Surgeon Perception of Risk and Benefit in the Decision to Operate

Greg D. Sacks, MD, MPH,*†‡ Aaron J. Dawes, MD,*†‡ Susan L. Ettner, PhD,‡§
Robert H. Brook, MD, ScD,‡§|| Craig R. Fox, PhD,§¶# Melinda Maggard-Gibbons, MD, MSHS,*†
Clifford Y. Ko, MD, MS, MSHS,*† and Marcia M. Russell, MD*†

Objective: To determine how surgeons' perceptions of treatment risks and benefits influence their decisions to operate.

Background: Little is known about what makes one surgeon choose to operate on a patient and another chooses not to operate.

Methods: Using an online study, we presented a national sample of surgeons (N = 767) with four detailed clinical vignettes (mesenteric ischemia, gastrointestinal bleed, bowel obstruction, appendicitis) where the best treatment option was uncertain and asked them to: (1) judge the risks (probability of serious complications) and benefits (probability of recovery) for operative and nonoperative management and (2) decide whether or not they would recommend an operation.

Results: Across all clinical vignettes, surgeons varied markedly in both their assessments of the risks and benefits of operative and nonoperative management (narrowest range 4%–100% for all four predictions across vignettes) and in their decisions to operate (49%–85%). Surgeons were less likely to operate as their perceptions of operative risk increased [absolute difference (AD) = –29.6% from 1.0 standard deviation below to 1.0 standard deviation above mean (95% confidence interval, CI: –31.6, –23.8)] and their perceptions of nonoperative benefit increased [AD = –32.6% (95% CI: –32.8,–28.9)]. Surgeons were more likely to operate as their perceptions of operative benefit increased [AD = 18.7% (95% CI: 12.6, 21.5)] and their perceptions of nonoperative risk increased [AD = 32.7% (95% CI: 28.7, 34.0)]. Differences in risk/benefit perceptions explained 39% of the observed variation in decisions to operate across the four vignettes.

Conclusions: Given the same clinical scenarios, surgeons' perceptions of treatment risks and benefits vary and are highly predictive of their decisions to operate.

Keywords: behavioral science: perception, risk and benefit, surgical decision making, variations

(*Ann Surg* 2016;264:896–903)

From the *Department of Surgery, David Geffen School of Medicine, University of California Los Angeles, CA; †VA Greater Los Angeles Healthcare System, Los Angeles, CA; ‡Department of Health Policy and Management, Fielding School of Public Health, University of California, Los Angeles, CA; §Department of Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA; ||RAND Corporation, Los Angeles, CA; ¶Anderson School of Management, University of California, Los Angeles, CA; and #Department of Psychology, College of Letters and Sciences, University of California, Los Angeles, CA.

Reprints: Greg D. Sacks, MD, MPH, Department of Surgery, UCLA David Geffen School of Medicine, 10833 Le Conte Ave. 72–227 CHS, Los Angeles, CA 90095; E-mail: gsacks@mednet.ucla.edu.

Disclosure: Two of the authors (G. D. S and A. J .D.) received support from the Robert Wood Johnson/Veterans Affairs Clinical Scholars program.

G. D. S. had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. He conducted and is responsible for the data analysis.

The authors declare no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/16/26406-0896
DOI: 10.1097/SLA.0000000000001784

Knowing when to operate and when not to operate is widely considered a fundamental skill for surgeons to master.[1–4] However, decisions whether or not to operate vary substantially between surgeons, as evidenced by the widespread and largely unexplained regional variation in surgical utilization rates.[5–7] Even when patient characteristics are held constant, in the form of clinical vignettes, there remains substantial disagreement among surgeons concerning the role for surgical intervention.[8–11] This persistently unexplained variation has led some researchers to conclude that much of the observed differences in the use of surgery may be attributable to differences in physician judgment about the indications for surgery.[6] However, the basis of such judgment, and its link to surgical decision making has received little attention in the literature to date.

The decision to operate on a patient, like many other clinical decisions, often falls in a discretionary gray area in which the best treatment option is unclear. According to normative decision theory, treatment decisions under uncertainty should be based on an evaluation for each available treatment option of: (i) the probabilities of possible outcomes; and (ii) the relative attractiveness or unattractiveness (ie, the utilities) of these outcomes.[1,12–15] Judging the likelihood of negative outcomes (risks) and the likelihood of desirable outcomes (benefits) is therefore an essential component of surgical decision making. Attempting to understand variation in the use of surgery, we hypothesized that surgeons may vary in their clinical decision making because of, in part, differences in how they perceive the risks and benefits of operative and nonoperative management. That is, we expected that surgeons' decision to operate would be negatively associated with their judgment of operative risk and nonoperative benefit and positively associated with their judgment of operative benefit and nonoperative risk.

To test this hypothesis, we presented a national sample of surgeons with detailed vignettes that were designed to target clinical scenarios without a clearly dominant treatment choice. We then asked surgeons to judge the risks and benefits of operative and nonoperative management and to rate their likelihood of recommending an operation. Understanding how surgeons make this critical decision will create new opportunities for improving patient-centered informed consent, educating surgical trainees on the process of clinical decision making, and reducing unnecessary variations in care.

METHODS

Study Sample

We recruited surgeons via email to participate in an online study. Eligible participants included all members of the American College of Surgeons who had either completed or were currently enrolled in a general surgery residency program. We sent a recruitment email in October 2014, followed by a reminder to nonrespondents in December 2014. As an incentive, individuals were able to participate in a Continuing Medical Education activity after study completion and were invited to enroll in a raffle for one of four laptop computers. Additional research questions (not presented in this study) required participants to be randomized to five subgroups. We excluded three of these subgroups from our current analysis (N = 1113) because they were either exposed to supplemental information that may have biased their responses (eg, data from a risk calculator) or they were not asked to provide responses for all pertinent variables (see Appendix 1, http://links.lww.com/SLA/B24 for full study protocol). The two included groups of surgeons differed only in that they were asked the same questions in a different order. This study was approved by the RAND institutional review board and all participants provided informed consent.

Study Format

We asked participants to review four clinical vignettes that a general surgeon would manage (Fig. 1) and assess the risks and benefits of operative and nonoperative management. The vignettes, developed by a panel of practicing surgeons, were clinical scenarios in which there was a clear diagnosis but no clearly dominant treatment option. We refined the vignettes and the accompanying questions to ensure clinical relevance using multiple rounds of pilot testing with a separate sample of surgeons (n = 26).

Variables

For each vignette, we asked participants to make four judgments: (i) risk of operative management, (ii) benefit of operative management, (iii) risk of nonoperative management, and (iv) benefit of nonoperative management (these predictions are subsequently referred to as risk/benefit parameters and were always elicited in this order). For risks of both operative and nonoperative management, participants were asked to judge the probability (on a percent scale from 0 to 100) that the patient would suffer a serious complication within 30 days. Serious complications were defined in accordance with the American College of Surgeons National Surgical Quality Improvement Program and include the occurrence of at least one of the following events within 30 days of the decision to operate or not operate: cardiac arrest, myocardial infarction, pneumonia, progressive renal insufficiency, acute renal failure, pulmonary embolism, deep vein thrombosis, systemic sepsis, respiratory failure, and urinary tract infection. Risks of operating also included return to the operating room, deep incisional or organ space surgical site infection, or wound disruption. For benefits, surgeons were asked to judge the probability that a patient would recover from their underlying condition within 30 days. Recovery was defined as the patient being free of the immediate threats of the surgical disease process and back to a reasonable level of baseline health. These definitions were available to participants throughout the study to allow for more standardized responses.

To measure surgeons' decisions whether or not to operate, we asked them how likely they were to recommend an operation based on the clinical information provided (5-point scale: 1 "very unlikely", 2 "unlikely", 3 "neutral", 4 "likely", 5 "very likely"). For some analyses, we dichotomized this variable into operate (4 and 5) versus not operate (1, 2, and 3). For these purposes, we considered a "neutral" response to be more analogous to not operating since neutrality in this case is more aligned with inaction than action.

Statistical Analyses

We calculated the response rate in accordance with American Association for Public Opinion Research Standard Definitions Committee.[16] All study participants provided demographic and practice information. To test for nonresponse bias, we compared demographics of all study participants with the general population of surgeons using data from the Association of American Medical Colleges (AAMC).[17] We also compared demographics for participants who responded to the first recruitment email with those who failed to respond to the first email but responded to the second reminder email. We performed these statistical comparisons using χ^2tests.

We examined the distribution of each estimated risk/benefit parameter and calculated relevant summary statistics. We also examined the distribution of surgeons' likelihood of recommending an operation for each clinical vignette.

We first explored the relationship between each risk/benefit parameter and the likelihood of recommending an operation using Spearman rank correlation. Then, to detect the association between surgeons' risk and benefit judgments across all four vignettes, we pooled the data for all of the clinical vignettes, using the surgeon-vignette as the unit of observation. Each observation therefore represented a unique surgeon making four risk/benefit judgments and a single decision to recommend an operation for a unique vignette. Each surgeon thereby appeared in four separate observations. For this analysis, we used a hierarchical logistic regression model to predict the decision to recommend an operation on a binary scale (yes/no) from all four risk/benefit parameters. This model included a random intercept for the surgeon and a dummy variable for each clinical vignette. Because some surgeons were presented the questions in one of two different orders (Appendix 1, http://links.lww.com/SLA/B24; surgeons were randomly assigned to assess risks and benefits either before or after rating their likelihood of recommending an operation), we also included a dummy variable corresponding to the surgeon's study arm. Using this model, we calculated the marginal effect for each risk/benefit parameter. Specifically, we calculated the probabilities that a surgeon judging 1.0 standard deviation below the parameter mean and 1.0 standard deviation above the parameter mean would recommend an operation. We subtracted these two values to obtain an absolute difference between the two groups (risk difference) and calculated the associated confidence intervals using bias-corrected bootstrapping with 1000 repetitions. To calculate what proportion of the variation in the decision to operate was explained by differences in assessed risks and benefits, we used a hierarchical linear regression model that was identical in specification to our logistic model, except with a continuous outcome variable (decision to recommend an operation on a 5-point scale).

Sensitivity Analyses

We performed additional analyses to check the robustness of our results by: (i) adding surgeon demographic variables (number of years in practice, fellowship training, practice type, annual operative volume, etc.) to the model; (ii) excluding surgeons with neutral responses (data for (i) and (ii) not shown); and (iii) analyzing each clinical vignette separately (Appendix 2, http://links.lww.com/SLA/B24). None of these three variations changed our results or conclusions. All statistical analyses were performed using Stata/IC, Version 13.1.

RESULTS

A total of 1880 surgeons participated (13.4% adjusted response rate) and 767 were included in this current study (Table 1). Most participants had completed residency training (84.6%) and the remaining participants were currently enrolled in a residency program. Respondents most commonly worked in academic centers (34.4%), and were males (72.9%) and White (72.0%). Respondents to the first email were demographically similar to respondents to the second email (data not shown) and both groups were similar to the overall population of surgeons according to data from the AAMC, with differences in sex and race/ethnicity being modest but statistically significant (Table 1).

Case 1: Mesenteric ischemia

You consult on a 75 year old male recently admitted to the ICU with chest pain and new-onset heart failure. He was initiated on medical therapy and his ejection fraction improved from 25% to 40%. Yesterday he complained of abdominal pain and a CT scan suggested mesenteric ischemia. He was made NPO and given antibiotics. He now reports moderate diffuse abdominal pain that is getting worse.

PMH: hypertension, insulin dependent diabetes type II (15 year duration)
PSH: none
Meds: insulin (20 units/day), amlodipine
Social history: current smoker (50 pack-year), lives independently

Today:
Vitals: HR 105, BP 98/60, O_2 saturation 95%
Physical exam: Laying in bed, appears somewhat uncomfortable
Crackles at both lung bases, abdomen firm, moderate tenderness in left lower quadrant, no rebound or guarding
Urine output (past 6 hours): 0.3cc/kg/hr
Labs:
WBC 13,600/µL (9,400/µL on admission)
HCO3 21 mEq/L
Creatinine 2.3 mg/dL (baseline 1.1 mg/dL)
Glucose 230
Albumin 2.4 g/dL
Lactate 2.1 mmol/L (normal 0.5-2.2 mmol/L)
CT scan (yesterday): small areas of pneumatosis in the mid-jejunum, no portal venous gas

Case 2: Gastrointestinal bleed

You consult on an 83-year-old male in the ICU who was admitted 2 days ago with bright red blood per rectum. He was treated 3 months ago for bleeding diverticula of descending colon with embolization. On this admission, his hemoglobin was 5.6 g/dL and he has received 8 units of packed red blood cells (as well as fresh frozen plasma/platelets). RBC nuclear scan and angiogram did not identify location of the bleed. Colonoscopy found fresh clots in descending colon only. He passed a moderate amount of bright red blood per rectum one hour ago.

PMH: chronic obstructive pulmonary disease (FEV1 60%), insulin dependent diabetes
Meds: insulin, aspirin, prednisone 10 mg/day, inhaled budesonide
Social history: current smoker (40 pack-year), ambulates with a walker, resides in a skilled nursing facility.

Vitals: HR 116, BP 94/54, O_2 saturation 94% on 4L nasal cannula
Pt. is resting comfortably in the bed
Physical exam: abdomen soft, non-tender, rectal exam with fresh blood clots, unable to visualize bleeding source on anoscopy
Urine output (past 6 hours): 0.5cc/kg/hr

Labs: (now)
Hgb 7.0 g/dL
Platelets 130,000/µL
PT 12.5 seconds
INR 1.3
Creatinine 2.2 mg/dL (baseline 2.0 mg/dL)
Albumin 2.0 g/dL

Case 3: Small bowel obstruction
A 68-year-old female elementary school teacher presented to the emergency room 3 days ago with her first bowel obstruction. She complained of 1 day of abdominal pain, bloating, bilious emesis, and obstipation. A nasogastric tube aspirated 800cc of bilious contents in the emergency room.

PMH: hypertension, hyperlipidemia (both well controlled)
PSH: hysterectomy (20 years ago)

On admission:
Vitals: normal
Physical exam: resting in bed, appeared uncomfortable
minimal, diffuse abdominal tenderness and distention
Labs: **WBC** 11,200/µL
HCO3 21 mEq/L
Creatinine 1.8 mg/dL (baseline 0.9 mg/dL)
Albumin 3.0 g/dL
Lactate 1.5 mmol/L (normal 0.5-2.2 mmol/L)

CT scan: markedly dilated small bowel, air fluid levels, no clear transition point but distal ileum is decompressed, air present in rectum, small amount of free fluid in pelvis, and no pneumatosis.

Now, hospital day three:
Her pain and abdominal exam are unchanged since admission. Her daily nasogastric tube output has been 1,100mL, 850mL, then 750mL, and she has passed no flatus or bowel movement. Laboratory results are unchanged except WBC 10,600/µL. Urine output is 0.3cc/kg/hr.

Case 4: Appendicitis
A 19-year-old otherwise healthy female college student presents to the emergency room complaining of 3 days of right lower quadrant pain. She had intermittent fevers (to 101°F), vomited twice, and has no appetite.

PMH: none
PSH: none
Vitals: T 101.3°F, HR 112, BP 131/74
Physical exam: Laying in bed, appears uncomfortable
Soft abdomen with moderate right lower quadrant tenderness to palpation with rebound.
Labs: **WBC** 16,400/µL
Urinalysis normal
Pregnancy test negative

Pelvic ultrasound: Normal ovaries and fallopian tubes bilaterally.
CT scan: Large (6 cm) phlegmon in the right lower quadrant adjacent to cecum with extensive stranding extending posteriorly to retroperitoneum, moderate free fluid in pelvis, no abscess. The appendix is not well visualized.

FIGURE 1. Description of clinical scenarios included in study.

Across the four vignettes, surgeons varied markedly in their assessments of the risks and benefits of both operative and nonoperative management (Fig. 2). For example, for the appendicitis vignette, surgeons varied considerably in their judgment of the likelihood of serious complication (mean 24%, SD 21%, range 0%–100%) or recovery (mean 86%, SD 19%, range 1%–100%) after operative management and the likelihood of serious complication (mean 31%, SD 27%, range 0%–100%) or recovery (mean 68%, SD 29%, mean 0%–100%) after nonoperative management.

Surgeons also varied markedly in their decisions to recommend an operation (Fig. 3). Although most surgeons recommended an operation for the small bowel obstruction (SBO) case (84%), there

TABLE 1. Characteristics of Study Participants and Available Data on General Population of Surgeons

	Study Participants°		General Population		
	Number	%	Number	%	P
Total	767		48,635		
Sex					<0.001
Male	550	72.9	39,614	81.5	
Female	205	27.2	8,991	18.5	
Race					<0.001
White	543	72.0	25,485	52.4	
Asian	100	13.3	5544	11.4	
Hispanic/Latino	73	9.7	2334	4.8	
Black	13	1.7	2140	4.4	
Hawaiian/PI	1	0.1		–	
American Indian/Alaskan Native	2	0.3	195	0.4	
Other	22	2.9	195	0.4	
Missing	–	–	12,791	26.3	
Level of training					
Attending	646	84.6	–	–	
Resident	118	15.5	–	–	
Practice type					
Academic	223	34.7	–	–	
Private	185	28.8	–	–	
Community	174	27.1	–	–	
VA	18	2.8	–	–	
County	13	2.0	–	–	
Military	8	1.2	–	–	
Other	22	3.4	–	–	
Fellowship*					
None	274	43.2	–	–	
ACS/ICU/Burns	88	13.9	–	–	
Other	284	43.0	–	–	
Practice general surgery*					
Routine	452	70.7	–	–	
Occasional	97	15.2	–	–	
Rarely	90	14.1	–	–	
Cases per year*					
<50	21	2.8	–	–	
51–100	54	7.1	–	–	
101–250	197	26.0	–	–	
251–500	294	38.8	–	–	
>500	75	9.9	–	–	
Residency graduation year*					
2010 or later	123	16.4	–	–	
2000–2010	171	22.7	–	–	
Before 2000	342	45.5	–	–	
Residency years completed§					
0 (first year)	15	12.9	–	–	
1	17	14.7	–	–	
2	20	17.2	–	–	
3	38	32.8	–	–	
4	26	22.4	–	–	

ACS indicates acute care surgery; ICU, intensive care unit; PI, pacific islander; VA, veterans affairs.
*Applies only to surgeons who have completed a residency.
§Applies only to surgeons currently enrolled in a general surgery residency. Numbers that do not add up to the total are the result of missing data ° Data on general population of surgeons obtained from the Association for American Medical Colleges.

was more disagreement for the other cases [67% recommended an operation for mesenteric ischemia, 54% for gastrointestinal bleed (GIB), and 49% for appendicitis].

In unadjusted analyses, there was a significant but weak (Spearman $\rho < 0.4$) correlation between the (binary) decision to recommend an operation and judgments of operative risks and benefits. There was a more moderate correlation (Spearman $\rho > 0.4$) between the decision to recommend an operation and judgments of nonoperative risks and benefits (Appendix 3, http://links.lww.com/SLA/B24). These correlations suggest that surgeons are more likely to recommend an operation when they believe the risks of operating are lower, the benefits of operating are higher, the risks of not operating are higher, and the benefits of not operating are lower, as one would expect.

When all four risk/benefit judgments were included in a regression model, there remained a significant association between all four judgments of risk and benefit and the (binary) decision to recommend an operation (Table 2). Surgeons were significantly less likely to recommend an operation as their judgment of the risks of operating increased. Averaged across all vignettes, mean judged

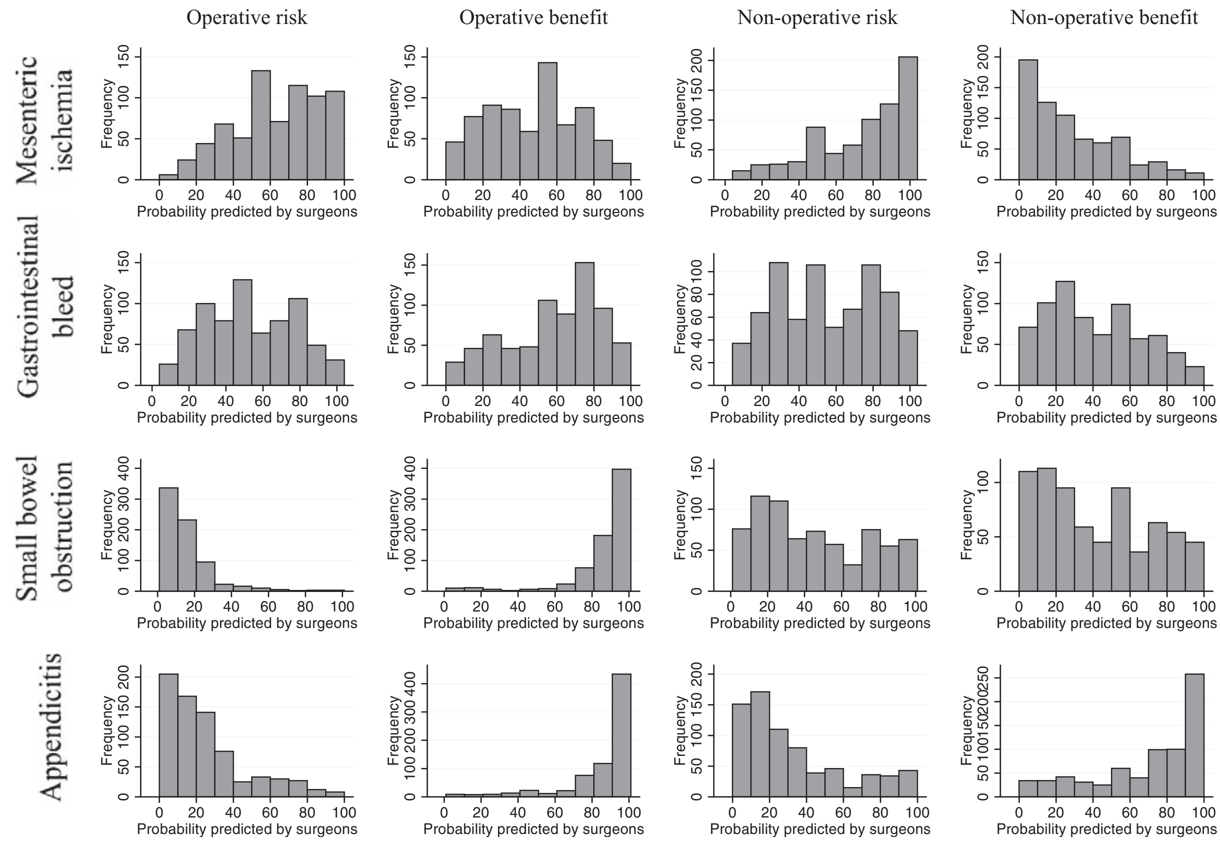

FIGURE 2. Variation in surgeons' judgment of operative and nonoperative risk and benefit. Footnote: Each histogram shows the number of surgeons (n = 767 total) predicting each value of risk or benefit. The x-axis, which is divided into intervals of 10, refers to surgeons' judged likelihood of risk (the occurrence of serious complication) or benefit (patient recovery) following operative or nonoperative management. Risks represent the surgeons' judged probability that a patient would suffer a serious complication, which was defined in accordance with the American College of Surgeons National Surgical Quality Improvement Program and include the occurrence of at least one of the following within 30 days of the decision to operate or not operate: cardiac arrest, myocardial infarction, pneumonia, progressive renal insufficiency, acute renal failure, pulmonary embolism, deep vein thrombosis, systemic sepsis, respiratory failure, and urinary tract infection. Risks of operating also included return to the operating room, deep incisional, or organ space surgical site infection, or wound disruption. Benefits represent the surgeons' judged probability that a patient would recover, which was defined as the patient being free of the immediate threats of the surgical disease process and back to a reasonable level of baseline health within 30 days.

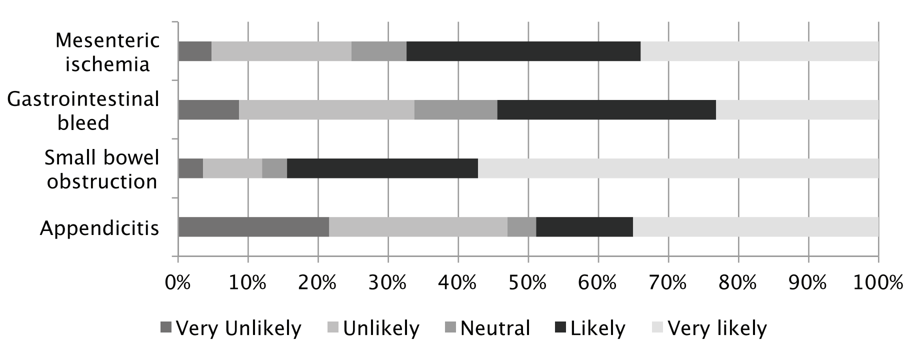

FIGURE 3. Variation in surgeon response to ''how likely are you to recommend an operation?''. Each bar is broken up into proportion of surgeons (n = 767 total) who responded 1 ''Very unlikely'' 2 ''Unlikely'' 3 ''Neutral'' 4 ''Likely'' 5 "Very likely''.

TABLE 2. Association Between Surgeon Perception of Treatment Risks and Benefits and the Decision to Recommend an Operation (n = 767 surgeons)

Risk Benefit Parameter	Surgeons' Judged Probability, Mean (Standard Deviation)	Probability That Surgeon at 1 Standard Deviation Below Mean Recommends an Operation	Probability That Surgeon at 1 Standard Deviation Above Mean Recommends an Operation	Absolute Difference in Probability That Surgeon Recommends an Operation*	95% Confidence Interval
Operative risk	38.6 (28.4)	78.4	48.8	−29.6	(−31.6, −23.8)
Operative benefit	68.4 (28.2)	55.0	73.7	18.7	(12.6, 21.5)
Nonoperative risk	51.5 (31.0)	48.3	81.0	32.7	(28.7, 34.0)
Nonoperative benefit	44 (30.4)	81.6	49.0	−32.6	(−32.8, −28.9)

*Absolute difference (risk difference) calculated by subtracting probability that surgeon at 1 standard deviation above the mean recommended an operation from the probability that surgeon a 1 standard deviation below the mean recommended an operation.

Model predicts the surgeon's decision to recommend an operation, when controlling for each surgeon's judgment of the risks and benefits of operative and nonoperative management. Model also includes dummy variable for each of the four clinical vignettes and for the experimental arm to which the surgeon was randomly assigned.

operative risk was 38.6% (SD = 28.4%). Surgeons judging operative risk at 1.0 SD below the mean (38.6%–28.4% = 10.2%) would have a 78.4% predicted probability of recommending an operation, whereas surgeons judging operative risks at 1.0 SD above the mean (38.6%+28.4% = 67.0%) would have a 48.8% predicted probability of recommending an operation. Thus, surgeons judging operative risk at 1.0 SD above the mean would be 29.6% points less likely to recommend an operation than surgeons judging operative risk at 1.0 SD below the mean [95% CI −31.6, −23.8)]. Conversely, surgeons were significantly more likely to recommend an operation as their judgment of the benefits of operating increased [AD = 18.7% (95% CI 12.6, 21.5)]. Surgeons were also significantly more likely to recommend an operation as their judgment of the risks of not operating increased [AD = 32.7% (95% CI 28.7, 34.0)]. Finally, surgeons were significantly less likely to recommend an operation as their judgment of the benefits of not operating increased [AD = −32.6% (95% CI −32.8, −28.9)].

In sum, differences in surgeons' judgments of the risks and benefits of operating and not operating explained a substantial amount of the observed variation in the decision to recommend an operation ($R^2 = 0.39$).

DISCUSSION

Using a large national sample of surgeons and four common clinical scenarios for which there was no dominant treatment option, we found wide variation in surgeons' perception of treatment risks and benefits. Moreover, these marked differences in surgical judgment strongly predicted surgeons' clinical decisions. As hypothesized, surgeons were more likely to recommend an operation if they perceived the risks of operating to be low and the benefits of operating to be high and were more likely to recommend against an operation if they perceived the risks of nonoperative management to be low and its benefits to be high. Collectively, differences in how surgeons perceive these risks and benefits explained 39% of the variation in surgeons' decision to recommend an operation.

Previous research has identified regional variations in the utilization rates of surgical procedures that are otherwise unexplained by differences in disease prevalence or other clinical markers.[5–7] In fact, such variation in the use of surgery is still present, even when patient characteristics are held constant in the form of clinical vignettes.[8–11] Until now, the reasons for this variation in clinical decision making have remained largely unknown. By analyzing surgeons' decision-making process using a series of clinical vignettes, we have provided evidence that variation in decision-making may be largely attributable to differences in how surgeons perceive the risks and benefits of operating and not operating. Considering these differences, surgeons appear, at least on average, to choose treatments that are aligned with their expectations for which treatment optimizes the patient's utility by maximizing the benefits and minimizing the harms. What varies then is surgeons' judgment of the likelihood of the possible treatment outcomes.

Normative decision theory stipulates that treatment decisions should be based on an evaluation of the probability of different outcomes for each treatment option along with their relative attractiveness or unattractiveness.[1,12–15] To the extent that variation in treatment decisions is driven by differences in (perceived) patient preferences over possible outcomes (eg, differences in how patients might feel about a long and painful recovery), then the observed variation in the use of surgery may be justifiable. However, such differences in treatment decisions become more difficult to rationalize if they are based on unrealistic variation in judged probabilities of possible outcomes, as we observed in this study. Actual surgical complication rates differ only modestly from one surgeon to the next after accounting for patient characteristics and statistical noise.[18] Therefore, addressing the substantial differences in surgeons' probability assessments of treatment outcomes may offer a novel approach to reducing some of the observed variations in care. This said our study measured only the perceived probability of outcomes and not their relative attractiveness, which will be an important topic for future research.

Assessment of patients' operative risk has been a hallmark of surgical practice for almost half a century[19] and is now a feature of hospital benchmarking algorithms and pay-for-performance programs.[20,21] Surgeons have even taken the lead in incorporating such formal risk assessment into routine clinical decision making.[22] But compared with the advances made in assessing a patient's operative risk, far less work has been done on the concept of operative benefit. In fact, for many procedures, the concept of operative benefit lacks a well-accepted definition, making it difficult to incorporate such a measure into routine clinical care. Nevertheless, our findings suggest that surgeons do incorporate their judgment of this outcome into their clinical decisions. This highlights a need to establish a practical measure for operative benefit so that it can meaningfully inform clinician judgment, and can be used when counseling patients. Such a measure will likely differ depending on the specific operation but will need to encompass the postoperative outcomes that are most important to patients.[23]

Our findings further demonstrate that the decision to recommend an operation depends not only on how surgeons judge the risks and benefits of operating, but also on how they judge the risks and benefits of not operating. In fact, our findings suggest that decisions on whether or not to operate were related more closely to judgments about nonoperative outcomes than operative outcomes. This said, defining and quantifying the risks and benefits associated with nonoperative management is much more challenging than defining and quantifying operative risks and benefits, primarily because of the paucity of data on patients who undergo medical management of potentially surgical conditions. This is not surprising given that it is difficult to use administrative data to accurately identify patients who were managed nonoperatively but might otherwise have been managed surgically. Furthermore, surgical registries typically track outcomes for patients who undergo surgery and do not keep track of patients managed nonoperatively.[24] However, in light of the observed marked variation in surgeons' perceptions of nonoperative outcomes and how strongly these perceptions predict treatment decisions, it will be important moving forward to develop ways to reliably measure such outcomes.

There are several limitations to our study. First, surgeons may respond differently in hypothetical clinical scenarios compared with real practice settings. However, in at least two studies, responses to clinical vignettes do appear to be reliable proxies for real-world decisions.[25,26] In addition, vignette-based studies have also revealed important biases in clinical decision making related to patient's race and sex, which may also reflect real clinical phenomena.[27] Second, our results leave room for omitted variables that may contribute to decisions whether or not to operate beyond the influence of judged outcome probabilities, such as local culture, financial incentives, and risk preferences. Furthermore, patients' treatment preferences were not included in this study and we are therefore unable to determine whether differences in perceived patient preference contributed to the observed variation in decision making. Third, as with any vignette-based study, it is unclear the extent to which our findings generalize beyond the particular clinical vignettes included in our study. Fourth, although we have emphasized the association between judged risks and benefits and the decision to operate, we cannot rule out the possibility that causality runs in both directions. That is, it is possible that some surgeons inferred risks and benefits from their clinical decisions rather than the other way around. However, we note that correlations between judgments of risks and benefits of operating versus not operating are relatively low (median Spearman correlation for operative versus nonoperative risks = 0.34; median Spearman correlation for operative versus nonoperative benefits = 0.20), which appears to mitigate some, though not all, of the concern for reverse causality. Our findings thus point to an association but do not prove that risk/benefit perceptions causally influence decisions. Future studies might incorporate experimental designs that independently manipulate levels of (operative and nonoperative) risk and benefit in order to make a more convincing case for such a causal link. Finally, a relatively small proportion of recruited surgeons completed our study, which may limit the generalizability of our findings. However, it is typical for online studies to have low response rates, particularly among surgeons.[28] Furthermore, our data include a large and fairly representative sample, which may partly allay concern about the generalizability of our findings.[29]

We believe that our study has important implications for research and practice. First, our results provide preliminary evidence that judgments of both operative and nonoperative risk and benefit may jointly influence surgeons' treatment recommendations. This framework may open new avenues for further evaluating surgical decision-making and suggest that these components of surgical judgment might be incorporated into surgical training. Specifically, future research should focus on how surgical decision making is influenced not only by differences in risk perception, but also by differences in risk tolerance and how surgeons view the relative attractiveness or unattractiveness of different outcomes (ie, utilities). Furthermore, this research will need to also move beyond 30-day outcomes and include other important outcomes that matter most to patients, including long-term outcomes, functional status, and quality of life.[23] Second, assuming surgeons' disparate risk and benefit assessments in this study reflect their true clinical impressions, our findings suggest that patients who are counseled for surgery may receive different information about risks and benefits of available treatments that differ markedly depending on which surgeon they consult. Use of objective data, when available, may therefore be an effective way to reduce some of this variation in care.[22] Such data could be shared with patients to ensure that their decisions are based on accurate probabilities so that they can make better informed decisions that are congruent with their values and preferences.[30,31]

CONCLUSIONS

In a series of clinical vignettes, surgeons vary in their perceptions of the risks and benefits associated with operative and nonoperative management. These differences in risk and benefit assessments were closely associated with surgeons' treatment decisions and explained a large proportion of the observed variance in decision making. Surgical decision making may be enhanced by collecting data on risks and benefits for all available treatment options so that those data can one day be explicitly incorporated into the decision-making process.

ACKNOWLEDGMENTS

The authors thank David Hoyt, Carlos Pellegrini, and Patricia Turner, and the American College of Surgeons for their tremendous support on this project and help in recruiting study participants, Sandra Berry for her expertise in designing the online study instrument, David Asch for his assistance in formulating and developing a testable research hypothesis, and Thomas Rice for his support and revisions to the manuscript.

REFERENCES

1. Eddy DM. Clinical decision making: from theory to practice. Anatomy of a decision. *JAMA*. 1990;263:441–443.
2. Hall JC, Ellis C, Hamdorf J. Surgeons and cognitive processes. *Br J Surg*. 2003;90:10–16.
3. Yule S, Flin R, Paterson-Brown S, et al. Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery*. 2006;139:140–149.
4. Szatmary P, Arora S, Sevdalis N. To operate or not to operate? A multi-method analysis of decision-making in emergency surgery. *Am J Surg*. 2010;200:298–304.
5. Wennberg J, Gittelsohn. Small area variations in health care delivery. *Science*. 1973;182:1102–1108.
6. Birkmeyer JD, Reames BN, McCulloch P, et al. Understanding of regional variation in the use of surgery. *Lancet*. 2013;382:1121–1129.
7. The Dartmouth Institute of Health Policy. Practice C. *The Dartmouth Atlas of Healthcare*. Hanover, NH: Dartmouth Medical School; 2013.
8. Wilson NP, Wilson FP, Neuman M, et al. Determinants of surgical decision making: a national survey. *Am J Surg*. 2013;206:970–978.
9. Rutkow IM, Starfield BH. Surgical decision making and operative rates. *Arch Surg*. 1984;119:899–905.
10. Rutkow IM, Gittelsohn AM, Zuidema GD. Surgical decision making. The reliability of clinical judgment. *Ann Surg*. 1979;190:409–419.
11. Rutkow IM. Surgical decision making the reproducibility of clinical judgment. *Archives of Surgery*. 1982;117:337–340.
12. Hunink MGM, Weinstein MC, Wittenberg E, et al. *Decision Making in Health and Medicine*. Cambridge: Cambridge University Press; 2014.
13. Birkmeyer JD, Birkmeyer NO. Decision analysis in surgery. *Surgery*. 1996;120:7–15.
14. Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med*. 1975;293:229–234.

15. Winterfeldt Von D, Edwards W. *Decision Analysis and Behavioral Research.* Cambridge: Cambridge University Press; 1986.
16. The American Association for Public Opinion Research. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 7th ed. AAPOR; 2011.
17. Association of American Medical Colleges. *Diversity in the Physician Workforce: Facts & Figures 2014.* Available at: aamcdiversityfactsandfigures.org. Accessed February 1, 2016.
18. Shih T, Cole AI, Al-Attar PM, et al. Reliability of surgeon-specific reporting of complications after colectomy. *Ann Surg.* 2015;261:920–925.
19. Neuman MD, Bosk CL. What we talk about when we talk about risk: refining surgery's hazards in medical thought. *Milbank Q.* 2012;90:135–159.
20. Dimick JB, Osborne NH, Hall BL, et al. Risk-adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg.* 2010;210:503–508.
21. Kansagara D, Englander H, Salanitro A, et al. Risk prediction models for hospital readmission: a systematic review. *JAMA.* 2011;306:1688–1698.
22. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. *J Am Coll Surg.* 2013;217:833–842.
23. Schwarze ML, Brasel KJ, Mosenthal AC. Beyond 30-day mortality: aligning surgical quality with outcomes that patients value. *JAMA Surgery.* 2014;149: 631–632.
24. Birkmeyer JD, Shahian DM, Dimick JB, et al. Blueprint for a New American College of Surgeons: National Surgical Quality Improvement Program. *J Am Coll Surg.* 2008;207:777–782.
25. Peabody JW, Luck J, Glassman P, et al. Comparison of vignettes, standardized patients, and chart abstraction: a prospective validation study of 3 methods for measuring quality. *JAMA.* 2000;283:1715–1722.
26. Cutler D, Skinner J, Stern AD, Wennberg D. Physician beliefs and patient preferences: a new look at regional variation in health care spending. National Bureau of Economic Research Working Paper Series. August 2013. Available at: http://www.nber.org/papers/w19320.pdf?new_window = 1. Accessed on June 25, 2015.
27. Schulman KA, Berlin JA, Harless W, et al. The effect of race and sex on physicians' recommendations for cardiac catheterization. *N Engl J Med.* 1999;340:618–626.
28. Thoma A, Cornacchi SD, Farrokhyar F, et al., Evidence-Based Surgery Working Group. How to assess a survey in surgery. *Can J Surg.* 2011;54: 394–402.
29. Livingston EH, Wislar JS. Minimum response rates for survey research. *Arch Surg.* 2012;147:110.
30. Nabozny MJ, Kruser JM, Steffens NM, et al. Constructing high-stakes surgical decisions: it's better to die trying. *Ann Surg.* 2016;263:64–70.
31. Pecanac KE, Kehler JM, Brasel KJ, et al. It's Big Surgery. *Ann Surg.* 2014;259:458–463.