



Published in final edited form as:

Surgery. 2016 February ; 159(2): 375–380. doi:10.1016/j.surg.2015.06.043.

Reliability of Hospital Cost Profiles in Inpatient Surgery

Tyler R. Grenda, MD^{1,2}, Robert W. Krell, MD^{1,2}, and Justin B. Dimick, MD, MPH^{1,2}

¹Department of Surgery, University of Michigan, Ann Arbor, MI

²Center for Healthcare Outcomes and Policy, University of Michigan, Ann Arbor, MI

Abstract

Background—With increased policy emphasis on shifting risk from payers to providers through mechanisms such as bundled payments and Accountable Care Organizations, hospitals are increasingly in need of metrics to understand their costs relative to peers. However, it is unclear whether Medicare payments for surgery can reliably compare hospital costs.

Methods—We used national Medicare data to assess patients undergoing colectomy, pancreatectomy, and open incisional hernia repair from 2009–2010 (n= 339,882 patients). We first calculated risk-adjusted hospital total episode payments for each procedure. We then used hierarchical modeling techniques to estimate the reliability of total episode payments for each procedure and explored the impact of hospital caseload on payment reliability. Finally, we quantified the number of hospitals meeting published reliability benchmarks.

Results—Mean risk-adjusted total episode payments ranged from \$13,262 (standard deviation [SD] \$14,523) for incisional hernia repair to \$25,055 (SD \$22,549) for pancreatectomy. The reliability of hospital episode payments varied widely across procedures and depended on sample size. For example, mean episode payment reliability for colectomy (mean caseload: 157) was 0.80 (SD 0.18), while for pancreatectomy (mean caseload: 13) mean reliability was 0.45 (SD 0.27). Many hospitals met published reliability benchmarks for each procedure. For example, 90% of hospitals met reliability benchmarks for colectomy, 40% for pancreatectomy, and 66% for incisional hernia repair.

Conclusions—Episode payments for inpatient surgery are a reliable measure of hospital costs for commonly performed procedures, but are less reliable for lower volume operations. These findings suggest that hospital cost profiles based on Medicare claims data may be used to benchmark efficiency especially for more common procedures.

INTRODUCTION

In an effort to control healthcare costs in the United States, there has been increasing emphasis on making providers more accountable for quality and costs. In the current policy environment, as providers accept risk and shift to hospital-based surgical practices, they are increasingly in need of metrics to understand their costs relative to peers. These metrics

Corresponding Author: Tyler R. Grenda, MD, Center for Healthcare Outcomes and Policy, University of Michigan, 2800 Plymouth Road, Building 16, Room 016-100N-12, Ann Arbor, MI 48109-2800, (O): 734-763-0147, (F): 734-998-7473, tgrenda@med.umich.edu.

could assist hospitals in two specific ways. First, they could help hospitals identify areas where they have high costs due to poor quality, as there is a well-established link between poor quality (e.g. surgical complications) and costs.¹ Absent of quality problems, hospitals could use these metrics to identify areas of inefficiency where they utilize more resources than other hospitals to achieve similar outcomes.

Nonetheless, it is unclear whether Medicare payments for surgery can be used to compare hospital costs in this regard. One key metric to assess performance may be the reliability of estimates of payments around an episode of surgery. An episode includes all inpatient care, as well as readmissions, related to the surgical procedure within a defined period of time around the index operation. The concept of reliability is similar to statistical power in clinical trials, reflecting how confidently one can discriminate between providers. It has been recognized that many clinical outcomes have low reliability due to small sample sizes and rare clinical event rates.^{2, 3} While the reliability of outcome measures such as 30-day mortality and complications for surgical procedures have been investigated, the reliability of total episode payments, a measure that is associated with each episode of care, has not yet been explored.³⁻⁶

A better understanding of surgical episode payment reliability would provide additional information on how confidently this proposed quality metric could be used to compare providers, establishing an additional metric for assessing episode efficiency. This has important implications in helping Accountable Care Organizations (ACO) and hospitals interested in bundled payments understand potential liabilities and opportunities for improvement under these newer payment strategies. In this context, we used national Medicare data to explore the reliability of risk-adjusted total episode payments for 3 major surgical procedures. After profiling hospital surgical episode payments, we assessed the impact of hospital case volume on payment reliability. We then compared hospital surgical episode reliability levels to commonly accepted benchmarks used for other outcomes.

METHODS

Data source and study population

For the present study, we used 2009–2010 Medicare Provider Analysis and Review files, which include claims for services provided to Medicare beneficiaries admitted to certified inpatient hospitals. Using relevant *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM codes), we identified adults aged 65–99 undergoing inpatient colectomy, pancreatectomy, and open incisional hernia repair procedures. We selected these procedures because they are common in the Medicare population, display marked variation in hospital caseloads, and have potential for significant expenses related to perioperative care. To limit hospitals with extremely low caseloads, we excluded hospitals with less than or equal to the bottom 1% of caseload for each procedure.

Outcomes

Our primary outcome was risk-adjusted 30-day total episode hospital payments. To define total payments, we linked claims from Centers for Medicare and Medicaid Services (CMS)

files relevant to the particular surgical episode to each patient's records. Our evaluation included payments for all services from the date of hospital admission for the index procedure to 30 days from the date of discharge from the hospital. Using established methods, we price-standardized all payments, adjusting for regional differences in payments for Medicare services.⁷ To assess total episode payments, we estimated price-standardized facility-related Medicare payments for each patient, encompassing diagnosis-related group (DRG) payments, readmissions within 30 days of discharge, and outlier payments when applicable.⁸

Analysis

We first calculated hospital risk-adjusted total episode payments for each procedure, using linear regression models with the log-transformed price-standardized 30-day payment as the dependent variable. All risk-adjustment models included patient age, sex, race, urgency of operation, median ZIP-code income and 29 patient comorbidities identified using methods defined by Elixhauser et al. as covariates.^{9, 10} Hospital risk-adjusted total episode payments were then estimated by exponentiation of the mean log-transformed hospital adjusted payment.

Calculation of Reliability

Conceptually, reliability is analogous to power calculations in clinical trials used to avoid type II error (failing to detect a 'true' difference between groups). Mathematically, it is a quantification of an outcome measure's signal-to-noise ratio, where "signal" is variation due to true differences for a specific measure, and "noise" is variation attributable to measurement error. Since reliability is calculated using a signal-to-noise ratio, minimizing noise will result in a higher proportion of signal (i.e. higher reliability), or "true" differences in an outcome such as episode payments in this study. Reliability is measured on a scale of 0 to 1, with 1 indicating perfect reliability.¹¹

To estimate hospital payment reliability, we used hierarchical linear regression models, accounting for hospital-level random effects, with log-transformed standardized payments as the dependent variable. Hierarchical regression is a statistical method that utilizes empirical Bayes theorem to model variation at the hospital level. This method has been previously used to estimate the reliability of other surgical outcomes.³ We estimated "signal" using the hospital-level random intercept variance (i.e. variation in payments at the hospital level) in the hierarchical model after adjusting for case-mix. We estimated each hospital's "noise" by calculating the standard error of its predicted payments. We then calculated each hospital's reliability as [signal/(signal + noise)].

While outcome reliability is heavily influenced by sample size, in this case hospital caseload, higher reliability does not necessarily correlate with quality.^{2, 3} Rather, it conveys how confidently a provider can compare its outcome measures to peers, and thus determine their position relative to standards or quality benchmarks, depending on the outcome being assessed. To explore the impact of hospital caseload on payment reliability, we grouped hospitals into quartiles based on their caseload over the 2-year period. We then compared the reliability of total episode payments across quartiles for each procedure. Finally, we

quantified the proportion of hospitals meeting published reliability thresholds (0.7 and 0.5) for each procedure separately.^{12, 13}

All statistical tests are two-sided with a p-value of <0.05 considered significant. All statistical analyses were performed using STATA release 13 (StataCorp, College Station, TX). This study was judged to be exempt from human subject review by the University of Michigan Institutional Review Board.

RESULTS

Our study cohort included 339,882 patients in 3,434 hospitals who underwent pancreatectomy, colectomy, or incisional hernia repair. Patient characteristics as well as risk-adjusted total payments are shown in Table 1. Caseload varied widely across procedures with a mean of 157 cases for colectomy, 13 for pancreatectomy, and 83 for incisional hernia repair. The mean risk-adjusted total payment for colectomy was \$20,229 (standard deviation [SD], \$18,337), \$13,262 (SD \$14,523) for incisional hernia repair, and \$25,055 (SD \$22,549) for pancreatectomy.

Figure 1 shows procedure-specific total payment reliability as well as the overall range of hospital reliability in relation to common reliability benchmarks. Both overall payment reliability and the range of hospital reliability depended upon the procedure. For the most common procedure (colectomy), the median reliability level for all hospitals was highest (0.87, interquartile range [IQR] 0.58–0.96) and the range of hospital reliability was narrowest (Figure 1). In contrast, the median reliability for the least common procedure (pancreatectomy) was lowest (0.41, IQR 0.22–0.77) and range of hospital reliability levels broadest.

Table 2 shows total payment reliability levels for each procedure, stratified by quartiles of hospital caseloads. For each procedure, hospitals with higher caseloads had higher reliability levels for total episode payments. For example, the mean reliability of episode payments for colectomy ranged from 0.58 (SD 0.13) for hospitals in the lowest caseload quartile to 0.96 (SD 0.01) for hospitals in the highest caseload quartile. There were similar trends when assessing payment reliability levels across different hospital caseloads for pancreatectomy and incisional hernia repair, though the absolute payment reliability levels were lower than those for colectomy (Table 2).

Figure 2 shows the proportion of hospitals meeting common reliability benchmarks (0.5 and 0.7) for each procedure.^{12, 13} Most hospitals (74.9%) met a reliability benchmark of 0.7 (i.e. considered excellent) for colectomy. Fewer hospitals met the same benchmark for incisional hernia repair (44.4%) or pancreatectomy (21.8%). The same trend was observed when using a reliability benchmark of 0.5.

DISCUSSION

We found that episode payments for inpatient surgery are a reliable measure of relative costs for many hospitals. As detailed in this study, reliability for many hospitals met commonly accepted thresholds. Furthermore, hospital cost profiles may have higher reliability levels

than other commonly reported clinical outcomes such as mortality or complications. These findings suggest that using Medicare claims data to create hospital cost profiles would allow providers to reliably compare their costs relative to other hospitals, identifying episodes of similar surgical care that are inefficient relative to peers. This will be of particular value as hospitals attempt to understand their liabilities and opportunities for improvement under new policies that shift financial risk to providers (i.e. bundled payments and Accountable Care Organizations).

Most prior work assessing the reliability of quality measures has focused on clinical outcomes. These studies have demonstrated that clinical outcomes generally have low reliability largely due to small caseloads (i.e. sample size) at most hospitals and low rates of most adverse events.^{2, 3} As a result, an emphasis has been placed on considering reliability when selecting outcome measures.^{4, 5, 14–16} In contrast to other clinical outcome measures, we have demonstrated that episode payments are more reliable and many hospitals meet commonly accepted reliability benchmarks, even for less common procedures such as pancreatectomy. This is due to the fact that every hospital admission has an associated episode payment, whereas adverse clinical outcomes, (i.e. morbidity or mortality) occur in only a small subset of admissions. While many hospitals meet reliability levels, several will not achieve these benchmarks due to low case volume. A novel method that uses advanced modeling techniques to adjust for measurement error (i.e. reliability-adjustment) would be one solution for addressing measurement problems for less common procedures. These are similar to approaches used by Centers for Medicare and Medicaid Services and their Hospital Compare public reporting of mortality and readmissions.

While the reliability of physician cost profiles has previously been evaluated across several different medical diagnoses, this has remained unexplored in major surgical procedures.¹² Adams and colleagues evaluated cost profile reliability for select surgical specialties (e.g. vascular surgery), but did not investigate specific surgical procedures. Our study goes beyond this prior work by providing an analysis aimed at identifying the reliability of individual surgical procedures. Because most policies are constructed to evaluate specific procedures (e.g. hip replacement, coronary artery bypass grafting), measures of episode costs at the procedure-level are essential in the current policy environment for helping hospitals identify areas of risk and opportunities for improvement. Furthermore, this metric has the ability to justify comparisons at multiple levels, including individual providers. However, this may be limited by case volume at the physician level, as many providers may not have large enough caseloads to produce a reliable outcome measure.

This study has important limitations. First, our study was limited to Medicare patients, which does not represent the entire population of patients undergoing these surgical procedures. Nonetheless, many of the policies for which hospitals will be rated are designed for Medicare patients. Thus, the reliability of these metrics is important in the context of these policies. Second, we used administrative data, which has recognized limitations with capturing patient comorbidities and risk factors. However, the aim of this study was specifically to evaluate the reliability of these metrics. Since reliability is primarily a function of provider case volume and variation across hospitals, the details of risk-adjustment are much less important in the context of this specific analysis. We do recognize

that when episode payments are used in the policy context, it will be important to optimize risk-adjustment profiles to capture and adjust for different levels of severity of illness.

These cost metrics would be particularly useful given the current changes in the policy landscape. The Patient Protection and Affordable Care Act (PPACA) includes several policy changes that shift financial risk from payers to providers.^{17, 18} For example, the Center for Medicare and Medicaid Innovation has launched a pilot program for bundled payments where hospitals accept a fixed price for all hospital services regardless of whether they experience complications and require additional resources.¹⁹ Accountable Care Organizations (ACO) advance this model in which the ACO accepts a degree of financial responsibility for care, in exchange for potential financial incentives for achieving quality standards and savings benchmarks through the Shared Savings Program.^{20, 21} As a result, there will be a demand for metrics that can assist hospitals in adapting to these key policy changes.

This study has important implications given recent changes in healthcare policy. Both Medicare and the private sector are redesigning payment systems to shift risk to providers as evidenced by Medicare's Shared Savings Program, which creates Accountable Care Organizations that are responsible for managing healthcare costs of beneficiaries, and the Center for Medicare and Medicaid Innovation's (CMMI) bundled payments program.²²⁻²⁴ Under both of these programs, risk for episodes of care with associated high costs will shift from the payer to the provider. In this environment, hospitals could use cost metrics directly in two ways: They can identify areas of high cost due to poor quality care, as well as, highlight areas of inefficiency where they use more resources than other providers to achieve the similar outcomes. Thus, the use of these metrics will facilitate quality improvement in areas of greatest liability and recognition of what their most efficient processes are for emulation. Because these policies aim to encourage the delivery of high quality and low cost care, benefits of these programs will only be realized by those providers that meet quality and savings benchmarks. Measures of episode costs that identify areas for improvement will be paramount to assist those participating in these programs to capitalize on the potential benefits of these policies.

SUMMARY

The findings of our study demonstrate that Medicare payments around an episode of surgery provide a reliable measure of relative financial performance, especially among common procedures, such as colectomy. Episode payments are less reliable for lower volume procedures, such as pancreatectomy. These metrics will be important for surgical leaders to understand how their hospital performs relative to peers in the current policy environment, where financial risk is shifting to providers through mechanisms such as ACOs and bundled payments.

Acknowledgments

Disclosures: Dr. Justin Dimick is a consultant and equity owner of ArborMetrix, Inc. – an Ann Arbor-based healthcare analytics and information technology firm. ArborMetrix, Inc. was not involved, in whole or in part, in the collection or analysis of any data presented herein. Dr. Robert W Krell received payment from Blue Cross/Blue Shield of Michigan for data entry unrelated to the submitted work.

Funding: Tyler R. Grenda, MD is supported by the Agency for Healthcare Research and Quality grant 2T32HS000053. The funding source had no role in the design or conduct of the study, or the acquisition, analysis, or interpretation of the data; or in the drafting or review of the manuscript.

REFERENCES

1. Birkmeyer JD, Gust C, Dimick JB, et al. Hospital quality and the cost of inpatient surgery in the United States. *Ann Surg.* 2012; 255(1):1–5. [PubMed: 22156928]
2. Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality: the problem with small sample size. *JAMA.* 2004; 292:847–851. [PubMed: 15315999]
3. Krell RW, Hozain A, Kao LS, et al. Reliability of risk-adjusted outcomes for profiling hospital surgical quality. *JAMA Surg.* 2014; 149:467–474. [PubMed: 24623045]
4. Dimick JB, Staiger DO, Birkmeyer JD. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res.* 2010; 45:1614–1629. [PubMed: 20722747]
5. Hashmi ZG, Dimick JB, Efron DT, et al. Reliability adjustment: a necessity for trauma center ranking and benchmarking. *J Trauma Acute Care Surg.* 2013; 75:166–172. [PubMed: 23940864]
6. Kao LS, Ghaferi AA, Ko CY, et al. Reliability of superficial surgical site infections as a hospital quality measure. *J Am Coll Surg.* 2011; 213:231–235. [PubMed: 21622011]
7. Gottlieb DJ, Zhou W, Song Y, et al. Prices don't drive regional Medicare spending variations. *Health Affairs.* 2010; 29:537–543. [PubMed: 20110290]
8. Birkmeyer JD, Gust C, Baser O, et al. Medicare payments for common inpatient procedures: implications for episode-based payment bundling. *Health Serv Res.* 2010; 45:1783–1795. [PubMed: 20698899]
9. Elixhauser A, Steiner C, Harris DR, et al. Comorbidity measures for use with administrative data. *Med Care.* 1998; 36:8–27. [PubMed: 9431328]
10. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care.* 2004; 42:355–360. [PubMed: 15076812]
11. Adams JL. *The Reliability of Provider Profiling: A Tutorial.* Santa Monica, CA: RAND Corporation; 2009.
12. Adams JL, Mehrotra A, Thomas JW, et al. Physician Cost Profiling — Reliability and Risk of Misclassification. *N Engl J Med.* 2010; 362:1014–1021. [PubMed: 20237347]
13. Scholle SH, Roski J, Adams JL, et al. Benchmarking physician performance: reliability of individual and composite measures. *Am J Manag Care.* 2008; 14:833–838. [PubMed: 19067500]
14. Dimick JB, Ghaferi AA, Osborne NH, et al. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg.* 2012; 255:703–707. [PubMed: 22388108]
15. Osborne NH, Ko CY, Upchurch GR Jr, et al. The impact of adjusting for reliability on hospital quality rankings in vascular surgery. *J Vasc Surg.* 2011; 53:1–5. [PubMed: 21093202]
16. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg.* 2013; 217:336–346. [PubMed: 23628227]
17. Kocher RP, Adashi EY. Hospital Readmissions and the Affordable Care Act. *JAMA.* 2011; 306:1794–1795. [PubMed: 22028355]
18. [Accessed December 18, 2013] Patient Protection and Affordable Care Act. Available at: <http://www.hhs.gov/healthcare/rights/law/index.html>.
19. Chernew M. Bundled payment systems: can they be more successful this time. *Health Serv Res.* 2010; 45:1141–1147. [PubMed: 20831714]
20. Fisher ES, Shortell SM. Accountable care organizations. *JAMA.* 2010; 304:1715–1716. [PubMed: 20959584]
21. Berwick DM. Launching Accountable Care Organizations — The Proposed Rule for the Medicare Shared Savings Program. *N Engl J Med.* 2011; 364:32.
22. Bennett AR. Accountable care organizations: principles and implications for hospital administrators. *J Healthc Manag.* 2012; 57:244–254. [PubMed: 22905603]

23. Ginsburg PB. Spending to save--ACOs and the Medicare Shared Savings Program. *N Engl J Med.* 2011; 364:2085–2086. [PubMed: 21612463]
24. [Accessed April 2, 2014] Medicare Shared Savings Program. Available at: <http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/index.html?redirect>.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

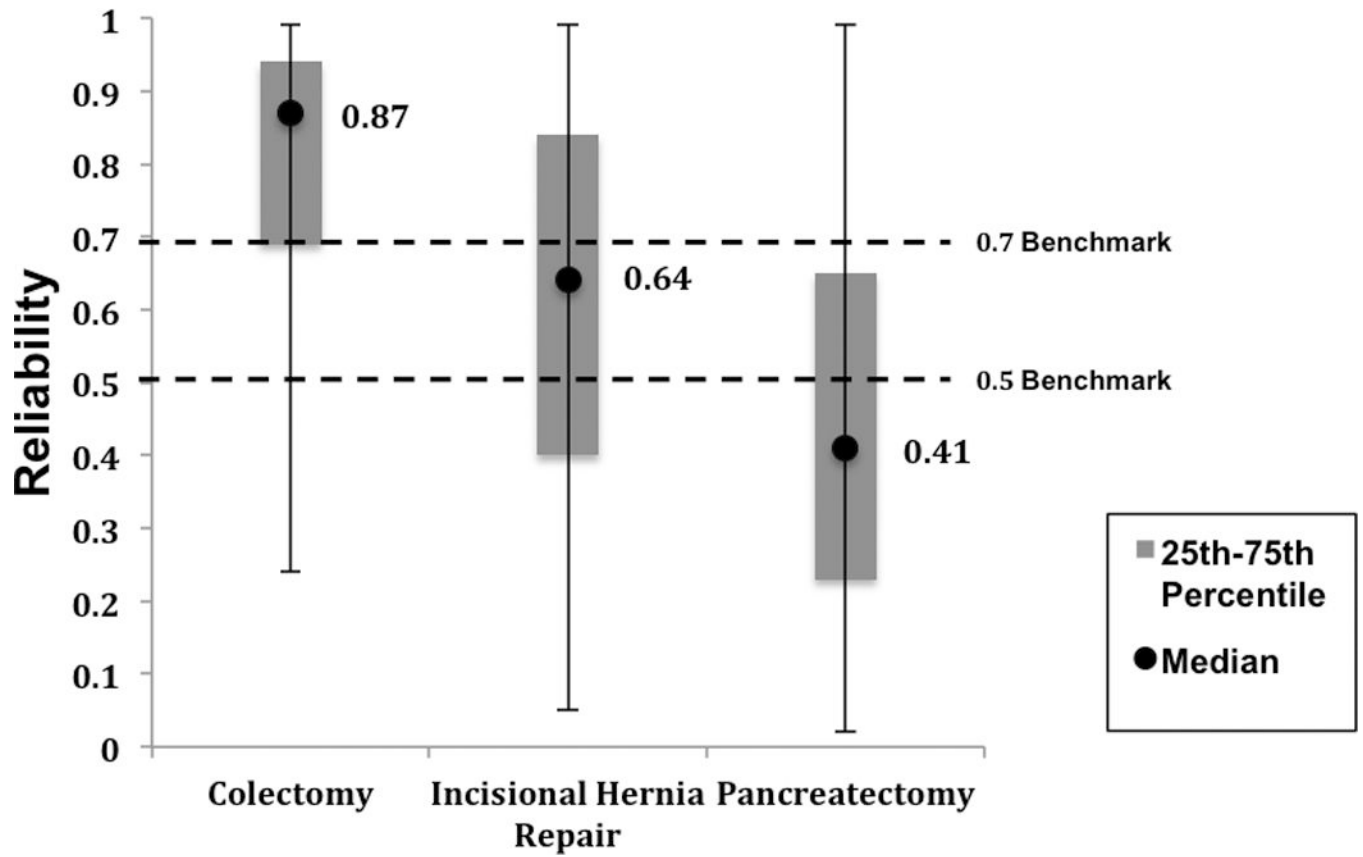


Figure 1. Reliability of total episode payments for all hospitals by procedure. (Black circles: median procedure-specific episode payment reliability. Gray boxes: interquartile range. Error bars: total range of payment reliability for all hospitals performing the specified procedure.)

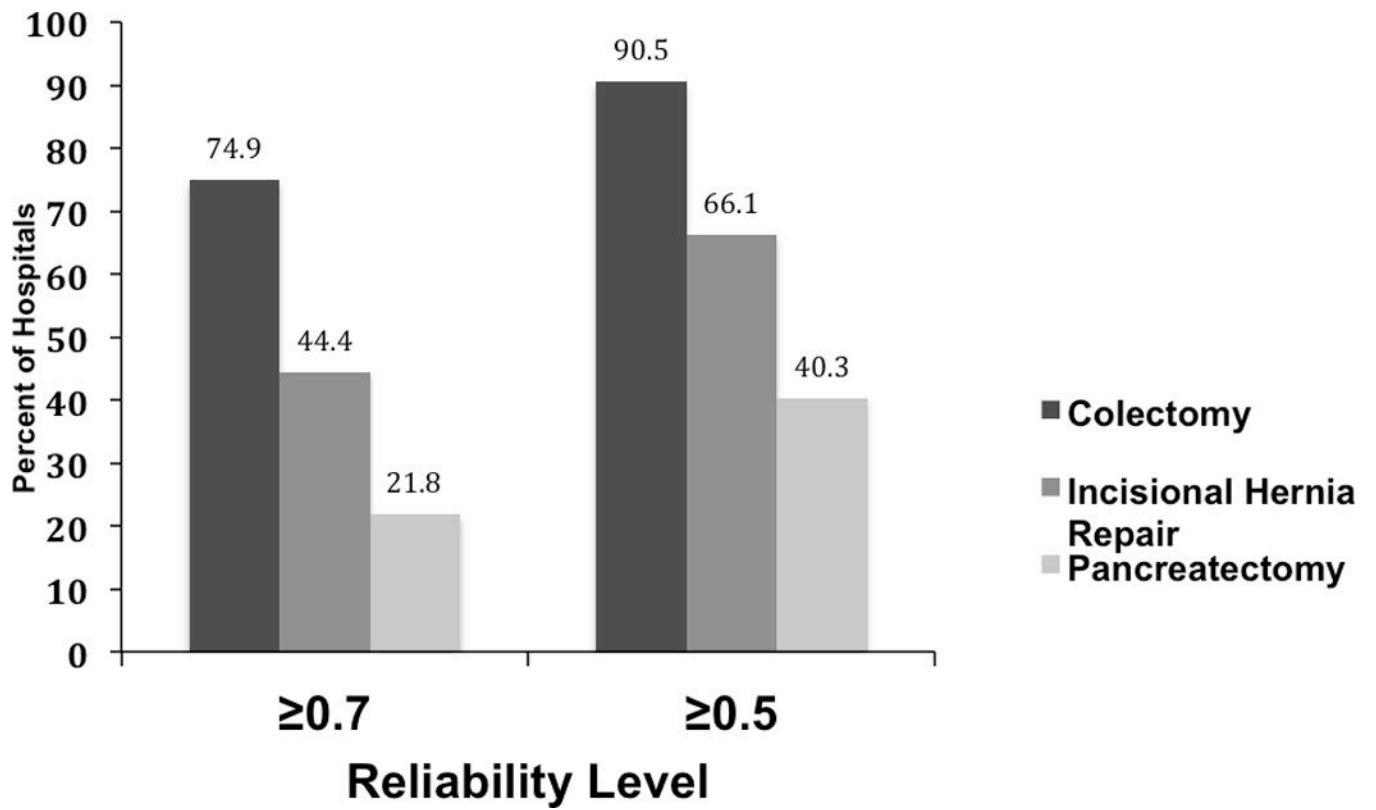


Figure 2. Percentage of hospitals meeting published reliability benchmarks (≥ 0.5 and ≥ 0.7) by procedure.

Table 1

Baseline characteristics, outcomes, and total episode payments for Medicare patients undergoing inpatient surgery procedures in 2009–2010.

	Procedure		
	Colectomy	Hernia Repair	Pancreatectomy
No. Patients	221,977	109,490	8,415
No. Hospitals	2,671	2,673	619
Caseload (mean, SD)	157.0 (115.6)	83.8 (67.2)	13.5 (21.9)
Patient characteristics			
Age (mean, SD)	76.5 (7.4)	74.8 (7.1)	74.0 (5.8)
Female sex (%)	58.7	60.4	49.9
White race (%)	87.0	87.8	87.2
3 Comorbidities (%)	32.1	31.3	33.4
Emergent case (%)	33.1	30.6	7.2
Outcome variables			
Length of stay (med, IQR)	8.0 (5,14)	5.0 (3, 9)	11.0 (8, 18)
Overall complication rate (%)	31.7	21.6	30.2
Total Episode Payments			
Risk-adjusted (mean, SD)	\$20,229(\$18,337)	\$13,262(\$14,523)	\$25,055(\$22,549)

(SD: standard deviation; IQR: interquartile range)

Table 2

Reliability of risk-adjusted total episode payments across hospital caseloads.

	Overall	Quartile of Hospital Caseload		
		Lowest Quartile (<25%)	Middle Quartiles (25–75%)	Highest Quartile (>75%)
Colectomy				
Caseload (mean, SD)	157.0 (115.6)	21.6 (7.7)	60.0 (14.6)	168.3 (81.1)
Reliability (mean, SD)	0.80 (0.18)	0.58 (0.13)	0.86 (0.04)	0.96 (0.01)
Incisional Hernia Repair				
Caseload (mean, SD)	83.8 (67.2)	10.3 (3.7)	28.0 (7.0)	85.3 (46.0)
Reliability (mean, SD)	0.61 (0.25)	0.32 (0.12)	0.64 (0.09)	0.88 (0.05)
Pancreatectomy				
Caseload (mean, SD)	13.5 (21.9)	2.4 (0.49)	6.0 (1.7)	32.4 (30.1)
Reliability (mean, SD)	0.45 (0.27)	0.22 (0.15)	0.37 (0.12)	0.77 (0.14)

(SD: standard deviation)