```
Updates from last version:
* Replaced screenshot with up-to-date version
* Clarified language describing readmission measure to indicate
30 days from discharge
```

Assessing surgeon-level risk of patient harm during elective surgery for public reporting

Olga Pierce, Marshall Allen
ProPublica

Whitepaper as of

August 4, 2015

Olga Pierce and Marshall Allen are journalists at ProPublica, a non-profit organization dedicated to journalism in the public interest. This analysis was done by ProPublica in consultation with Sebastien Haneuse, Karen Joynt and Ashish Jha of the Harvard T. H. Chan School of Public Health, Marty Makary of Johns Hopkins University School of Medicine, dozens of other researchers, surgeons and other practicing physicians, and hundreds of patients who have been harmed while receiving medical care.

Abstract

Background Patients undergoing elective surgical procedures sometimes suffer avoidable, serious complications like infections, blood clots and hemorrhage. However, little is currently known in American medicine about the relative performance of surgeons when it comes to minimizing these types of harm. While practices such as peer review may identify individual egregious errors, there is little infrastructure in place to identify patterns of complications over time, or across separate facilities. Overall, it's estimated that fewer than 1% of surgical outcomes are being measured[1], leaving patients, the medical community, and even surgeons themselves unaware of how their patients fare collectively over time. Most surgeons do not know how their peers outside their own facility perform. Tracking and publicly reporting the relative performance of surgeons could give medical providers, hospital leaders, and regulators a powerful tool – and an incentive – to improve the safety of patients undergoing surgery.

Methods Using administrative data, we identified all Medicare patients who underwent one of eight common elective, low-risk surgical procedures over a 5-year period.We then undertook to fairly compare surgeons and hospitals, without penalizing those with the most complex patients. To do so, we used two measures to identify harm to elective surgical patients: in-hospital mortality, and readmission within 30 days of discharge with a diagnosis identified by experts as a likely complication of surgery. A mixed effects model with hospital and surgeon random effects was then used to risk-adjust each surgeon's raw complication rate.

Results We found that aggregate rates of harm were quite low. None of the procedures had an average death/readmission rate over 5 percent. However, there was substantial variation within hospitals and between surgeons. Contrary to conventional wisdom that there are 'good' and 'bad' hospitals, no hospital performed in the worst quartile by our central measure across all eight of our procedures. Only one hospital was in the best quartile across all procedures. The best-performing surgeons had risk-adjusted rates of harm about 50 percent less than average. The worst performers had risk-adjusted rates as high as three times the average. Often this variation was in-hospital: multiple surgeons performing the same type of procedure at the same hospital with widely divergent rates of harm. About 2000 hospitals (half of those in the data) had top- and bottom-quintile surgeons performing the same procedure. Low-performing surgeons were also unexpectedly dispersed. Two out of three hospitals in the analysis had at least one bottom-quintile performing surgeon. Finally, a comparison of the standard deviations of hospital and surgeon random effects found that surgeon performance accounts for more of the variability of performance between hospitals than hospital-wide performance on a given procedure.

Conclusions There is substantial variation between surgeons in the rates of harm their patients suffer resulting from surgery, which cannot be attributed to patients' health, or differences in hospital overall performance. Identifying positive outlier surgeons can help to identify best practices. Identifying negative outliers offers an opportunity for intervention and improvement.

1 Introduction

Transparency in health care, particularly with regard to safety and quality, is crucial if patients are to make informed decisions about where they seek care, and if providers are to continuously improve the care they provide[2]. To date, however, little information on the relative performance of surgeons performing routine elective procedures is readily accessible to patients and health care providers in the U.S. The goal of this project is to provide patients, and the health care community, with reliable and actionable data points, at both the level of the surgeon and the hospital, in the form of a publicly available online searchable database.

Note: the analysis in this paper covers all entities with NPI numbers, as they appear in Medicare data. The public database presents slightly different figures as it excludes the less common case of organizations with NPI numbers.

1.1 The critical role of surgeons

This analysis comes at a watershed moment for performance reporting: until recently surgeons could not be identified in Medicare data, the most frequently used administrative data. Now surgeon identifiers in the data are unencrypted, and it is possible to attach surgeries – and patient outcomes – to surgeons by name.

While administrative data has been shown to be as reliable as clinical data for some reporting purposes[3], what neither administrative nor clinical data can do is give us 100% certainty about how the complications described in this analysis came about.

But while a physician may not be to blame for the negative outcome of a surgery, s/he can still be considered responsible. The American College of Surgeons adopted a statement in 1996[4] stating that surgeons are responsible for the overall episode of care in each case, from the diagnosis, to the operation, to managing the post-operative care and rehabilitation, including any complications.
“The best interest of the patient is thus optimally served,” reads the statement, “because of the surgeon’s comprehensive knowledge of the patient’s disease and surgical management.”

2 Methods

2.1 Data

We obtained the Medicare 100% Standard Analytic Files for years 2009-2013. The data is administrative, submitted to Medicare by hospitals for billing purposes. Each record in the data represents a patient’s hospital visit, indicating when and where the patient was admitted, what diagnoses were present, which procedures were performed, and who was the operating surgeon. The data, however, conceals the identity of each patient. Because one of our metrics is 30-day readmissions, we excluded December 2013 from our analysis time window.

This data does not represent 100% of patients, only those who are enrolled in fee-for-service Medicare. However, Medicare pays for a substantial portion of all hospital care provided in the United States, and quality measures based on this data are often treated as indicative of the overall standard of care of providers.

2.2 Identifying Patient Cohort

Evaluating surgeon performance from administrative data presents some challenges. Many surgeons (but not all) perform relatively low volumes of surgeries. Harm to patients can be a relatively rare event, which makes it more difficult to statistically model.

Still, comparing surgeons on a level playing field is important. Penalizing surgeons who take on complex cases, or operate under difficult hospital conditions is unfair, and also limits the utility of reported data for patients. For these reasons, we first set out to identify the most uniform patient cohort possible, a strategy that has been successful in previous studies[5].

2.2.1 Identifying procedures for analysis

The surgeries included in our analysis are elective procedures, typically performed on relatively healthy patients. Medical literature shows undergoing them without death or serious complications is a reasonable expectation. They are also some of the most commonly performed procedures in Medicare.

ICD-9 Code	Procedure	N
51.23	Laparoscopic cholecystectomy	201,351
60.5	Radical prostatectomy	78,763
60.29	Transurethral prostatectomy (TURP)	73,752
81.02	Cervical fusion of the anterior column, anterior technique	52,972
81.07	Lumbar and lumbosacral fusion of the posterior column, posterior technique	106,689
81.08	Lumbar and lumbosacral fusion of the anterior column, posterior technique	102,716
81.51	Total hip replacement	494,576
81.54	Total knee replacement	1,190,631
Total		2,376,851

Table 1: Procedures included in analysis and 5-year count of index admissions in our patient cohort

2.2.2 Identifying index admissions

To identify index procedures – the individual surgeries for which we analyzed the outcomes – we used the ICD-9 procedure codes shown in Table 1. We then further standardized the patient cohort by:

- Excluding patients whose cases didn't appear to be elective, either because they were admitted through the emergency room, or because the field indicating elective surgeries in the data was not coded to indicate an elective procedure.
- Choosing only patients who were admitted by a physician or clinician, and excluding those transferred in from another facility, such as a nursing home or correctional facility.
- Choosing only patients who had one of the most common principal diagnoses indicating the procedure. For example, a patient who underwent total knee replacement with a diagnosis of 715.96 - *Osteoarthrosis* would

be included, whereas a patient with a principal diagnosis of 821.23 - *Closed fracture of the femur* would be excluded. (See Appendix A for a full list of included diagnosis codes per procedure.)

- Further filtering principal diagnoses in consultation with experts, who advised us on which codes indicated cases that would be inappropriate to compare to more standard cases.
- Excluding bilateral hip and knee replacements, and knee and hip revision surgeries as opposed to primary replacements.

2.3 Measures

Our goal in identifying complications was to avoid, as much as possible, relying on the type of hospital billing coding that may be inconsistent from one facility to another. Not every hospital codes every field in the same way, so it is important to focus only on the most concrete and reliably consistent fields.

In particular, Present on Admission coding, used often by Medicare and other ratings groups, has been shown to be unreliable for identifying hospital-acquired conditions[6,7]. To this end, we relied on two measures often referred to as 'hard endpoints', because they are not subject to coding interpretation.

We also decided to not include cases in our analysis where patients went to the emergency department but were not admitted to the hospital, because physicians and other experts advised us that these cases often do not represent a serious complication.

Once our index admissions were identified, we used each patient's unique identifier to scan for one of two potential negative outcomes:

Death Patient died during initial hospital stay.

Complication Patient was discharged alive, but was admitted to a hospital within 30 days of discharge with a principal diagnosis indicating a negative surgical outcome as determined by a panel of expert reviewers.

We first identified all cases where a patient died during their initial hospital stay, flagged those cases as deaths, and then excluded them from our readmission analysis. This occurred to 3,470 patients in our analysis.

Then we identified all cases where patients returned to a hospital and were readmitted within 30 days of discharge. If patients were readmitted more than once in 30 days, a rare occurrence, we considered only the first readmission.

When we had identified all readmissions, we compiled the more than 2,000 principal diagnoses coded as causes of readmission. We chose to use the principal diagnosis field only for two reasons. First, since the principal diagnosis is the reason determined at the end of a hospital stay for the stay itself it indicates a level of seriousness of the complication. A 'post-surgical pain' principal diagnosis, for example, indicates that the pain was the cause of the return to the hospital, and not incidental to some other health problem. Hospitals are also largely paid based on the principal diagnosis and thus incorrectly reporting that code could have serious repercussions[8].

For the eight procedures, we consulted a panel of at least five doctors and surgeons, including specialists in each procedure, to review each of the codes. Reviewers were asked to indicate whether each principal diagnosis was likely to be a complication related to the index surgery. For each procedure we were able compile a consensus list of codes. (For a full list of codes determined to indicate surgical complications, see Appendix B.)

We then used the list of codes indicating complications to select only the readmissions that were likely related to surgery, to the exclusion of other readmissions that were either clearly scheduled (like V57.89 - *Care involving other specified rehabilitation procedure* or not related to surgery (like 584.9 - *Acute kidney failure*). Our panels wrestled with some codes, like 486 - *Pneumonia*. While pneumonia can be the result of improperly administered anesthesia, for example, it is also a condition common in elderly patients who have not had surgery at all. In such cases, we chose to give surgeons the benefit of the doubt and did not count this as a complication.

Screening out hospital admissions that did not appear to be surgical complications reduced the total readmissions by more than 50%, as shown in Table 3.

Complications	64,367
Non-complications	76,190
All-cause readmission total	140,557

Table 2: Fewer than half of readmissions were determined to be complications of surgery

2.3.1 Complications

Complication type	N	Example
Infection	13,899	998.59 - *Postoperative infection*
Clot	7,732	415.11 - *Iatrogenic pulmonary embolism*
Reaction	5,164	996.6 - *Infection and inflammatory reaction due to internal joint prosthesis*
Mechanical	4,850	996.47 - *Mechanical complication of prosthetic joint implant*
Sepsis	4,702	03.89 - *Septicemia*
Bone	3,535	996.44 - *Peri-prosthetic fracture around prosthetic joint*
Death	3,470	
Hematoma	3,168	998.12 - *Hematoma complicating a procedure*
Wound	2,793	998.2 - *Accidental puncture or laceration during a procedure*
Hemorrhage	2,698	998.11 - *Hemorrhage complicating a procedure*
Pain	2,169	338.18 - *Acute postoperative pain*
Digestive	1,957	997.49 - *Digestive system complications*
C.diff	1,843	00.845 - *Intestinal infection due to Clostridium difficile*
Misc. Comp.	1,531	787.22 - *Dysphagia, oropharyngeal phase*
Vascular	1,159	997.2 - *Surgical complications of the peripheral vascular system*
Inflammation	931	604.99 - *Orchitis, epididymitis, and epididymo-orchitis, no mention of abscess*
Seroma	673	998.13 - *Seroma complicating a procedure*
Fever	520	780.62 - *Postprocedural fever*
Urinary	486	997.5 - *Surgical complications of the urinary tract*

Table 3: The 20 most frequent types of complication

Any complication that warrants readmission to a hospital can be considered serious. But there is also further information to be gleaned from the hospital visits resulting from complications. The average length of stay was five days, more than one would expect if a patient was simply hospitalized as a precaution. And 910 patients died after readmission for a complication. The total cost to Medicare for the readmission hospital stays alone was $654 million.

2.3.2 Summary of hospital and surgeon unadjusted complication rates

Hospital and surgeon unadjusted rates have similar distributions: a large number of observations clustered at or near zero, with a long right tail.

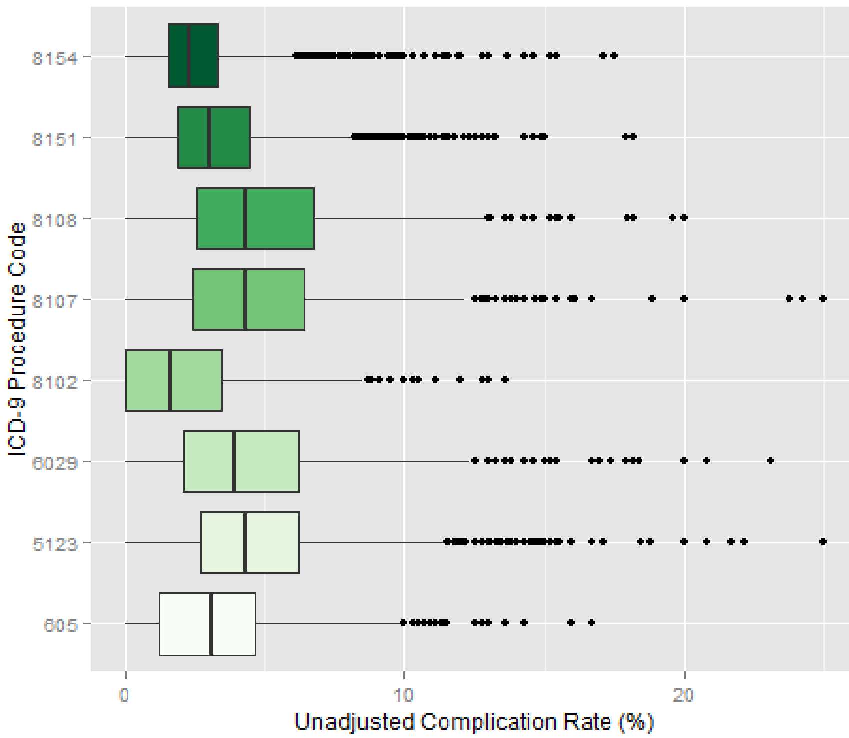


Figure 1: **Hospital Unadjusted Complication Rates**

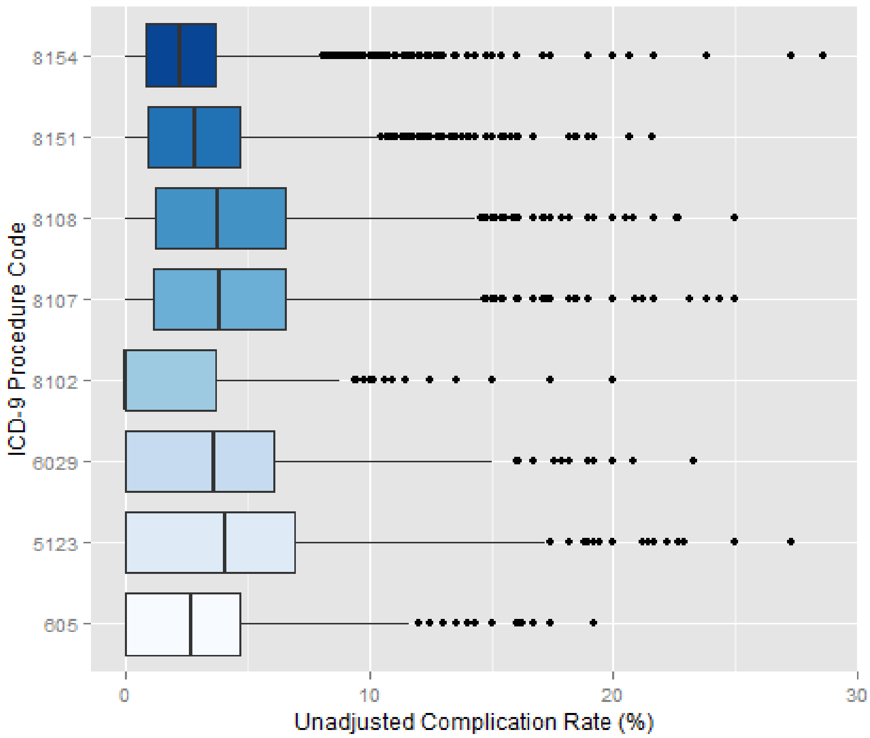


Figure 2: **Surgeon Unadjusted Complication Rates**

2.4 Model results

In order to adjust for differences in case mix we fit a series of logistic-Normal generalized linear mixed models[9], using the lme4 package in R version 3.1.310. This model includes the following fixed effects:

Age The age of the patient at time of surgery in 5-year categories. The Age 65-69 category was reference.

Sex The sex indicated in the patient's record. This variable is excluded for the two prostate procedures.

Health Score A score reflecting risk associated with a patient's comorbidities. This continuous, ratio-scale variable was calculated using VanWalraven's[11] technique to create an index of the Elixhauser comorbidities[12] for each patient.

Multilevel spinal fusion For the spinal fusion surgeries, this dichotomous variable indicates whether or not the procedure was a fusion of more than two vertebrae. Such procedures are considered more complex.

Pancreatitis Diagnosis For the cholecystectomy procedure, this dichotomous variable indicates if the patient was diagnosed with pancreatitis, making the procedure more complex.

Robot-assisted A dichotomous variable indicating whether a radical prostatectomy was performed with the assistance of a surgical robot.

We did not adjust for race, in order to avoid concealing racial disparities in health care. Centers for Medicare and Medicaid Services analyses typically do not risk adjust based on race.

We calculated a risk-adjusted complication rate for each surgeon across all hospitals where s/he performed a particular procedure. Each surgeon and hospital has a unique risk-adjusted rate for each procedure. There is no overall hospital or surgeon score.

The models for each of the procedures, for both hospitals and surgeons, had statistically significant likelihood-ratio test p-values at the 95% significance level relative to their corresponding intercept-only models.

2.4.1 Model results for surgeons

Variable	81.07	81.08	81.02	81.51	81.54	51.23	60.29	60.5
(Intercept)	-3.39*	-3.43*	-4.01*	-3.84*	-3.80*	-3.29*	-3.52*	-3.58*
Age 70-74	0.04	0.15*	0.07	0.21*	0.11*	0.18*	0.21*	0.17*
Age 75-79	0.11*	0.21*	0.26*	0.34*	0.25*	0.27*	0.14.	0.28*
Age 80-84	0.24*	0.32*	0.69*	0.51*	0.41*	0.33*	0.26*	0.07
Age 85+	0.25*	0.17	0.90*	0.66*	0.58*	0.61*	0.65*	1.37*
Sex	0.06	0.11*	-0.53	0.02	-0.28*	-0.10*	NA	NA
Health Score	-0.00	-0.00	-0.00	-0.00	-0.00.	-0.00	0.01.	0.01
Multilevel spinal fusion	0.50*	0.46*	0.56*	NA	NA	NA	NA	NA
Pancreatitis Diagnosis	NA	NA	NA	NA	NA	0.07	NA	NA
Robot-assisted	NA	NA	NA	NA	NA	NA	NA	-0.05
Hospital ranef σ	0.46	0.36	0.55	0.35	0.39	0.28	0.29	0.42
Surgeon ranef σ	0.16	0.26	0.13	0.27	0.25	0.18	0.26	0.14

Table 4: Model Coefficients. (Significance codes for interpreting p-values for the coefficients: levels between 0.0-0.05: *, levels between 0.05-0.10: .)

2.4.2 Model results for hospitals

Variable	81.07	81.08	81.02	81.51	81.54	51.23	60.29	60.5
(Intercept)	-3.28*	-3.31*	-3.97*	-3.77*	-3.76*	-3.21*	-3.48*	-3.49*
Age 70-74	0.04	0.10*	0.15 .	0.21*	0.10*	0.17*	0.21*	0.12*
Age 75-79	0.10*	0.15*	0.32*	0.33*	0.24*	0.23*	0.21*	0.17*
Age 80-84	0.19*	0.25*	0.75*	0.50*	0.39*	0.35*	0.36*	-0.10
Age 85+	0.22*	0.19*	0.78*	0.66*	0.58*	0.59*	0.70*	1.21*
Sex	0.06*	0.10*	-0.51*	0.03	-0.28*	-0.16*	NA	NA
Health Score	0.00	0.00	0.00	0.00	0.00	0.00	0.01*	0.00
Multilevel spinal fusion	0.52*	0.49*	0.58*	NA	NA	NA	NA	NA
Pancreatitis Diagnosis	NA	NA	NA	NA	NA	0.07 .	NA	NA
Robot-assisted	NA	NA	NA	NA	NA	NA	NA	-0.04
Hospital ranef σ	0.30	0.30	0.38	0.33	0.34	0.19	0.29	0.27

Table 5: Model Coefficients for Hospitals. (Significance codes for interpreting p-values for the coefficients: levels between 0.0-0.05: *, levels between 0.05-0.10: .)

2.5 Patient health

A notable feature of our model is that the small but significant effect of the Health Score (per-procedure AUC of .57-.63) essentially disappears when age and the hospital and surgeon effects are included in the model. This shows that the quality of care is likely a more important factor determining patient outcomes.

2.6 Income Sensitivity analysis

A sensitivity analysis was conducted using each hospital's SSI rate as a proxy for patient socioeconomic status. The rate expresses what portion of a hospital's care to Medicare patients goes to patients who qualify for Supplemental Security Income, a federal welfare program which requires very low income and assets for eligibility[13].

$$SSIrate = \frac{MedicareSSIdays}{TotalMedicaredays}$$

Including the additional variable had little effect, perhaps because the metric refers to a hospital's patients as a whole, not the specific group of patients undergoing elective surgical procedures. It it also possible that some of the effect often attributed to patient socioeconomic status is actually due to the effect of the hospitals and surgeons who provide them care.

3 Findings

The main metric to be used for surgeon and hospital comparison is the Adjusted Complication Rate, which is calculated by applying the modeled surgeon random effect to the entire patient population for a procedure in our analysis, presuming that the surgeon is operating at an average hospital (this is achieved by setting the hospital random effect to 0).

3.1 Distributions

Below is a figure illustrating the distribution of surgeon adjusted complication rates for the 81.54 procedure, total knee replacement. The shape is typical of all our procedures. A shoulder-and-neck shaped distribution, where a steep upslope at the far left end of the distribution gives way to a relatively flat section in the middle, which then becomes a steep upslope at the far right end. In this plot, the x-axis is surgeons ranked by adjusted rate, and the y-axis is the Adjusted Complication Rate. The light-gray shaded area around each surgeon's ACR is the 95% confidence interval, based on posterior variance. The dark-gray area around each ACR is the 1σ confidence interval. (For the distributions of the remaining procedures, see Appendix C.)

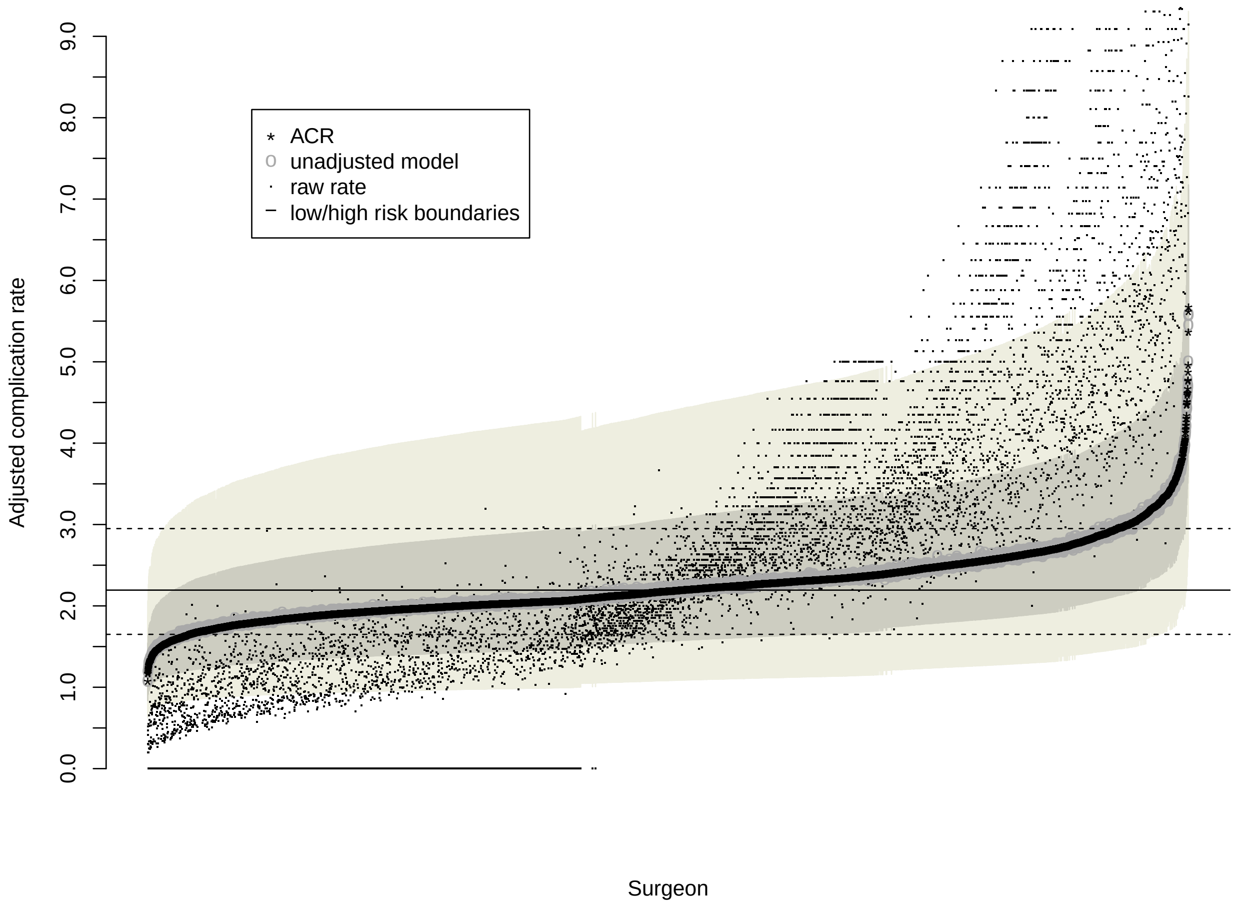


Figure 3: **Adjusted, unadjusted and raw complication rates for each surgeon performing procedure 81.54 total knee replacement. N=9,601 surgeons.**

A key feature of this figure is that it illustrates the high degree of shrinkage that takes place when the raw complication rates are adjusted. While raw rates ranged from 0.0 to 29%, the Adjusted Complication Rate goes from 1.1 to 5.7%. The small difference between the unadjusted and adjusted model results for most surgeons shows that the shrinkage is largely a result of modeling in the first place, not due to adjusting for case mix.

This shrinkage is another piece of the measured approach we are taking: we are taking care not to unfairly characterize surgeons and hospitals.

There were 1,983 surgeons who had no complications at all on this procedure, some performing more than 100 surgeries. Still, the model adjusts their rate up to 1.1% or more. This is a result of the thought experiment implicit in the analysis: What would happen if a given surgeon operated on all of the 1.2 million patients in our data who underwent a knee replacement? The Adjusted Complication Rate does not directly represent a surgeon's past outcomes. It is an assessment of how s/he would perform at a hypothetical average hospital, on a standardized patient pool.

Procedure	Low-risk <=	Medium-risk<>	High-risk >=	Min	Max	Average
51.23	3.9	4.0 - 5.1	5.2	3.5	6.4	4.4
60.5	2.3	2.4 - 3.7	3.8	1.8	7.6	2.9
60.29	3.4	3.5 - 4.4	4.5	3.3	5.3	3.8
81.02	1.4	1.5 - 2.3	2.4	1.2	6.3	1.9
81.07	3.0	3.1 - 4.9	5.0	2.2	12.2	4.1
81.08	3.5	3.6 - 4.9	5.0	2.8	7.5	4.1
81.51	2.2	2.3 - 3.4	3.5	1.4	5.6	2.8
81.54	1.6	1.7 - 2.9	3.0	1.1	5.7	2.3

Table 6: Procedure and ACR summary. Low-risk and high-risk cut points are determined by the slope of the procedure's ACR distribution.

Procedure	Total surgeons	With ACR	Low-risk	Medium-risk	High-risk	High %
51.23	21,479	3,068	143	2,734	191	6.2%
60.5	5,093	1,038	69	881	88	8.5%
60.29	7,898	1,079	19	1,031	29	2.7%
81.02	5,624	728	41	598	89	12.2%
81.07	6,214	1,815	120	1,414	281	15.5%
81.08	6,136	1,758	192	1,390	176	10.0%
81.51	13,414	5,790	219	5,076	495	8.5%
81.54	18,029	9,601	358	8,560	683	7.1%

Table 7: Number of surgeons in each risk category by procedure

3.2 In-hospital variation

It was extremely rare for a hospital to be high- or low-performing in each of our eight procedures. In fact this happened at only a handful of the 299 hospitals that met the volume requirement in all eight procedures. This suggests that a given hospital's performance can vary greatly depending on the surgery.

But not only was there great variation between hospital departments, there was a great deal of variation between surgeons in the same hospital departments. At 1,567 hospitals, there were top-quintile and bottom-quintile performers doing the same procedure.

3.2.1 Comparing random effects

One way to evaluate contributions of surgeon variation and hospital variation to overall hospital performance is to rerun the hospital risk-adjustment model above, but this time adding a surgeon random effect. The relative sizes of the random effects standard deviations, or σ, give us an indication of how much of the overall variation between hospitals' performance in each procedure is attributable to hospital-wide characteristics, and how much is attributable to the surgeons themselves.

Below is a table of the hospital and surgeon random effects σ for each procedure, all significant at 95% confidence level. In every case, the surgeon σ is greater than the hospital σ, in some cases much larger. In the case of procedure 60.5, for example, the surgeon σ is more than four times the hospital σ. This means that overall, hospitals (as isolated from the surgeons who perform surgeries there) are relatively similar. When hospitals have different adjusted complication rates, it is mainly due to variation in surgeon performance.

Likewise, a patient's choice of hospital will, in general, have less impact on his or her risk of readmission or death than choice of surgeon.

	81.07	81.08	81.02	81.51	81.54	51.23	60.59	60.5
Surgeon ranef σ	.51	.39	.68	.39	.4	.29	.3	.42
Hospital ranef σ	.13	.21	.23	.24	.26	.16	.25	.1

Table 8: The standard deviations of the surgeon and hospital random effects (in log-odds) from the hospital model above, with a surgeon random effect added

3.3 Limitations

We have made every effort to level the playing field for proper comparison of surgeons, but there is information that administrative data simply cannot tell us. Coding can be idiosyncratic between hospitals, and some hospitals clearly code many more comorbidities than others. No data can really take us inside the operating room to give us full detail of what happened to a patient and why. While we believe our conclusions offer generalizable insight into variations in performance at hospitals, our conclusions are based on a subset of surgeries, and are limited by access to data only reflecting the portion of a surgeon's patients who are age 65 and older and enrolled in fee-for-service Medicare.

3.4 The critical importance of access to state data

Data is kept at the state level that includes patients of all ages, regardless of their insurance status. This data could be used to offer groundbreaking insights into the important question of how to keep patients safe at hospitals. Yet this data is restricted. In some states this data is not made available for research at all, sometimes because it is under control of the state hospital association. In other states, the price is thousands or tens of thousands of dollars for each year of data, a prohibitive cost for most researchers. In most states, the data is so redacted that analysis of the type seen in this paper is impossible. Providing an encrypted unique patient identifier, as Medicare does, faciliates an important longitudinal look at the effects of medical care on a patient. Finally, only a handful of states include physician names in the data. In all other states that provide all-payer data, researchers may manage to identify poorly performing – or even dangerous – surgeons or other physicians but cannot know who they are, and thus help them improve, and inform the public.

4 Discussion

In low-risk, elective surgeries there are surgeons whose patients experience complications at much higher rates than others. This variation can't be explained by measurable differences in case mix, or the performance of hospitals in which the surgeon performs procedures. While we can't conclude from this data why exactly we see this variation, there are some indications:

Skill A University of Michigan study found that a surgeon's skill, as assessed by peers, was by far the most predictive of complications after bariatric surgery[14].

Adherence to best practice Studies have found evidence that surgeons sometimes do not adhere to best practice[15].

Extremely low volume Some doctors in our data appeared to have extremely low surgical volumes. Research indicates this can translate into higher complication rates. While some of these low procedure counts may simply reflect surgeons who rarely operate on Medicare patients, we were able to confirm using state data that some genuinely perform only a few procedures per year[16].

Active involvement in entire episode of care As journalists, our information-gathering process differs from that of more traditional academic data analysts in important ways. When there is an apparent pattern, we supplement our data findings as much as possible with reporting. In case studies of high- and low-performing surgeons by our measures, another pattern emerged, albeit anecdotally: surgeons with extremely low rates of complications expressed a sense of personal responsibility toward patients, in the lead-up to surgery, in the recovery period immediately after, and in the long-term, far after the patient exits the hospital doors. Patients who suffered complications after procedures performed by surgeons with high rates of complications often described limited interactions with their surgeon, especially after surgery. Colleagues of these surgeons often confirmed this phenomenon.

4.1 On Transparency

The evidence in this analysis of extensive in-hospital variation in surgeon performance raises questions about whether reporting outcomes simply at the hospital level provides enough transparency. Surgeon-level data about patient outcomes:

- Can inform the practice of surgeons who want to improve their performance.
- Allows appropriate intervention by administrators, quality improvement coordinators and peer review committees at hospitals.
- Supplements complaint-based processes, like those of state medical boards and specialty boards, which have the important job of protecting the public.
- Provides patients with information they need to make informed decisions.

5 Conclusion

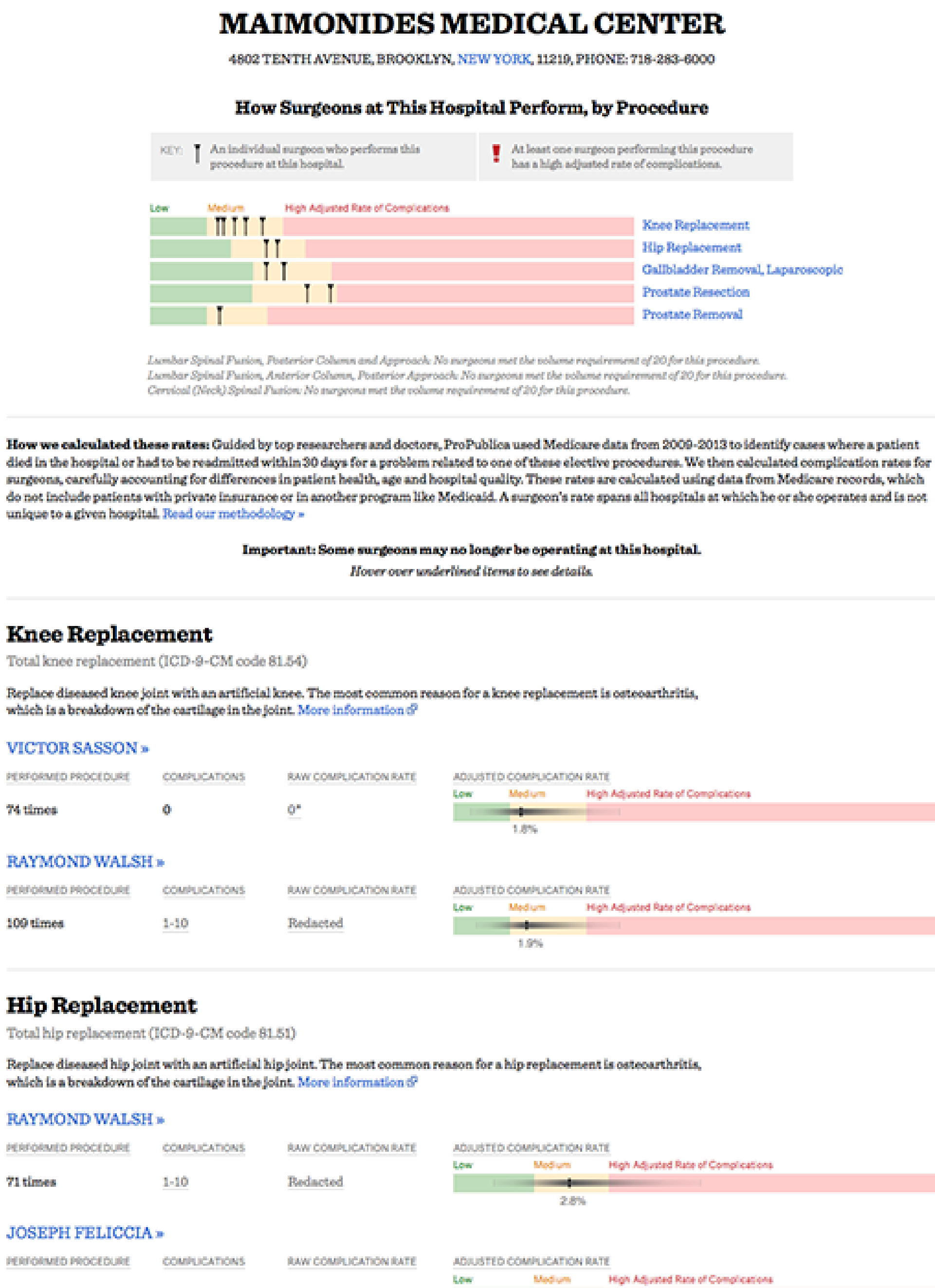
We analyzed 2,376,851 instances of eight common, low-risk elective procedures from five years of administrative data. We then calculated the risk-adjusted rates of death and surgery-related readmission of hospitals and surgeons. Surgeon performance on this measure shows substantial variation, even between surgeons operating at the same hospital. Once the effect of surgeons is isolated, the performance of hospitals is relatively similar. While most current reporting is focused at the hospital level, there are important benefits to reporting adjusted surgeon-level complication rates.

6 Reporting

We will be reporting results for every surgeon and hospital that appears in our data for a given procedure in an online searchable database that is open and free to the public. We are, however, constrained by patient privacy protections in our data use agreement with CMS. Hospitals and surgeons that did fewer than 20 of a procedure in our data window will be denoted "low volume." If surgeons or hospitals did 20 or more of a procedure and had greater than 0, but fewer than 11 complications, we will report their ACR,confidence interval and procedure count. If surgeons or hospitals did 20 or more of a procedure and had 11 or more complications, or zero complications, we will also report their raw complication count and rate.

Below is a recent screen capture of sample results from an imaginary hospital to give the reader a sense of what the database may look like. The bars at the top show the Adjusted Complication Rate for each of the surgeons at said hospital across all procedures. The green, yellow, and red areas on these bars are determined by the shape of the distribution of adjusted surgeon complication rates for each procedure.

Underneath the hospital-level summary are surgeon names accompanied by their adjusted complication rates and 95% confidence intervals. The intensity of the gray color in each interval reflects the distribution of probability within the interval: most intense around the point estimate, fainter as the area under the probability curve becomes smaller approaching the bounds of the interval.
In a separate part of the database, we will allow patients to determine the best combination of doctor and hospital in their geographic area.



MAIMONIDES MEDICAL CENTER

4802 TENTH AVENUE, BROOKLYN, NEW YORK, 11219, PHONE: 718-283-6000

How Surgeons at This Hospital Perform, by Procedure

KEY: An individual surgeon who performs this procedure at this hospital.
At least one surgeon performing this procedure has a high adjusted rate of complications.

Low Medium High Adjusted Rate of Complications

Knee Replacement
Hip Replacement
Gallbladder Removal, Laparoscopic
Prostate Resection
Prostate Removal

Lumbar Spinal Fusion, Posterior Column and Approach: No surgeons met the volume requirement of 20 for this procedure.
Lumbar Spinal Fusion, Anterior Column, Posterior Approach: No surgeons met the volume requirement of 20 for this procedure.
Cervical (Neck) Spinal Fusion: No surgeons met the volume requirement of 20 for this procedure.

How we calculated these rates: Guided by top researchers and doctors, ProPublica used Medicare data from 2009-2013 to identify cases where a patient died in the hospital or had to be readmitted within 30 days for a problem related to one of these elective procedures. We then calculated complication rates for surgeons, carefully accounting for differences in patient health, age and hospital quality. These rates are calculated using data from Medicare records, which do not include patients with private insurance or in another program like Medicaid. A surgeon's rate spans all hospitals at which he or she operates and is not unique to a given hospital. Read our methodology »

Important: Some surgeons may no longer be operating at this hospital.
Hover over underlined items to see details.

Knee Replacement

Total knee replacement (ICD-9-CM code 81.54)

Replace diseased knee joint with an artificial knee. The most common reason for a knee replacement is osteoarthritis, which is a breakdown of the cartilage in the joint. More information

VICTOR SASSON »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE
74 times	0	0*	1.8%

RAYMOND WALSH »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE
109 times	1-10	Redacted	1.9%

Hip Replacement

Total hip replacement (ICD-9-CM code 81.51)

Replace diseased hip joint with an artificial hip joint. The most common reason for a hip replacement is osteoarthritis, which is a breakdown of the cartilage in the joint. More information

RAYMOND WALSH »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE
71 times	1-10	Redacted	2.8%

JOSEPH FELICCIA »

PERFORMED PROCEDURE	COMPLICATIONS	RAW COMPLICATION RATE	ADJUSTED COMPLICATION RATE

Figure 4: **Reporting for a sample hospital and surgeon. Live website at projects.propublica.org/surgeons/**

7 References

1. Lyu, H., Cooper, M., Patel, K., Daniel, M., & Makary, M. A. (2015). Prevalence and Data Transparency of National Clinical Registries in the United States.*Journal for Healthcare Quality.* [Epub ahead of print.]

2. National Patient Safety Foundation's Lucian Leape Institute. *Shining a Light: Safer Health Care Through Transparency.* Boston, MA: National Patient Safety Foundation; 2015.http://www.npsf.org/transparency.

3. Krumholz, H. M., Lin, Z., Drye, E. E., Desai, M. M., Han, L. F., Rapp, M. T., ... & Normand, S. L. T. (2011). An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Circulation: Cardiovascular Quality and Outcomes*, 4(2), 243-252.

4. American College of Surgeons. Statement adopted February 1996. Available at: https://www.facs.org/about-acs/statements/25-perioperative .

5. Dimick, J. B., Staiger, D. O., Baser, O., & Birkmeyer, J. D. (2009). Composite measures for predicting surgical mortality in the hospital. *Health Affairs*, 28(4), 1189-1198.

6. Khanna, R. R., Kim, S. B., Jenkins, I., El-Kareh, R., Afsarmanesh, N., Amin, A., ... & White, R. H. (2015). Predictive value of the present-on-admission indicator for hospital-acquired venous thromboembolism. *Medical care*, 53(4), e31-e36.

7. Goldman, L. E., Chu, P. W., Osmond, D., & Bindman, A. (2011). The Accuracy of Present on Admission Reporting in Administrative Data. *Health services research*, 46(6pt1), 1946-1962.

8. Centers for Medicare and Medicaid Services. Medicare Claims Processing Manual Chapter 23 - Fee Schedule Administration and Coding Requirements, page 1-2. Available at: http://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/downloads/clm104c23.pdf

9. Normand, S.-L. T., Glickman, M. E., and Gatsonis, C. A. (1997). Statistical methods for profiling providers of medical care: issues and applications. *JAMA*, 92(439):803–814.

10. Bates D, Maechler M, Bolker B and Walker S (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, ¡URL: http://CRAN.R-project.org/package=lme4¿.

11. van Walraven, C., Austin, P. C., Jennings, A., Quan, H., & Forster, A. J. (2009). A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. *Medical care*, 626-633.

12. Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical care*, 36(1), 8-27.

13. Centers for Medicare and Medicaid Services. http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/dsh.html Accessed Jun. 17, 2015.

14. Birkmeyer, J. D., Finks, J. F., O'Reilly, A., Oerline, M., Carlin, A. M., Nunn, A. R., ... & Birkmeyer, N. J. (2013). Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine*, 369(15), 1434-1442.

15. Merkow, R. P., Bilimoria, K. Y., Sohn, M. W., Oh, E. H., Sellers, M. M., Paruch, J. L., ... & Bentrem, D. J. (2014). Adherence with postdischarge venous thromboembolism chemoprophylaxis recommendations after colorectal cancer surgery among elderly Medicare beneficiaries. *Annals of surgery*,260(1), 103-108.

16. Lau, R. L., Perruccio, A. V., Gandhi, R., & Mahomed, N. N. (2012). The role of surgeon volume on patient outcome in total knee arthroplasty: a systematic review of the literature. *BMC musculoskeletal disorders*, 13(1), 250.