Impact of a Risk Calculator on Risk Perception and Surgical Decision Making

A Randomized Trial

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

Objective: The aim of this study was to determine whether exposure to data from a risk calculator influences surgeons' assessments of risk and in turn, their decisions to operate.

Background: Little is known about how risk calculators inform clinical judgment and decision-making.

Methods: We asked a national sample of surgeons to assess the risks (probability of serious complications or death) and benefits (recovery) of operative and nonoperative management and to rate their likelihood of recommending an operation (5-point scale) for 4 detailed clinical vignettes wherein the best treatment strategy was uncertain. Surgeons were randomized to the clinical vignettes alone (control group; n = 384) or supplemented by data from a risk calculator (risk calculator group; n = 395). We compared surgeons' judgments and decisions between the groups.

Results: Surgeons exposed to the risk calculator judged levels of operative risk that more closely approximated the risk calculator value (RCV) compared with surgeons in the control group [mesenteric ischemia: 43.7% vs 64.6%, $P < 0.001$ (RCV = 25%); gastrointestinal bleed: 47.7% vs 53.4%, $P < 0.001$ (RCV = 38%); small bowel obstruction: 13.6% vs 17.5%, $P < 0.001$ (RCV = 14%); appendicitis: 13.4% vs 24.4%, $P < 0.001$ (RCV = 5%)]. Surgeons exposed to the risk calculator also varied less in their assessment of operative risk (standard deviations: mesenteric ischemia 20.2% vs 23.2%, $P = 0.01$; gastrointestinal bleed 17.4% vs 24.1%, $P < 0.001$; small bowel obstruction 10.6% vs 14.9%, $P < 0.001$; appendicitis 15.2% vs 21.8%, $P < 0.001$). However, averaged across the 4 vignettes, the 2 groups did not differ in their reported likelihood of recommending an operation (mean 3.7 vs 3.7, $P = 0.76$).

Conclusions: Exposure to risk calculator data leads to less varied and more accurate judgments of operative risk among surgeons, and thus may help inform discussions of treatment options between surgeons and patients. Interestingly, it did not alter their reported likelihood of recommending an operation.

Keywords: risk calculator, risk perception, surgical decision-making, variations in care

(*Ann Surg* 2016;264:889–895)

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

C.Y.K. is employed by the American College of Surgeons. The other authors report no conflicts of interest.

G.D.S. and A.J.D. received support from the Robert Wood Johnson Clinical Scholars program.

Funding sources: Two of the authors (G.D.S. and A.J.D.) received support from the Robert Wood Johnson Clinical Scholars program.

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).

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

ISSN: 0003-4932/16/26406-0889
DOI: 10.1097/SLA.0000000000001750

All surgical procedures confer some possibility of a significant negative outcome. Depending on the specific procedure or patient profile, these outcomes range from minor complications to severe adverse events, such as major disability or death. The challenge in making clinical decisions lies not only in determining what adverse events are possible, but also in estimating how likely each is to occur.[1–3] To do so, surgeons depend in part on the published literature[4]; however, these data are often aggregated over populations of heterogeneous patients, making it difficult to assess risk for individual cases. Thus, to differentiate risk for individual patients, surgeons must rely on their own intuition and experience and as a result, surgeons can vary markedly in their assessments.[5–10]

To assist surgeons in assessing an individual patient's operative risk (defined here as the probability of an adverse event from surgery), the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) developed a surgical risk calculator, and made it available as an online tool.[11] The calculator uses national registry data to estimate the likelihood of postoperative complications as a function of the procedure type, patient demographics, and a patient's current health state. When tested on a large national sample of patients, the calculator exhibited high accuracy across a wide range of operations.[10] Although the calculator was not designed to determine the appropriateness of an operation for a particular patient, it can inform both surgeons and patients on operative risk as part of a shared decision-making process.

Although risk calculators have been developed for not only surgery but also other medical specialties,[12–15] their impact on physicians' clinical assessments has not yet been tested. We therefore conducted an experimental study using a national sample of surgeons to determine how data from a risk calculator influences surgeons' risk judgments and their associated treatment decisions. We hypothesized that exposure to risk calculator data would lead to judgments that more closely approximate the risk calculator values and vary less between surgeons. We further conjectured that exposure to the risk calculator data might, in turn, influence surgeons' decision to operate.

METHODS

Participant Recruitment and Incentives

We recruited study participants using an email invitation sent to all general surgeon members of the American College of Surgeons

(ACS) who had either completed or were currently enrolled in a general surgery residency program. We sent an initial email in October 2014, and sent a reminder to nonrespondents in December 2014. To encourage participation, we linked the study to an online Continuing Medical Education activity and offered to enroll respondents in a raffle to win 1 of 4 laptop computers. Participants were demographically comparable with the national population of surgeons and included surgeons from each of the 50 states, although our sample had a greater proportion of female and white surgeons compared with national averages (Appendix 1, http://links.lww.com/SLA/B11). The RAND Institutional Review Board reviewed and approved the study and all participants provided informed consent.

Study Design

We designed 4 detailed clinical vignettes that we expected would yield differences in opinion concerning whether or not an operation was indicated. Each vignette described an urgent inpatient general surgery scenario [mesenteric ischemia, gastrointestinal bleed (GIB), small bowel obstruction (SBO), appendicitis; Appendix 2, http://links.lww.com/SLA/B11] and provided clinical information needed to make a treatment decision. The vignettes were developed by a panel of general surgeons and were pilot-tested using a separate sample of surgeons ($n = 26$) and revised iteratively to ensure clinical relevance. During this pilot testing, we modified the scenarios to ensure substantial variation in surgeons' treatment recommendations.

For each vignette, we asked participants to make 4 sequential judgments: risks of operating, benefits of operating, risks of nonoperative management, and benefits of nonoperative management. For risks, we asked surgeons to assess the likelihood (on a scale from 0% to 100%, wherein only integer responses were allowed) of 2 separate adverse outcomes—death and serious complication—within 30 days of either operating or not operating on the patient. Because the actual probability of death for the SBO and appendicitis cases is exceedingly low (<1%), we asked participants to predict only the likelihood of serious complications for these cases. For benefits, we asked surgeons to assess the probability that a patient would recover within 30 days. Definitions of serious complication and recovery were available to participants throughout the study (Table 2 footnote). Finally, we asked surgeons how likely they would be to recommend an operation for each vignette (5-point rating scale: 1 "very unlikely," 2 "unlikely," 3 "neutral," 4 "likely," 5 "very likely"). We presented the vignettes in a random order for each participant.

Randomization

We randomized participants to the clinical vignettes either alone (control group) or alongside data from the ACS-NSQIP risk calculator (risk calculator group; Appendix 3, http://links.lww.com/SLA/B11). Before the study, we entered the relevant clinical and demographic details for each clinical vignette into the risk calculator and obtained the predicted probability of death (for the mesenteric ischemia and GIB vignettes) and serious complication (for all 4 cases), which we provided to the risk calculator group. For example, for the mesenteric ischemia vignettes, surgeons in the risk calculator group were informed: "The American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) has developed a risk calculator that estimates the chance of an unfavorable outcome (such as a complication or death) within 30 days after surgery. The estimates are calculated using data from a large number of patients who had a similar surgical procedure. The ACS-NSQIP risk calculator predicts the following probabilities for this patient after surgery: Death: 12%, Serious complication: 25%." Other than this additional text, the 2 study arms were identical. Randomization between participants was performed automatically by the online software package, so that each participant saw all 4 vignettes either with or without the risk calculator data.

Statistical Analysis

We first assessed the effectiveness of our randomization by comparing demographic information provided by participants between the risk calculator and control groups using χ^2 tests. We compared unadjusted risk/benefit judgments as well as the likelihood of recommending an operation (5-point Likert scale) between the 2 groups using Mann-Whitney U tests. We also compared variation between the 2 groups in predictions using the F-test for homogeneity of variances. As each clinical vignette required additional hypothesis testing, we adjusted the threshold for statistical significance using a Bonferroni correction, thereby requiring a P value <0.0125.

We then tested the risk calculator's average effect across all vignettes by pooling the data for all of the clinical vignettes and using the surgeon-vignette as the unit of observation. We used separate models to compare surgeons' assessment of operative risk, operative benefit, nonoperative risk, nonoperative benefit, as well as their stated likelihood of recommending an operation. For these analyses, we used hierarchical linear regression models that included a random intercept for the surgeon, dummy variables corresponding to the 4 clinical vignettes, and another dummy variable for the surgeon's random group assignment. Using this latter variable, we calculated the marginal effects of the risk calculator for each outcome of interest and calculated both the absolute (percentage point) and relative (percent) difference between the 2 groups. Statistical analyses were performed using Stata/IC, Version 13.1 (Stata Corp, College Station, TX).

RESULTS

A total of 779 surgeons participated in the experiment, of which 384 were randomly assigned to the control group and 395 to the risk calculator group. The majority of participants had completed residency training (86.4%), were male (73.5%), and worked in private practice (50.4%). Before the study, 55.5% of the participants were aware of the ACS-NSQIP risk calculator and of these, over half reported having used the calculator in their practice (44.9% occasionally and 5.9% routinely). There were no statistically significant differences between the risk calculator and control groups on demographic characteristics (Table 1).

Surgeons in both the risk calculator and control groups varied considerably in their judgments of operative risk; however, surgeons in the risk calculator group clustered their predictions around the risk calculator value that had been provided to them (Fig. 1). Judged probabilities of operative risk (serious complication) among surgeons exposed to the risk calculator were significantly lower from those of surgeons in the control group, and more closely approximated the risk calculator value (RCV) [means: mesenteric ischemia (RCV = 25%): 43.7% vs 64.6%, $P < 0.001$; gastrointestinal bleed (RCV = 38%): 47.7% vs 53.4%, $P < 0.001$; small bowel obstruction (RCV = 14%): 13.6% vs 17.5%, $P < 0.001$; appendicitis (RCV = 5): 13.4% vs 24.4%, $P < 0.001$; Table 2]. Surgeons exposed to the risk calculator also varied less in their assessments of operative risk than those in the control group (standard deviations: mesenteric ischemia 20.2% vs 23.2%, $P = 0.01$; gastrointestinal bleed 17.4% vs 24.1%, $P < 0.001$; small bowel obstruction 10.6% vs 14.9%, $P < 0.001$; appendicitis 15.2% vs 21.8%, $P < 0.001$).

Interestingly, although the risk calculator influenced surgeons' assessments of risks following an operation, it did not significantly change their likelihood of recommending an operation. For all 4 vignettes, the average response to the question "how likely are you to recommend an operation as opposed to pursuing medical

TABLE 1. Surgeon Characteristics According to Random Exposure to Risk Calculator Data

	Total (n = 779)	Control Group (n = 384)	Risk Calculator Group (n = 395)	*P*
	n (%)			
Level of training				0.321
Attending	673 (86.4)	327 (85.2)	346 (87.6)	
Resident	106 (13.6)	57 (14.8)	49 (12.4)	
Practice type				0.323
Academic	325 (42.2)	150 (39.1)	175 (44.2)	
Private	388 (50.4)	201 (52.3)	187 (47.2)	
Other	57 (7.4)	33 (8.6)	34 (8.6)	
Sex				0.925
Female	203 (26.5)	100 (26.3)	103 (26.6)	
Male	564 (73.5)	280 (73.7)	284 (73.4)	
Fellowship*				0.400
None	280 (35.9)	134 (34.9)	146 (36.9)	
ACS/ICU/burns	101 (12.9)	45 (11.7)	56 (14.1)	
Other	399 (51.2)	205 (53.4)	194 (49)	
Residency graduation year*				0.364
2010 or later	136 (20.5)	66 (20.3)	70 (20.6)	
2000–2010	172 (25.9)	92 (28.3)	80 (23.6)	
Before 2000	356 (53.6)	167 (51.4)	189 (55.8)	
Residency years completed†				0.115
0 (first year)	13 (12.3)	5 (8.8)	8 (16.3)	
1	14 (13.2)	10 (17.5)	4 (8.2)	
2	29 (27.4)	14 (24.6)	15 (30.6)	
3	36 (34)	17 (29.8)	19 (38.8)	
4	14 (13.2)	11 (19.3)	3 (6.1)	
Aware of risk calculator before study?				0.233
Yes	406 (55.5)	190 (53.2)	216 (57.6)	
No	326 (44.5)	167 (46.8)	159 (42.4)	
Use risk calculator in clinical practice?‡				0.676
Never	199 (49.1)	97 (51.1)	102 (47.4)	
Occasionally	182 (44.9)	81 (42.6)	101 (47)	
Routinely	24 (5.9)	12 (6.3)	12 (5.6)	

ACS, acute care surgery; ICU, intensive care unit; PI, Pacific Islander; VA, Veterans Affairs. Some of the numbers do not add up to the total sample size because of missing data.
*Applies only to surgeons who have completed a residency.
†Applies only to surgeons currently enrolled in a general surgery residency.
‡Applies only to surgeons who were aware of the risk calculator before the study.

management?'' was similar between surgeons in the risk calculator group and the control group [Fig. 2; mesenteric ischemia (3.8 vs 3.7, $P = 0.59$), gastrointestinal bleed (3.4 vs 3.6, $P = 0.05$), small bowel obstruction (4.4 vs 4.3, $P = 0.45$), appendicitis (3.3 vs 3.1, $P = 0.26$)].

Averaged across all 4 vignettes in a hierarchical regression model, surgeons exposed to the risk calculator judged lower risks of operating (−10.4 percentage points, $P < 0.001$; Table 3) than those in the control group. Furthermore, although the risk calculator provides no information on the benefits of operating or the risks and benefits of not operating, surgeons exposed to risk calculator data also judged lower risks of not operating (−5.6 percentage points, $P < 0.001$) and higher benefits of not operating (+3.4 percentage points, $P = 0.005$) compared with surgeons in the control group. There was also a nonsignificant trend toward higher operative benefits judged by surgeons in the risk calculator group (+2.2 percentage points, $P = 0.059$). Across all 4 vignettes, there was no significant difference between the 2 groups in surgeons' self-reported likelihood that they would recommend an operation (3.7 vs 3.7, $P = 0.76$).

DISCUSSION

In this experiment, we found that surgeons exposed to data from a risk calculator subsequently estimated lower probabilities of adverse surgical outcomes that more closely approximated risk calculator values compared with surgeons who were not exposed to such data. Although we had anticipated that the reductions in perceived risk among surgeons exposed to risk calculator data might, in turn, lead to an increased tendency for these surgeons to recommend an operation, we observed no such tendency. One potential explanation is that exposure to the risk calculator was also associated with significantly lower assessments of nonoperative risk and significantly higher assessments of nonoperative benefit. These associations were unexpected, given that the risk calculator provides data only for the risks of operating. However, it appears that this concomitant shift in judged probabilities of nonoperative outcomes offset the effect of reducing perceived operative risk.

Accurate assessment of a patient's operative risk has historically been limited by the lack of reliable data.[4] As a result, surgeons frequently rely on their own intuition when predicting both the magnitude and likelihood of potential adverse outcomes. Previous studies have examined these intuitions by comparing surgeons' own estimates of risk parameters with those derived from validated surgical risk assessment tools. These studies have drawn contradictory conclusions, with some suggesting high concordance between surgeons' probability estimates of adverse outcomes and those derived from risk assessment tools,[16–18] and others revealing systematic over- or underestimation by physicians compared with the output of these risk assessment tools.[19,20]

Our study further reveals both the strengths and limitations of surgeons' intuitive predictions. For 2 assessments, surgeons in both

TABLE 2. Surgeon Judgment of Risks and Benefits for Operate and Nonoperative Management With and Without Risk Calculator

Risk/Benefit Parameter	Risk Calculator Value	Control Group (n = 384)	Risk Calculator Group (n = 395)	P*	Control Group (n = 384)	Risk Calculator Group (n = 395)	P*
	Percent	Mean, %			Standard Deviation, %		
Mesenteric ischemia							
Operative management							
Death	12	42.5	23.2	<0.001	24.1	15.8	<0.001
Serious complication	25	64.6	43.7	<0.001	23.2	20.2	0.01
Recovery	—	42.8	54.7	<0.001	23.3	23.7	0.80
Nonoperative management							
Death	—	66.1	55.6	<0.001	28.3	30.2	0.23
Serious complication	—	75.8	65.5	<0.001	24.4	27.5	0.02
Recovery	—	26.3	31.0	<0.001	22.9	24.1	0.37
Gastrointestinal bleed							
Operative management							
Death	26	27.7	26.7	0.05	20.2	12.8	<0.001
Serious complication	38	53.4	47.7	<0.001	24.1	17.4	<0.001
Recovery	—	56.0	54.2	0.36	23.7	23.9	0.85
Non-operative management							
Death	—	43.4	42.7	0.92	26.8	24.6	0.10
Serious complication	—	59.3	56.9	0.15	25.8	24.8	0.44
Recovery	—	35.8	38.4	0.08	23.6	23.1	0.66
Small bowel obstruction							
Operative management							
Serious complication	14	17.5	13.6	<0.001	14.9	10.6	<0.001
Recovery	—	85.7	83.6	0.32	18.0	24.0	<0.001
Non-operative management							
Serious complication	—	49.7	46.5	0.13	29.1	29.7	0.72
Recovery	—	36.2	40.4	0.06	27.0	28.8	0.24
Appendicitis							
Operative management							
Serious complication	5	24.4	13.4	<0.001	21.8	15.2	<0.001
Recovery	—	85.5	86.2	0.02	20.6	24.4	<0.001
Nonoperative management							
Serious complication	—	32.2	26.1	<0.001	26.8	26.0	0.58
Recovery	—	67.3	69.1	0.20	29.1	29.9	0.57

*P-value for mean represents the significant of Mann Whitney U tests comparing the risk calculator and no risk calculator groups. The P-value for the standard deviation reflects a comparison between the same groups using an F-test for homogeneity of variances.

Serious complication 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, urinary tract infection and for operative management also included return to the operating room, deep incisional or organ space surgical site infection, or wound disruption.

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 within 30 days.

the risk calculator and control groups aligned well with the risk calculator values (risk of death in the GIB vignette and the risk of complication in the SBO vignette; Table 2). However, for all other assessments, surgeons' estimates diverged, often substantially, from the corresponding risk calculator value, even among those who had been exposed to the risk calculator data. Given the ACS-NSQIP risk calculator's good prediction capability (c-statistic 0.944 for mortality and 0.816 for morbidity),[10] future research might focus on developing protocols for enhancing the impact of risk calculators on clinical judgment (eg, provide surgeons with compelling evidence of the tool's accuracy).

Our study goes beyond comparing surgeon risk assessments with an objective benchmark and is the first to our knowledge that examines the effect of risk calculators on surgeons' perception of risk. For all of our vignettes, it appears that surgeons in our control condition tended to overestimate the likelihood of adverse operative outcomes so that those who received risk calculator data tended to provide lower probabilities of such outcomes. More interestingly, surgeons in the risk calculator group also predicted slightly higher benefits of operating when averaged across all scenarios. This may be explained in part by what behavioral scientists call the affect heuristic,[21] whereby decisions and judgments (including those of risk and benefit) are driven in part by people's emotional response to an activity.[22–25] The affect heuristic is a plausible explanation here because if surgeons base their decision to operate on their affective response to operating (ie, how they feel), then exposing them to risk calculator data—showing the risks of operating are lower than they thought—would tend to make operating seem more favorable and thereby also increase surgeons' perception of the benefits of operating.[25] By this explanation, the risk calculator appeared to make surgeons feel more optimistic about both the risks and the benefits of operating. It would seem that the role of affect in surgical decision-making warrants further investigation.

Given that surgeons exposed to risk calculator data judged the risks of operating to be lower, one might expect that they would in turn be more likely to recommend an operation. Interestingly, we found that surgeons in the risk calculator group were no more likely to recommend an operation than surgeons in the control group. There are several possible explanations for this finding. First, the differences in judged operative risk between the 2 groups, although

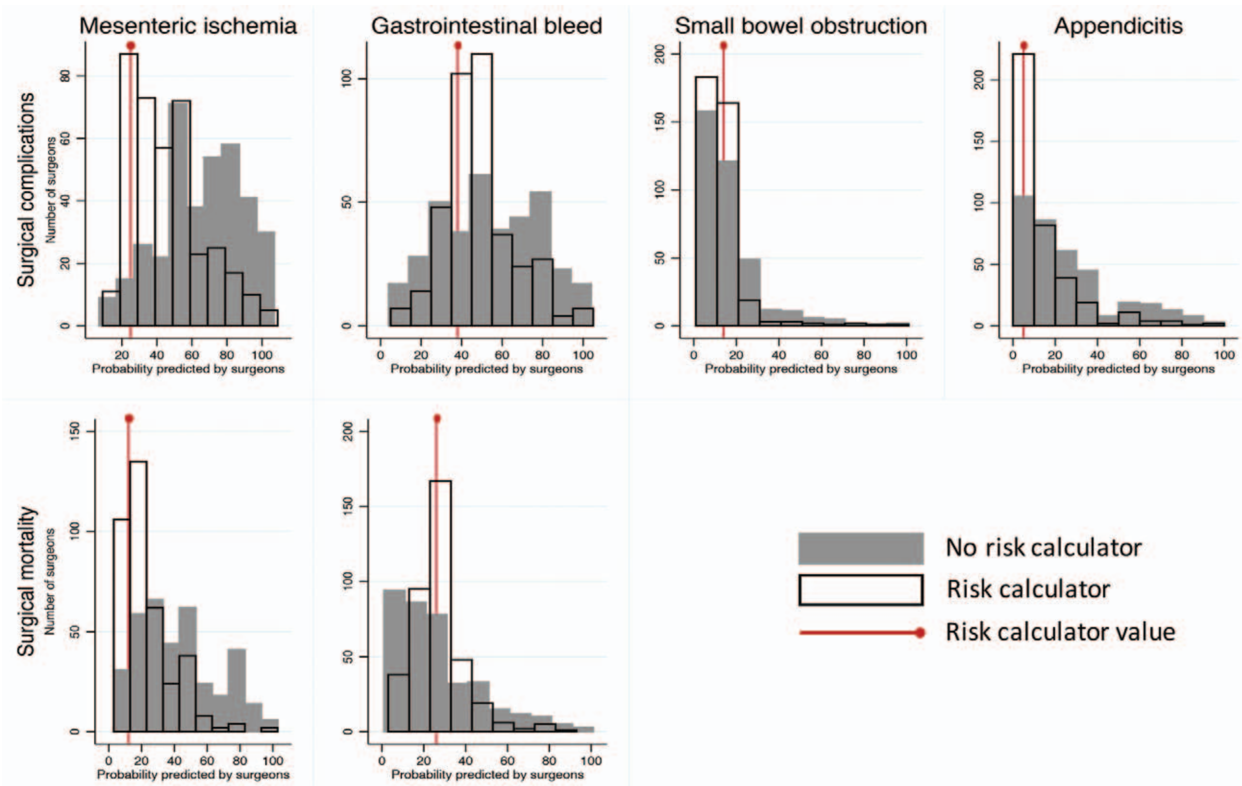

FIGURE 1. Surgeon's judgment of serious complication and mortality with and without risk calculator data. Red line signifies value from risk calculator, risk of mortality was exceedingly low for the small bowel obstruction and appendicitis cases and therefore surgeons were not asked to make these predictions. 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 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, urinary tract infection, return to the operating room, deep incisional or organ space surgical site infection, or wound disruption.

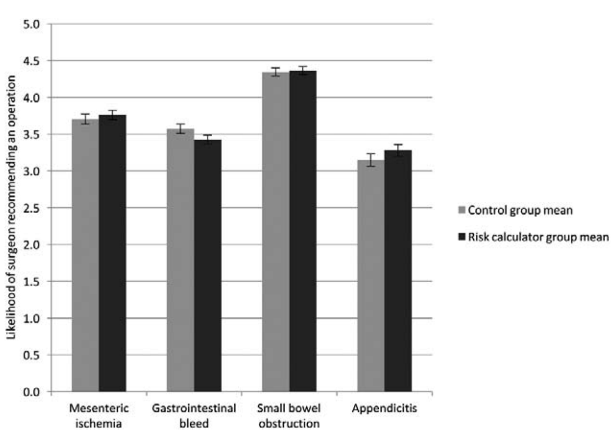

FIGURE 2. Surgeon decision to recommend an operation with and without risk calculator. Likelihood of surgeon recommending an operation measured by surgeon response to the question ''how likely are you to recommend an operation at this time?'' as answered on a 5-point Likert scale 1 ''very unlikely,'' 2 ''unlikely,'' 3 ''neutral,'' 4 ''likely,'' 5 ''very likely,'' P values for comparison between groups using Mann-Whitney test: mesenteric ischemia ($P = 0.59$), gastrointestinal bleed ($P = 0.05$), small bowel obstruction ($P = 0.45$), appendicitis ($P = 0.26$).

statistically significant, may not have been sufficiently meaningful in a clinical sense to give rise to a change in the recommended treatment. Moreover, it is possible that surgeons' treatment decisions are independent of their reported risk estimates. We find these explanations unlikely, however, because previous research has documented a strong association between surgeons' decisions to operate and their judgments of the risks and benefits of both operative and nonoperative management[9] (see Appendix 4, http://links.lww.com/SLA/B11 for validation of this finding using this sample). Second, it appears that when risk calculator data suggested that the risks of operating are lower than what surgeons expected, this led to not only lower judgments of operative risk (and higher judgments of benefit) but also lower judgments of nonoperative risk (and higher judgments of benefit). This finding may be partly attributable to what behavioral scientists call an ''anchoring effect,'' which describes an excessive reliance on an external or internal starting value. In this case, externally provided risk estimates may have caused surgeons to consider information consistent with those values, even for the seemingly unrelated nonoperative risk assessments.[26–28] Finally, it is also possible that surgeons tend to base their assessments of risks and benefits on their clinical decisions rather than making clinical decisions based on assessments of risks and benefits (ie, reverse causality). By this explanation, surgeons (perhaps non-consciously) might have modified their judgments of both operative and nonoperative management to be consistent with the intuitive decision that they had made based on the vignette alone, before seeing the

TABLE 3. Adjusted Effect of Risk Calculator on Surgeon Judgments and Clinical Decisions Across All 4 Clinical Vignettes

Surgeon response	Control Group (n = 384)	Risk Calculator Group (n = 395)	Absolute Difference Between Groups	Relative Difference Between Groups	P^*
			Percent		
Operative management					
Serious complication, %	39.9	29.6	−10.4	−26.0	<0.001
Recovery, %	67.5	69.6	2.2	3.2	0.059
Non-operative management					
Serious complication, %	54.3	48.7	−5.6	−10.3	<0.001
Recovery, %	41.5	44.9	3.4	8.1	0.005
Likelihood of recommending an operation, 1–5 Likert scale†	3.7	3.7	0.0	0.4	0.76

*P value comparing risk calculator and control group calculated from separate hierarchical linear regressions to account for clustering of predictions and decisions between surgeons. Regressions included dummy variables for each clinical vignette as well as dummy variable for experimental arm assignment.

†5-point Likert scale: 1 "very unlikely," 2 "unlikely," 3 "neutral," 4 "likely," 5 "very likely,"

calculator data. These changes resulted in a compensatory effect whereby changes in judgments of operative risks and benefits were balanced by symmetric changes in the judgments of nonoperative risks and benefits, thereby leaving the relative favorability of operating unchanged.

It is important to note that our study does not speak to whether surgeons made "right" or "wrong" treatment decisions. In fact, we deliberately designed the vignettes so that no particular treatment decision was obviously correct. Absent a clearly dominant treatment option, decisions like these may best be made on a patient-by-patient basis, wherein the risks and benefits of each treatment are clearly communicated to patients and a shared decision-making approach is used to ensure that the treatment choice is aligned with the patient's preferences and values.

Toward this end, the risk calculator offers a useful way to ensure that patients are aware of the most accurate estimates for the likelihood of adverse events following surgery. Our results suggest that the calculator does, in fact, help surgeons make more accurate and less varied estimates of operative risk, thereby enabling them to provide patients with more accurate information during the informed consent process. We note that the informed consent process can often be cursory and may not always include an explicit discussion of the relevant risks and benefits,[29,30] even though knowledge of such information can significantly influence patients' decisions.[2,31] Routine use of the risk calculator may therefore be a useful means of facilitating accurate communication of the risks of surgery to patients.

We acknowledge that there are several limitations to our study. First, given our use of hypothetical vignettes, it is unclear whether our results would extend to judgments and decisions in clinical environments, wherein surgeons must simultaneously take into account the clinical, social, financial, and legal implications of their actions. Second, our results may or may not generalize to elective operations or operations for other surgical specialties. Future research will be needed to determine whether our findings apply to these other clinical settings. It is also possible that our results would not generalize to surgeons who declined to participate in our study, although it is unclear how willingness to complete this study would bias our results. Finally, it is possible that the impact of risk calculator data on subsequent risk assessments reflected for some participants a demand effect to cooperate with the experimenter rather than sincere change in beliefs.

These limitations notwithstanding, we believe our study has several important theoretical and practical implications. Most normative accounts of decisions under uncertainty posit that surgical decisions ought to depend on the probabilities of both surgical and nonsurgical outcomes. The present study suggests that by exposing surgeons to objective data, it is possible to influence how they perceive these risks and benefits. However, our findings also suggest that providing data on only the risks of an operation may be insufficient to change surgeons' treatment decisions. Future surgical decision-making studies might test enhanced tools that also provide data concerning nonoperative outcomes to determine whether such enriched information is more effective in influencing surgeons' treatment decisions. Further research is also needed to investigate the most effective way to present and communicate such data to patients so that they adequately understand it and can meaningfully participate in treatment decisions that best accord with their own preferences.[32]

CONCLUSIONS

Accurate assessment of treatment risks and benefits is a necessary input to good clinical decision-making. In this study, we found that providing surgeons with objective data from a well-validated risk calculator resulted in improved and less varied judgments of operative risks that more closely approximate the risk calculator values. However, these data did not alter surgeons' decision whether or not to recommend an operation. As the popularity of such tools increases, further study will be necessary to determine the effect of risk calculators in clinical settings.

ACKNOWLEDGMENTS

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

REFERENCES

1. Eddy DM. The challenge. *JAMA*. 1990;263:287.
2. Pecanac KE, Kehler JM, Brasel KJ, et al. It's big surgery. *Ann Surg*. 2014;259:458–463.
3. Barry MJ, Edgman-Levitan S. Shared decision making—pinnacle of patient-centered care. *N Engl J Med*. 2012;366:780–781.
4. Neuman MD, Bosk CL. What we talk about when we talk about risk: refining surgery's hazards in medical thought. *Milbank Quarterly*. 2012;90:135–159.
5. McClellan M, Brook RH, Appropriateness of care. A comparison of global and outcome methods to set standards. *Med Care*. 1992;30:565–586.

6. Eddy DM. Variations in physician practice: the role of uncertainty. *Health Aff (Millwood)*. 1984;3:74–89.
7. Eddy DM. Clinical decision making: from theory to practice. Anatomy of a decision. *JAMA*. 1990;263:441–443.
8. 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.
9. Sacks GD, Dawes AJ, Ettner SL, et al. Surgeon perception of risk and benefit in the decision to operate. *Ann Surg*. 2016;000:000–000.
10. 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;.e1-3.
11. ACS-NSQIP Risk Calculator. American College of Surgeons National Surgical Quality Improvement Program. Available at: http://riskcalculator.facs.org/. Accessed on May 5, 2015.
12. Ramanan B, Gupta PK, Gupta H, et al. Development and validation of a bariatric surgery mortality risk calculator. *J Am Coll Surg*. 2012;214:892–900.
13. Cohen ME, Bilimoria KY, Ko CY, et al. Development of an American College of Surgeons National Surgery Quality Improvement Program: morbidity and mortality risk calculator for colorectal surgery. *J Am Coll Surg*. 2009;208:1009–1016.
14. Kumbhani DJ, Wells BJ, Lincoff AM, et al. Predictive models for short- and long-term adverse outcomes following discharge in a contemporary population with acute coronary syndromes. *Am J Cardiovasc Dis*. 2013;3:39–52.
15. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation*. 2014;129:S49–73.
16. Markus PM, Martell J, Leister I, et al. Predicting postoperative morbidity by clinical assessment. *Br J Surg*. 2005;92:101–106.
17. Hartley MN, Sagar PM. The surgeon's 'gut feeling' as a predictor of postoperative outcome. *Ann Royal Coll Surg Engl*. 1994;76:277–278.
18. Hobson SA, Sutton CD, Garcea G, et al. Prospective comparison of POSSUM and P-POSSUM with clinical assessment of mortality following emergency surgery. *Acta Anaesthesiol Scand*. 2007;51:94–100.
19. Farges O, Vibert E, Cosse C, et al. 'Surgeons intuition' versus 'prognostic models'. *Ann Surg*. 2014;260:923–930.
20. Morse SB, Haywood JL, Goldenberg RL, et al. Estimation of neonatal outcome and perinatal therapy use. *Pediatrics*. 2000;105:1046–1050.
21. Slovic P, Finucane ML, Peters E, et al. The affect heuristic. *Eur J Operat Res*. 2007;177:1333–1352.
22. Zajonc RB. Feeling and thinking: preferences need no inferences. *Am Psychol*. 1980;35:151–175.
23. Fischhoff B, Slovic P, Lichtenstein S, et al. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sci*. 1978;9:127–152.
24. Slovic P, Kraus N, Lappe H, et al. Risk perception of prescription drugs: report on a survey in Canada. *Can J Public Health*. 1991;82:S15-20–S38-44.
25. Finucane ML, Alhakami A, Slovic P, et al. The affect heuristic in judgments of risks and benefits. *J Behav Decis Making*. 2000;13:1–17.
26. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*. 1974;185:1124–1131.
27. Strack F, Mussweiler T. Explaining the enigmatic anchoring effect: mechanisms of selective accessibility. *J Pers Soc Psychol*. 1997;73:437–446.
28. Chapman G, Johnson E. Anchoring, activation, and the construction of values. *Organ Behav Hum Decis Process*. 1999;79:115–153.
29. Kalet A, Roberts JC, Fletcher R. How do physicians talk with their patients about risks? *J Gen Intern Med*. 1994;9:402–404.
30. Braddock CH, Edwards KA, Hasenberg NM, et al. Informed decision making in outpatient practice: time to get back to basics. *JAMA*. 1999;282:2313–2320.
31. Nabozny MJ, Kruser JM, Steffens NM, et al. Constructing high-stakes surgical decisions: it is better to die trying. *Ann Surg*. 2015;1:64–70.
32. Lloyd A, Hayes P, Bell PR, et al. The role of risk and benefit perception in informed consent for surgery. *Med Decis Making*. 2001;21:141–149.

