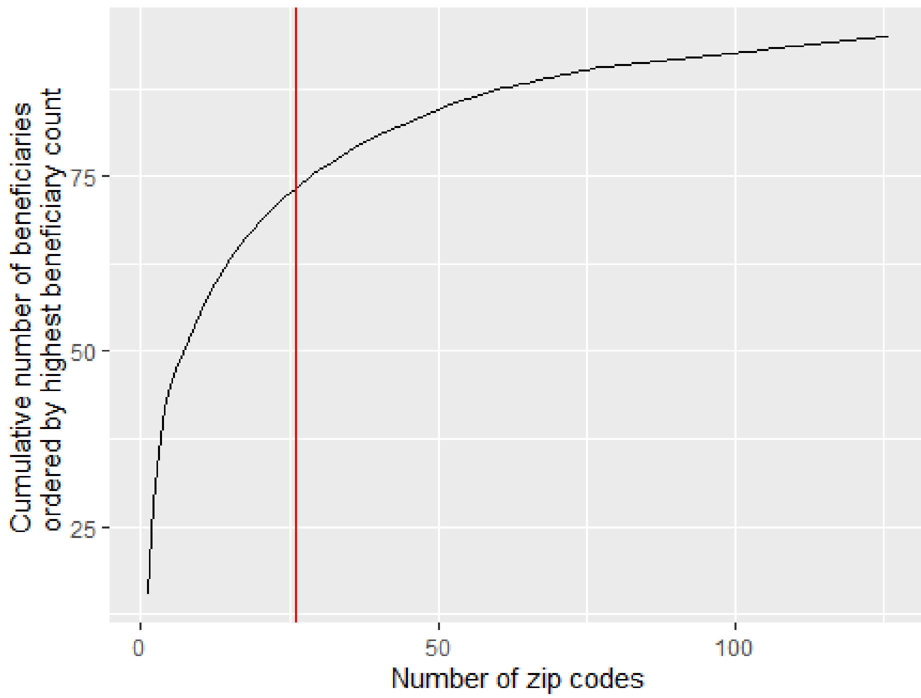
We used CMS's HCRIS data set to be able to compare charity care spending and Medicaid revenue across hospital types; however, this data set also has potential limitations. Hospitals are not required to have the cost reports audited by independent accounting firms and only some reports are audited by the federal government. Therefore, the charity care amount on these forms can be subject to inaccuracies or misrepresentations (source: Bai et al, 2021, Health Affairs). Charity care offered by hospital physicians is not always captured in HCRIS, which may underestimate charity care spending by hospitals with a salaried-physician model. Some hospitals may not report their revenue from Medicaid Managed Care programs to CMS; for these hospitals, their share of Medicaid revenue will be underestimated.

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 to the health of communities, 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 community area.

We defined the community area by using the zip codes of the hospital's Medicare patient population, sorted by the number of patients each zip code supplied. We then defined the radius of this area as the distance to zip codes whose contribution to the total patient population dropped significantly (see Figure 2 below).



Defining the catchment area radius: The red line is the zip code 'turning point'. The maximum distance across zip codes to the left of this line is the catchment area radius.

In other words, zip codes outside of the radius contributed very few if any individuals to the hospital's patient population, while zip codes inside the radius contributed the vast majority of patients (median of 87.6% of patients across hospitals). All people living within the defined radius were deemed to be potential patients of that hospital, and thus defined the population in the denominator of the inclusivity score.

We calculated the demographics by using the American Community Survey data on income and education as proxies for socioeconomic status, 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 compared to the demographics of the population within the zip codes inside the defined radius.

To calculate the demographic scores for people in the community area 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 the community area zip codes. 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 the contribution of all zip codes beyond the point at which 50% of a hospital's patients had come.

We created the hospital score by using the Census zip code demographic data of the actual patients, weighted by contribution to the total hospital's patients, and without a distance attenuation. We then compared the community 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 community area. For example, if a zip code had an 80% population of low-income earners and 20% high-income earners, we assume 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% 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. Conversely, we would not be able to know if the 20% of high income earners was completely absent from the hospital's patient population. 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 rewards hospitals that effectively "over-serve" communities with lower average income and education attainment and higher minority populations. That's by design. Hospitals whose catchment area demographics and patient demographics are reasonably aligned receive a mid-range score in the percentile ranking of the inclusivity scores. For example, a hospital may be in a catchment area that is all very wealthy and their entire patient demographics might reflect this demographic fact. In such a situation it will be difficult to improve upon a middling score. We did not include a racial inclusivity score for hospitals where most people in the community area were the same race (defined as the probability of selecting two persons that are the same race being greater than 95%).

For full details, please see our paper on inclusivity at *MedRxiv*.

VALUE

The value category was based on two components: Cost efficiency (a new measure this year) and avoiding overuse, which are weighted in a ratio of 3:2 respectively.

NEW THIS YEAR

- A new component of cost efficiency has been added to the value category. Due to this addition, the Value category is weighed slightly higher than last year in the overall Social Responsibility ranking.
- Adjustments have been made to the avoiding overuse algorithm to improve accuracy.

AVOIDING OVERUSE

This component includes rates of overuse of 12 low-value medical services (see Table below)

Table: Overuse definitions for 12 low-value services

Name of low-value service	Description of service	When is it overuse?	How is overuse measured?
Arthroscopic knee surgery	Surgery to remove damaged cartilage or bone in the knee using an arthroscope (tiny camera)	Overuse when it's for patients with osteoarthritis or for "runner's knee" (damaged cartilage). Excluding patients with meniscal tear.	Overuse measured as the proportion of arthroscopic knee surgery that met our criteria for overuse.
Carotid artery imaging for fainting	A test to screen for carotid (neck) artery disease. Includes CT, Magnetic resonance angiography, and duplex ultrasound	Overuse for patients where syncope (fainting) is the primary diagnosis on the claim and no history of syncope in the past two years. Exclusions for stroke/TIA, retinal vascular occlusion/ischemia, nervous and musculoskeletal symptoms.	Measured as the proportion of patients who came to the hospital with fainting but no other symptoms of serious disease and received carotid artery imaging.
Carotid endarterectomy	Procedure to remove plaque buildup from a carotid (neck) artery in a patient to prevent stroke	Overuse when performed on female patients without stroke symptoms or history of stroke.	Measured as the proportion of carotid endarterectomies that met our criteria for overuse, out of all the CEAs performed.
Coronary artery stenting	Procedure to place a stent or balloon in a coronary artery	Overuse when performed on patients with stable heart disease (not having a heart attack or	Measured as the proportion of coronary stents that met criteria

		unstable angina). Excluding patients with current and past diagnosis of unstable angina as well as patients having a heart attack.	for overuse, out of all the stents placed.
EEG for fainting	A test of the electrical activity of the brain	Overuse for patients where syncope (fainting) is the primary diagnosis on the claim and no history of syncope in the past two years.	Measured as the proportion of patients who fainted but no other symptoms of serious disease who received an EEG.
EEG for headache	A test of the electrical activity of the brain	Overuse for patients with headache as the primary diagnosis on the claim and no history of headache in the past two years. Also exclusions for epilepsy and recurrent seizures, convulsions, and abnormal involuntary movements.	Measured as the proportion of patients who came to the hospital with headache but no other symptoms of serious disease who received an EEG.
Head imaging for fainting	A CT scan or MRI of the head	Overuse for patients where syncope (fainting) is the primary diagnosis on the claim and no history of syncope in the past two years. Exclusions for epilepsy or convulsions, cerebrovascular diseases including stroke/TIA and subarachnoid hemorrhage, head or face trauma, altered mental status, nervous and musculoskeletal system symptoms, including gait abnormality, meningismus, disturbed skin sensation, speech deficits, personal history of stroke/TIA.	Measured as the proportion of patients who came to the hospital with fainting but no other symptoms of serious disease and received an MRI or CT scan.
Hysterectomy	Surgical removal of the uterus	Overuse for patients except malignancy and carcinoma in situ.	Measured as the proportion of hysterectomies that met our criteria for overuse, out of all the hysterectomies performed.
Inferior vena cava filter (IVC)	Procedure to place a filter (a medical device) in the large vein in the abdomen to prevent blood clots from moving to the lungs	Overuse for all patients	Measured as the number of times an IVC filter was overused, as proportion of total hospital volume.
Renal artery stenting	Procedure to place a stent or balloon in the renal (kidney) artery in a patient with high blood pressure or cholesterol (plaque) buildup in the artery	Overuse when done for hypertension or plaque buildup. Excluding patients that had diagnosis of fibromuscular dysplasia of renal artery	Measured as the number of times a renal artery stent or balloon was overused, as a proportion of total hospital volume.

		(abnormal twisting of the blood vessels)	
Spinal fusion/ laminectomy	Procedure to remove part of a spinal vertebra or fuse vertebrae together	Overuse for patients with low-back pain, excluding patients with radicular symptoms, herniated disc, radicular pain, scoliosis; also excluding prior 2 occurrences within 30 days of radiculopathy, sciatica, or lumbago.	Measured as the proportion of spinal fusion or laminectomy procedures that met our criteria for overuse, out of all the spinal fusions done.
Vertebroplasty	Procedure to inject cement into the vertebrae to relieve pain from spinal fractures	Overuse for patients with spinal fractures caused by osteoporosis. Excluding claims with bone cancers, myeloma, or hemangioma.	Measured as the proportion of patients that came in with spinal fractures caused by osteoporosis who received vertebroplasty.

We chose these services from the overuse literature. Renal stenting and inferior vena cava filters have been shown in high-quality clinical trials to be ineffective and are nearly always considered overuse. The remaining interventions are considered overuse 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 100% Medicare claims datasets (MEDPAR and outpatient) to count instances when these services were used. Hospitals without a capacity to perform a service, as reflected in their claims history, were excluded from the rating for that particular service. Hospitals without capacity to perform at least four services were excluded entirely from the overuse ratings. Hospitals with capacity to perform fewer than eight services were also excluded if two of those services were renal stent or EEG for headache, because of the very low volume of these two services. Renal stent and EEG for headache were considered low volume because among the 12 services we examined, these two had the lowest instances of overuse across all hospitals in our national sample.

To calculate overuse rates for the 12 services, we used the total patient volume as the denominator for those services which are low-value in most cases (renal stenting and inferior vena cava filter). For the remaining services where there was some benefit in certain circumstances, we used a service-specific (for the procedures) or diagnosis-specific denominator (for tests and imaging). We used a reliability adjustment on these rates so hospitals with smaller denominator volumes had their rates shifted towards the overall mean.

Before combining these rates into one metric, we standardized them using a minimum-maximum transformation (so they were between zero and one). We then calculated the overuse score as the weighted sum of these 12 standardized values. The weights were determined by the count of total low-value services nationally in our data set. If a hospital had no capacity for a service, we redistributed this weight to their other service results.

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 necessarily 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.

We tried 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 to perform the service.

COST EFFICIENCY

The cost efficiency component measures the clinical outcomes hospitals achieve over the cost of care. This metric encompases two details: 30-day mortality and cost, and 90-day mortality and cost.

Calculating 30-day and 90-day episode costs

We measured 30- and 90-day total, standardized Medicare payments for patients hospitalized in 2016 to 2018. We excluded any hospitalizations that were transfers from another hospitalization, had denied Medicare payments, if patients left against medical advice, or where the primary payer was not Medicare.

For each hospitalization, we found the claim payment amount in all claims within 30 or 90 days from the admission date. These claims included: inpatient, outpatient, carrier, skilled nursing facility, home health agencies, durable medical equipment and hospice claims. We excluded any claims where Medicare denied the payment.

We prorated any claims that started but did not finish in the 30- or 90-day period after the index hospitalization. For example, if a patient had another inpatient visit starting on day 29 after their first hospitalization, and finishing on day 31, then only two-thirds of this inpatient claim payment would be included in the patient's total 30-day payment.

Medicare adjusts their payment amounts to hospitals and other providers based on various geographic factors. To account for this, we calculated standardized payments using the Virtual Research Data Center's public use files of 2016 to 2018 Hospital Referral Regions (HRR) standardized ratio tables for patients over 65. These tables have separate values for each claim type (inpatient, outpatient, etc.). Our standardized payment amount was the hospital's HRR standardized payment value for the claim year divided by the HRR actual payment value, multiplied by the claim payment amount.

A hospitalization's 30-day and 90-day standardized payments were the total sum of the standardized payments across each claim type.

Risk-standardized payments

We risk adjusted the 30-day and 90-day standardized payments for each hospitalization using hierarchical logistic regression models. The response variable in the model was the episode standardized payment per survival day where survival day was the number of days the patient survived in the 30-day or 90-day episode. Model predictions provided the risk-standardized payment per survival day with hospital effects (predicted) and without hospital effects (expected).

The risk-standardized payment per survival day for each hospitalization was multiplied by the number of survival days to get the predicted and expected episode cost for each hospitalization.

We then calculated the mean risk-standardized predicted cost (P) and expected cost (E) for each hospital. A hospital's risk-standardized payment (RSP) is the hospital's P/E ratio multiplied by the national average episode cost. We calculated 30-day and 90-day RSP for each hospital using this method.

Cost efficiency metric

Our goal for the cost efficiency score was to reward hospitals with low mortality rates and low costs, and give the lowest scores to hospitals with high mortality rates and high costs. We also decided to bias our scores to give hospitals with high costs and low mortality a higher score than hospitals with low costs and high mortality. This is because we believe that if there is a trade-off between costs and mortality, we should favor better mortality rates compared to lower costs.

In order to operationalize this metric, we mapped the respective 30-day and 90-day risk standardized mortality rates and risk-standardized payments on a cartesian plane. We transformed the mortality rates and payments using a min-max transformation, so the range of values of the two variables were equal.

We then created a point on this plane that represented the ideal (most cost efficient) hospital, with the lowest mortality rate and payment value. We then used vectors to calculate the distance and angles between every single hospital in the data set and this ideal hospital using polar coordinates.

We then multiplied these two values, the distance and the angle, between a hospital's results and the best, theoretical hospital to generate our cost efficiency metric. We included the angle in the cost efficiency metric to ensure that if there were two hospitals with an equal distance from the ideal hospital on the payment-mortality plane, hospitals with lower mortality would receive a better score than hospitals with higher mortality. Larger angles reflected higher mortality rates, while smaller angles higher payments.

Cost efficiency limitations

We included Medicare payments and not payments from other payers, such as patient contributions or other insurers. This means we might be underestimating the true costs of some patient episodes.

Since our data includes Medicare beneficiaries and standardized costs, we cannot examine price variation as part of our metric. A hospital might be highly cost efficient because they have low readmissions and avoid unnecessary care, but they may charge high prices to non-Medicare patients.

Our cost standardization method is specific to the HRR. Some HRRs are quite large, and there might be more specific adjustments made within these regions not accounted for in our standardization approach.

While mortality and cost are adjusted for underlying patient risk, it is likely that some environmental and social factors that impact patient outcomes may not be accounted for in our risk adjustment. That means hospitals caring for the poorest and sickest patients may appear to do worse on mortality and cost (see *clinical outcomes limitations* for more).

OUTCOMES

Our outcomes 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 outcomes 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.

NEW THIS YEAR

- Our clinical outcomes measures include a new adjustment for hospitals having disproportionately sicker or healthier patients by including patient risk mix within the model. This means that the proportion of high-risk patients at a hospital is taken into account when looking at clinical outcomes.
- Our clinical outcomes measures include an adjustment for hospital case mix, based on the proportion of cases that fall within certain Diagnosis Related Groups (DRG) such as cardiac or orthopedic cases.
- For clinical outcomes, we added a patient-level indicator for end-stage renal disease to the algorithm, to better adjust for patient risk.
- In 2020, the clinical outcomes component included a metric for 1-year mortality. We have removed this metric in 2021 to avoid penalizing hospitals for community conditions outside of their control.

CLINICAL OUTCOMES

Clinical outcomes were composed of risk-standardized rates of mortality and readmission, weighted in a 80:20 ratio respectively. Mortality included rates of in-hospital, 30-day, and 90-day mortality, which were weighted in a ratio of 30:30:20 respectively. We chose these mortality endpoints to cover measurements in CMS' inpatient quality reporting programs as well as a more extended period 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 2016 through 2018. 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), an 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)

Risk-standardized mortality and readmission

We risk adjusted the mortality and readmission rates for each hospitalization using hierarchical logistic regression models. Model predictions provided the risk-standardized mortality or readmission with hospital effects (predicted) and without hospital effects (expected) for each hospitalization.

For each hospital, we then calculated the predicted (P) and expected mortality (E) based on all of its hospitalizations. A hospital's risk-standardized mortality (RSMR) is the hospital's P/E ratio multiplied by the national observed mortality rate. We calculated in-hospital, 30-day, and 90-day risk-standardized mortality using this method and also applied the same method for 7 and 30-day readmission.

In addition to the patient conditions in RSI, we included model effects to account for differences in hospital volume, case mix, and patient risk mix. At both the patient and hospital level, we included model effects for dual eligibility, and at the patient level an indicator for end stage renal disease.

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 unrelated to the quality of their care. In this year's implementation, we included additional model effects to adjust for hospitals having disproportionately sicker or healthier patients.

For more details, please see our paper in the journal *Medical Care*.

PATIENT SAFETY

For patient safety we used well-established indicators provided by CMS on its Hospital Compare website for hospitalizations, such as rates of pressure ulcers, accidental punctures, and central intravenous line infections (our data included 2017 to 2019 to cover admissions in 2018). We included the CMS composite measure (PSI-90), which comprises 10 separate indicators of patient safety, as well as 5 hospital acquired infection (HAI) measures. We included a reliability adjustment for the HAI measures using the reported numerator and denominator counts from Hospital Compare. For a patient safety overall score, hospitals had to have had at least three of the PSI-90 or HAI results. For more detail and a listing of the measures used, please see the CMS webpage on hospital acquired conditions.

PATIENT SATISFACTION

CMS Hospital Compare was also the source for our patient satisfaction ranking. CMS uses the annual Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey to give a rating of patient experience across 10 factors. We took the average of the 10 linear mean scores of these factors published on the 2018 Hospital Compare site, which also reports a percentage of patients with each summary response. The linear mean scores for each component are adjusted for patient-mix and survey-mode by CMS.

We chose to include 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 10 measures used, please see the CMS webpage on patient experience.

PUTTING IT TOGETHER

NEW THIS YEAR

- This year, hospitals' Social Responsibility grades are based on hospitals' grades in each Tier 2 category, rather than their raw scores. This means that hospitals ranked at the very top must have A grades in outcomes, value, and equity.
- This year, the Tier 2 and Tier 3 grades and Tier 4 stars are based on underlying hospital performance rather than relative performance (how other hospitals did). Previously, hospitals only received 5 stars or A grades if they scored in the top percentile of all hospitals. This year, if many hospitals all perform well on a certain metric, they could all receive 5 stars or A grades.

GRADES, STARS, AND RANKINGS

Tier 4

The lowest tier, tier 4, includes 42 details, presented for each hospital as a star rating. For each detail, we divide the range of results into five equal bins. Hospitals in the top bin receive five stars, the second bin four stars and so on.

In last year's rankings (LIHI 1.0), star ratings from 1-5 stars were given based on relative performance compared to other hospitals (ie. the top 20% of hospitals received five stars and the bottom 20% of hospitals received one star). Now the hospital star ratings reflect their underlying performance, rather than hospitals' relative performance based on other hospitals. This means that if many hospitals receive scores at the high end of the score range, they all will receive five stars.

Tier 3

These 42 details are rolled up into eight components in tier 3: pay equity, community benefit, inclusivity, avoiding overuse, cost efficiency, clinical outcomes, patient safety and patient satisfaction. The methods for calculating each of these components is detailed in the relevant methods section. For each component, we explore the distribution of the results and assign grade values based on set cut-off values. Assuming this distribution is a normal distribution, we set the cut-offs so approximately 25% of hospitals receive an A, 40% receive a B, 20% receive a C and 15% receive a D. These percentages can deviate from the actual grade counts, as the component values are sometimes not normally distributed.

Last year, the top 8.33% of hospitals received an A+ for Tier 3, the second 8.33% received an A, and so on. Now the for Tier 3 grades are based on the underlying scores, so all hospitals within the top range of scores receive an A.

Tier 2

These eight components are then rolled up into three categories for tier 2: equity, value, and outcomes. Equity includes inclusivity, community benefits and pay equity (weighted 40, 40 and 20% respectively). Value includes cost efficiency and rates of overuse 12 procedures and tests (weighted 60 and 40% respectively). Outcomes include clinical outcomes, patient safety and patient satisfaction (weighted 62.5, 25 and 12.5% respectively).

To roll up these components for the Tier 2 ranking, we first take the 'grade point average' (GPA) of the component grades within each category. Similar to a college GPA value, we assign a 4 to an A, 3 to a B, 2 to a C, and 1 to a D. The GPA of the category is the weighted average of these values. Hospitals with a weighted GPA of 3.3 or higher get an A for Tier 2; a GPA of 2.7 or higher receives a B; a GPA of 1.8 receives a C; while anything less than this receives a D overall.

Last year, the top 8.33% of hospitals based on their Tier 2 score received an A+, the second 8.33% received an A, and so on. Now the Tier 2 grades more closely reflects hospitals' overall performance across all Tier 3 components.

To assign rankings for Tier 2, hospitals are sorted first by their grade and weighted GPA. Then within grades, hospitals are sorted by the weighted sum of their component scores.

Tier 1

The Lown Social Responsibility GPA is the weighted sum of a hospital's grades in the three categories: equity, value and outcomes (weighted 30, 30 and 40%, respectively). We used the same cut-offs to assign grades described above for the category grades.

The Lown Social Responsibility rankings are determined by first sorting the hospitals by their Tier 2 grades, then their GPA, and then the weighted sum of their Tier 2 scores.

In LIHI 1.0, the rankings and grades of Tier 1 were dependent on the weighted sum of the category scores. This year we decided to use the category grades first to determine the final rank and grade of the hospital. This now means that the very top-ranked hospitals for Social Responsibility must have A grades in outcomes, value, and equity.

We dropped hospitals from the Social Responsibility ranking if we did not have a clinical outcome and cost efficiency result due to the sampling used in our clinical outcomes modelling. We removed 699 hospitals without clinical outcomes or cost efficiency scores, leaving 3,010 hospitals with rankings for Social Responsibility. The results for these hospitals on other metrics are still visible.

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." Our goal was to see how these systems compared against each other within the various tiers of the Lown Index. We only classified hospitals that were selected for our ranking into systems. A system may have additional hospitals that were not included in the Lown Index.

For all metrics except patient satisfaction, we consolidated hospital component scores to the system level by calculating an average of each hospital metric across the system of hospitals weighted by annual average patient volume from 2016 to 2018. To calculate patient satisfaction scores at the system level, we computed a hospital average weighted by number of completed surveys recorded within the 2018 Hospital Compare dataset.

Hospital systems limitations

We used a weighted average across all hospitals within a system to calculate systems scores. Hospitals with higher patient volume are weighted higher within our systems rollup. We could have, alternatively, summed the numerator and denominators for all metrics within each system and calculated a system score that arguably could have reflected the culture of a system. However, we found that this approach meant that the system scores were most dependent on the hospitals with the largest patient volumes, and results from smaller volume hospitals made little impact on the system results. Our weighted average approach combines the results of individual hospitals, and therefore is likely a closer reflection of the combination of individual hospital's cultures opposed to the system culture as a whole. Finally, when ranking systems by state, the system will appear in that state's rankings if a system includes at least one hospital in that state.

ABOUT THIS WHITE PAPER

This white paper is part of a series analyzing specific metrics in the Lown Institute Hospitals Index for Social Responsibility. 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. 2021 Methodology of the Lown Institute Hospitals Index for Social Responsibility. Needham, MA: The Lown Institute. 2021.