

BRIEF REPORTS

Hospital Patient Safety Grades May Misrepresent Hospital Performance

Wenke Hwang, PhD^{1*}, Jordan Derk, BS², Michelle LaClair, MPH¹, Harold Paz, MD, MS³

¹Department of Public Health Sciences, Division of Health Services Research, Penn State University College of Medicine, Hershey, Pennsylvania; ²Department of Public Health Sciences, Penn State University College of Medicine, Hershey, Pennsylvania; ³Department of Medicine and Public Health Sciences, Penn State University College of Medicine, Hershey, Pennsylvania.

Nationally, there is strong interest in measuring hospital performance in patient safety. The Leapfrog Group uses a survey, along with other data sources, to calculate patient safety scores for 2600 hospitals across the United States. Under this methodology, every hospital is assigned 1 of 5 letter grades (A, B, C, D, F) depending on how the hospital stands in safety performance relative to all other hospitals. The results have been widely marketed and disseminated to employers, payors, and the public. Leapfrog strongly encourages employers and payors to negotiate hospital reimbursement rates based on the safety grade the hospital receives. Leapfrog's effort to develop a standardized method to provide patient safety information should be commended. However, less than one-half of the 2600 hospitals participated in the Leapfrog survey. For those nonparticipating hospitals, certain safety measures were absent

and alternative measures were used to calculate the safety score. A sample of the nation's most prestigious hospitals (n = 35) was drawn from the *U.S. News & World Report's* "Best Hospitals." Overall, the group of participating hospitals (n = 18) received an average grade of A (mean safety score = 3.165), whereas the group of nonparticipating hospitals received an average grade of B (mean safety score = 3.012). These nonparticipating hospitals were rescored using the methodology for participating hospitals. The results show that the majority of nonparticipating hospitals would have received a better safety grade. This demonstrates a potential shortcoming of Leapfrog's method and its tendency to discriminate against nonparticipating hospitals. *Journal of Hospital Medicine* 2014;9:111–115. © 2014 Society of Hospital Medicine

The Institute of Medicine (IOM) reported over a decade ago that between 44,000 and 98,000 deaths occurred every year due to preventable medical errors.¹ The report sparked an intense interest in identifying, measuring, and reporting hospital performance in patient safety.² The report also sparked the implementation of many initiatives aiming to improve patient safety.³ Despite these efforts, there is still much room for improvement in the area of patient safety.⁴ As the public has become more aware of patient safety issues, there has been an increased demand for information on hospital safety. The Leapfrog Group, a leading organization that examines and reports on hospital performance in patient safety, cites the IOM report as providing the focus that their newly formed organization required.⁵

Using 26 national measures of safety, The Leapfrog Group calculates a numeric Hospital Safety Score for over 2,600 acute care hospitals in the United States.⁶ The primary data used to calculate this score are collected through the Leapfrog Hospital Survey, the

Agency for Healthcare Research and Quality, the Centers for Disease Control and Prevention, and the Centers for Medicare and Medicaid Services (CMS). The American Hospital Association's (AHA) Annual Survey is used as a secondary data source as necessary. The Leapfrog Group conducts the survey annually, and substantial efforts are put forth to invite hospital administrators to participate in the survey. Participation in the Leapfrog survey is optional and free of charge.

Leapfrog recently moved a step further in their evaluation of hospital safety by releasing the Hidden Surcharge Calculator to enable employers to estimate "the hidden surcharge they pay for their employees and dependents because of hospital errors."⁷ The calculation depends largely on the letter grade (A–F) that the hospital received from Leapfrog's Hospital Safety Score. For example, Leapfrog estimated a commercially insured patient admitted to a hospital with a grade of C or lower would incur \$1845 additional cost per admission than if the same patient was admitted to a hospital with a grade of A.⁷ The Leapfrog group encourages employers and payers to use this information to adjust benefits structures so that employees are discouraged from using hospitals that receive lower hospital safety scores. Leapfrog also encourages payers to negotiate lower reimbursement rates for hospitals with lower hospital safety scores.

The accuracy of Leapfrog's hospital safety grades warrants attention because of the methodology used to score hospitals that do not participate in the Leapfrog Survey. One common barrier that prevents

*Address for correspondence and reprint requests: Wenke Hwang, PhD, Department of Public Health Sciences, Division of Health Services Research, Penn State University College of Medicine, 600 Centerview Drive, Suite 2200, Hershey, PA 17033; Telephone: 717-531-7070; Fax: 717-531-4359; E-mail: whwang@phs.psu.edu

Additional Supporting Information may be found in the online version of this article.

Received: September 10, 2013; Revised: November 22, 2013; Accepted: December 2, 2013

2014 Society of Hospital Medicine DOI 10.1002/jhm.2139

Published online in Wiley Online Library (Wileyonlinelibrary.com).

hospitals from participating is the amount of effort required to complete the annual survey, including extensive inputs from hospital executives and staff. According to Leapfrog, 4 to 6 days are required for a hospital to compile the necessary survey data.⁸ Leapfrog estimates a 90-minute commitment for the hospital chief executive officer or designated administrator to enter the information into the online questionnaire. This is a significant commitment for many hospitals. As a result, among the approximately 2600 acute care hospitals covered by Leapfrog's 2012 to 2013 safety grading, only 1100 (or 42.3%) actually participated in the Leapfrog hospital survey. This limits Leapfrog's ability to provide accurate scores and assign fair safety grades to many hospitals.

METHODS

Leapfrog Hospital Safety Score

Leapfrog's designated Hospital Safety Score is determined by 26 measures. The set of safety measures and their relative weight are determined by a 9-member Leapfrog expert panel of patient safety experts.⁹ The hospital safety score is divided equally into 2 domains of safety measures: process/structural and outcomes.⁶ The process measures "represent how often a hospital gives patients recommended treatment for a given medical condition or procedure," whereas structural measures "represent the environment in which patients receive care."¹⁰ The process/structural measures include computerized physician order entry (CPOE), intensive care unit (ICU) physician staffing (IPS), 8 Leapfrog safety practices, and 5 surgical care improvement project measures. The outcome measures "represent what happens to a patient while receiving care." The outcomes domain includes 5 hospital-acquired conditions and 6 patient safety indicators. A score is assigned and weighted for each measure. All scores are then summed to produce a single number denoting the safety performance score received by each hospital. Every hospital is assigned 1 of 5 letter grades depending on how the hospital's numeric score stands in safety performance relative to all other hospitals. The letter grade A denotes the best hospital safety performance, followed in order by letter grades B through F. The cutoffs for A and B grades represent the first and second quartile of hospital safety scores. The cutoff for the C grade represents the hospitals that were between the mean and 1.5 standard deviations below the mean. The cutoff for the D grade represents the hospitals that were between 1.5 and 3.0 standard deviations below the mean. F grades indicate safety scores more than 3.0 standard deviations below the mean.¹¹

Nonparticipating Hospitals

The Leapfrog Survey contributes values for 11 of the 26 measures utilized to calculate the Hospital Safety Score. The score of a nonparticipating hospital will not reflect 8 of these 11 measures. For the 3 remain-

ing measures, CPOE, IPS, and central line-associated blood stream infection, secondary data from the AHA Survey, AHA Information Technology Supplement Survey, and CMS Hospital Compare were used as proxies, respectively (Table 1). The use of a proxy effectively limits the maximum score attainable by nonparticipating hospitals. For instance, 2 of these 3 measures, CPOE and IPS, are calculated on different scales depending on hospital survey participation status. For CPOE, nonparticipating hospitals are limited to a maximum of 65 out of 100 points; for IPS, they are limited to 85 out of 100 points.⁶ Because the actual weight for each of these proxy measures is increased for nonparticipating hospitals in the calculation of the final score, their effective impact is exacerbated. The weight of CPOE and IPS measures in the overall weighted score are increased from 6.1% and 7.0% to 11.0% and 12.6%, respectively.

Study Sample

We examined the Leapfrog safety grades for "top hospitals," as ranked by *U.S. News & World Report*. Included in this sample were the top 15 ranked hospitals in each of the specialties, excluding those specialties whose ranks are based solely on reputation. Hospitals ranked in more than 1 specialty were only included once in the sample. This resulted in a final study sample of 35 top hospitals. Eighteen of these top hospitals participated in the Leapfrog Survey, whereas 17 did not.

Utilizing Leapfrog's spring 2013 methodology,⁶ the Hospital Safety Scores for the 35 top hospitals were calculated. The mean safety score for the 18 participating hospitals was then compared with the mean score for the 17 nonparticipating hospitals. Finally, the safety scores for each of the 17 nonparticipating hospitals, listed in Table 2, were estimated as if they had participated in the Leapfrog Survey. To do this, we assumed that the 17 nonparticipating hospitals could each earn average scores for the CPOE, IPS, and 8 process/structural Leapfrog measures as received by their 18 participating counterparts.

RESULTS

Out of these 35 top hospitals, those that participated in the Leapfrog Survey generally received higher scores than the nonparticipants (Table 2). The group of participating hospitals received an average grade of A (mean safety score, 3.165; standard error of the mean [SE], 0.081), whereas the nonparticipating hospitals received an average grade of B (mean safety score, 3.012; SE, 0.047). These grades were consistent whether mean or median scores were used.

To further examine the potential bias against nonparticipating hospitals, the safety scores for each of the 17 nonparticipating hospitals were estimated as if they had participated in the Leapfrog Survey. The letter grade of this group increased from an average of B

TABLE 1. Data Sources for the Patient Safety Score: Survey Participants Versus Nonparticipants

	Participants	Nonparticipants
Process/structural measures (50% of score)		
Computerized Physician Order Entry	2012 Leapfrog Hospital Survey	2010 IT Supplement (AHA)
ICU Physician Staffing (IPS)	2012 Leapfrog Hospital Survey	2011 AHA Annual Survey
Safe Practice 1: Leadership Structures and Systems	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 2: Culture Measurement, Feedback, and Intervention	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 3: Teamwork Training and Skill Building	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 4: Identification and Mitigation of Risks and Hazards	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 9: Nursing Workforce	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 17: Medication Reconciliation	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 19: Hand Hygiene	2012 Leapfrog Hospital Survey	Excluded
Safe Practice 23: Care of the Ventilated Patient	2012 Leapfrog Hospital Survey	Excluded
SCIP-INF-1: Antibiotic Within 1 Hour	CMS Hospital Compare	CMS Hospital Compare
SCIP-INF-2: Antibiotic Selection	CMS Hospital Compare	CMS Hospital Compare
SCIP-INF-3: Antibiotic Discontinued After 24 Hours	CMS Hospital Compare	CMS Hospital Compare
SCIP-INF-9: Catheter Removal	CMS Hospital Compare	CMS Hospital Compare
SCIP-VTE-2: VTE Prophylaxis	CMS Hospital Compare	CMS Hospital Compare
Outcome measures (50% of score)		
HAC: Foreign Object Retained	CMS HACs	CMS HACs
HAC: Air Embolism	CMS HACs	CMS HACs
HAC: Pressure Ulcers	CMS HACs	CMS HACs
HAC: Falls and Trauma	CMS HACs	CMS HACs
Central Line-Associated Bloodstream Infection	2012 Leapfrog Hospital Survey	CMS HAIs
PSI 4: Death Among Surgical Inpatients With Serious Treatable Complications	CMS Hospital Compare	CMS Hospital Compare
PSI 6: Collapsed Lung Due to Medical Treatment	CMS Hospital Compare	CMS Hospital Compare
PSI 12: Postoperative PE/DVT	CMS Hospital Compare	CMS Hospital Compare
PSI 14: Wounds Split Open After Surgery	CMS Hospital Compare	CMS Hospital Compare
PSI 15: Accidental Cuts or Tears From Medical Treatment	CMS Hospital Compare	CMS Hospital Compare

NOTE: Abbreviations: AHA, American Hospital Association; CMS, Centers for Medicare and Medicaid Services; DVT, deep vein thrombosis; HACs, hospital-acquired conditions; HAIs, healthcare-associated infections; ICU, intensive care unit; INF, infection; IPS, ICU physician staffing; IT, Information Technology; PE, pulmonary embolism; PSI, patient safety indicators; SCIP, Surgical Care Improvement Project; VTE, venous thromboembolism; *Based on publicly available Leapfrog methodology, accessed September 2013.

TABLE 2. Leapfrog Hospital Safety Grades of U.S. News & World Report's Top Hospitals

Participants	Leapfrog Grade	Nonparticipants	Leapfrog Grade
Brigham and Women's Hospital, Boston, MA	A	Abbott Northwestern Hospital, Minneapolis, MN	A
Duke University Medical Center, Durham, NC	A	Barnes-Jewish Hospital/Washington University, St. Louis, MO	C
Massachusetts General Hospital, Boston, MA	B	Baylor University Medical Center, Dallas, TX	C
Mayo Clinic, Rochester, MN	A	Cedars-Sinai Medical Center, Los Angeles, CA	C
Methodist Hospital, Houston, TX	A	Cleveland Clinic, Cleveland, OH	C
Northwestern Memorial Hospital, Chicago, IL	A	Florida Hospital, Orlando, FL	B
Ronald Reagan UCLA Medical Center, Los Angeles, CA	D	Hospital of the University of Pennsylvania, Philadelphia, PA	A
Rush University Medical Center, Chicago, IL	A	Indiana University Health, Indianapolis, IN	A
St. Francis Hospital, Roslyn, NY	A	Mount Sinai Medical Center, New York, NY	B
St. Joseph's Hospital and Medical Center, Phoenix, AZ	B	New York-Presbyterian Hospital, New York, NY	C
Stanford Hospital and Clinics, Stanford, CA	A	NYU Langone Medical Center, New York, NY	A
Thomas Jefferson University Hospital, Philadelphia, PA	C	Ochsner Medical Center, New Orleans, LA	A
UCSF Medical Center, San Francisco, CA	B	Tampa General Hospital, Tampa, FL	C
University Hospitals Case Medical Center, Cleveland, OH	A	University of Iowa Hospitals and Clinics, Iowa City, IA	C
University of Michigan Hospitals and Health Centers, Ann Arbor, MI	A	University of Kansas Hospital, Kansas City, KS	A
University of Washington Medical Center, Seattle, WA	C	UPMC, Pittsburgh, PA	B
Vanderbilt University Medical Center, Nashville, TN	A	Yale-New Haven Hospital, New Haven, CT	B
Wake Forest Baptist Medical Center, Winston-Salem, NC	A		

NOTE: Abbreviations: NYU, New York University; UCLA, University of California Los Angeles; UCSF, University of California San Francisco; UPMC, University of Pittsburgh Medical Center.

(mean safety score, 3.012; SE, 0.047) to an average of A (mean safety score, 3.216; SE, 0.046). Among the 17 nonparticipating hospitals, 15 showed an increase in safety score, of which 8 hospitals rescored a change

in score significant enough to receive 1 or 2 letter grades higher (Table 3). Only 2 hospitals had slight decreases in safety score, without any impact on letter grade.

TABLE 3. Estimated Safety Scores and Letter Grades for the 17 Nonparticipants Rescored as Participants

Hospital	Original Score (Grade)	Estimated Score* (Grade)
Abbott Northwestern Hospital, Minneapolis, MN	3.17 (A)	3.44 (A)
Barnes-Jewish Hospital/Washington University, St. Louis, MO	2.83 (C)	3.11 (B)
Baylor University Medical Center, Dallas, TX	2.90 (C)	3.25 (A)
Cedars-Sinai Medical Center, Los Angeles, CA	2.92 (C)	3.30 (A)
Cleveland Clinic, Cleveland, OH	2.76 (C)	2.78 (C)
Florida Hospital, Orlando, FL	2.98 (B)	3.38 (A)
Hospital of the University of Pennsylvania, Philadelphia, PA	3.29 (A)	3.26 (A)
Indiana University Health, Indianapolis, IN	3.14 (A)	3.37 (A)
Mount Sinai Medical Center, New York, NY	3.01 (B)	3.02 (B)
New York-Presbyterian Hospital, New York, NY	2.76 (C)	3.15 (A)
NYU Langone Medical Center, New York, NY	3.26 (A)	3.30 (A)
Ochsner Medical Center, New Orleans, LA	3.19 (A)	3.59 (A)
Tampa General Hospital, Tampa, FL	2.86 (C)	3.05 (B)
University of Iowa Hospitals and Clinics, Iowa City, IA	2.70 (C)	3.00 (B)
University of Kansas Hospital, Kansas City, KS	3.29 (A)	3.35 (A)
UPMC, Pittsburgh, PA	3.04 (B)	3.24 (A)
Yale-New Haven Hospital, New Haven, CT	3.10 (B)	3.08 (B)

NOTE: Abbreviations: ICU, intensive care unit; NYU, New York University; UPMC, University of Pittsburgh Medical Center.

*Average scores for the following measures were substituted for missing or incomplete data: computerized physician order entry; ICU physician staffing; Safe Practice 1: Leadership Structures and Systems; Safe Practice 2: Culture Measurement, Feedback, and Intervention; Safe Practice 3: Teamwork Training and Skill Building; Safe Practice 4: Identification and Mitigation of Risks and Hazards; Safe Practice 9: Nursing Workforce; Safe Practice 17: Medication Reconciliation; Safe Practice 19: Hand Hygiene; Safe Practice 23: Care of the Ventilated Patient.

We applied the same methods to test the top 17 Honor Roll Hospitals as designated by *US News & World Report*; among them, half are participating hospitals and another half nonparticipating hospitals. One hospital, Johns Hopkins Hospital was not scored by Leapfrog because no relevant Medicare data are available for Leapfrog to calculate its safety score. For this reason, Johns Hopkins was excluded from our comparison. The results persist even with this smaller sample of top hospitals. The group of 8 participating hospitals had an average grade of A (mean safety score, 3.145; SE, 0.146), whereas another 8 nonparticipating hospitals received an average grade of B (mean safety score, 3.011; SE, 0.075).

DISCUSSION

The Leapfrog Group's intent to provide patient safety information to patients, physicians, healthcare purchasers, and hospital executives should be commended. However, the current methodology may disadvantage nonparticipating hospitals. The combination of lower maximum scores and increased weight of the CPOE and IPS scores may result in a lower hospital safety score than is justified. Nonparticipating hospitals may also face more intensive pressure from employers and payors to lower their reimbursement rates due to the newly released Leapfrog Hidden Surcharge Calculator.

Leapfrog acknowledges that “the more data points a hospital has to be scored on, the better its opportunity to achieve a higher score.”⁸ This justification may lead to bias against nonparticipating hospitals. On the other hand, it is possible that hospitals with good safety records are more likely to participate in the Leapfrog Survey than those with poorer safety records. Without detailed nonresponse analysis from Leapfrog, it is impossible to know if there is a selection bias. Regardless, the Leapfrog result can subsequently misguide the payment rate negotiation between insurers and hospitals.

With this consideration in mind, Leapfrog should explicitly acknowledge the limitations of its methodology and consider revising it in future studies. For example, Leapfrog could only report on those measures for which there are data available for both participating and nonparticipating hospitals. Pending this revision, every effort must be made to distinguish between participating and nonparticipating hospitals. The outcomes of Leapfrog's hospital safety grades are made available online to consumers without distinguishing between participating and nonparticipating hospitals. The only method to differentiate the categories is to examine the data sources in detail amid a large volume of data. It is unlikely that consumers comparing hospital safety grades will take note of this caveat. Thus, Leapfrog's grading system can drastically misrepresent many nonparticipating hospitals' patient safety performances.

This study of The Leapfrog Group's Hospital Safety Score is not without limitations. The small sample utilized in this study limited the power of statistical testing. The difference in mean scores between participating and nonparticipating hospitals is not statistically significant. However, The Leapfrog Group uses specific numerical cutoff points for each letter grade classification. In this classification system statistical significance is not considered when assigning hospitals with different letter grades. It was clear that nonparticipating hospitals were more likely to receive lower letter grades than participating hospitals.

The small sample also posed challenges when attempting to account for missing data when comparing participating hospitals versus nonparticipating hospitals. Although a multiple imputation approach may have been ideal to address this, the small sample size coupled with the large amount of missing data (58% of hospitals did not participate in the Leapfrog Survey) led us to question the accuracy of this approach in this situation.¹² Instead, a crude, mean imputation approach was utilized, relying on the assumption that nonresponding hospitals had the same mean performance as responding hospitals on those domains where data were missing. In this study, we purposely selected a sample of hospitals from *U.S. News & World Report's* top hospitals. We believe the mean imputation approach, although not perfect, is

appropriate for this sample of hospitals. Future study, however, should examine if hospitals that anticipated lower performance scores would be less likely to participate in the Leapfrog Survey. This would help strengthen Leapfrog's methodology in dealing with nonresponsive hospitals.

Disclosures: Harold Paz is the CEO of Penn State Hershey Medical Center, which did not participate in the Leapfrog Survey. The authors have no financial conflicts of interest to report.

References

1. Kohn LT, Corrigan J, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 2000.
2. Stelfox HT, Palmisani S, Scurlack C, Orav EJ, Bates DW. The "To Err is Human" report and the patient safety literature. *Qual Saf Health Care*. 2006;15(3):174–178.
3. Clancy CM, Scully T. A call to excellence. *Health Aff (Millwood)*. 2003;22(2):113–115.
4. US Department of Health and Human Services. Adverse events in hospitals: national incidence among Medicare beneficiaries. Available at: <http://oig.hhs.gov/oei/reports/oei-06-09-00090.pdf>. Published November 2010. Accessed on August 2, 2013.
5. The Leapfrog Group. The Leapfrog Group—fact sheet 2013. Available at: <https://leapfroghospitalssurvey.org/web/wp-content/uploads/Fsleapfrog.pdf>. Accessed October 9, 2013.
6. The Leapfrog Group. Hospital Safety score scoring methodology. Available at: http://www.hospitalssafetyscore.org/media/file/Hospital-SafetyScore_ScoringMethodology_May2013.pdf. Published May 2013. Accessed June 17, 2013.
7. The Leapfrog Group. The Hidden Surcharge Americans Pay for Hospital Errors 2013. Available at: http://www.leapfroggroup.org/employers_purchasers/HiddenSurchargeCalculator. Accessed August 2, 2013.
8. The Leapfrog Group. 2013 Leapfrog Hospital Survey Reference Book 2013. <https://leapfroghospitalssurvey.org/web/wp-content/uploads/reference.pdf>. Published April 1, 2013. Accessed June 17, 2013.
9. Austin JM, D'Andrea G, Birkmeyer JD, et al. Safety in numbers: the development of Leapfrog's composite patient safety score for U.S. hospitals [published online ahead of print September 27, 2013]. *J Patient Saf*. doi: 10.1097/PTS.0b013e3182952644.
10. The Leapfrog Group. Measures in detail. Available at: <http://www.hospitalssafetyscore.org/about-the-score/measures-in-detail>. Accessed June 17, 2013.
11. The Leapfrog Group. Explanation of safety score grades. Available at: http://www.hospitalssafetyscore.org/media/file/ExplanationofSafetyScoreGrades_May2013.pdf. Published May 2013. Accessed June 17, 2013.
12. Sterne JA, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.