



Surgeon specialization and operative mortality in United States: retrospective analysis

Nikhil R Sahni,^{1,2} Maurice Dalton,³ David M Cutler,^{1,3} John D Birkmeyer,^{4,5} Amitabh Chandra^{3,6}

¹Department of Economics, Harvard University, Cambridge, MA, USA

²McKinsey and Company, Boston, MA, USA

³National Bureau of Economic Research, Cambridge, MA, USA

⁴Dartmouth-Hitchcock Health System, Lebanon, NH, USA

⁵Dartmouth Institute for Health Policy and Clinical Practice, Lebanon, NH, USA

⁶Harvard Kennedy School, Cambridge, MA, USA

Correspondence to: N R Sahni, 280 Congress Street #1100, Boston, MA 02210, USA
nikhilsahni@fas.harvard.edu

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2016;354:i3571
<http://dx.doi.org/10.1136/bmj.i3571>

Accepted: 24 June 2016

ABSTRACT

OBJECTIVE

To measure the association between a surgeon's degree of specialization in a specific procedure and patient mortality.

DESIGN

Retrospective analysis of Medicare data.

SETTING

US patients aged 66 or older enrolled in traditional fee for service Medicare.

PARTICIPANTS

25 152 US surgeons who performed one of eight procedures (carotid endarterectomy, coronary artery bypass grafting, valve replacement, abdominal aortic aneurysm repair, lung resection, cystectomy, pancreatic resection, or esophagectomy) on 695 987 patients in 2008-13.

MAIN OUTCOME MEASURE

Relative risk reduction in risk adjusted and volume adjusted 30 day operative mortality between surgeons in the bottom quarter and top quarter of surgeon specialization (defined as the number of times the surgeon performed the specific procedure divided by his/her total operative volume across all procedures).

RESULTS

For all four cardiovascular procedures and two out of four cancer resections, a surgeon's degree of specialization was a significant predictor of operative mortality independent of the number of times he or she performed that procedure: carotid endarterectomy (relative risk reduction between bottom and top quarter of surgeons 28%, 95% confidence interval 0% to 48%); coronary

artery bypass grafting (15%, 4% to 25%); valve replacement (46%, 37% to 53%); abdominal aortic aneurysm repair (42%, 29% to 53%); lung resection (28%, 5% to 46%); and cystectomy (41%, 8% to 63%). In five procedures (carotid endarterectomy, valve replacement, lung resection, cystectomy, and esophagectomy), the relative risk reduction from surgeon specialization was greater than that from surgeon volume for that specific procedure. Furthermore, surgeon specialization accounted for 9% (coronary artery bypass grafting) to 100% (cystectomy) of the relative risk reduction otherwise attributable to volume in that specific procedure.

CONCLUSION

For several common procedures, surgeon specialization was an important predictor of operative mortality independent of volume in that specific procedure. When selecting a surgeon, patients, referring physicians, and administrators assigning operative workload may want to consider a surgeon's procedure specific volume as well as the degree to which a surgeon specializes in that procedure.

Introduction

Hundreds of studies have shown that surgeons with higher volumes have better outcomes across a variety of procedures.¹⁻⁶ Researchers have identified several factors contributing to this association, including experience and technical skill.⁷⁻¹⁰ The ease of measuring this volume-outcomes relation, coupled with the strength of the association, makes it a powerful way to ascertain surgeons' quality.

At the same time, the degree to which a surgeon specializes in a specific procedure may be as important as the number of times that he or she performs it.¹¹⁻¹⁴ A surgeon who specializes in one operation may have better outcomes owing to muscle memory built from repetition, higher attention and faster recall as a result of less switching between different procedures, and knowledge transfer of outcomes for the same procedure performed in different patients.^{8,15-18} If this specialization hypothesis holds true, a surgeon performing 20 procedures of which all 20 are valve replacements (denoting 100% specialization in the procedure) would have lower operative mortality rates than a surgeon who performs 100 operations of which 40 are valve replacements (denoting 40% specialization in the procedure). In contrast, the volume-outcomes hypothesis would suggest that selecting the surgeon who performs 40 valve replacements would lead to superior outcomes for patients. To the best of our knowledge, no study has described a statistical association between a surgeon's degree of specialization in a specific procedure and patients' mortality.

Our objective was to test the hypothesis of a specialization-outcomes relation independent of a surgeon's

WHAT IS ALREADY KNOWN ON THIS TOPIC

Cohort studies and systematic reviews have shown that surgeons with higher volumes have better outcomes across a variety of procedures up to a certain threshold

In other sectors such as airlines, banking, and automobile assembly, specialization has been shown to have a positive association with outcomes

Previous work has looked at the relation between outcomes and self reported specialty or board certification but not the association between a surgeon's degree of specialization in a specific procedure and mortality

WHAT THIS STUDY ADDS

A statistically significant relative risk reduction in operative mortality was observed between the bottom and top quarters of surgeon specialization for six procedures

The relative risk reduction in operative mortality due to surgeon specialization was greater than that due to procedure specific volume for five procedures

The specialization-outcomes relation accounted for some portion, if not all, of the volume-outcomes relation

The observed specialization-outcomes relation suggests a new, easily measured metric of surgeons' quality that builds on the volume-outcomes relation to inform the way healthcare is organized and delivered

volume in that specific procedure. We looked at the same eight procedures originally studied for the volume-outcomes relation—carotid endarterectomy, coronary artery bypass grafting, valve replacement, abdominal aortic aneurysm repair, lung resection, cystectomy, pancreatic resection, and esophagectomy—to estimate the specialization-outcomes relation.¹ We determined surgeons' specialization by using US Medicare data rather than a surgeon's self reported specialty or board certification.^{6,9,19} For each procedure, we compared risk adjusted 30 day mortality between surgeons who performed the same volume of the specific procedure but varied in the degree of specialization in that procedure.

Methods

We examined all eight conditions that were studied by Birkmeyer et al: four cardiovascular procedures (carotid endarterectomy, coronary artery bypass grafting, valve replacement, and abdominal aortic aneurysm repair) and four cancer resections (lung resection, cystectomy, pancreatic resection, and esophagectomy).¹ We identified all patients undergoing one of these procedures and the associated surgeons in the Medicare Inpatient file from 2008 to 2013, the latest year of data available to us at the time of this study.

Defining surgeons' volume and specialization

For a given inpatient claim, we defined a procedure by using the ICD-9 (international classification of diseases, ninth revision) procedure code listed in the principal procedure field. We attributed each surgery to the surgeon listed in the operating physician field of the inpatient claim. We defined "total operative volume" (v) as all procedures attributed to a surgeon and "procedure specific volume" (v_i) as the number of cases attributed to a surgeon for the specific procedure being examined (see appendix for code list).^{1,2} We defined surgeon specialization (s_i) as procedure specific volume divided by total operative volume across all procedures ($s_i=v_i/v$).

Using National Plan and Provider Enumeration System data, we identified each surgeon's self reported specialty and included only surgeons in clinically appropriate specialties (see appendix for provider taxonomy codes).²⁰ We also excluded surgeons with fewer than three consecutive years of available inpatient claims.

We divided surgeons into quarters based on their specialization. We also divided surgeons into quarters based on procedure specific volume and verified that a similar volume-outcomes relation existed in our data as in the original research (see appendix figure A1).¹

Study population

We included patients aged 66 or older who were continuously enrolled in traditional fee for service Medicare starting 12 months before the admission month through four months after the index admission or death.²¹ We required that at least one year had passed before we identified a new procedure as a separate case for the same patient.

We further limited patients on the basis of specific diagnoses and other procedures performed (see appendix for ICD-9 codes). To avoid potential adverse mortal-

ity effects due to delays in performing surgery, we limited patients to those with procedures performed within three days of admission.^{22,23} We used SAS version 9.3 to build the study cohort.

Patients' characteristics

For each patient, we compiled age, sex, and race data from the Medicare Master Beneficiary Summary file. Using the method developed by Krumholz et al, which mapped the 189 hierarchical condition categories developed by Pope et al for the Centers for Medicare and Medicaid Services (CMS) into 17 groups (cerebrovascular disease, chronic liver disease, chronic obstructive pulmonary disease, dementia, diabetes, dialysis, hypertension, major psychiatric disorder, metastatic cancer, paralysis, peripheral vascular disease, pneumonia, protein calorie malnutrition, renal failure disease, stroke, substance abuse, trauma), we measured the comorbidity profile for each patient based on all Medicare Provider Analysis and Review (MedPAR) claims occurring 365 days before the index hospital admission, not including the index hospital admission event.^{24,25} For hospitals, we used hospital listings in the 2010 Dartmouth Atlas to identify academic medical centers.²⁶

Analysis and outcome measures

The outcome variable of interest was death occurring within 30 days of the initial hospital admission date. We used a multilevel mixed logit model to examine the specialization-outcomes relation. We controlled for age, sex, race, year of surgery, comorbidity profile, day of the week, procedure type (see appendix for details on groupings), days between admission and surgery, and whether the hospital was an academic medical center. We included surgeon random effects to account for unobserved characteristics of surgeons, such as a surgeon's technical skill or the experience of the surgeon, as well as repeated observations for the same surgeon. We also controlled for procedure specific volume in quarters.

We included hospital random effects to account for hospital specific factors, such as staffing ratios, the hospital's financial health, and electronic health record systems, that may mediate the relation between specialization and mortality. Including these effects allowed us to interpret the estimated relation between surgeons' specialization and outcomes as the opportunity to improve mortality through superior matching of patients to surgeons within a hospital, as opposed to matching patients to surgeons in different hospitals. We converted odds ratios to relative risks by using standard methods as developed by Zhang and Yu and specified for Stata by Cummings.^{27,28} We estimated the relative risk reduction from greater specialization as the difference in mortality between an average surgeon in the bottom quarter versus top quarter of surgeon specialization. We also estimated the absolute risk reduction and the number needed to treat (NNT).

Sensitivity analysis

We estimated a second set of regressions with surgeons' specialization and procedure specific volume as

continuous rather than ordinal variables. This specification has the benefits of estimating only one coefficient for surgeon specialization and being less subject to small sample sizes, but it is more restrictive than the specification with indicator variables for the degree of specialization. We used Stata 14 for all analyses and regressions.

Patient involvement

Our interest in this study was driven by finding ways to improve outcomes for surgical patients. Following previous research, we chose our main measure of 30 day mortality to focus on a patient's survival after surgery.¹⁶ Countries such as the United States and United Kingdom have also begun to report this measure so that patients can assess surgeons' quality.^{29,30} No patients were involved in the design of this study.

Results

Descriptive statistics

We examined 39 157 344 Medicare fee for service inpatient claims between 2008 and 2013. Our study cohort consisted of 695 987 patients operated on by 25 152 surgeons. Tables 1 and 2 summarize the characteristics of surgeons and patients for each of the eight procedures arranged by surgeon specialization. For most of the procedures, we found no clinically important association between surgeons' specialization and patients' age, sex, race, or comorbidity profile (see appendix for details on comorbidities). Average surgeon specialization for a procedure ranged from 6% for esophagectomy to 40% for coronary artery bypass grafting. Within a procedure, the difference in specialization among the least specialized and most specialized surgeons ranged from 43 percentage points (0.1-43%) for esophagectomy to 94 percentage points (0.1-94%) for coronary artery bypass grafting.

By definition, surgeons' specialization might increase as a result of greater procedure specific volume (numerator) or lesser total operative volume across all procedures (denominator). In certain procedures, such as abdominal aortic aneurysm repair and cystectomy, increased surgeon specialization was associated with greater procedure specific volume and relatively stable total operative volume. For other procedures, such as coronary artery bypass grafting and esophagectomy, we observed similar trends in procedure specific volume but also lesser total operative volume. A greater percentage of procedures were performed at academic medical centers in the higher compared with lower quarters of surgeon specialization for all procedures except carotid endarterectomy and coronary artery bypass grafting.

Tables 3 and 4 illustrate the extent to which a surgeon's degree of specialization was related to the number of times he or she performed that specific procedure. Overall, we found that surgeons in the top quarter of procedure specific volume varied in their specialization in the same procedure. For example, for carotid endarterectomy nearly two thirds of surgeons in the top quarter of procedure specific volume were also in the top quarter of surgeon specialization, but for coronary artery bypass grafting less than 50% of surgeons in the

top quarter of procedure specific volume were in the top quarter of surgeon specialization.

Relative risk reduction by quarters

We observed a statistically significant reduction in risk adjusted 30 day operative mortality between the bottom and top quarters of surgeon specialization for all four cardiovascular procedures and two of the four cancer resections independent of procedure specific volume (fig 1). Among the four cardiovascular procedures, the relative risk reduction in mortality from greater specialization ranged from 15% for coronary artery bypass grafting to 46% for valve replacement (table 5). Among the four cancer resections, the relative risk reduction in mortality ranged from 28% for lung resection to 48% for esophagectomy. Among the six procedures for which we found a statistically significant relative risk reduction, the absolute risk reduction ranged from 0.3% (NNT=334 patients) for carotid endarterectomy to 2.8% (NNT=36 patients) for abdominal aortic aneurysm repair. For the two cancer resections (pancreatic resection and esophagectomy) for which we did not find a statistically significant effect, the numbers of patients sampled were the smallest among all eight procedures reviewed (less than 11 000 patients for each procedure).

Specialization-outcomes relation versus volume-outcomes relation

Because we included procedure-specific volume in our regressions, we were able to estimate the relative risk reduction in mortality as a result of selecting a surgeon in the top quarter of surgeon specialization compared with a surgeon in the top quarter of procedure specific volume. For two of the four cardiovascular procedures (carotid endarterectomy and valve replacement) and three of the four cancer resections (lung resection, cystectomy, and esophagectomy), we observed a greater relative risk reduction due to surgeon specialization than due to procedure specific volume (see appendix for full regression results).

To estimate the role of surgeon specialization on the volume-outcomes relation, we estimated the relative risk reduction between the bottom and top quarters of procedure specific volume with and without accounting for surgeon specialization (table 6). When we excluded surgeon specialization from the model, we found a statistically significant association between procedure specific volume and mortality for all four cardiovascular procedures and for two of the four cancer resections (lung resection and pancreatic resection). When we included surgeon specialization, the observed volume-outcomes relation decreased from 9% for coronary artery bypass grafting up to 100% for cystectomy.

Sensitivity analysis

We ran a second version of the regressions with surgeon specialization and procedure specific volume as continuous variables. For all four cardiovascular procedures and two of the four cancer resections (lung resection and esophagectomy), we observed a statistically significant association between surgeon specialization and mortality (see appendix for full results). For carotid

Table 1 | Characteristics of surgeons and patients by quarter of surgeon specialization: cardiovascular procedures

Characteristics	All patients	Quarter of specialization			
		Bottom	Second	Third	Top
Carotid endarterectomy					
Surgeons:					
No (%)	5505 (100)	1377 (25.0)	1376 (25.0)	1377 (25.0)	1375 (25.0)
Average % specialization (range)	19.2 (0.1-80.8)	2.7 (0.1-4.3)	7.3 (4.3-9.6)	13.1 (9.6-16.5)	26.4 (16.5-80.8)
Carotid endarterectomy volume*	20.8	2.4	7.3	14.4	28.6
Total operative volume*	107.1	88.2	99.4	109.6	109.0
Patients:					
No (%)	183792 (100)	6546 (3.6)	24 313 (13.2)	53 001 (28.8)	99932 (54.4)
Age, years	75.9	75.6	75.8	75.7	76.0
No (%) female	78748 (42.8)	2788 (42.6)	10 423 (42.9)	22 659 (42.8)	42 878 (42.9)
No (%) white	173771 (94.5)	5926 (90.5)	22 434 (92.3)	49 873 (94.1)	95 538 (95.6)
No of comorbidities†	1.13	1.29	1.22	1.17	1.07
No (%) performed at AMC	11 302 (6.1)	491 (7.5)	2182 (9.0)	3427 (6.5)	5202 (5.2)
Coronary artery bypass grafting					
Surgeons:					
No (%)	2745 (100)	687 (25.0)	694 (25.3)	678 (24.7)	686 (25.0)
Average % specialization (range)	40.0 (0.1-93.9)	16.4 (0.1-21.9)	27.9 (22.0-33.3)	39.3 (33.4-45.3)	56.3 (45.3-93.9)
Coronary artery bypass grafting volume*	38.5	19.9	32.3	40.0	47.2
Total operative volume*	101.5	119.2	116.5	101.9	85.2
Patients:					
No (%)	170 803 (100)	17 847 (10.4)	41 996 (24.6)	51 801 (30.3)	59 159 (34.6)
Age, years	73.9	73.9	73.8	73.9	73.8
No (%) female	50 246 (29.4)	5219 (29.2)	12 147 (28.9)	15 112 (29.2)	17 768 (30.0)
No (%) white	157 121 (92.0)	16 153 (90.5)	38 516 (91.7)	48 118 (92.9)	54 334 (91.8)
No of comorbidities†	0.75	0.80	0.76	0.74	0.74
No (%) performed at AMC	16 150 (9.5)	3002 (16.8)	4159 (9.9)	4151 (8.0)	4838 (8.2)
Valve replacement					
Surgeons:					
No (%)	2627 (100)	657 (25.0)	657 (25.0)	658 (25.0)	655 (24.9)
Average % specialization (range)	35.5 (0.1-92.6)	6.7 (0.1-9.5)	13.9 (9.5-17.8)	23.9 (17.8-30.0)	50.4 (30.0-92.6)
Valve replacement volume*	37.0	8.0	15.5	22.6	53.1
Total operative volume*	103.5	117.4	111.8	95.3	103.4
Patients:					
No (%)	135 118 (100)	7557 (5.6)	21 240 (15.7)	34 106 (25.2)	72 215 (53.4)
Age, years	76.3	75.4	75.7	76.2	76.6
No (%) female	57 837 (42.8)	3205 (42.4)	9234 (43.5)	14 482 (42.5)	30 916 (42.8)
No (%) white	127 955 (94.7)	7003 (92.7)	19 885 (93.6)	32 288 (94.7)	68 779 (95.2)
No of comorbidities†	0.81	0.92	0.83	0.81	0.79
No (%) performed at AMC	29 089 (21.5)	273 (3.6)	937 (4.4)	2440 (7.2)	25 439 (35.2)
Abdominal aortic aneurysm repair					
Surgeons:					
No (%)	4679 (100)	1170 (25.0)	1176 (25.1)	1164 (24.9)	1169 (25.0)
Average % specialization (range)	11.1 (0.1-66.7)	0.9 (0.1-1.6)	3.3 (1.6-4.5)	6.9 (4.6-9.0)	15.7 (9.0-66.7)
Abdominal aortic aneurysm repair volume*	10.9	0.9	3.6	7.2	15.1
Total operative volume*	102.7	97.6	110.8	105.7	99.6
Patients:					
No (%)	88 365 (100)	2734 (3.1)	11 527 (13.0)	25 020 (28.3)	49 084 (55.5)
Age, years	76.5	76.2	76.4	76.5	76.5
No (%) female	20 166 (22.8)	746 (27.3)	2771 (24.0)	5729 (22.9)	10 920 (22.2)
No (%) white	83 265 (94.2)	2510 (91.8)	10 581 (91.8)	23 536 (94.1)	46 638 (95.0)
No of comorbidities†	0.98	0.97	1.05	1.01	0.95
No (%) performed at AMC	12 143 (13.7)	76 (2.8)	547 (4.7)	1433 (5.7)	10 087 (20.6)

Surgeons were into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Numbers may not add up to 100% owing to rounding. AMC=academic medical center.

*All surgeon volume numbers reported per annum.

†Number of comorbidities as mapped by Krumholz et al from 189 hierarchical condition categories developed by Pope et al for Centers for Medicare and Medicaid Services into 17 groups.^{24 25}

Table 2 | Characteristics of surgeons and patients by quarter of surgeon specialization: cancer resections

Characteristics	All patients	Quarter of specialization			
		Bottom	Second	Third	Top
Lung resection					
Surgeons:					
No (%)	3616 (100)	904 (25.0)	905 (25.0)	906 (25.1)	901 (24.9)
Average % specialization (range)	28.6 (0.1-89.1)	1.5 (0.1-2.5)	4.5 (2.5-6.2)	10.4 (6.3-15.0)	41.0 (15.1-89.1)
Lung resection volume*	23.4	1.6	5.1	10.4	32.6
Total operative volume*	89.5	102.9	113.2	101.1	80.3
Patients:					
No (%)	85966 (100)	2548 (3.0)	9067 (10.5)	20725 (24.1)	53626 (62.4)
Age, years	74.3	74.4	74.1	74.2	74.3
No (%) female	43102 (50.1)	1191 (46.7)	4341 (47.9)	10057 (48.5)	27513 (51.3)
No (%) white	79315 (92.3)	2368 (92.9)	8319 (91.8)	19191 (92.6)	49437 (92.2)
No of comorbidities†	0.98	1.16	1.10	1.03	0.93
No (%) performed at AMC	16034 (18.7)	70 (2.7)	116 (1.3)	478 (2.3)	15370 (28.7)
Cystectomy					
Surgeons:					
No (%)	3319 (100)	835 (25.2)	830 (25.0)	826 (24.9)	828 (24.9)
Average % specialization (range)	16.2 (0.2-72.9)	1.0 (0.2-1.5)	2.2 (1.5-2.9)	4.1 (2.9-5.4)	23.0 (5.4-72.9)
Cystectomy volume*	7.1	0.4	0.8	1.3	10.2
Total operative volume*	40.2	37.3	35.1	33.0	43.1
Patients:					
No (%)	18438 (100)	1153 (6.3)	1910 (10.4)	3182 (17.3)	12193 (66.1)
Age, years	75.4	75.9	75.7	75.9	75.2
No (%) female	3578 (19.4)	283 (24.5)	443 (23.2)	647 (20.3)	2205 (18.1)
No (%) white	17257 (93.6)	1068 (92.6)	1781 (93.2)	2957 (92.9)	11451 (93.9)
No of comorbidities†	0.91	1.00	0.99	0.90	0.90
No (%) performed at AMC	5887 (31.9)	75 (6.5)	110 (5.8)	250 (7.9)	5452 (44.7)
Pancreatic resection					
Surgeons:					
No (%)	1612 (100)	403 (25.0)	404 (25.1)	404 (25.1)	401 (24.9)
Average % specialization (range)	21.9 (0.1-69.8)	0.4 (0.1-0.7)	1.4 (0.7-2.2)	5.0 (2.2-8.3)	28.9 (8.4-69.8)
Pancreatic resection volume*	11.5	0.3	0.9	2.7	15.1
Total operative volume*	54.2	76.3	66.1	54.5	51.6
Patients:					
No (%)	10191 (100)	455 (4.5)	720 (7.1)	1629 (16.0)	7387 (72.5)
Age, years	74.4	74.6	74.9	74.3	74.4
No (%) female	5159 (50.6)	227 (49.9)	360 (50.0)	864 (53.0)	3708 (50.2)
No (%) white	9173 (90.0)	403 (88.6)	644 (89.4)	1421 (87.2)	6705 (90.8)
No of comorbidities†	1.26	1.14	1.16	1.28	1.28
No (%) performed at AMC	4443 (43.6)	16 (3.5)	60 (8.3)	235 (14.4)	4132 (55.9)
Esophagectomy					
Surgeons:					
No (%)	1049 (100)	263 (25.1)	262 (25.0)	262 (25.0)	262 (25.0)
Average % specialization (range)	6.3 (0.1-43.4)	0.4 (0.1-0.6)	1.1 (0.6-1.6)	2.6 (1.6-3.7)	10.1 (3.8-43.4)
Esophagectomy volume*	3.9	0.4	0.9	1.9	6.0
Total operative volume*	73.5	101.0	82.9	73.5	66.5
Patients:					
No (%)	3314 (100)	296 (8.9)	465 (14.0)	732 (22.1)	1821 (54.9)
Age, years	72.9	72.6	72.8	72.8	72.9
No (%) female	663 (20.0)	53 (17.9)	103 (22.2)	128 (17.5)	379 (20.8)
No (%) white	3128 (94.4)	275 (92.9)	435 (93.5)	691 (94.4)	1727 (94.8)
No of comorbidities†	1.23	1.25	1.34	1.36	1.15
No (%) performed at AMC	1278 (38.6)	10 (3.4)	53 (11.4)	184 (25.1)	1031 (56.6)

Surgeons were into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Numbers may not add up to 100% owing to rounding.

AMC=academic medical center.

*All surgeon volume numbers reported per annum.

†Number of comorbidities as mapped by Krumholz et al from 189 hierarchical condition categories developed by Pope et al for Centers for Medicare and Medicaid Services into 17 groups.^{24,25}

Table 3 | Distribution of surgeons by quarter of surgeon specialization: cardiovascular procedures

Quarter of surgeon specialization	Mean, %	Quarter of procedure specific volume, No (%)			
		Bottom	Second	Third	Top
Carotid endarterectomy (n=5505 surgeons)					
Bottom	2.7	1054 (73)	298 (22)	25 (2)	0 (0)
Second	7.3	265 (18)	623 (46)	435 (32)	53 (4)
Third	13.1	64 (4)	283 (21)	593 (44)	437 (32)
Top	26.4	54 (4)	145 (11)	293 (22)	883 (64)
Total	–	1437 (100)	1349 (100)	1346 (100)	1373 (100)
Coronary artery bypass grafting (n=2745 surgeons)					
Bottom	16.4	390 (57)	202 (29)	77 (11)	18 (3)
Second	27.9	129 (19)	213 (31)	219 (32)	133 (20)
Third	39.3	93 (14)	155 (23)	204 (30)	226 (33)
Top	56.3	76 (11)	118 (17)	189 (27)	303 (45)
Total	–	688 (100)	688 (100)	689 (100)	680 (100)
Valve replacement (n=2627 surgeons)					
Bottom	6.7	438 (66)	175 (26)	40 (6)	4 (1)
Second	13.9	122 (18)	273 (41)	217 (34)	45 (7)
Third	23.9	69 (10)	151 (23)	260 (40)	178 (27)
Top	50.4	31 (5)	70 (10)	129 (20)	425 (65)
Total	–	660 (100)	669 (100)	646 (100)	652 (100)
Abdominal aortic aneurysm repair (n=4679 surgeons)					
Bottom	0.9	919 (76)	246 (21)	5 (0)	0 (0)
Second	3.3	191 (16)	573 (48)	367 (32)	45 (4)
Third	6.9	47 (4)	264 (22)	501 (44)	352 (31)
Top	15.7	45 (4)	105 (9)	271 (24)	748 (65)
Total	–	1202 (100)	1188 (100)	1144 (100)	1145 (100)

Surgeons were divided into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to a surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Numbers may not add up to 100% owing to rounding.

Table 4 | Distribution of surgeons by quarter of surgeon specialization: cancer resection

Quarter of surgeon specialization	Mean, %	Quarter of procedure specific volume, No (%)			
		Bottom	Second	Third	Top
Lung resection (n=3616 surgeons)					
Bottom	1.5	730 (74)	165 (20)	9 (1)	0 (0)
Second	4.5	183 (19)	435 (52)	262 (29)	25 (3)
Third	10.4	54 (5)	174 (21)	429 (48)	249 (28)
Top	41.0	22 (2)	65 (8)	198 (22)	616 (69)
Total	–	989 (100)	839 (100)	898 (100)	890 (100)
Cystectomy (n=3319 surgeons)					
Bottom	1.0	691 (57)	121 (17)	18 (3)	5 (1)
Second	2.2	349 (29)	297 (42)	142 (24)	42 (5)
Third	4.1	128 (11)	224 (32)	281 (48)	193 (24)
Top	23.0	51 (4)	62 (9)	141 (24)	574 (71)
Total	–	1219 (100)	704 (100)	582 (100)	814 (100)
Pancreatic resection (n=1612 surgeons)					
Bottom	0.4	347 (68)	52 (15)	4 (1)	0 (0)
Second	1.4	136 (27)	181 (53)	85 (24)	2 (1)
Third	5.0	26 (5)	95 (28)	209 (58)	74 (19)
Top	28.9	3 (1)	14 (4)	63 (17)	321 (81)
Total	–	512 (100)	342 (100)	361 (100)	397 (100)
Esophagectomy (n=1049 surgeons)					
Bottom	0.4	204 (70)	54 (20)	5 (2)	0 (0)
Second	1.1	69 (24)	129 (47)	62 (25)	2 (1)
Third	2.6	13 (4)	75 (27)	120 (48)	54 (23)
Top	10.1	4 (1)	16 (6)	64 (25)	178 (76)
Total	–	290 (100)	274 (100)	251 (100)	234 (100)

Surgeons were divided into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to a surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Numbers may not add up to 100% owing to rounding.

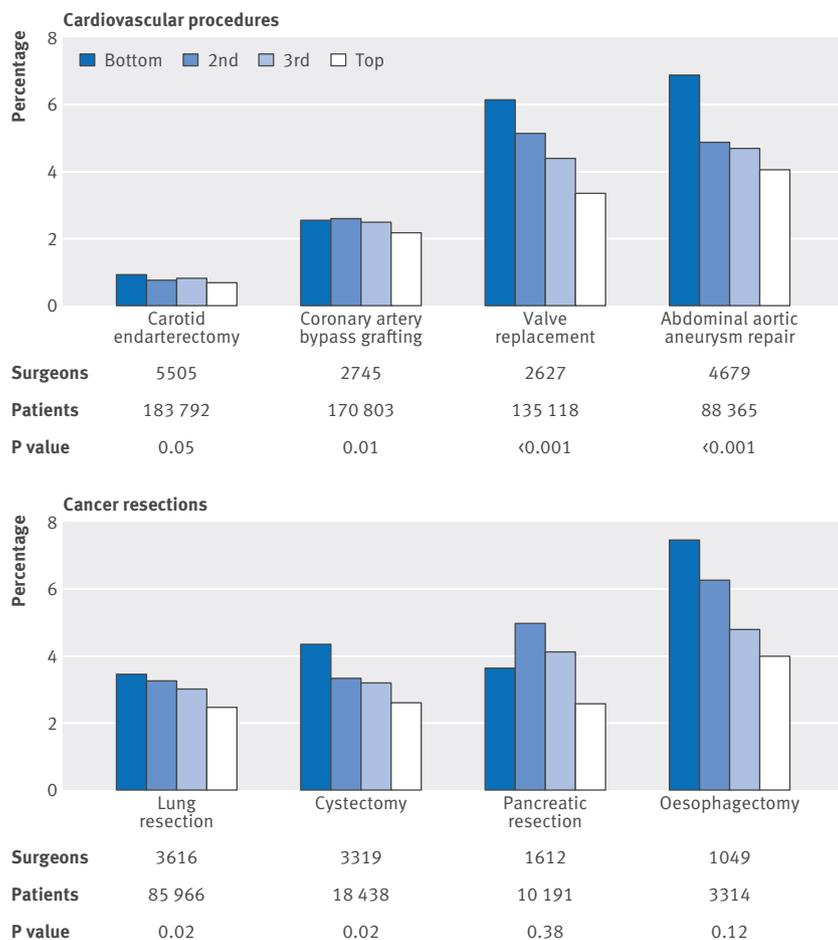


Fig 1 | Risk adjusted and volume adjusted 30 day operative mortality rate by quarter of surgeon specialization. Mortality was defined as rate of death occurring within 30 days of initial hospital admission date. Surgeons were divided into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Multilevel mixed logit regression was run, controlled for procedure specific volume; hospital where procedure was performed; age, sex, and race of patient; year of surgery; comorbidity profile; day of week; procedure type; days between admission and surgery; and whether hospital was academic medical center. P values were estimated to test for difference between bottom and top quarters within surgeon specialization. Average surgeon specialization for bottom/second/third/top quarter by procedure was: carotid endarterectomy 2.7%/7.3%/13.1%/26.4%; coronary artery bypass grafting 16.4%/27.9%/39.3%/56.3%; valve replacement 6.7%/13.9%/23.9%/50.4%; abdominal aortic aneurysm repair 0.9%/3.3%/6.9%/15.7%; lung resection 1.5%/4.5%/10.4%/41.0%; cystectomy 1.0%/2.2%/4.1%/23.0%; pancreatic resection 0.4%/1.4%/5.0%/28.9%; and esophagectomy 0.4%/1.1%/2.6%/10.1%

endarterectomy and abdominal aortic aneurysm repair, we also observed a statistically significant association between procedure specific volume and mortality.

Discussion

Our objective in this study was to quantify a possible association between a surgeon's degree of specialization in a specific procedure and patients' mortality. For six of the eight procedures examined (carotid endarterectomy, coronary artery bypass grafting, valve replacement, abdominal aortic aneurysm repair, lung resection, and cystectomy), we found surgeon specialization to be an important

predictor of mortality. Given the well documented volume-outcomes relation in healthcare, we also controlled for the number of times the surgeon performed the specific procedure. Our results showed that the observed specialization-outcomes relation was independent of the surgeon's volume in that specific procedure.

The volume-outcomes relation is generally attributed to "learning by doing," which shows diminishing returns after a certain level of "doing." We found a similar result for certain procedures such as abdominal aortic aneurysm repair, suggesting some minimal cut-off threshold of surgeon specialization. However, for procedures such as valve replacement, our results showed a linear trend from the bottom to top quarter of surgeon specialization, suggesting the absence of diminishing returns from specialization.

Furthermore, for five procedures (carotid endarterectomy, valve replacement, lung resection, cystectomy, and esophagectomy), the relative risk reduction in mortality from selecting a surgeon in the top quarter of surgeon specialization was greater than that from selecting a surgeon in the top quarter of procedure specific volume. Additionally, surgeon specialization accounted for at least some portion (if not all) of the observed volume-outcomes relation.

Several factors may explain the observed association between specialization and outcomes. Repetition of tasks has been shown to improve mortality and could be manifested in surgeons as muscle memory and dexterity.^{8,15} Greater familiarity with a medical device and its components has also been associated with better survival rates.¹⁸ Reduced mortality has been linked to performance of the same procedure under varying patient related circumstances, allowing a surgeon to transfer relevant knowledge and skills between patients. Furthermore, focusing on a single procedure reduces the cognitive demands of switching tasks.^{8,11,17} These potential mechanisms might result from both greater volume for a specific procedure (for example, task repetition) and from less total operative volume across all procedures (for example, academic research). We did not explore these mechanisms, as this was beyond the scope of this paper.

Strengths and limitations of study

Our study should be interpreted in the context of its limitations. One potential limitation may be unobserved choices by surgeons in selecting patients. For example, in a given hospital, a surgeon who specializes in only coronary artery bypass grafts may be performing these surgeries on unobservably healthier patients because he or she is less skilled than a colleague who also performs other procedures. However, given the relative comparability of patients' characteristics between quarters of surgeon specialization, this explanation is unlikely to be true.

A second limitation may be unobserved characteristics of surgeons, such as a surgeon's technical skill, the age and experience of the surgeon, or a less specialized surgical training program from which the surgeon did not become particularly skilled in a specific procedure. We attempted to account for these factors by including surgeon random effects.

Table 5 | Relative risk reduction in 30 day operative mortality from surgeon specialization and procedure specific volume

Procedure	Comparison between bottom and top quarter of surgeon specialization		Comparison between bottom and top quarter of procedure specific volume		No of patients
	Relative risk reduction in mortality, % (95% CI)	P value	Relative risk reduction in mortality, % (95% CI)	P value	
Cardiovascular procedures					
Carotid endarterectomy	28 (0 to 48)	0.05	18 (-19 to 44)	0.30	183792
Coronary artery bypass grafting	15 (4 to 25)	0.01	35 (24 to 44)	<0.001	170803
Valve replacement	46 (37 to 53)	<0.001	22 (7 to 34)	0.006	135118
Abdominal aortic aneurysm repair	42 (29 to 53)	<0.001	74 (68 to 79)	<0.001	88365
Cancer resections					
Lung resection	28 (5 to 46)	0.02	4 (-32 to 30)	0.80	85966
Cystectomy	41 (8 to 63)	0.02	-16 (-72 to 23)	0.48	18438
Pancreatic resection	29 (-52 to 68)	0.38	39 (-22 to 70)	0.16	10191
Esophagectomy	48 (-17 to 77)	0.12	-77 (-261 to 20)	0.15	3314

Mortality was defined as rate of death occurring within 30 days of initial hospital admission date. Surgeons were divided into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to a surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Multilevel mixed logit regression was run, controlled for procedure specific volume; hospital where procedure was performed; age, sex, and race of patient; year of surgery; comorbidity profile; day of week; procedure type; days between admission and surgery; and whether hospital was academic medical center. P values were estimated to test for difference between bottom and top quarters within surgeon specialization and procedure specific volume.

Table 6 | Volume-outcomes relation accounted for by surgeon specialization

Procedure	Relative risk reduction between bottom and top quarters of procedure volume, % (95% CI)		Percentage of volume-outcomes relation accounted for by surgeon specialization
	Regression: procedure specific volume only	Regression: procedure specific volume and surgeon specialization	
Cardiovascular procedures			
Carotid endarterectomy	34 (12 to 51)	18 (-19 to 44)	47
Coronary artery bypass grafting	38 (28 to 47)	35 (24 to 44)	9
Valve replacement	46 (38 to 54)	22 (7 to 34)	53
Abdominal aortic aneurysm repair	82 (80 to 85)	74 (68 to 79)	11
Cancer resections			
Lung resection	24 (5 to 40)	4 (-32 to 30)	84
Cystectomy	18 (-10 to 39)	-16 (-72 to 23)	100
Pancreatic resection	56 (35 to 71)	39 (-22 to 70)	30
Esophagectomy	-8 (-83 to 37)	-77 (-261 to 20)	--

Mortality was defined as rate of death occurring within 30 days of initial hospital admission date. Surgeons were divided into four groups of equal size based on procedure specific volume and surgeon specialization. Procedure was defined as ICD-9 procedure code listed in principal procedure field. Procedure specific volume was defined as number of cases attributed to surgeon for specific procedure and total operative volume as all procedures attributed to a surgeon. Surgeon specialization was defined as procedure specific volume divided by total operative volume across all procedures. Multilevel mixed logit regression was run, controlled for procedure specific volume; hospital where procedure was performed; age, sex, and race of patient; year of surgery; comorbidity profile; day of week; procedure type; days between admission and surgery; and whether hospital was academic medical center. P values were estimated to test for difference between bottom and top quarters within surgeon specialization and procedure specific volume.

Other limitations relate to the data involved. We generated our results on the basis of only the Medicare fee for service population. Although using these data guarantees that all patients were similarly insured and enabled us to measure outcomes effectively, we cannot factor in non-Medicare cases to measure total operative volume, procedure specific volume, and surgeon specialization, and nor can we account for the surgery team. We attempted to account for this discrepancy by using hospital random effects, which accounted for differences in payer mix across hospitals but not those across physicians. The restrictions of claims data, including unmeasured case mix differences and inaccurate coding, also limited our study. These limitations are also present in the volume-outcomes literature.

Areas for further study

Several aspects of these results warrant further study. We examined eight procedures; additional results for

other procedures would be valuable to determine external validity. Further studies could examine whether spillover effects exist from specializing in similar surgeries to the specified procedure, the effect of team continuity and specialization, and the relation to morbidity metrics and episode costs.^{16,31} Furthermore, the balance of surgeon specialization and department specialization could be studied to determine optimal surgical case distribution to maximize outcomes for patients. Finally, specialization could be examined in non-surgical settings, such as the degree of a primary care physician's panel of patients with a specific chronic condition.

Conclusion and policy implications

Our findings may have implications for policy makers, administrators, physicians, and patients, especially as surgeons' specialization is measurable using data available to health systems. Policy makers considering how to improve the quality of rural or smaller hospitals in which

surgeons cannot meet minimal volume thresholds could use surgeon specialization to assign patients to surgeons. At larger facilities, administrators determining case distribution across surgeons might consider not only a given surgeon's volume in that procedure but also his or her degree of specialization. A physician might use a measure of surgeon specialization to refer his or her patient to the most appropriate surgeon, possibly improving patients' outcomes.³² Finally, if these data are made available to a patient, he or she could choose a surgeon who specializes in the relevant procedure to possibly improve his or her chance of survival. We have been careful not to suggest that surgeons should specialize more, as that would require establishment of a causal relation. In all of our examples, surgeons' specialization is used as a surrogate for surgeons' quality, much like procedure specific volume is used as a proxy for surgeons' quality. Whether the degree of specialization causally improves surgical quality remains a topic for future work.

We thank William Gordon and Atheendar Venkataramani for their clinical contributions. We also thank Ishani Ganguli for her clinical contributions and editing of the manuscript.

Contributors: NRS contributed to the literature review, study question and design, and data analysis and interpretation; prepared the first draft of the report; and contributed to subsequent versions. MD, JDB, DMC, and AC contributed to study design, data analysis and interpretation, and drafting of the final report. All authors approved the final version. NRS is the guarantor.

Funding: This report represents independent research funded in part by the National Institute on Aging (NIA) for JDB (P01-AG019783-07S1), DMC (P01-AG031098), and AC (P01-AG005842-24 and P01-AG19783). The views expressed are those of the authors and not necessarily those of the NIA. The funders had no role in study design, in the analysis or interpretation of data, in the writing of the report, or in the decision to submit the paper for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organization for the submitted work other than the NIA as above; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not needed.

Data sharing: No additional data available.

Transparency declaration: The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>.

- Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med* 2003;349:2117-27. doi:10.1056/NEJMsa035205.
- Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002;346:1128-37. doi:10.1056/NEJMsa012337.
- Reames BN, Ghaferi AA, Birkmeyer JD, Dimick JB. Hospital volume and operative mortality in the modern era. *Ann Surg* 2014;260:244-51. doi:10.1097/SLA.0000000000000375.
- Finks JF, Osborne NH, Birkmeyer JD. Trends in hospital volume and operative mortality for high-risk surgery. *N Engl J Med* 2011;364:2128-37. doi:10.1056/NEJMsa1010705.
- Elixhauser A, Steiner C, Fraser I. Volume thresholds and hospital characteristics in the United States. *Health Aff (Millwood)* 2003;22:167-77. doi:10.1377/hlthaff.22.2.167.

- Chowdhury MM, Dagash H, Pierro A. A systematic review of the impact of volume of surgery and specialization on patient outcome. *Br J Surg* 2007;94:145-61. doi:10.1002/bjs.5714.
- Maruthappu M, Gilbert BJ, El-Harasis MA, et al. The influence of volume and experience on individual surgical performance: a systematic review. *Ann Surg* 2015;261:642-7. doi:10.1097/SLA.0000000000000852.
- KC D, Staats BR. Accumulating a portfolio of experience: the effect of focal and related experience on surgeon performance. *Manuf Serv Oper Manag* 2012;14:618-33doi:10.1287/msom.1120.0385.
- Bilimoria KY, Phillips JD, Rock CE, Hayman A, Prystowsky JB, Bentrem DJ. Effect of surgeon training, specialization, and experience on outcomes for cancer surgery: a systematic review of the literature. *Ann Surg Oncol* 2009;16:1799-808. doi:10.1245/s10434-009-0467-8.
- Birkmeyer JD, Finks JF, O'Reilly A, et al. Michigan Bariatric Surgery Collaborative. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2013;369:1434-42. doi:10.1056/NEJMsa1300625.
- Staats BR, Gino F. Specialization and variety in repetitive tasks: evidence from a Japanese bank. *Manage Sci* 2012;58:1141-59. doi:10.1287/mnsc.1110.1482.
- Tsikritsis N. The effect of operational performance and focus on profitability: a longitudinal study of the U.S. airline industry. *Manuf Serv Oper Manag* 2007;9:506-17doi:10.1287/msom.1060.0133.
- Huckman RS, Zinner D. Does Focus Improve Operational Performance? Lessons from the Management of Clinical Trials. *Strateg Manag J* 2008;29:173-93doi:10.1002/smj.650.
- Fisher ML, Iltner CD. The impact of product variety on automobile assembly operations: empirical evidence and simulation analysis. *Manage Sci* 1999;45:771-86doi:10.1287/mnsc.45.6.771.
- Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: results of a randomized, double-blinded study. *Ann Surg* 2002;236:458-63, discussion 463-4. doi:10.1097/00000658-200210000-00008.
- Clark JR, Huckman RS. Broadening focus: spillovers, complementarities, and specialization in the hospital industry. *Manage Sci* 2012;58:708-22doi:10.1287/mnsc.1110.1448.
- Bohn RE, Lapre MA. *Accelerated Learning by Experimentation. Learning Curves: Theory, Models, and Applications.* CRC Press, Taylor and Francis, 2011.
- Ramdas K, Saleh K, Stern S, Liu H. New Joints More Hip? Learning in the Use of New Components. 2012. http://faculty.london.edu/kramdas/New%20joints%20More%20Hip_080612.pdf.
- Rhee D, Papandria D, Yang J, et al. Comparison of pediatric surgical outcomes by the surgeon's degree of specialization in children. *J Pediatr Surg* 2013;48:1657-63. doi:10.1016/j.jpedsurg.2012.12.048.
- Centers for Medicare and Medicaid Services. National Plan and Provider Enumeration System. <https://nppes.cms.hhs.gov/NPPES/Welcome.do>.
- Chandra A, Dalton MA, Holmes J. Large increases in spending on postacute care in Medicare point to the potential for cost savings in these settings. *Health Aff (Millwood)* 2013;32:864-72. doi:10.1377/hlthaff.2012.12.1262.
- Moja L, Piatti A, Pecoraro V, et al. Timing matters in hip fracture surgery: patients operated within 48 hours have better outcomes. A meta-analysis and meta-regression of over 190,000 patients. *PLoS One* 2012;7:e46175. doi:10.1371/journal.pone.0046175.
- Shoda N, Yasunaga H, Horiguchi H, et al. Risk factors affecting inhospital mortality after hip fracture: retrospective analysis using the Japanese Diagnosis Procedure Combination Database. *BMJ Open* 2012;2:e000416. doi:10.1136/bmjopen-2011-000416.
- Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev* 2004;25:119-41.
- Krumholz HM, Normand ST, Galusha DH, et al. Risk-adjustment models for AMI and HF 30-day mortality. 2007. <http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1163010421830>.
- The Dartmouth Atlas of Health Care. Hospital discharges and post-acute care: hospital-specific data. <http://www.dartmouthatlas.org/tools/downloads.aspx?tab=41>.
- Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA* 1998;280:1690-1. doi:10.1001/jama.280.19.1690.
- Cummings P. Methods for estimating adjusted risk ratios. *Stata J* 2009;9:175-96.
- Centers for Medicare and Medicaid Services. Hospital compare datasets. <https://data.medicare.gov/data/hospital-compare>.
- National Health Service. Performance of hospitals in England. <https://www.nhs.uk/service-search/performance/results?resultsViewId=1015>.
- Huckman RS, Pisano GP. The firm specificity of individual performance: evidence from cardiac surgery. *Manage Sci* 2006;52:473-88doi:10.1287/mnsc.1050.0464.
- Schwartz LM, Woloshin S, Birkmeyer JD. How do elderly patients decide where to go for major surgery? Telephone interview survey. *BMJ* 2005;331:821. doi:10.1136/bmj.38614.449016.DE.

© BMJ Publishing Group Ltd 2016

Appendix