An assessment of data and methodology of online surgeon scorecards

Linda W. Xu, MD, Amy Li, BA, Christian Swinney, BA, Maya Babu, MD, MBA, Anand Veeravagu, MD, Stacey Quintero Wolfe, MD, Brian V. Nahed, MD, MSc, and John K. Ratliff, MD, in affiliation with the Council of State Neurosurgical Societies

OBJECTIVE Recently, 2 surgeon rating websites (Consumers’ Checkbook and ProPublica) were published to allow the public to compare surgeons through identifying surgeon volume and complication rates. Among neurosurgeons and orthopedic surgeons, only cervical and lumbar spine, hip, and knee procedures were included in this assessment.

METHODS The authors examined the methodology of each website to assess potential sources of inaccuracy. Each online tool was queried for reports on neurosurgeons specializing in spine surgery and orthopedic surgeons specializing in spine, hip, or knee surgery. Surgeons were chosen from top-ranked hospitals in the US, as recorded by a national consumer publication ranking system, within the fields of neurosurgery and orthopedic surgery. The results were compared for accuracy and surgeon representation, and the results of the 2 websites were also compared.

RESULTS The methodology of each site was found to have opportunities for bias and limited risk adjustment. The end points assessed by each site were actually not complications, but proxies of complication occurrence. A search of 510 surgeons (401 orthopedic surgeons [79%] and 109 neurosurgeons [21%]) showed that only 28% and 56% of surgeons had data represented on Consumers’ Checkbook and ProPublica, respectively. There was a significantly higher chance of finding surgeon data on ProPublica (p < 0.001). Of the surgeons from top-ranked programs with data available, 17% were quoted to have high complication rates, 13% with lower volume than other surgeons, and 79% had a 3-star out of 5-star rating. There was no significant correlation found between the number of stars a surgeon received on Consumers’ Checkbook and his or her adjusted complication rate on ProPublica.

CONCLUSIONS Both the Consumers’ Checkbook and ProPublica websites have significant methodological issues. Neither site assessed complication occurrence, but rather readmissions or prolonged length of stay. Risk adjustment was limited or nonexistent. A substantial number of neurosurgeons and orthopedic surgeons from top-ranked hospitals have no ratings on either site, or have data that suggests they are low-volume surgeons or have higher complication rates. Consumers’ Checkbook and ProPublica produced different results with little correlation between the 2 websites in how surgeons were graded. Given the significant methodological issues, incomplete data, and lack of appropriate risk stratification of patients, the featured websites may provide erroneous information to the public.

https://thejns.org/doi/abs/10.3171/2016.7.SPINE16183

KEY WORDS orthopedic surgery; surgeon ratings; Consumers’ Checkbook; ProPublica; website

In July 2015, 2 surgeon rating websites, Consumers’ Checkbook and ProPublica, were released to the public. The goal of each site was to provide patients with the surgical outcomes for common elective surgeries considered relatively low risk based on data released from the Centers for Medicare and Medicaid Services (CMS).25 These 2 websites estimate surgeon case volume and complication rates, grading surgeons against their colleagues. Providing patients with easy online access to performance data of surgeons is one means to allow patients to make more informed decisions. Performance metrics are increasingly used by hospitals and surgeons to promote bet-
ter practices and outcomes. Press releases heralding the release of the websites offered identification of individual surgeons with higher than expected complication occurrences. The implication was that these named surgeons were providing lower quality care than their colleagues.

There are critiques of these approaches that have been reported. A methodological critique of the ProPublica Surgeon Scorecard was recently published by Friedberg et al. at RAND Health, which highlighted potentially significant problems in the way ProPublica chose to rate surgeons. These problems include issues with their “adjusted complication rates,” which do not incorporate complications occurring without readmission, during the index hospitalization (when the surgery took place), or beyond the 30 days postdischarge. In addition, ProPublica ignores hospital-to-hospital variation in their analysis, assuming instead that the surgeon performs at a “hypothetical average hospital,” which may undermine comparisons between physicians operating in different facilities.

Both Consumers’ Checkbook and ProPublica have created surgeon rating algorithms based on methodologies that are not peer reviewed and have not been subject to appropriate scientific scrutiny. While there have been critiques of these methods, no study has quantitatively analyzed the ratings produced by these websites. In this study, we aim to both qualitatively review the methods in which surgeons are rated and quantitatively determine the utility of these surgeon rating websites when used to assess a group of surgeons in the fields of spinal neurosurgery and orthopedic surgery.

Methods

Methodology Assessment

We reviewed the methodology of each site via their published methods sections. Appendices outlining risk adjustment strategies, when present, were reviewed. We relied on reporting by the individual sites on their publicly accessible websites for methodology descriptions.

Hospital Selection

To identify a set of surgeons for this assessment, we selected spinal neurosurgeons and orthopedic surgeons from the top 20 hospitals as designated by the 2015–2016 US News and World Report in Neurology & Neurosurgery and Orthopedics, respectively. Review of only large, top-rated institutions helped standardize our comparison. While our findings may translate into the community setting as well, our study design, which omitted surgeons at smaller centers for the sake of standardization, prevents this from being definitively concluded.

In neurosurgery, the following 20 hospitals were included in our analysis: Mayo Clinic, Massachusetts General Hospital, Johns Hopkins Hospital, UCSF Medical Center, New York-Presbyterian University Hospital of Columbia and Cornell, Brigham and Women’s Hospital, UCLA Medical Center, Cleveland Clinic, NYU Langone Medical Center, Northwestern Memorial Hospital, Barnes-Jewish Hospital/Washington University, Emory University Hospital, Hospitals of the University of Pennsylvania-Penn Presbyterian, Mount Sinai Hospital, Rush University Medical Center, University of Pittsburgh Medical Center, Cedars-Sinai Medical Center, University of Kansas Hospital, IU Health Academic Health Center, and St. Joseph’s Hospital and Medical Center. We excluded Houston Methodist Hospital (ranked no. 16) in this study because the hospital did not specify spine specialty neurosurgeons on their website. We replaced this with St. Joseph’s Hospital and Medical Center (ranked no. 21) in our analysis.

In orthopedic surgery, the following 20 hospitals were included: Hospital for Special Surgery, Mayo Clinic, Cleveland Clinic, Massachusetts General Hospital, Hospital for Joint Diseases (NYU Langone Medical Center), Rush University Medical Center, Brigham and Women’s Hospital, Santa Monica-UCLA Medical Center and Orthopedic Hospital, Northwestern Memorial Hospital, New England Baptist Hospital, University of Pittsburgh Medical Center, Cedars-Sinai Medical Center, UCSF Medical Center, Thomas Jefferson University Hospital, University of California Davis Medical Center, Beaumont Hospital, Barnes-Jewish Hospital/Washington University, University of Iowa Hospital and Clinics, Abbott Northwestern Hospital, and Johns Hopkins Hospital.

Neurosurgeon and Orthopedic Surgeon Selection

From the top 20 hospitals in the 2015–2016 US News and World Report, we identified spinal neurosurgeons and orthopedic surgeons. The methodology of Consumers’ Checkbook (http://www.checkbook.org/surgeonratings) and ProPublica (https://projects.propublica.org/surgeons) included only spine surgery and hip or knee replacement surgery. Therefore, we limited our selection of neurosurgeons to spine surgeons only; we included only orthopedic surgeons who specialized in spine, hip, or knee procedures. Specification of spine surgery instead of including general neurosurgeons was deemed necessary to maximize the accuracy of our findings. While the exact distribution of spine procedures cannot be definitively determined for such a large sample size, it can be assumed that those designated as spine surgeons primarily perform spine-related procedures. The same cannot be definitively said for general neurosurgeons.

Consumers’ Checkbook only analyzed data sets of surgeons practicing between January 1, 2009, and December 31, 2012. ProPublica only examined surgeons between 2009 and 2013. To maximize the chances that our study surgeons would be included in both of the surgeon rating websites, we limited our selection of surgeons to those who had finished residency or fellowship no later than 2010. Data on year of residency graduation was obtained from each program’s website. A total of 510 surgeons were included in this study, with 401 orthopedic surgeons (79%) and 109 neurosurgeons (21%).

Consumers’ Checkbook

The Consumers’ Checkbook website was searched for the individual surgeons’ records between August 12, 2015, and August 23, 2015. The Consumers’ Checkbook surgeon rating website provides the surgeon’s full name, sex, location, hospital affiliations, contact information, and educational background, including board certification and
medical school. The website assigns each surgeon a number of stars (3 to 5) based on how confident Consumers' Checkbook is of the surgeon's “better-than-average” outcomes. A 5-star rating represents the highest confidence that the surgeon’s results are better than that of the average surgeon in the US. A 3-star rating represents the lowest confidence of the surgeon’s outcomes. When a surgeon is not given stars, it is either because the surgeon’s results are not better than average or there are not enough data on the surgeon. Hence, no surgeon is awarded fewer than 3 stars. In patients without a star rating, the website does not distinguish between lack of data or data indicating that the surgeon’s results are below or equivalent to average. Consumers’ Checkbook also rates the surgeon’s operative volume compared with other doctors in their database: the volume of surgeries is listed as lower, medium, or higher volume.

**ProPublica**

The ProPublica website was searched for the individual surgeons’ records between August 12, 2015, and August 23, 2015. The ProPublica surgeon rating website provides the following output for each surgeon: surgeon’s full name, hospital affiliation, and complications related to the types of surgery performed. ProPublica provides the number of times the surgeon performed each operation, the reported number of complications, and the adjusted complication rate. The adjusted complication rate is further divided into low, medium, and high in comparison with other surgeons for the same operation. In the event of multiple operations listed, the adjusted complication rates were averaged for each surgeon.

**Statistical Analysis**

Data for each selected surgeon were collected from each website. The total number of operations performed was calculated by adding separate surgical types listed on ProPublica. An average complication rate was calculated using the separate complication rates listed per procedure. A chi-square test was used to compare availability of data between the 2 websites. A Student t-test was used to compare complication rates on ProPublica and star rating on Consumers’ Checkbook, and estimates of surgeon volume between websites.

**Results**

**Consumers’ Checkbook**

Using CMS inpatient data from January 1, 2009, to December 31, 2012, adverse outcomes following 18 surgical procedures were recorded for each surgeon from the Consumers’ Checkbook website. Adverse outcomes were defined as inpatient mortality, prolonged risk-adjusted length of stay (LOS), 90-day postdischarge mortality, and 90-day readmission. Logistic regression was used to estimate the probability of an adverse outcome and used to adjust the measured adverse outcome rate to create a risk-adjusted rate for each surgeon. This logistic regression included considerations for age, sex, year of the procedure, certain selected other diagnostic codes the patient also carried, and combinations with other procedures at the time of intervention. The algorithms of how prolonged risk-adjusted LOS or the logistic regression are calculated are not provided by Consumers’ Checkbook for review.

**ProPublica**

Using CMS inpatient data from 2009 to 2013, adverse outcomes following 8 surgical procedures were recorded for each surgeon from the ProPublica website. Complications were defined as death during initial hospital stay and 30-day readmission for a principal diagnosis suggestive of negative surgical outcome by a panel of reviewers. Complications were then adjusted for using logistic regression models for age, sex, health score based on comorbidities, and multilevel spinal procedures for spine surgeries.

**Issues in Consumers’ Checkbook and ProPublica Methodology**

Both surgeon rating websites claim to provide surgical outcomes for each surgeon in comparison with his or her colleagues: Consumers’ Checkbook in the form of a star grading system, and ProPublica in the form of their adjusted complication rate classification. However, neither website calculates complications through accepted standards by surgeons or the general public.22,23,27

Consumers’ Checkbook uses the term “complications” but does not actually measure true surgical complications, instead using LOS, readmissions, and deaths as proxies for occurrence of complications after surgery.4 ProPublica uses the term “complications” when they are really referring to readmissions to the hospital and death.24 The literature reports a high number of complications occurring during the patient’s initial hospitalization rather than subsequent patient encounters.3,18,32,34,35,38,39 For readmissions, the ProPublica site reviewed only the principal diagnosis as the cause for the readmission. Approximately half of readmissions in the ProPublica effort were considered to be secondary to postoperative complications. The site does not capture complications that occurred after a procedure during a patient’s initial, index hospitalization or complications that are managed on an outpatient basis.

Consumers’ Checkbook also does not directly assess perioperative complications, only examines LOS as compared with predicted LOS, examines only 90-day mortality, and uses prolonged LOS as a proxy for complication occurrence. The site does not explain how it developed a predicted LOS for each patient and does not exclude other reasons for prolonged LOS that may not be related to surgeon factors or occurrence of perioperative adverse events.

**Risk Adjustment**

Consumers’ Checkbook reports using a risk-adjusted LOS to measure surgeon outcomes, but they were vague about the details used to create the risk adjustment.5 ProPublica uses a “health score,” which is a modification of the Elixhauser comorbidity score, as an indicator of preoperative risk.24,33 This measure was developed to combine the Charlson comorbidity system with the Elixhauser classification system to measure a patient’s medical risk for in-hospital deaths.33 The Charlson comorbidity score may not be a reliable predictor of complication risk, especially in patients undergoing spine surgery.56
Using the described “health score” as a predictor of complications in surgical patients is not confirmed in the literature. ProPublica found little merit in their risk-adjustment approach: there was minimal correlation between a patient’s health score and postoperative complication occurrence, with an area under the curve of 0.57–0.63 per procedure.24

Consumers’ Checkbook Scoring
We searched a total of 510 surgeons in the Consumers’ Checkbook surgeon rating website: 401 orthopedic surgeons (79%) and 109 neurosurgeons (21%). Consumers’ Checkbook only found information on 145 surgeons (28%), with 72% of surgeons having no surgical information or listing on the website. Two surgeons had a profile but no stars listed. Of the surgeons found, 21 (14%) had a 5-star rating, 10 (7%) had a 4-star rating, and 114 (79%) had the lowest 3-star rating. The average number of stars assigned was 3.36. Concerning operative volume, 105 surgeons (73%) had a higher volume, 20 surgeons (14%) had a medium volume, and 18 surgeons (13%) had a lower volume of surgeries compared with other doctors in the same field.

ProPublica Scoring
We searched the same 510 surgeons in the ProPublica surgeon rating website. ProPublica had an entry for 413 surgeons (81%) on the website, but only 286 (56%) of these surgeons had any surgical data listed in their profile. The median number of surgeries listed per surgeon with data was 119 (range 20–1546 surgeries). The median adjusted complication rate was 2.8 (range 1.2–7.5 complications). Of the surgeons with data, different complication rates were given per elective procedure examined: 25 surgeons (9%) had a low adjusted complication rate, 270 (94%) a medium adjusted complication rate, and 50 (17%) a high adjusted complication rate for at least 1 listed procedure.

Comparing the Websites
The median complication rate of 2.8%, ranging from 1.2% to 7.5%, likely vastly underestimates the true complication rate after spine and joint surgeries compared with published series of 1%–10% of morbidity for spine procedures and up to 7%–9% for hip and knee replacements.6,7,12,13,15,19,20 Ninety-four surgeons (18.4%) were not represented on either surgeon rating website. The relative likelihood of finding a surgeon’s name on ProPublica was 2.8 times higher than Consumers’ Checkbook (95% confidence interval [CI] 2.5–3.3, p < 0.001). The relative risk of finding a surgeon’s data (not just the name) on ProPublica was 2.4 times higher than Consumers’ Checkbook (95% CI 2.1–2.8, p < 0.001).

We compared ProPublica’s adjusted complication rate categories and Consumers’ Checkbook’s outcome grading (3, 4, or 5 stars) to analyze how well the 2 websites correlated with each other in evaluating the same surgeons. Surgeons with the lowest 3-star rating on Consumers’ Checkbook had an average adjusted complication rate of 3.2% on ProPublica. Surgeons with a 4-star rating were assigned an average of 3.1% for the adjusted complication rate on ProPublica. Finally, surgeons with the best 5-star rating on Consumers’ Checkbook had a lower average adjusted complication rate of 2.6% on ProPublica. There was a trend toward decreasing complication rate as stars increased, but we found no statistical significance in the correlation (p = 0.79 between 3 and 4 stars, p = 0.23 between 4 and 5 stars; Fig. 1).

We compared the operative volume estimated by both Consumers’ Checkbook and ProPublica for surgeons. The actual number of cases listed on ProPublica was compared with the lower, medium, or higher volume classification given to surgeons on Consumers’ Checkbook. There was a statistically higher average number of cases of “higher volume” surgeons compared with “medium volume” surgeons (p = 0.01). No significant difference in average cases reported on ProPublica existed between “medium volume” and “lower volume” surgeons (p = 0.19, Fig. 2).

Discussion
A valid assessment of surgeon-specific outcome requires high-quality input data, accepted definitions of complications and preoperative risks, and adequate sample sizes.2 The Consumers’ Checkbook and ProPublica surgeon rating websites harbor severe methodological shortcomings with all 3 of these issues. The 2 sites claim to record surgeon complications and to offer a window into the clinical performance of surgeons nationwide; however, the output of these sites, as assessed in this study, is of limited value.

Administrative Data
The 2 sites use ICD-9 procedure codes to define their patient cohorts. There are substantial opportunities for
bias with this approach. The procedure codes combine a wide variety of procedures, with a wide variety of different complexities and different propensities for readmission risk. Spinal reconstructive procedures have a much higher risk of readmission and a higher risk of perioperative adverse event occurrence when compared with less extensive 1- or 2-level spinal fusions.

With this metric alone, surgeons performing reconstructive procedures would appear to have higher rates of complications/readmissions, and hence to have poorer results with this gross metric. Thus, for the spine surgery patient cohort, the 2 approaches actually analyze both low- and high-risk procedures. More than 2000 patients in the lumbar fusion (ICD-9 code 81.07) patient group are undergoing treatment with a primary diagnosis of scoliosis. For a Medicare-aged population, those procedures are likely reconstructive and not low risk. The researchers do not describe different complication rates in the different diagnosis categories they review.

Definition of Complication and Risk Adjustment

The definition of a complication used by each website is not representative of the experiences of surgeons and patients. A focus only on complications that produce a readmission or death likely misses the majority of complications occurring in this patient population. Neither website captures any other complications that occur without hospital readmission, perhaps the most common postoperative complications for elective spine or joint surgery. Use of prolonged LOS as a proxy for complication occurrence is not a validated approach and not a direct assessment of complication occurrence. Both websites provide inadequate information to patients who may consider complications of a procedure to be more than just those severe enough to render readmission, death, or a prolonged LOS.

Focusing on complications that require readmission and on patient deaths in isolation may present a potentially significant sampling error that exposes these models to substantial bias. Ignoring complications that occur during a patient’s index hospitalization likely ignores the majority of complications that a patient may suffer. The literature reports complication rates in lumbar fusion surgery conservatively of approximately 20%, with higher rates in older patients and in patients with significant comorbidities. Some procedures may have complication rates of up to 50%; this means the approximately 5% rate of readmissions focused on by the ProPublica report misses 75% or more of perioperative complications.

The ProPublica and Consumers’ Checkbook methodologies offer poor or no risk adjustment in each surgeon assessment. Patient comorbidities and medical conditions are important predictors and confounders of complication occurrence after surgery and should be adjusted appropriately when measuring surgeon outcomes. The “health score” reported as part of the ProPublica site has not been validated for use in spine surgery procedures and had low predictive accuracy. This poor predictive accuracy may be evidence of a weakness in the model’s risk stratification; the ProPublica conclusion that patient comorbidities have no impact on complication occurrence is not supported by the literature. It has been suggested that logistic regression in general poorly identifies outliers and reduces variability observed from random noise, and that hierarchical models or multilevel modeling is a more statistically valid method, but no validated studies have used these models to assess real-world surgeon data.

Sample Size

Deaths and readmissions in elective procedures are infrequent. Therefore, the rare readmission can drive a change in rating category for a surgeon, especially when both websites only account for data within a short 4-year (Consumers’ Checkbook) or 5-year period (ProPublica). This small sample of readmitted patients may not be sufficiently powered to predict overall complication incidence and focuses on a small subset of severe events, not a surgeon’s overall clinical care.

The small numbers of readmissions per surgeon mean that only a few readmissions may drive a change in score. The provided data for St. Joseph’s Hospital in Phoenix, Arizona, reveal that 3 surgeons have high adjusted rates of readmissions/deaths. However, the overall number of readmissions/deaths for all surgeons from St. Joseph’s is less than 10 per surgeon, and hence not reported. No raw readmission/death rate is reported for these surgeons. Since the total number of readmissions/deaths for each surgeon is redacted, further analysis is impossible. It is clear that, when low-, medium-, and high-risk surgeons are each having fewer than 10 readmissions/deaths over a 5-year data sample, perhaps only a single readmission could drive a change in risk category for a single surgeon. Statistically, a large number of procedures, possibly 2–10 times the average number of procedures performed by each surgeon, could be necessary to detect someone with true differences in complication rate. Thus, given these small sample sizes, the websites cannot truly predict which surgeons are either above or below average.
Findings for Surgeons From Top-Ranked Hospitals

The flaws in the methodologies of both surgeon rating websites may explain the inconsistent representation of neurosurgeons and orthopedic surgeons from top-ranked hospitals in the US, as well as a lack of significant agreement in the grading of surgeons between websites.

In this study, we focused on how surgeons from top-ranked hospitals in the fields of neurosurgery and orthopedic surgery were represented in the 2 recently released surgeon rating websites. For Consumers’ Checkbook, we found that 72% of surgeons were not represented on their website. For ProPublica, 19% of the surgeons were not found on the website. Even among surgeons with entries in ProPublica by name, we saw no surgical data listed 31% of the time. Even among surgeons who had been practicing for years at top-ranked hospitals, only 9% had any procedures performed listed as low complication on ProPublica and 79% were listed as 3 stars, the lowest number of stars that can be listed on Consumer Checkbook.

We tested the internal validity between the 2 websites in evaluating the same surgeons. When displaying complication outcomes in comparison with their colleagues, ProPublica uses an adjusted complication rate and Consumers’ Checkbook a star grading system. We found no statistical correlation between the number of stars a surgeon receives on Consumers’ Checkbook and the surgeon’s adjusted complication rate ranking on ProPublica. Operative volume estimates of surgeons between the 2 websites showed a greater correlation, but as only a few elective cases were included in evaluation of surgeon volume, this may not actually reflect true surgeon volume if the surgeon performs a great number of nonlisted procedures or nonelective cases.

The websites we assessed are advertised as being able to identify the best surgeons in the country compared with their peers. From a consumer perspective, an ideal surgeon would likely have low complication rates per ProPublica and high volume and 5 stars on Consumers’ Checkbook. Of our list of the 510 surgeons in these fields, only 4 surgeons fit these criteria. Our data call into question how information may be interpreted by consumers and whether it can truly be used to inform the choice of which surgeon to seek out for a procedure.

Implications of Reporting Surgeon Outcomes

Although it is difficult to predict how reported surgeon outcomes will affect the community of neurosurgeons and orthopedic surgeons, referring physicians, and patients, our best-case study thus far is with cardiothoracic surgery, the first specialty to start public reporting. In the 1980s to 1990s, a set of data outlining the surgical mortality rate following coronary artery bypass grafting (CABG) was released to the public. At both the individual surgeon and hospital level, those with reported low mortality rates saw rises in their market shares in the years following reporting and surgeons were able to charge more for their services.

Thus, it follows that those listed with high mortality rates or with no data must have had a reduction in the overall patient marketplace. In fact, 1 study showed those in the lowest quartile of performance were 3.5 times more likely to stop practicing. Given the possibility of real financial and career fallout, more than 60% of cardiothoracic surgeons reported less willingness to operate on the sickest patients and had refused at least 1 ill patient in the prior calendar year because of fear of its effect on mortality rate. At the same time, patients who were candidates for aortic repair, which is not reported, with similar comorbidities as CABG candidates, were significantly more likely to receive operations. Reporting also encourages “gamming” of the statistics by exaggerating preoperative morbidity to change risk assessments, changing case classification to fall outside of the reported case category, and transferring patients to long-term care facilities if a death is anticipated. As distribution and penetration of online information and quality reports are not equal among all populations, 1 study showed after CABG reports were released, patients of higher socioeconomic classes were treated by surgeons with lower mortality rates while patients of lower socioeconomic classes were treated by surgeons with higher mortality rates. Furthermore, an unexpected consequence of reporting surgeon outcomes is decreased access to cases for trainees, as any complication would be reported under the primary surgeon’s record. While outcome reporting has been shown to increase interest in quality improvement activities, there has been no study showing that these reports actually change surgeon mortality rates. Thus, despite the uncertain validity and meaning of reported surgeon outcomes, there can be certain financial, patient access, and training issues that follow.

Other Possible Metrics

Given the problems of validity with current publicly available performance measures, the question remains: what would be an accurate way to provide surgeon feedback and help the public make informed choices? Any valid surgeon reporting would require accurate input data, accepted definitions and assessments of complications and risks, and adequate sample size. The National Surgical Quality Improvement Program (NSQIP) has been a curated data set collected by sampling patient outcomes of hospital systems and services and producing reports for private review, but not open to the public. This has been effective in improving outcomes of participating hospital systems, but given random sampling of procedures, does not aim to provide data on specific surgeons. In a study of whether these data could be extrapolated out to provide feedback to specific surgeons, the NSQIP was found to have recorded data on 22%–35% of any individual surgeon’s case volumes and that disagreement in the data from the NSQIP and data from the total surgeon’s case volume could vary by up to 29%. Therefore, while the NSQIP does resolve the issue of quality of data, it does not completely eliminate the bias of small sample size and risk adjustment. Not surprisingly, most surgeons oppose the use of the NSQIP to produce surgeon-specific measures and do not support public release of these data. Alternatively, it perhaps is more logical to focus on hospital-based metrics rather than individual statistics. Individual surgeon statistics falsely attribute outcome to a surgeon alone, whereas the real outcome for a patient relies heavily on hospital, staff, and resources. Attempting to rank individual surgeons may remain futile as outliers remain rare.
and the real statistical difference between surgeons may, in fact, not be as significant as the public may assume.\textsuperscript{30}

**Conclusions**

The goals of transparency in reporting outcomes may be important in helping patients attain a high quality of care from surgeons. For these data to be valid, however, the approach must include high-quality data input, accepted definitions of complication and risk, and adequate sample size. The Consumers’ Checkbook and ProPublica surgeon rating websites harbor severe methodological issues with all of the basic necessities for valid reporting. A substantial number of spinal neurosurgeons and orthopedic surgeons from top-ranked hospitals who specialize in spine, hip, or knee surgery have no ratings on either site and the real statistical difference between surgeons may, in fact, not be as significant as the public may assume.\textsuperscript{30}

**Acknowledgments**

This work was supported by a Congress of State Neurological Surgeons Socioeconomic Fellowship to Dr. Xu.

**References**


33. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ: A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med Care* 47:626–633, 2009


**Disclosures**

Dr. Veeravagu has direct stock ownership in Precision Oncology.

**Author Contributions**

Conception and design: Ratliff. Acquisition of data: Xu, Li, Swinney. Analysis and interpretation of data: Xu. Drafting the article: Xu, Li. Critically revising the article: all authors. Reviewed submitted version of manuscript: all authors. Approved the final version of the manuscript on behalf of all authors: Ratliff. Statistical analysis: Xu. Study supervision: Ratliff.

**Supplemental Information**

**Previous Presentations**

Portions of this work were presented in abstract form as proceedings at the AANS Annual Scientific Meeting in Chicago, Illinois, on May 2, 2016.

**Correspondence**

John Ratliff, Department of Neurosurgery, Stanford University School of Medicine, 300 Pasteur Dr., Stanford, CA 94305-5327. email: jratliff@stanford.edu.