UCLA
UCLA Electronic Theses and Dissertations

**Title**
Surgeon Perception of Risk and Benefit in the Decision to Operate Under Uncertainty

**Permalink**
https://escholarship.org/uc/item/5r91k4sc

**Author**
Sacks, Greg

**Publication Date**
2016

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Los Angeles

# Surgeon Perception of Risk and Benefit in the Decision to Operate Under Uncertainty

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Health Policy and Management

by

Greg Sacks

2016

© Copyright by

Greg Sacks

2016

ABSTRACT OF THE DISSERTATION

Surgeon Perception of Risk and Benefit

in The Decision to Operate

Under Uncertainty

by

Greg Sacks

Doctor of Philosophy in Health Policy and Management

University of California, Los Angeles, 2016

Professor Susan Ettner, Chair

For decades, researchers have noticed widespread variation in the use of surgical procedures across the United States. Yet despite extensive study, the underlying reasons for this variation remains poorly understood since it seems unrelated to traditional measures of supply and demand, such as the prevalence of disease, the number of available physicians, or measurable differences in patient preference. Instead, some researchers have attributed the persistently unexplained variation to differences in physician judgment on the role of surgery. However, the precise composition of this judgment, how it might differ from one physician to the next, and how such judgment translates into clinical decision-making remain unknown. In this dissertation, I explore the variation in surgeons' decisions to operate by examining how they perceive the risks and benefits of two opposing treatment options: to operate or not to operate.

Using data collected from a national study where surgeons provide their opinions on a series of detailed clinical vignettes, I document wide variation both in how surgeons perceive these risks and benefits and in their associated clinical decisions. I then go on to demonstrate that the variation in perception is closely aligned with surgeons' clinical decisions, thus revealing the substantial consequences of differences in perception. In fact, accounting only for differences in surgeons' perceptions of the risks and benefits of operating and not operating, I was able to explain a total of 36% of the variation in surgeons' clinical decisions. The dissertation then examines whether surgeons' perception of risk and benefit are associated with their experience or general attitude towards risk. Finally, I examine the results of an experiment where surgeons were, by random assignment, provided data from a surgical risk calculator. Although exposure to risk calculator data strongly influenced surgeons' reported perception of risk, it did not alter their reported likelihood of recommending an operation. One possible reason for this latter finding is that the risk calculator also appeared to influence surgeons' judgments of non-operative risks and benefits, despite the fact that the calculator presented no data on these outcomes. Presented together with a detailed conceptual model, the results of this dissertation offer a useful framework for the future study of surgical decision-making and to promote the idea that clinical decisions can be both data-driven and patient-centered.

The dissertation of Greg Sacks is approved.

Robert H. Brook

Thomas H. Rice

Melinda Aileen Gibbons

Susan Louise Ettner, Committee Chair

University of California

2016

TABLE OF CONTENTS

**INTRODUCTION** ..... **1**
1.1 DECISION TO OPERATE IS IMPORTANT ..... 1
1.2 MOST DECISIONS ARE IN THE GREY ..... 2
1.3 VARIATION IN THE DECISION TO OPERATE ..... 3
1.4 REASONS FOR VARIATION ..... 4
1.5 TRADEOFFS BETWEEN RISK AND BENEFIT ..... 6

**2 LITERATURE REVIEW** ..... **11**
2.1 HOW SURGEONS *SHOULD* MAKE THE DECISION TO OPERATE ..... 11
*2.1.1 Descriptive models* ..... *12*
*2.1.2 Prescriptive models* ..... *14*
2.2 HOW SURGEONS *DO* MAKE DECISIONS ..... 17
*2.2.1 Qualitative data* ..... *17*
*2.2.2 Quantitative* ..... *20*
2.3 HOW SURGEONS *MIGHT* MAKE DECISIONS ..... 34
*2.3.1 Uncertainty to decision in three steps* ..... *34*
*2.3.2 Risk prediction* ..... *35*
*2.3.3 Risk interpretation* ..... *37*
*2.3.4 Risk response* ..... *39*
2.4 SUMMARY ..... 41
2.5 CONTRIBUTION TO LITERATURE ..... 42

**3 CONCEPTUAL MODEL** ..... **44**
3.1 DEFINING THE OUTCOME ..... 44
3.2 WEIGHING THE RISKS AND BENEFITS ..... 46
3.3 PATIENT FACTORS ..... 49
*3.3.1 Clinical factors* ..... *49*
*3.3.2 Non-clinical factors* ..... *50*
3.4 SURGEON FACTORS ..... 51
*3.4.1 Thoroughness of prediction* ..... *52*
*3.4.2 Surgeon risk attitude* ..... *53*
*3.4.3 Confidence* ..... *54*
*3.4.4 Affect towards operating* ..... *54*
*3.4.5 Prediction skill* ..... *56*
3.5 ENVIRONMENTAL FACTORS ..... 59
*3.5.1 Local practice culture* ..... *59*
*3.5.2 Surgeon incentives* ..... *60*
*3.5.3 External forces* ..... *61*
3.6 QUESTIONS AND HYPOTHESES ..... 62

**4 METHODS** ..... **67**
4.1 DATA SOURCE ..... 67
*4.1.1 Study sample* ..... *67*
*4.1.2 Study design* ..... *68*
4.2 MEASUREMENT MODEL/VARIABLE DEFINITIONS ..... 73
*4.2.1 Surgeon factors* ..... *73*
*4.2.2 Patient factors* ..... *78*
*4.2.3 Environmental factors* ..... *79*
4.3 STATISTICAL ANALYSIS ..... 80
5.1 DESCRIPTIVE STATISTICS ..... 90
5.2 REGRESSION RESULTS ..... 93

5.3 SUMMARY ..... 116
**6 DISCUSSION ..... 119**
6.1 INTERPRETATION OF FINDINGS ..... 121
*6.1.1 Perceived risks and benefits ..... 121*
*6.1.2 Risk attitude ..... 125*
*6.1.3 Experience ..... 126*
*6.1.4 Risk calculator ..... 127*
6.2 LIMITATIONS ..... 133
*6.2.1 Threats to internal validity ..... 133*
*6.2.2 Threats to external validity ..... 136*
6.3 IMPLICATIONS AND FUTURE WORK ..... 137
**7 APPENDIX ..... 149**
**8 REFERENCES ..... 161**

## Table of Figures

FIGURE 3.1 FACTORS INFLUENCING THE SURGEON'S INITIAL DECISION TO RECOMMEND AN OPERATION ..... 46
FIGURE 4.1 FLOW DIAGRAM OF STUDY PARTICIPANTS ..... 72
FIGURE 4.2 CONCEPTUAL MODEL WITH CORRESPONDING EMPIRICAL MEASUREMENT PROXIES ..... 74
FIGURE 5.1 VARIATION IN SURGEONS' JUDGMENT OF OPERATIVE AND NON-OPERATIVE RISK AND BENEFIT ..... 91
FIGURE 5.2 VARIATION IN SURGEON RESPONSE TO "HOW LIKELY ARE YOU TO RECOMMEND AN OPERATION?" ..... 92
FIGURE 5.3 DISTRIBUTION OF RISK ATTITUDE SCORES FOR SURGEONS IN STUDY SAMPLE ..... 92
FIGURE 5.4 ASSOCIATION BETWEEN RISK ATTITUDE AND DECISION TO RECOMMEND AN OPERATION AT DIFFERENT VALUES OF PERCEIVED RELATIVE RISK ..... 96
FIGURE 5.5 ASSOCIATION BETWEEN RISK ATTITUDE AND DECISION TO RECOMMEND AN OPERATION AT DIFFERENT VALUES OF PERCEIVED RELATIVE BENEFIT ..... 97
FIGURE 5.6 ASSOCIATION BETWEEN SURGEON RISK ATTITUDE AND PERCEIVED LEVEL OF ABSOLUTE OPERATIVE RISK ..... 98
FIGURE 5.7 ASSOCIATION BETWEEN SURGEON EXPERIENCE AND PERCEPTION OF ABSOLUTE OPERATIVE RISK ..... 103
FIGURE 5.8 ASSOCIATION BETWEEN SURGEON EXPERIENCE AND PERCEPTION OF ABSOLUTE OPERATIVE BENEFIT ..... 105
FIGURE 5.9 ASSOCIATION BETWEEN ANNUAL OPERATIVE VOLUME AND THE PERCEIVED LEVEL OF ABSOLUTE OPERATIVE BENEFIT ..... 106
FIGURE 5.10 SURGEON'S JUDGMENT OF OPERATIVE RISK WITH AND WITHOUT RISK CALCULATOR DATA ..... 108
FIGURE 5.11 SURGEON DECISION TO RECOMMEND AN OPERATION WITH AND WITHOUT RISK CALCULATOR ..... 112
FIGURE 5.12 EFFECT OF ONLY ASKING SURGEONS TO ASSESS OPERATIVE OUTCOMES ON THE DECISION TO RECOMMEND AN OPERATION BY PERCEIVED LEVEL OF ABSOLUTE OPERATIVE RISK ..... 114
FIGURE 5.13 EFFECT OF ONLY ASKING SURGEONS TO ASSESS NON-OPERATIVE OUTCOMES ON THE DECISION TO RECOMMEND AN OPERATION BY PERCEIVED LEVEL OF ABSOLUTE NON-OPERATIVE RISK ..... 115
FIGURE 5.14 EFFECT OF ONLY ASKING SURGEONS TO ASSESS NON-OPERATIVE OUTCOMES ON THE DECISION TO RECOMMEND AN OPERATION BY PERCEIVED LEVEL OF ABSOLUTE NON-OPERATIVE RISK ..... 116

## Table of Tables

TABLE 4.1 PEARSON'S ADAPTED JACKSON PERSONALITY INDEX ..... 73
TABLE 4.2 LIST OF VARIABLES AND RELEVANT STUDY COHORTS INCLUDED IN EACH REGRESSION ..... 82
TABLE 4.3 SUMMARY OF REGRESSION NOTATION ..... 85
TABLE 5.1 ASSOCIATION BETWEEN SURGEON PERCEPTION OF RELATIVE RISKS AND BENEFITS OF OPERATING AND THE DECISION TO RECOMMEND AN OPERATION ..... 94

TABLE 5.2 ASSOCIATION BETWEEN SURGEONS' DECISION TO OPERATE AND THEIR PERCEPTION OF TREATMENT RISKS AND BENEFITS ..... 99
TABLE 5.3 ASSOCIATION BETWEEN SURGEON EXPERIENCE AND THE LEVEL OF PERCEIVED ABSOLUTE OPERATIVE RISK ..... 102
TABLE 5.4 ASSOCIATION BETWEEN SURGEON EXPERIENCE AND THE LEVEL OF PERCEIVED ABSOLUTE OPERATIVE BENEFIT ..... 104
TABLE 5.5 SURGEON JUDGMENT OF RISKS AND BENEFITS FOR OPERATIVE AND NON-OPERATIVE MANAGEMENT WITH AND WITHOUT RISK CALCULATOR ..... 107
TABLE 5.6 ADJUSTED EFFECT OF RISK CALCULATOR ON SURGEON JUDGMENTS AND CLINICAL DECISIONS ACROSS ALL FOUR CLINICAL VIGNETTES ..... 109
TABLE 5.7 EFFECTS OF ROUTINE USE OF RISK CALCULATOR ON PREDICTION SKILL ..... 111
TABLE 5.8 EFFECTS OF RISK CALCULATOR ON SURGEONS' DECISION TO RECOMMEND AN OPERATION ..... 113
TABLE 5.9 SUMMARY OF EVIDENCE FOR HYPOTHESES ..... 117

## Table of Appendices

APPENDIX 1 COMPARISON OF FIRST-WAVE RESPONDENTS, SECOND-WAVE RESPONDENTS, AND THE GENERAL POPULATION OF SURGEONS ..... 149
APPENDIX 2. FULL QUESTIONNAIRE FROM ONLINE STUDY ..... 150
APPENDIX 3. COMPARISON OF DEMOGRAPHIC AND CLINICAL CHARACTERISTICS ACROSS RANDOMLY ASSIGNED STUDY ARMS ..... 158
APPENDIX 4 COMPARISON OF LISTWISE DELETION AND MULTIPLE IMPUTATION FOR REGRESSION 1 ..... 159
APPENDIX 5. ASSOCIATION BETWEEN SURGEON PERCEPTION OF RELATIVE RISKS AND BENEFITS OF OPERATING AND THE DECISION TO RECOMMEND AN OPERATION ..... 160

ACKNOWLEDGMENTS

I am filled with both appreciation and gratitude to those who have accompanied me on this journey through the past few years of discovery. In particular, I would like to thank Susan Ettner for her support, mentorship, and unparalleled ability to think through difficult conceptual and analytical problems. Every step of the way, she went above and beyond her role as my thesis advisor and committee chair. Robert Brook deserves both credit and my gratitude for being the first person to throw his support behind this research project and for encouraging me to tackle a complicated research question, even when it might have been expedient to direct my energy elsewhere. I would like to thank Thomas Rice for his contribution as a committee member and for his thoughtful guidance at every stage of the project. I am particularly indebted to my surgeon mentors, Melinda Maggard-Gibbons, Marcia Russell, and Clifford Ko, for their support, guidance, and countless hours spent meticulously tweaking details of this study and making sure it maintained its clinical relevance.

I am also extremely thankful to Craig Fox for his participation in this research. Although not a committee member, his expertise and guidance were pivotal in the development of this research, starting with the big picture conception of research questions all the way down to the minutiae of editing grammar and revising prose for the related manuscripts. I am fortunate to have Carlos Pellegrini as my mentor and role model in surgery. Without his support, and that of David Hoyt and Patricia Turner at the American College of Surgeons, the research project that gave rise to this dissertation would never have materialized. I am also grateful to David Asch for his assistance in formulating and developing testable research hypotheses, Sandra Berry for her expertise in designing the online study instrument, Joe Hines, my former residency program director, for his continued support and allowing me extra time away from residency to complete

my dissertation. A big thank you also goes out to Lisa Rosenbaum for helping me refine my fledgling research ideas and for encouraging me to persevere when obstacles arose along the way. I would also like to thank Craig Peskind for his thoughtful reflections on his encounters with surgeons, without which I might not have stumbled on my overarching research questions. It has been a privilege to have my friend and colleague, Aaron Dawes, as a companion during both the highs and lows of conducting this research. I would also like to extend my sincere thanks to the leadership of the Robert Wood Johnson/Veterans Affairs Clinical Scholars program for creating a fellowship program that allowed me to dream big and pursue my research goals.

I am forever grateful to my mother, to whom I owe a special debt. Her love, support, and emotional strength are the pillars on which my life has been built. Lastly, my greatest thanks goes to my wife, intellectual accomplice, and best friend, Christina Penfield, for her endless love and support. Her reflective contributions are deeply embedded throughout this document.

Vita

**Education and Training**

| | |
|---|---|
| 09/2002-03/2006 | BA, University of California, Santa Barbara, *Highest Honors* (Creative Studies/Biochemistry) |
| 09/2009-05/2010 | MPH, Harvard School of Public Health |
| 08/2006-06/2011 | MD, University of California, Irvine School of Medicine |
| 06/2011-06/2019 | General Surgery Residency, University of California, Los Angeles, Department of Surgery |
| 07/2013-06/2016 | Robert Wood Johnson Clinical Scholars Fellow, University of California, Los Angeles |

**Bibliography**

*Peer-Reviewed Publications*


1. **Sacks GD**, Hill C, Rogers SO. Insurance status and hospital discharge disposition after trauma: inequities in access to postacute care. *J Trauma*. 2011;71(4):1011–1015.
2. **Sacks GD**, Panchmatia JR, Marino M, Hill C, Rogers SO. The effect of operative timing on functional outcome after isolated spinal trauma. *J Trauma*. 2011;71(6):1668–1672.
3. **Sacks GD**, Diggs BS, Hadjizacharia P, Green D, Salim A, Malinoski DJ. Reducing the rate of catheter-associated bloodstream infections in a surgical intensive care unit using the Institute for Healthcare Improvement central line bundle. *Am J Surg*. 2014;207(6):817–823.
4. **Sacks GD**, Dawes AJ, Russell MM, Lin AY, Maggard-Gibbons M, Winograd D, Chung HR, Tomlinson J, Tillou A, Shew SB, Hiyama DT, Cryer HG, Brunicardi FC, Hiatt JR, Ko C. Evaluation of Hospital Readmissions in Surgical Patients: Do Administrative Data Tell the Real Story? *JAMA Surg.* 2014;149(8):759
5. **Sacks GD**, Chung K, Jamil K, Garg M, Dunn JCY, DeUgarte DA. Surgical salvage of acquired lung lesions in extremely premature infants. *Pediatr Surg Int*. 2014;30(5):573–576.
6. Dawes AJ, **Sacks GD**, Russell MM, Lin AY, Maggard-Gibbons M, Winograd D, Chung HR, Tillou A, Hiatt JR, Ko C. Preventable Readmissions to Surgical Services: Lessons Learned and Targets for Improvement. *J Am Coll Surg*. 2014;219(3):382–389.
7. Lawson EH, Louie R, Zingmond DS, **Sacks GD**, Brook RH, Hall BL, Ko CY. Using Both Clinical Registry and Administrative Claims Data to Measure Risk-Adjusted Surgical Outcomes. *Ann Surg*. 2014:1.
8. Stey AM, Russell MM, Ko CY, **Sacks GD**, Dawes AJ, Gibbons MM. Clinical registries and quality measurement in surgery: A systematic review. *Surgery*. 2015;157(2):381–395.
9. Dawes AJ, Hemmelgarn M, Nguyen DK, **Sacks GD**, Clayton S, Cope J, Ganz PA, Maggard-Gibbons M. Are primary care providers prepared to care for survivors of breast cancer in the safety net? *Cancer*. 2015;121(8):1249–1256.
10. Dawes AJ, **Sacks GD**, Cryer HG, Preston C, Gorospe D, Cohen M, McArthur D, Russell MM, Maggard-Gibbons M, Ko CY. Intracranial pressure monitoring and inpatient mortality in severe traumatic brain injury: a propensity-score matched analysis. *J Trauma Acute Care Surg.* 2015 Mar;78(3):492-502.
11. **Sacks GD**, Lawson EH, Dawes AJ, Gibbons MM, Zingmond DS, Ko CY. Which patients require more care after hospital discharge? An analysis of post-acute care use for patients undergoing surgery. *J Am Coll Surg.* Feb 2015;220(6):1113–1121.e2.

12. Snyder E, Amado V, Jacobe M, **Sacks GD**, Bruzoni M, Mapasse D, DeUgarte DA. General surgical services at an urban teaching hospital in Mozambique. *Journal of Surgical Research* 2015;198(2):340–345.
13. **Sacks GD**, Shannon E, Dawes A, Rollo J, Nguyen D, Russell M, Ko CY, Shekelle P, Gibbons MM. Teamwork, communication and safety climate: a systematic review of interventions to improve surgical culture. *BMJ Quality and Safety* 2015;24(7):458–467.
14. Wu JX, Dawes AJ, **Sacks GD**, Brunicardi FC, Keeler EB. Cost-effectiveness of non-operative management versus laparoscopic appendectomy for acute uncomplicated appendicitis. *Surgery* 2015;158(3):712–721.
15. **Sacks GD**, Lawson EH, Dawes AJ, Russell M, Gibbons MM, Zingmond DS, Ko CY. Relationship Between Hospital Performance on a Patient Satisfaction Survey and Surgical Quality. *JAMA Surgery* 2015;150(9):858–864.
16. Dawes AJ, **Sacks GD**, Cryer HG, Gruen JP, Preston C, Gorospe D, Cohen M, McArthur D, Russell MM, Maggard-Gibbons M, Ko CY. Inter-hospital variation in mortality for patients with severe traumatic brain injury: does compliance with evidence-based guidelines matter? *JAMA Surgery*.
17. **Sacks GD**, Lawson EH, Hines OJ. Morbidity and Mortality Conference 2.0. *Annals of Surgery.* 2015; 262(2):228-9
18. Steiner CA, Maggard-Gibbons M, Raetzman SO, Barrett ML, **Sacks GD**, Owens PL. Return to Acute Care Following Ambulatory Surgery. *JAMA* 2015;314(13):1397–1399.
19. **Sacks GD**, Hiatt JR. Weekday or Weekend Discharge—Does It Make a Difference? *JAMA Surgery* 2015:1–1.
20. Stark A, Maciel RC, Sheppard W, **Sacks GD**, Hines OJ. An early warning score predicts risk of death after in-hospital cardiopulmonary arrest in surgical patients. *Am Surg.* 2015; 81(10):916-921(6)
21. **Sacks GD**, Lawson EH, Dawes AJ, Weiss R, Russell M, Brook RH, Zingmond DS, Ko CY. Variation in Hospital Use of Postacute Care After Surgery and the Association With Care Quality. *Med Care*. 2016;54(2):172–179.
22. Stark AP, **Sacks GD**, Rochefort MM, Donahue TR, Reber HA, Tomlinson JS, Dawson DW, Eibl G, Hines OJ. Long-term survival in patients with pancreatic ductal adenocarcinoma. *Surgery*. February 2016.
23. **Sacks GD**, Dawes AJ, Ettner SL, Brook RH, Fox CR, Russell M, Ko CY, Gibbons MM. The effects of a risk calculator on surgical decision-making: a randomized trial. *Annals of Surgery*. In press.
24. **Sacks GD**, Ulloa JG, Shew SB. Is there a relationship between hospital volume and patient outcomes in gastroschisis repair? *Journal of Pediatric Surgery*. In press.
25. **Sacks GD**, Dawes AJ, Ettner SL, Brook RH, Fox CR, Gibbons MM, Ko CY, Russell M. Surgeon perception of risk and benefit in the decision to operate. *Annals of Surgery*. In press.
26. Yule S, **Sacks GD**, Maggard-Gibbons M. Innovative Approaches for Modifying Surgical Culture. *JAMA Surgery.* In press.

# Introduction

## 1.1 Decision to operate is important

The field of medicine has evolved tremendously over the past centuries. Gone are the days of bloodletting, mercury treatments, and trephination, replaced instead by modern technologies ranging from robotics to gene sequencing and organ transplantation. The ability to cure disease and alleviate suffering is greater today than at any time in history. Yet, the delivery of medical care remains a profoundly human endeavor, relying on the discretion and judgment of healthcare providers, most commonly physicians. As such, physicians often retain substantial control over treatment decisions, thereby holding considerable sway over each patient's healthcare fate. This dynamic is arguably no more evident than in the field of surgery, where clinical decisions can be of the highest stakes, sometimes even weighing in the balance the outcomes of life and death.

The types of clinical decisions made by surgeons are innumerable, ranging from the mundane to the critical. Whether to feed a patient after surgery, to order an imaging study, or to place a drainage catheter during an operation are just some of these routine clinical decisions. Yet, of all the clinical decisions in surgery, one stands out in particular as not only the most intricate and challenging, but also the most consequential for patients. This decision, steeped in complexity spanning the clinical, the social, the economic, and even the ethical, arguably lies at the heart of what it means to be a surgeon. That is the decision to operate on a patient.

## 1.2 Most decisions are in the grey

Some decisions in surgery are relatively straightforward. Resecting an early colon cancer in a young healthy patient or repairing a strangulated groin hernia in a middle-aged healthy man, for example, are decisions that are far from controversial. Most if not all surgeons in these situations would make the same decision to operate. However, straightforward decisions like these in surgery are far from the norm. According to John Birkmeyer, John Wennberg, and other leading healthcare researchers, "many decisions about surgery reside within a large grey area of clinical discretion, bounded by comparatively small tails of clearly appropriate and inappropriate indications for intervention."[1] The unambiguous decision to operate, like the unambiguous decision to not operate, is far less frequent than the decision that is hampered by conflicting data and incomplete knowledge.

All clinical decisions, including the decision to operate or not operate on a patient, are fundamentally decisions that are made under uncertainty. Uncertainty exists because it is not possible to know for sure what will happen after a test is ordered or a treatment is administered. Test results can help clinch a diagnosis but they can also be misleading. Medications and procedures can cure disease but they can also cause side effects and complications. Without knowing with certainty whether such events will occur, the best that can be done is to judge the likelihood of these outcomes and use those probabilities to inform our choices. In so doing, clinical decisions of all varieties become tradeoffs between the likelihood of desired events (benefits) and the likelihood of undesired events (risks).

Even clinical scenarios that are considered routine can be complicated by extraneous circumstances that shift the decision into the grey zone of uncertainty. A ruptured abdominal aortic aneurysm is as clear an indication for an operation as any. Yet, some surgeons may pause

and even consider not operating if the aneurysm belongs to an elderly patient with a poor quality of life and advanced dementia. For these surgeons, putting such a patient through the long process of recovery, not to mention the immediate life-threatening nature of the operation, only serves to prolong suffering. Thus, contrary to what most people believe, straightforward cases where surgery is clearly indicated or not indicated are not overly common, and it is often very difficult to determine when an operation should be performed. Most situations instead require surgeons to consider a wide array of diverse information before they reach a conclusion on whether or not to operate.

## 1.3 Variation in the decision to operate

Given this level of complexity, it is no surprise that for many operations, surgeons appear to differ greatly in their use of surgery. Such differences have been well documented by the abundant literature exploring regional variations in care. By dividing the United States into hospital referral areas, researchers—primarily those at the Dartmouth Atlas—have documented repeatedly and extensively how providers in one region perform surgery at different rates to those in other areas. For example, the incidence of radical prostatectomy in Salt Lake City was found to be more than three times that in San Francisco.[2] Similarly, the incidence of lower extremity bypass in Baltimore is five times that in Temple, Texas. However, operations themselves are not performed on either cities or regions. They are performed on individual patients by individual surgeons. Research that exposes variations in care on the regional level is therefore really identifying differences that can ultimately be traced back to individual decision-making on whether or not to use a particular surgical intervention. Consequently, in order to truly explain regional variation in the use of surgery, it is necessary to develop a better understanding of the antecedent decision-making that drive the utilization rates for surgical

procedures. In short, simply describing variations in care is inadequate to fully comprehend the complex nature of surgical care patterns; the next required step is to explain why.

Understanding how and why surgeons vary in their use of surgery is of utmost importance to policy makers, patients, and surgeons alike. Surgical procedures are exceedingly common; a total of 51.4 million procedures are performed annually in the United States at a cost of several hundred billion dollars.[3] Understanding the factors that influence surgical use rates is therefore critical to any strategy that aims to reduce healthcare costs or ensure that healthcare resources are being allocated efficiently. For patients, the imperative is obvious. While an operation has the potential to cure disease, relieve pain, and improve quality of life, it can also be a negative, life-changing experience marked by pain, suffering, uncertainty, and often-dramatic financial loss. Before undergoing such an invasive procedure, patients need to be aware of any potential biases that may inform the surgeon's decision and also if there is any reasonable disagreement among surgeons on whether an operation is needed. Unnecessary variations in the delivery of surgical care may also be considered a threat to the integrity and scientific underpinnings of the surgical profession. In order to advance the body of scientific knowledge in surgery, there is a need to develop strategies to mitigate unnecessary practice variations and ensure that, whenever possible, a surgeon's decision to operate is based on strong, objective evidence and is supported by carefully considered reasoning.

## 1.4 Reasons for variation

Over the past several decades, many researchers have attempted, albeit with little success, to explain why such dramatic variation exists in the use of surgery. Despite many different approaches, much of this literature suggests that the differences in the use of surgery between regions cannot be explained by either the supply (number of physicians, use of diagnostic tests,

etc.) or the demand (prevalence of disease, patient preferences, etc.) of surgical services, [1] nor can it be explained by the relative overuse or underuse of surgery.[4] These studies have therefore concluded that similar patients appear to undergo remarkably different treatments simply based on the region in which care was delivered. The true cause of the observed regional variations in care remains a mystery.

Nevertheless, after discrediting many proposed theories for why variation exists, at least one group of well-known researchers have concluded that variations in the use of surgery arise primarily due to differences in physician beliefs about the indications for surgery.[1] Surgeons in high-use regions simply believe that surgery is indicated, while surgeons in low-use regions believe the opposite. This is not a trivial conclusion. For decades, researchers have been searching for clinical, epidemiologic, economic, and even political explanations for variations in care, and after disproving each of these possible explanations, the answer instead may lie somewhere within the surgeon mind, rooted in their possibly subjective beliefs about the necessary role of surgery in patient care.

But there are at least two limitations to concluding that surgical variation is the result of differences in surgeons' beliefs. First, as mentioned, is that it is founded largely on the exclusion of other feasible and tested theories and is not in and of itself supported by any evidence. There are no data, for example, to suggest that surgeons in high-use regions believe more in the role of surgery for the treatment of disease or that surgeons in low-use regions are more skeptical about its role. This gap between the use of surgery and surgeons' underlying beliefs has not yet been empirically bridged. Second, but intimately tied to the first problem, is the question of what is the nature of these differences in beliefs and on what are they based? For clinical scenarios where treatment variation exists, the question therefore still remains: what makes one surgeon

choose to operate on a patient, and another surgeon, when faced with the same patient choose not to operate?

## 1.5 Tradeoffs between risk and benefit

Whether or not it is acknowledged explicitly, there lies at the center of any decision, particularly those made in medicine, an assessment of risk and benefits. Choosing which university to attend, who to vote for in presidential elections, how to invest money for retirement, and even what type of dessert to eat after dinner, all require an individual to confront each option's risks and benefits. One may, for example, compare the benefits of eating a tasty chocolate cake against the risks associated with excessive caloric, fat, and sugar consumption. Or one may consider the health benefits of eating a fruit salad against the risks of leaving dinner unsatisfied. While other psychological processes, many of which rely on cognitive heuristics or shortcuts, also play a role in the decision (and which will be covered in more detail in subsequent chapters), the tradeoff between risk and benefit—at least one's perception of these concepts—is central to the decision. In surgery, these competing concepts are all the more consequential and often require consideration of a far wider array of possible outcomes for several different treatment options.

It follows then that for a surgeon, deciding between operating and not operating on a patient requires him/her to weigh the risks and benefits of operating against the risks and benefits of not operating, the latter typically accompanied by less invasive medical management. However, these *absolute* risks or *absolute* benefits cannot be considered in isolation but must also be considered in relation to one another, i.e. as *relative* risks and *relative* benefits. For example, a high-risk surgical candidate may have a 90% chance of dying either during or soon after an operation. Although a 90% chance of dying from an operation might seem high, it may

actually be lower than the chance of death were the patient not to undergo an operation. Similarly, the benefits of an operation must also be weighed in relation to the benefits of foregoing an operation. Deliberate evaluation of these four parameters (risks of operating, benefits of operating, risks of not operating, and benefits of not operating) therefore forms the basis of a careful and well-considered decision to operate.

Thinking explicitly about the concepts of risk and benefit in surgical decision-making is a surprisingly modern phenomenon. For much of the twentieth century, limited by the lack of reliable clinical data on which to make these assessments, surgeons found it extremely difficult to predict the risks of an operation. Explaining this predicament, a 1970 textbook author tellingly exclaimed, "these [data] are practically nonexistent.... because of these factors the accurate assessment of the operative risk for an individual case is impossible today. All we can do is guess."[5,6] Then, in 1977, Lee Goldman, while still a resident in internal medicine at Harvard, developed the first widely used tool to assess at least one aspect of a patient's risk for undergoing surgery. This tool, eponymously named the Goldman Index, used a short list of easily measurable patient factors to predict the likelihood that a patient would have a cardiac event postoperatively. Use of this tool soon became widespread throughout the surgical community, although likely more in surgical literature and textbooks than in routine clinical practice.

Since then, with the proliferation of large administrative datasets and prospective registries, it has become possible to more accurately assess operative risk. As recently as 2013, the American College of Surgeons developed a risk calculator to help surgeons and patients better understand the potential risks involved for patients undergoing an operation. The calculator draws on a very large dataset representing over one million patients undergoing a wide variety of operations from all over the country and uses an individual patient's specific risk

factors to estimate risk for a list of possible complications after surgery. This calculator has demonstrated strong predictive performance and is available as a free online tool.[7] The purpose of the risk calculator is to help surgeons more accurately assess the risks of an operation and to facilitate communication of those risks to patients. As such, it is worth noting that the calculator was not designed to dictate whether an operation should or should not be performed. Nevertheless, making such data available to surgeons and their patients at the point of care is a novel innovation that is among the first of its kind in medicine.

Although risk calculators have been developed for not only surgery but also other medical specialties,[8-11] their impact on physicians' clinical assessments has not yet been tested. There remains, therefore, several unanswered questions surrounding their use in clinical practice. For example, the risk calculator is intended to inform surgeons and patients about the risks of an operation, however, it is currently unknown whether use of a risk calculator actually influences surgeons' perceptions of risk. Furthermore, it remains unknown how use of such a tool might influence the decision to operate. The decision to recommend an operation hinges on more than just the risks of operating but also the benefits of operating, as well as the risks and benefits of not operating. Because, by design, the risk calculator only provides data on one of these parameters, it is possible that use of the calculator may have an unintended effect on clinical decision-making. As reported in a review article that tracks the history of how surgeons assess risk in surgical decision-making, focusing only on a narrow notion of risk (that is those pertaining to only to specific risks, such as the risks of death or surgical site infection and not the risk of worsening functional status, for example) may result in a more fragmented approach to surgical decision-making that is biased towards certain outcomes over others. "By defining operative risk as those end points for which prediction rules exist, physicians and clinical

researchers elevate a set of predictable outcomes over alternative end points such as changes in quality of life that, albeit difficult to predict, may nonetheless be important to individual patients."[5] Empirical evidence is therefore needed to determine how such tools might influence risk perception and clinical decision-making.

Studying surgical decision-making is inherently challenging. In clinical practice there are so many factors, ranging from clinical to social to economic, that influence the decision to operate that they can be too numerous to count. Therefore, any attempt to meaningfully compare two surgical decisions in clinical practice will be impeded by the inability to accurately measure and control for a comprehensive list of variables that might confound or influence the decision. Even if an analysis focused on a narrow decision for a single type of operation in a unique sample of patients and in a specific clinical setting, there would still be concern that certain clinical or non-clinical factors would escape unmeasured. In other words, it will always be exceedingly difficult to reach the conclusion that "all else is held equal."

In light of these obstacles, I designed a prospective study that presented surgeons with a series of clinical vignettes, which allowed me to hold each patient's clinical characteristics constant, thereby ensuring that any differences in surgeons' decisions or clinical judgments would be due to factors other than the patient's clinical condition. Most importantly, my goal was to investigate how surgeons perceive the risks and benefits of treating patients with either medical or surgical management and to examine how these perceptions are associated with their self-reported treatment decisions.

To collect the necessary data, I partnered with the leadership at the American College of Surgeons to recruit a national sample of surgeons to participate in this study in the Fall of 2014. The data obtained from that study forms the basis of this dissertation, the overarching objectives

of which are to develop a conceptual framework for how surgeons make the decision to operate, to examine how the concepts of risk and benefit inform the decision, and to evaluate how the use of a surgical risk calculator might affect surgeon's perceptions of risk and benefit and thereby inform their decision to operate.

# 2 Literature review

Of the ten rules for the redesign of health care outlined by the Institute of Medicine's *Crossing the Quality Chasm*, four reflect the need to optimize medical decision-making.[12,13] Yet, for surgery, there is very little evidence on how current decisions are made, much less how to improve them. Over the years, there have been only few attempts to create conceptual frameworks for surgeons to use in making complex surgical decisions. These frameworks range from those grounded in the principles of economics to those based more in the clinical realm. However, the empirical evidence to test these models is invariably lacking. What little data exist on how surgeons make the decision to operate arise from 3 main sources: those that use qualitative methods to discover from surgeons themselves how they make decisions; those that quantify the decisions made by analyzing surgical utilization rates; and those that quantify the decisions made by analyzing individual surgeon practices. In addition to reviewing these topics, this chapter will also summarize relevant evidence from the decision science literature that may offer useful concepts and theories with which to understand surgeon decision-making. Considered together, the existing literature can be divided into those studies that describe how surgeons *should* make the decision to operate, how surgeons actually *do* make the decision to operate, and finally, how surgeons *might* make this challenging clinical decision.

## 2.1 How surgeons *should* make the decision to operate

Considering the pervasiveness of surgical procedures in our healthcare system, one would think that a consensus would exist among surgeons and other healthcare professionals on exactly how a surgeon should approach the important decision of whether or not to recommend an operation. Yet, over the years, only a few authors have attempted to outline a conceptual

framework that describes how a surgeon might synthesize the myriad parameters that influence the decision outcome. These models range from the descriptive to the prescriptive. Descriptive models tend to explain the processes that potentially influence surgeons' decisions. Processes may be both extrinsic to the surgeon's immediate control (such as health policy and environmental factors) as well as intrinsic, reflecting the cognitive workings of the surgeon's mind. Although only backed by limited empirical evidence, these models shed light on the ways in which surgeons actually make decisions in clinical settings. Conversely, prescriptive models describe how surgeons *ought* to make decisions, using a normative framework with which to organize and synthesize often-disparate data points in order to reach the best possible decision. Yet these prescriptive models often have limited utility in a real clinical setting, due both to the lack of relevant data required to apply the models, as well as the fact that these models tend to prescribe a perfectly rational approach to decision-making, one that may not align with surgeons' and patients' natural decision-making tendencies.

### 2.1.1 Descriptive models

A good example of a descriptive model of surgeon decision-making was developed by Ira Rutkow, then a surgeon at The Johns Hopkins Hospital.[14] This model positions the surgeon as the final arbiter of the decision to perform surgery. The surgeon is then influenced by several intrinsic factors such as a predisposition towards surgical intervention, clinical knowledge, technical skill, as well as the surgeon's experience and training. In this model, the individual surgeon is also influenced by determinants at other levels, such as the professional organization level (including hospital content and staffing), as well as the society/community level (including societal philosophies, community expectations, and patient attitudes).

Another model, developed in the UK by Chand et al., places a surgeon's prediction of patient risk at the center of many different aspects of surgical care, including patient selection, informed consent, the level of medical care provided, and the ultimate patient outcomes.[15] This model, however, focuses only on clinical factors that influence risk prediction (such as the presence of symptoms, patient fitness, operative insult, and underlying pathology), while ignoring any non-clinical factors that might influence the decision to operate. These non-clinical reasons will be explained in more detail in the conceptual model chapter but include things such as the surgeon's experience, knowledge, and confidence, as well as the local practice culture of the institution in which care is delivered. Furthermore, this model does not offer any explanation for how surgeons might choose between the competing treatments of operating versus not operating.

A group of surgeons in Sweden have proposed a descriptive model for how surgeons make the decision to perform an appendectomy for appendicitis. This model, developed by Larsson et al., suggests that the decision to operate is formed by the interplay between the surgeon's medical assessment of the patient's condition as well as a set of contextual characteristics, such as organizational conditions, patient preferences, and other professional actors. When the surgeon's medical assessment of patient is ambiguous, surgeons are influenced more by contextual factors. This model, however, also excludes a clear description of how surgeons use their medical assessment to determine the possible treatment outcomes and how that determination is used to make a clinical decision. Existing descriptive models therefore all fall short of explaining how surgeons balance their evaluation of a patient's clinical condition with their clinical knowledge and experience, and incorporate all these parameters within different clinical contexts to make a clinical decision.

### 2.1.2 Prescriptive models

Descriptive models are useful tools with which to conceptualize the various factors influencing a surgeon's decision to recommend an operation. However, the few models that exist are lacking in any supporting empirical evidence. Further limitations of descriptive models include the inability to use the model to differentiate between different decisions and the fact that they offer no means with which to judge the accuracy, validity, or appropriateness of the decision outcome. To this end, there have been several proposed models that offer a prescriptive approach to surgical decision-making.

#### *2.1.2.1 Expected utility theory*

The oldest, and most well-known prescriptive model for decision-making, if not the most important theory in all of social science,[16] is expected utility theory. The modern version of expected utility was developed in 1738 by the Swiss scientist Daniel Bernoulli. Bernoulli proposed that people make decisions as a function of the likelihood that a decision will result in a desired outcome and the desirability of that expected outcome. For example, consider the choice between a 50% chance of winning $100 or a 100% chance of winning $40. Before Bernoulli, scientists of the day assumed that people would choose the former option since it offered a higher expected value ($50 vs. $40). However, Bernoulli noted that most people are inherently risk-averse and actually prefer the guaranteed $40. People are, in this case, willing to pay a premium of $10 to avoid the risky option.[17] Bernoulli's version of expected utility theory was the cornerstone of economists' view of human decision-making until the 1970's when the work of Daniel Kahneman and Amos Tversky elegantly debunked many of its underlying assumptions (more discussion on this topic in a later chapter).

It wasn't until 1959 that expected utility theory was first applied to medicine.[18] A landmark article published in *Science* applied the concepts of probability and value to clinical decisions relating to diagnosis and treatment and eventually gave rise to the birth of decision-analysis in health care, an approach to analyzing competing treatment options based on their expected value after taking into account different possible outcomes, such as the likelihood that each treatment will either work as intended or lead to various adverse outcomes. The first application of decision analysis in health care came in 1967 in a paper by Henschke and Flehinger's entitled "Decision Theory in Cancer Therapy," which tackled the then-controversial topic of performing a prophylactic neck dissection in the setting of head and neck cancer when no metastatic lesions are palpable in the lymph nodes.[19] Since then, decision analysis has been used more frequently in health care research, although much skepticism remains regarding its use in routine clinical practice.[20] In the field of surgery, decision analysis techniques have been used to support the use of carotid endarterectomy in asymptomatic patients[21] and the use of percutaneous coronary interventions in high-risk patients.[22]

Cost effectiveness analysis uses many of the same modeling techniques used in decision analysis, except that it explicitly includes monetary costs in the model. In 1979, the economist Mark Pauly extended the basic tenets of cost effectiveness analysis to the realm of surgical decision-making.[23] At that time, Pauly was concerned that the rates of "unnecessary surgery" reflected a wasteful allocation of healthcare resources and that in order to address the problem, a clear definition of unnecessary surgery was needed. Frustrated with the medical field's failure to come up with such a definition, he claimed that "medicine as a discipline cannot generate either the conceptual apparatus or the complete information set needed to arrive at a general definition" of what constitutes unnecessary surgery. Taking matters into his own hands, he applied the

principles of expected utility theory to surgical decision-making, proposing that the decision to operate should be based on the expected costs and benefits of operating or not operating. As an economist, Pauly weighed heavily the monetary costs of treatment, but also acknowledged the importance of including clinical costs, such as the risk of death or complications. The cumulative costs and benefits of operating or not operating are then multiplied by their likelihood of occurring and by a standardized weight that takes into account each outcomes desirability. An unnecessary operation therefore occurs when the expected costs of an operation outweigh the expected benefits. Pauly's paper was ambitious for its attempt to comprehensively identify all relevant parameters that should enter the decision and also for offering a prescriptive approach for how these parameters should inform the calculation of the best clinical decision.

Deviating from the formal algorithmic approach to clinical decision-making, David Eddy, the well-known physician and mathematician, offered another type of prescriptive model that applies to all types of clinical decisions, not just surgery. Eddy's model instead emphasizes the cognitive steps that must be taken before making a clinical decision.[24] According to this model, clinicians first must engage in an analytic process that incorporates both clinician knowledge, as well as the available evidence, to evaluate the benefits, harms, and costs of each treatment option. Once these parameters have been estimated, the next step involves comparing the desirability of each option by evaluating the relative benefits, harms, and resulting costs for each option and choosing whichever option is found to produce the "highest yield." While this model offers a normative approach to decision-making, Eddy concedes that by requiring a clinician to forecast outcomes, interpret data, and analyze evidence, the model depends largely on a clinician's subjective judgment. By deliberately considering the possible outcomes associated with each treatment option, Eddy's model, much like the adaptations of expected utility theory,

are consequentialist models in that they favor the choice that yields, on average, the most favorable consequence.[25]

## 2.2 How surgeons *do* make decisions

Ideally what would follow next is a parallel discussion of the empirical evidence to support or refute the various descriptive and prescriptive models of surgical decision-making. However, despite its central role in clinical care—in fact, one study noted that surgeons think decision-making is a surgeon's most important skill[26]—there is a notable lack of high-quality evidence with which to analyze how surgeons make decisions. Instead, what is known about how surgeons make decisions is derived from 3 types of data. First, several studies have used qualitative methods, including surveys, focus groups, and semi-structured interviews to gather information directly from surgeons on the thinking behind their clinical decisions. Second, by analyzing the utilization rates of different surgical procedures, it is possible to evaluate variations in care to indirectly expose different elements of surgeons' decisions. And finally, few studies have attempted to directly measure how surgeons actually make decisions using standardized patient scenarios. Some of these studies also attempt to highlight how surgeons' predictions of the risks and benefits of treatments influence the decision to operate.

### 2.2.1 Qualitative data

One way to gain insight into how surgeons make decisions is to ask them. Indeed, several studies have employed qualitative methods such as focus groups and semi-structured interviews to do just that. One study revealed four themes that surgeons use to inform their decision to operate.[27] These themes include 1) choosing to operate when it will improve the patient's outcome and minimize the amount of harm caused, 2) choosing to operate if it will halt the progression of the disease, 3) choosing to operate if it is possible to use operative intervention to

secure an otherwise uncertain diagnosis, and 4) delay operating if the patient requires resuscitation first in order to optimize the patient's physiologic condition before operating. By revealing these 4 themes, this study reveals less about surgeon decision-making from its findings than from what it did not find. For example, no surgeon in the study reported using any systematic approach to decision-making. None mentioned a careful analysis of risks and benefits or any other calculation resembling an application of expected utility theory. In fact, when prompted to apply such concepts by predicting the outcomes for patients with and without an operation, surgeons differed widely in their predictions. Of note, surgeons who predicted better outcomes following an operation (compared to non-operative management) were more likely to choose to operate on the patient.

Some have claimed that surgical decision-making, much like decision-making in other medical specialties, has historically been dependent on the clinicians' 'gut feeling.'[15] Yet the results from the qualitative studies summarized here offer little evidence to suggest that 'gut feeling' is no longer a prevalent approach used by surgeons to make decisions. In fact, the study by Larsson et al. that was previously discussed in terms of its proposed descriptive decision-making model, provides some evidence to the contrary. By interviewing surgeons about their decision to operate on patients with appendicitis, these authors discovered that the decision to operate varied widely among surgeons and that surgeons frequently referred to idiosyncratic reasons for such variation. For example, one surgeon suggested that younger surgeons tend to operate more frequently in order to gain more operative experience. When asked the follow-up question, "do I understand you correctly here, you feel it's acceptable to perform more unnecessary surgery if you are young and on your way [to] getting more experience?" the

surgeon responded, “I think that’s OK.” The surgeon interviews described in this study offer little by way of careful clinical reasoning and instead focus on anecdote and varied judgments.

Leung and colleagues interviewed 39 academic surgeons at four university-affiliated hospitals in Canada about the factors that influence their operative decisions. The investigators identified factors that fit into three categories: “avowed”, “unavowed”, and “disavowed” factors. *Avowed* factors include those that are directly aligned with the ideals of the surgical profession, such as a fear of doing harm or acting in the patient’s interest. Avowed factors are also evident in another study that examined surgeon notes in the charts of patients who had been evaluated by a surgeon and found that the most common reasons cardiac surgeons chose not to operate were purely clinical, citing reasons such as patients being too high-risk, not meeting symptomatic or prognostic criteria for surgery, or technical challenges such as poor distal vessels.[28] The Leung study, however, also identified several *unavowed* factors that “were aligned with unacknowledged or undeclared principles but were considered somewhat necessary for surgeons to manage their multiple commitments and priorities.” These included consideration for colleagues or responding to patient pressures or pressures related to time or teaching requirements. Finally, surgeons also revealed *disavowed* factors that would benefit the surgeon, but not necessarily the patient. These included a concern for personal image or reputation, ego or overconfidence; pressure to fill the operating list; and monetary motivations. While it is unclear the degree to which each of these factors contributes to each clinical decision, it is apparent that surgeons themselves are much like other humans and respond to internal, sometimes selfish motives in addition to the deeply-held ideals such as benevolence and non-maleficence.

While limited, there is some evidence from other qualitative studies that surgeons do grapple, at least implicitly, with the concepts of treatment risk and benefits. For example,

surgeons report that they make decisions based on their expert evaluation of the patient's clinical condition and their expectations of the outcome of surgery.[29] In another study, a surgeon observed that with high-risk patients, "the risks you are willing to accept are much higher because the benefits are potentially higher."[30] Yet, there are also instances where risks and benefits are considered for purposes other than deciding what treatment is best for a patient. There are instances, for example, where surgeons may avoid operating on risky patients to protect their statistics. Nevertheless, given the challenges of navigating the inherent uncertainty of these clinical decisions, many surgeons also find it useful to consult other surgeons and physicians. Particularly in the case of cancer surgery, surgeons report that collective decision-making with other providers is desirable and helps improve decisions.[31]

### 2.2.2 Quantitative

The above studies all employ different types of qualitative methods, which allow researchers to develop a deeper understanding of human behaviors and possibly even glean the motivations and reasoning driving that behavior. However, especially in face-to-face interviews or focus groups, there is always the risk that surgeons might respond in a way that is socially desirable or in line with the researcher's expectations. To complement qualitative studies, it is therefore useful to identify a more realistic representation of how surgeons behave and make decisions in the real world.

One of the earliest attempts to quantify surgeon decision-making was done in the realm of surgical diagnosis. In the 1970's De Dombal and colleagues compared surgeon diagnostic accuracy with that of a computerized algorithm.[32] The authors found that the computer program outperformed surgeons in many diagnostic challenges, causing the researchers to question the ability of physicians to analyze clinical cases in a probabilistic fashion. Moving from diagnostic

to treatment decisions, researchers have thus far attempted to quantify surgical decision-making in two different ways. First, by comparing surgeons' utilization of surgical procedures, and second, by presenting surgeons with standardized patient scenarios and asking them anonymously how they would manage such patients in their clinical practice. These methods offer a direct means of quantitatively assessing how surgeons make the decision to operate.

#### *2.2.2.1 Surgical utilization rates*

As already mentioned, it is almost prohibitively difficult to study surgeon decision-making directly, largely stemming from the inability to control for all the variables that differ between one surgical decision and another. To bypass this obstacle, researchers have instead attempted to study surgeon decision-making indirectly by analyzing surgical utilization rates on the population level. By looking at aggregate surgical utilization rates, it is possible to control for various patient, surgeon, and health system factors that may differ across populations, thereby isolating the net effect of individual surgeon decisions. If in fact all confounding factors are adequately controlled for and surgical utilization rates still differ between populations, then the remaining variation is due to unmeasured differences in how surgeons make decisions. These differences would presumably cause some surgeons to use surgical interventions more or less frequently than others.

Consider, for example, a famous study conducted in the 1930's that identified a group of 1,000 New York school children, 61% of who had undergone tonsillectomy, an operation whose indiscriminate use had at that point in time been widely rebuked by professional societies. More interesting is the manipulation that followed, in which the researchers took the remaining 39% of children and presented them to school doctors. After evaluating the children, these doctors deemed 45% of them to require tonsillectomies. Once again, the researchers presented the

remaining children to another group of doctors, who this time referred another 46% for tonsillectomy. Finally, the last group of children, twice denied an operation, was evaluated by another group of doctors and once more, 44% of these children were referred for tonsillectomy. By the end of the study, if all children who had been referred for an operation actually underwent surgery, a total of 935 of the 1,000 children would have had their tonsils removed.[33]

The tonsillectomy study gave rise to an explosion of research on variations in health care. In recent decades, most of this type of research has focused on comparing surgical utilization rates between hospital referral regions (HRRs), which divide the country into catchment areas within which most patients receive their medical care. Analyzing utilization rates across HRRs has revealed widespread variations in practice for many different surgical procedures, sometimes by as much as ten-fold.[2,4,34-36] Particularly for preference-sensitive care—procedures for which the underlying evidence is weak or equivocal and is therefore most susceptible to individual patient preference—the use of surgical procedures vary tremendously. HRRs have emerged as the conventional norm for measuring geographic variations in care, however, similar variation has been documented on many other levels of analysis, including the surgeon,[37] school district,[38] health plan,[39] and even country.[35,40,41] Surprisingly, several studies have suggested that this variation is not necessarily a reflection of inappropriate use of surgery (overuse),[4,36,42] but may instead reflect significant underuse.[1]

Yet, the explanation for why geographic regions differ in their use of these services remains a mystery of modern health care. For decades, researchers have tried to explain such variations by attributing them to differences in disease prevalence, differences in referral rates to surgeons, or the distribution of providers in different areas. These attempts have been minimally successful, explaining away some of the variation due to surgeon characteristics,[43] different

screening and diagnostic strategies,[44,45] or differences in patient preference.[46] However, much of the regional variation in surgical utilization rates remains unexplained. After extensively reviewing the literature on this topic, one leading group of researchers concluded that regional variation is due primarily to differences in physician beliefs about the indications for surgery and also the extent to which surgeons incorporate patients' preferences into treatment decisions.[1] This conclusion gives rise to several questions. What is the basis for a surgeon's belief about the indications for surgery? How are these beliefs formed? And why is it that surgeons, when faced with the same patient, differ in their beliefs about the role for surgery as a possible treatment? The study of surgical utilization rates has given rise to these questions, but in order to find answers, a different research approach is needed, one that focuses more on the individual surgeon than on a group of surgeons, and one that analyzes directly and individual surgeon's decision-making process.

Although very few studies have examined these questions in detail, the literature does attempt to quantify surgeon decisions in at least two ways. One approach presents surgeons with standardized clinical vignettes and analyzes their decisions based on their global judgment on whether an operation is indicated. The second approach, which has been employed far less frequently, attempts to deconstruct the surgeon's global assessment into its corresponding parts. This typically involves analyzing how surgeons make decisions under uncertainty by evaluating the balance between risks and benefits for each possible treatment option. The following sections will discuss each approach in detail.

#### *2.2.2.2 Global judgments*

The global-judgment approach to decision-making refers to how a surgeon evaluates whether a surgery is appropriate based on an overall, global impression of the clinical scenario.

This type of decision-making is typically employed in situations where high-quality empirical evidence is lacking and all that surgeons have to base their decision on is a best guess based on years of training and experience with similar decisions. Global judgments often form the basis of specialty organization consensus statements, which are created to provide guidance for clinicians faced with difficult and controversial decisions. Whenever it is possible, these statements incorporate the best available evidence in combination with expert judgment. However, when data are not available, consensus is built on expert experience and judgment. By virtue of the fact that these types of decisions are based on an overarching 'gist' or global assessment, this type of decision-making process is by nature opaque and subject to substantial variability.[47]

Perhaps the most well-known studies of surgeons' global judgments on the decision to operate are those done using the RAND-UCLA Appropriateness Method (RUAM). In response to the already mentioned studies revealing widespread variations in care, as well as additional studies that demonstrated similar variations in the quality of such care, researchers at UCLA and the RAND Corporation collaborated to try to determine whether such variation is a reflection of overutilization, underutilization, or both. However, before such assessments can be made, a clear definition of the appropriate use of each type of procedure was needed. The RUAM assembles an expert panel of physicians, presents them with a wide range of standardized clinical scenarios, and asks them to judge the appropriateness of performing a specific procedure in each scenario. For many of these scenarios, high-quality clinical studies may not be available to determine the appropriateness of surgical intervention and therefore, panelists rely on their own expert global judgment. To reduce variation in judgment, the RUAM also uses a modified Delphi approach, which tends to increase the overall level of agreement and improve the likelihood that the group

reaches a consensus. Since its development, the RUAM has been used to create appropriateness criteria for a wide variety of procedures.[36,48-50]

Instead of studying a small number of panel members' global judgments, other studies have presented large number of surgeons with standardized clinical vignettes similar to those used by RUAM. For example, in the 1970's, Ira Rutkow used this method to question the validity of the second opinion programs that were being implemented across the country on both a voluntary and mandatory basis as a way to control the increasing number of surgical procedures and the associated growth in health care expenditures. Rutkow presented surgeons from a broad range of surgical specialties with clinical vignettes in which an operation was a possible treatment option and found that for all specialties, agreement among surgeons when surgery was indicated was nearly nonexistent.[51] Two years later, Rutkow presented the same clinical vignettes to the same group of surgeons and found, in addition to the widespread variation in the decision to operate *between* surgeons, similar variation existed *within* the same surgeon. For each clinical scenario, up to 41% of surgeons changed their mind on whether or not an operation was indicated over the two-year period.[52] Using similar methods, Rutkow et al. also found similar widespread disagreement between surgeons in a multinational study including surgeons in the United States, England, and Canada.[40] These findings caused Rutkow and his coauthors to conclude that "surgical decision-making is a semi-exact scientific process, and it is unreasonable to expect exact answers to clinical problems."[51]

One would think that decades' worth of research and clinical progress might have exacted consensus among surgeons on the indications for surgery. Yet, in 2013, Wilson et al. published the findings of their national survey, which presented 907 surgeons from around the country with 25 standardized clinical scenarios for common types of conditions treated by

general surgeons. Still, these researchers also found widespread variation between surgeons in their decision to operate. Even in a case that was designed as a negative control (intended for no surgeon to operate), the lack of consensus matched those of other cases. Interestingly, this study found that surgeons were less likely to operate in scenarios relevant to their own subspecialty than in scenarios outside of their specialty.[53] In unadjusted analyses, the authors also found a slightly lower tendency to operate among surgeons who were white and a higher tendency to operate among those who were fellowship trained. They also performed multivariate analysis and found that White and Hispanic surgeons were less likely to operate than surgeons of other races, after controlling for age, sex, fellowship, practice type, payment structure, financial incentives, geography, and operative volume. Other factors such as malpractice concern, operative volume, and the type of payment structure were not significantly associated with the tendency to operate. Of note, this study performed multivariate analysis without a clear rationale or conceptual model to explain the hypothesized relationships between these variables and the primary outcome. More importantly, the study only compared a surgeon's tendency to operate with specific structural characteristics related to the surgeon. The study did not, however, attempt to develop a mechanism by which these structural variables explain a surgeon's decision to operate. In order to investigate these mechanisms, it is necessary to consider other studies that go beyond analyzing surgeons' global judgments and instead evaluate the cognitive or intellectual underpinnings of their clinical decisions. Furthermore, all previous studies such as those by Wilson et al., and Rutkow et al. have used clinical vignettes that provide only a small amount of clinical information. The less data provided to surgeons, the greater the likelihood that surgeons will make certain clinical assumptions about the vignette. Therefore, it remains unclear whether variation in the decision to operate in these cases is due to variations in actual clinical judgment

or if it is simply due to variation in the assumptions made by surgeons to fill in omitted information.

#### *2.2.2.3 Risk and benefit*

The global judgment model of decision-making, as used by the RAND/UCLA Appropriateness Method, benefits from its familiarity and representativeness to the typical decision-making approach used on a day-to-day basis in clinical practice. However, the RUAM process has been criticized exactly because individual clinicians' decision-making processes tend to be so distinctive and varied that they remain opaque to outside observers, possibly even to the decision-makers themselves. Instead, critics claim, a process that evaluates decision-making needs to also carefully measure the different decision inputs. It is not enough to know *what* decision is made, it is necessary to know *how* that decision is made, and whether that process can stand up to external interrogation.

According to David Eddy, "the crux of any decision is to estimate the consequences of the available options."[54] With this in mind, Mark McClellan and Robert Brook performed a study using the same participants from a RUAM appropriateness panel on the use of carotid endarterectomy (CEA; a procedure used to remove plaque from the main artery supplying the brain).[47] This study aimed to understand exactly how these panelists evaluate the possible consequences of operating or not operating, and how those evaluations correlate with their ultimate judgment on appropriateness. To date, this study is the most relevant to the aims of this dissertation and its results will therefore be discussed in detail.

McClellan and Brook asked the eight expert panelists to judge the likelihood of various adverse events, such as death or stroke, with or without CEA. They were also asked to judge the likelihood that a patient would be healthy one year later, again with or without CEA. What

follows is a summary of the main findings. First, in addition to demonstrating variation in surgeon's global judgment on the appropriateness of the procedure, they found that surgeons varied widely in their estimates for the probability of both desired and undesired outcomes. They also found that surgeons' estimates for the risks associated with CEA were systematically lower than best available "true" estimates of these outcomes, as determined from literature review. However, surgeons tended to agree more in their prediction of risks associated with an operation than they did in their predictions following no operation. More importantly, the surgeons in this study were not always internally consistent in their predictions of risks and benefits. That is, as surgeons judged operations to be more appropriate, their judgments of the risks associated with operating were often greater than their judgments of the corresponding risks associated with not operating—a finding opposite to the expected association. In fact, for 60-70% of the panelists, the results from their global judgment of appropriateness differed from the gold standard approach of using decision analysis techniques (expected utility theory) to determine appropriateness, even when the probabilities entered into the decision analysis calculation were obtained directly from the same surgeon's predictions of various outcomes. Such lack of consistency did not apply equally across all surgeons. At least one surgeon in the study remained internally consistent in his (all panelists were male) determinations of appropriateness and his predictions of risks and benefits.

There are additional findings from this study that warrant discussion. Surgeons in this study tended to label CEA as inappropriate in any scenario in which the risks of treatment were high, regardless of whether the treatment involved operating or not operating. The surgeons therefore appeared to avoid high-risk patients, even if the risks of operating were *relatively* lower than the risks of not operating. However, the panelists' appropriateness judgments tended to

correlate more strongly with particular patient characteristics, such as age or illness severity, than it did with their outcome predictions. Finally, the study also showed, at least for the small sample of surgeons used, that most global judgments showed a stronger correlation with the estimates of good-health outcomes in the presence or absence of endarterectomy than with the impact of CEA itself on those outcomes. That is, CEA tended to be regarded as more appropriate when estimates of complication rates were relatively low, rather than when the likelihood of a beneficial impact of surgery was high. These many findings caused McClellan and Brook to conclude that "most clinicians do not seem to use a decision analysis model in making choices under conditions in which outcomes are probabilistic," and that "until such sophisticated databases can be developed and widely applied, [we] are likely to remain dependent on prior probabilities and implicit judgments." Based on these findings, then, it appears that the two approaches to decision-making—global judgment and decision analysis—often contradict each other, at least the surgeons included in this study.

Another group of researchers from Duke University performed a nearly identical study. Once more, these researchers assembled a RUAM panel to decide on appropriateness criteria for the same operation, CEA. This time, however, panelists were trained in the theory of probabilistic assessment and decision analysis and were also asked to predict outcomes following medical management, not just surgical management. This study's findings contradicted sharply with those of McClellan and Brook. The Duke study found a very high correlation between surgeons' global judgment and the output from decision models populated by the same surgeons' predictions.

In this study, Oddone et al.'s primary conclusion was that surgeons, on average as a group, make reliable global judgments about the appropriateness of surgery and that these

judgments correlate highly with the results of decision analysis models. However, they also noted several surgeons in their study who made judgments that were not always internally consistent and frequently contradicted findings from decision models. In fact, three of the nine surgeons in the study deemed operating appropriate for almost all conditions, regardless of which appropriateness method was used. The authors use these findings to support the claim that instead of debating the appropriateness of surgery, surgeons should instead focus more explicitly on reaching consensus on the individual components of the decision, i.e. the risks and the benefits involved.[55]

Both the McClellan and Brook study and the Duke study have important shortcomings that are relevant here. First, both studies rely on a small sample size of physicians, many of whom in the Duke study were not even surgeons (some were neurologists, radiologists, or family doctors, who supposedly have expertise in matters relating to carotid endarterectomy). The studies also only evaluated a single type of operation. These limitations greatly limit the generalizability of their findings. Second, the decision inputs that were considered were exceedingly narrow. Instead of focusing on all the possible outcomes—both positive and negative—the authors only made explicit the outcome of stroke, either major or minor, in their decision model. Without evaluating the likelihood that the procedure will actually achieve its intended goals as well as the other myriad adverse outcomes that may follow either operative or non-operative management, the study cannot adequately compare global judgment and decision analysis outcomes. What remains missing from the literature then is a widespread assessment of surgeon decision-making that uses a broad sample of surgeons and simultaneously assesses the ways in which surgeons predict the relevant positive and negative outcomes following surgery or medical management.

More contemporarily, there have been almost no efforts to evaluate how surgeons evaluate risks and benefits in their assessment of whether a patient needs an operation. One study, which was previously mentioned in the qualitative methods section, presented surgeons with 5 clinical vignettes and asked them to predict the outcome for a patient with and without an operation. As expected, surgeons that chose to operate were more likely to predict a higher likelihood of positive outcome with an operation than those that chose not to operate.[27] The investigators also assessed the surgeons' confidence with their predictions and found that confidence dramatically increased over the first three years of residency and thereafter plateaued all the way through to the maximum measured 30 years of experience. Finally, this study also found that novices and experts predicted outcomes similarly, but that novices were more likely to favor operating. This study, however, sampled only 22 surgeons from all levels of training. Also, they asked surgeons to predict the treatment outcome on one linear scale ranging from 0 to 100. Without measuring each possible outcome separately (risks and benefits) it is not possible to assess how each of these parameters independently influenced the decision to operate.

One final study presented a group of 60 general surgeons in New Zealand with clinical vignettes and asked them whether or not the described patient required surgical intervention. Once more, this study noted poor agreement between surgeons (r=0.48). However, this study went a step further to determine why these surgeons differed in their assessment. Using cluster analysis, the authors were able to group surgeons by 2 main characteristics: those that gave higher priority to a patient's future quality of life and to his/her specific diagnosis and those surgeons that placed more weight on the details of the treatment and the probability of future complications.[56] Together, these studies suggest that surgeons do in fact differ in how they judge risks and benefits in the context of surgical decision-making. However, in addition to relying on

small sample sizes, these studies make no attempt to link these judgments to the type of decisions made.

#### *2.2.2.3.1 Risk calculators*

Evident from these previously mentioned studies and similar studies from other areas of medicine,[54,57] is the fact that physicians appear to differ greatly from one another in their judgments of likelihood for various treatment outcomes. Although such disagreement is not routinely discussed and is, for the most part, not made explicit in professional dialogue or medical training, there have been many attempts to improve physicians' ability to predict adverse outcomes following treatment. Most notably has been the emergence of various risk categorization systems that stratify patients based on their risk profile. Many of these, such as the Injury Severity Score (ISS) in trauma, or the APACHE score in critical care medicine, are used primarily for research purposes. Other categorization systems, such as the American Society of Anesthesiologists (ASA) physical status classification system, are used routinely in both research and clinical practice.

To evaluate how accurately surgeons assess patient's surgical risk, several studies have compared surgeons' predictions to those calculated by various risk stratification systems. Results of these studies have been mixed. One study compared predictions by both surgeons and anesthesiologists with those of an objective prediction tool called the Portsmouth Physiological and Operative Severity Score for the Enumeration of Mortality and Morbidity (P-POSSUM). The authors found that clinical judgments of mortality for both surgeons and anesthesiologists compared favorably with the P-POSSUM scoring system.[58] This finding was confirmed by another similar but smaller study.[59]

However, when compared directly with actual patient outcomes, surgeons risk judgments appear less accurate. For example, a study asked five surgeons to judge the likelihood of patient morbidity and mortality soon after completing an operation and then compared those estimates with the patients' eventual outcomes. This study found that, on average, surgeons overestimated the risk of complications for elective surgery, but underestimated the risk of complications for emergency surgery.[60] Although the authors are careful to avoid criticism, they speculate that surgeons might have underestimated the risks of emergent surgery because most of the complications that were recorded occurred after emergency cases and therefore, surgeons might not have appreciated how common such complications actually were. However, an alternate, more critical explanation is proposed by the authors of an earlier study that found surgeons were not able to reliably identify high-risk patients as well as an independent physician who simply performed a thorough physical exam and reviewed a handful of blood tests.[61] Although published in 1986 and preceding much of the advances in surgical risk-assessment, these authors suggest that operative risk assessment should rely on more than just a surgeon's global judgment.

In recent years, there has been an emergence of many different types of risk calculators that are available to surgeons online that are intended to help surgeons predict the specific risks of an operation for an individual patient. Such calculators have been developed, for example, for use in bariatric surgery,[8] colorectal surgery,[9] and more recently a universal risk calculator that can be used more broadly for a wide variety of surgical procedures.[7] This latter calculator, developed by ACS-NSQIP, has been well validated. However, the influence that this or any other calculator has on surgeons' assessment of risk or how it may potentially impact their decision to recommend an operation has not yet been evaluated.

## 2.3 How surgeons *might* make decisions

The previous sections highlight the literature on several conceptual frameworks for surgical decision-making and provide an overview of both the qualitative and quantitative studies on this topic. There is, however, a notable lacking of a well-developed literature on the main interest of this dissertation—that is, a detailed explanation of the mechanisms underlying how surgeons make decisions in the face of clinical uncertainty. Therefore, to round out the relevant background on this topic, it is necessary to consider literature from the field of behavioral science, which over the past 4 decades in particular, has made tremendous progress towards understanding how people (not just surgeons) make difficult decisions. The final section of this literature review will therefore discuss some of this relevant literature, focusing on the overarching theories and concepts instead of a detailed summary of each theory's abundant empirical evidence. Wherever possible, however, these theories are supported by relevant evidence from healthcare.

### 2.3.1 Uncertainty to decision in three steps

As already mentioned, decisions under uncertainty involve a tradeoff between risks and benefits. Proposed here, as a means by which to consider a broad range of behavioral science literature, is a framework that divides the process of making a surgical decision in the face of uncertainty into 3 distinct steps. First, the level of risk and benefit involved in the decision must be *predicted* for the patient in question. After considering the context of the clinical encounter, the surgeon must then *interpret* the level of risk and consider how that risk applies to the patient in question. Finally, the surgeon must choose how to *respond* to the perceived level of risk, a process that results in the surgeon's decision to either operate or not operate. As the following sections reveal, errors in judgment are possible—perhaps even likely—at each step in this

process, the result being that individuals may follow slightly different decision paths and often reach different decisions.

### 2.3.2 Risk prediction

Predicting the level of risk and benefit for a decision is challenging. Even among experts, there is little consensus on the levels of risk associated with different hazards.[62] There are several possible theories to explain why risk prediction varies so significantly between people. One explanation is the *substitution effect*. As Daniel Kahneman, the Nobel Prize winning psychologist who popularized the field of decision science and behavioral economics, summarizes: "when faced with a difficult question, we often answer an easier one instead, usually without noticing the substitution."[16] Coming up with a precise estimate of risk is exceedingly difficult and experts may instead ask themselves simpler questions that allow them to reach an answer to the more challenging one indirectly. For surgeons, this may involve quantifying a level of risk in relation to a known entity, such as an average patient undergoing a particular operation, rather than a specific, individual patient undergoing the same operation.

Closely related to the substitution effect is another effect that biases individuals' estimates: the *anchoring effect*. This occurs when people consider a particular value for an unknown quantity and allow that value to influence their prediction, even when the value is unrelated to the problem in question. For example, people asked to guess the age at which Gandhi died are more likely to guess higher values if they are first asked if he was older than 114 than if they are first asked if he was older than 35.[63]

Of course, non-random numbers can also serve as prediction anchors.[64] If a surgeon is asked to predict the probability of an adverse outcome for a patient after being told that a colleague predicted a likelihood of 100%, the surgeon is more likely to predict a high number

than if the other surgeon predicted a likelihood of 10%. Similarly, providing surgeons with external, objective evidence about the risk of a certain event, using either evidence from the literature or validated risk calculators, can also anchor estimates. One study gave surgical trainees cognitive feedback after they were asked to predict the probability of operative mortality for a hypothetical patient. The feedback included anchoring predictions by giving values for very-high or very-low risk patients and also by giving the trainee input on whether he/she generally over- or under-estimated risk. After the intervention, participants improved the reliability of their predictions.[65]

As is evident by these previous examples, risk judgment in the real world does not occur in a vacuum. Surgeons apply their expertise to come up with reasonable estimates, but they are also influenced by a variety of external factors in their environment. Unsurprisingly, they are also guided by their own personal experience. In particular, there is another cognitive heuristic that explains how people incorporate their experience into their daily decisions. The *availability heuristic* states that information or events that come most easily to mind have a disproportionate effect on thinking and behavior. People tend to estimate that homicides are more frequent than suicides, despite the opposite being true, presumably due to the prevalence of homicide and other violent crime reports in the media. The frequency of these news stories and the emotional response they tend to generate results in people having more ready access to these events in their memory, causing them to think their prevalence is higher than it truly is. Negative outcomes in surgery, such as complications or death, similarly weigh heavily on surgeons' minds and can have remarkable effects on their decision-making. By artificially exposing a group of surgeons to an adverse outcome in a simulation, one study demonstrated that surgeons were more likely to change their behavior to avoid a similar complication from occurring again.[66] A surgeon that has

recently experienced a bad complication will therefore judge the risks of an operation to be higher than another surgeon who is not biased by such an experience.

The prediction of risk is thus informed by far more than just the nature of the choice options. Recent events, cognitive heuristics, and the way in which the choice is presented, all determine the level of risk predicted. Additionally, there is a large body of literature supporting the notion that the level of risk perceived (and predicted) is directly influenced by how a person feels about the idea or activity in question.[25] A person's *affect*, referring to the 'goodness' or 'badness' experienced in response to a stimulus, is highly correlated with the level of risk predicted. People who feel positively towards an activity tend to predict its benefits to be higher and its risks lower. Those with negative feelings about the activity predict the opposite. Also relevant from this body of literature is the strong negative correlation between perceived risks and perceived benefits. People tend to think that activities that are high in benefit, such as vaccines, antibiotics, and x-rays, are also relatively low in risk, while those who believe such activities are low in benefit tend to believe that they are high in risk.[67] In fact, when people are provided with data on the true risks of an activity and that risk is lower than they would otherwise expect, those subjects simultaneously change their estimates for the benefits, reporting the benefits to be higher than they originally thought.[67] These results have been further demonstrated among expert subjects in the fields of toxicology and finance.[68,69] Translating this research to the realm of surgery, one would expect surgeon's perception of risk and benefit to also be highly negatively correlated.

### 2.3.3 Risk interpretation

Considered in isolation, the absolute level of risk judged by a surgeon represents little more than a probability estimate. Even if completely accurate, a 78% chance of death, for

example, has no inherent meaning in and of itself. For a patient undergoing an elective hernia repair, this number is undoubtedly prohibitively high. Conversely, for a patient with a ruptured abdominal aortic aneurysm, this level of risk pales in comparison to the otherwise certain death that would result if the aneurysm were not repaired. A 78% chance of death following an operation in this scenarios is therefore permissively low. Risks and benefits are therefore inextricably tied to the context in which they arise and it is the job of the surgeon to interpret these values within each clinical encounter.

One theory has been supported by substantial empirical evidence to explain how clinicians interpret risk. 'Fuzzy trace theory' states that risk is interpreted not as a mathematical calculus based on absolute values of risk predicted, but instead on an overarching 'gist' representation of the risk.[70] According to this theory, a surgeon interprets the 78% chance of death as a gist of that number, such as "that's really high risk" or "that's really bad." The gist risk is thereby also encoded with affect. It is not that the 78% is understood or interpreted for its absolute value, but instead, in each clinical scenario, the surgeon makes a judgment and labels that value with an affective valence, such as high, medium, or low risk.

The interpretation of risk may also depend on the way in which risks are framed. For example, consider a patient who presents for a second opinion after being deemed too high risk a surgical candidate by another surgeon. The first surgeon saw the patient "framed" only by her clinical presentation, while the second surgeon may see the patient "framed" not only by her clinical presentation but also by the knowledge that she is unhappy with the first surgeon's treatment recommendation. This difference may influence the second surgeon to interpret the patient's risk profile as relatively lower, possibly making him/her more likely to offer an operation. The fact that patient encounters are susceptible to framing effects is well established

in the medical literature.[71-73] One well-known example demonstrated that clinicians' decisions differed depending on whether risks were framed as percent mortality or as percent survival, even when the numbers were equivalent.[74] Similarly, clinicians have been shown to have difficulty distinguishing between values of identical risk when they are framed either as absolute or relative risk.

Individuals are also largely insensitive to probability and have difficulty distinguishing and interpreting various levels of risk, particularly when events are rare. One's mental image for winning a state lottery, for example, is similar if the chance of winning is 1 in 10,000 as it is if the chance of winning is 1 in 10,000,000.[25] Small probabilities have a paradoxical effect on the human mind. People are much more willing to pay money to reduce the chance of an adverse event from 1% to 0% than they are to reduce the same event's probability from 49% to 48%.[25] In the field of surgery, where rare events must almost always be taken into consideration, this cognitive bias is likely to play a significant role in decision-making.

### 2.3.4 Risk response

Once the risk has been estimated and interpreted in context, the final step for a surgeon is to determine what level of risk is acceptable for both the surgeon and the patient. At this stage of navigating uncertainty, several other cognitive processes play a role. First is the well-established principle of loss aversion, a major component of the prospect theory of decision-making advanced by Amos Tversky and Daniel Kahneman.[75] This principle states that most people are, by nature, loss averse. Translating across many cultures and domains, people tend to tolerate higher levels of risk to avoid a loss than they are when seeking a comparable gain. For surgeon decision-making this theory suggests that surgeons may be more likely to make a risky choice if they are faced with an otherwise certain loss but may not make the risky choice if they are faced

with an otherwise certain gain. Because depending on the clinical scenario both operating and not operating could be considered to be a risky choice, it is difficult to know how loss aversion might influence surgeons' decisions to operate.

Further informing a surgeon's willingness to respond to a given level of risk is the fear that choosing the wrong option might induce regret at a future time. For medical interventions, this fear seems to predispose clinicians to preferentially offer invasive treatments.[76] Faced with concern that inaction will lead to regret, surgeons would therefore choose to accept higher levels of risk and operate on a patient. Regardless of the outcome, choosing to operate will prevent the surgeon from wondering in the future what might have happened if an operation had been performed.

One final trait that might influence whether or not a surgeon chooses to operate on a patient with a given level of risk is that surgeon's overall risk attitude. As established by an abundance of previous literature, an individual's behavior in risky situations depends largely on that individual's overall attitude towards risk, which tends to fall on a spectrum ranging from risk seeking to risk averse. While specific studies linking a surgeon's risk attitude to the decision to operate are lacking, risk attitude appears to inform decisions made under uncertainty in two important ways. First, risk attitude appears to influence the level of risk perceived, thereby influencing an individual's prediction of risk. More risk-seeking individuals may simply perceive their actions to be less risky than a risk averse individual would. Secondly, risk attitude also influences whether an individual will act on a given level of risk. [77] In relation to the 3 steps described here, risk attitude influences both risk prediction and risk response. Interestingly, an individual's attitude towards risk is associated with several personal characteristics, including demographic factors such as age and gender.[77,78]

Risk attitude has been shown to be an important predictor of clinical decisions in several areas of medicine. For example, in emergency medicine, risk-averse physicians are more likely to admit patients with chest pain to the hospital and also to order certain laboratory tests than their more risk-seeking colleagues.[79] In surgery, one study demonstrated an association between surgeon risk attitude and the likelihood of creating an ostomy to divert stool away from a tenuous connection between two intestinal ends that is at risk of leaking.[80] Another study demonstrated that surgeons' attitude towards risk differs between countries, a finding that they attribute to differences in the doctor-patient relationship and systems of medical education.[81] The associations between risk attitude and medical decision-making, however, may not apply equally to all areas of medicine.[82] Finally, in addition to the above mentioned cognitive processes, clinicians are also susceptible to other biases including hindsight bias,[83] status-quo bias,[84] and overconfidence,[85-88] all of which may play a role in influencing physician judgment and decision-making.

## 2.4 Summary

The study of surgical decision-making is still in its infancy and as a result, there is a paucity of data with which to understand how surgeons make the consequential decision to operate on a patient. Some authors have laid out conceptual models that offer both a descriptive and prescriptive framework within which decisions can be analyzed. However, none of these models have been validated by empirical data. Qualitative data provides some insight into how surgeons think about their decision-making process from their own perspective. Due to the exclusively subjective nature of such data, these qualitative studies may be biased towards an overly favorable representation of surgeons' decision-making. On the other hand, much of the quantitative data collected to date—from analyzing surgical utilization rates across regions to

comparing surgeons' decisions in standardized clinical scenarios—reveal that surgeons' decision-making processes vary from one surgeon to the next. A more developed literature can be drawn on from the field of behavioral science, which suggests that decisions made under uncertainty—like those in surgery—can be conceptualized as a tradeoff between the risks and benefits associated with each choice. In the face of such complexity, human beings have evolved certain heuristics, or mental shortcuts, with which to ease the cognitive strain associated with these decisions, thereby allowing individuals to rapidly and (mostly) effectively navigate the complexities of the surrounding world. As outlined in this chapter, the process of evaluating both the risks and benefits associated with a decision involves 3 distinct although possibly contemporaneous steps: risk prediction, risk interpretation, and risk response, each of which is susceptible to bias and error.

## 2.5 Contribution to literature

This dissertation will fill a significant void in the literature on surgical decision-making. First, it will attempt to document variation between surgeons in their decision to operate by using detailed clinical scenarios in which there is no obviously dominant treatment option. Second, it will explore how surgeons vary in how they perceive the risks and benefits associated with various treatment options, while holding all patient characteristics constant in a clinical vignette. Third, for the first time, this dissertation will examine, within a large sample of surgeons, the relationship between surgeons' perceptions of treatment risks and benefits and their decision to perform an operation, as well as determine how surgeons' demographic and clinical characteristics, such as age, gender, race/ethnicity, and experience influence their perception of risks and benefits. Fourth, it will also test for the first time the role of surgeons' overall risk attitude in the perception of risks and benefits and the effect of this personality trait on their

decision to operate. Finally, this dissertation will also explore for the first time how use of a risk calculator might influence surgeon judgment and decision-making in a clinical context. Ultimately, the goals of this dissertation are to outline a novel conceptual framework for surgical decision-making, provide empirical evidence to support this model, and to provide important evidence that may be used to guide future efforts aiming to reduce variation in surgical practice.

# 3 Conceptual model

## 3.1 Defining the outcome

A surgeon's decision to operate on a patient is influenced by a wide variety of factors. Before exploring these factors, however, it is necessary to define exactly what is meant by the phrase "the decision to operate." In surgical practice, the decision to operate can be fluid, differing from time to time, and from person to person. It is therefore often challenging to identify the exact point in time when a decision is made. Particularly in the emergent setting, clinical decisions are made in a dynamic and iterative fashion. A patient that one moment appears to be improving with medical management can rapidly take a turn for the worse, forcing the surgeon's hand to operate. Such is the nature of surgical disease. But even in such rapidly evolving clinical situations, there tends to be at least one discrete moment (often several moments) when a surgeon pauses to evaluate the available data and makes a decision to either stay the course with non-operative therapy or to recommend that the patient undergo an operation.

In the elective, outpatient setting, identifying the moment of the decision can be equally elusive. In this controlled environment, the decision typically takes place towards the end of a clinic consultation after a discussion of the likely diagnosis and treatment options. Here a surgeon typically offers a recommendation for or against an operation and communicates that recommendation to the patient, at which point, the patient may voice agreement or disagreement. Regardless of the actual treatment plan pursued, there is still a distinct moment when the surgeon makes the decision to recommend an operation that precedes the deliberative conversation. For

this reason, the outcome of interest that is most relevant here and most closely approximates the process I am attempting to describe is the surgeon's initial decision to recommend surgery.

In its simplest form, the decision to recommend surgery involves a choice between two competing options: to operate or to not operate on a patient. In almost all cases, choosing to not operate, or choosing to not recommend surgery, involves a simultaneous decision to pursue an alternative treatment option, typically referred to by shorthand phrases such as "medical management," "conservative therapy," or "non-operative management," all of which I will use interchangeably. In the context of surgical disease, medical management typically includes treatments to maintain physiologic homeostasis (e.g. fluids, oxygen) as well as treatments to hinder the threatening disease process (e.g. antibiotics).

The initial decision to recommend an operation therefore serves as the end point for the conceptual model laid out here. This decision is extremely complex and is influenced by a wide variety of factors. Discussion of these factors will begin with a description of the main predictor of the decision, a surgeon's perception of the relative risks and benefits of operating. This will be followed by a description of the patient factors, corresponding to those plotted in the center of the conceptual model in Figure 3.1, followed by the surgeon factors to the left, and finally the environmental factors to the model's right. Wherever relevant, a brief reference to testable hypotheses is mentioned in the corresponding section. However, a more detailed summary of all the included research questions and hypotheses are included at the chapter's end.

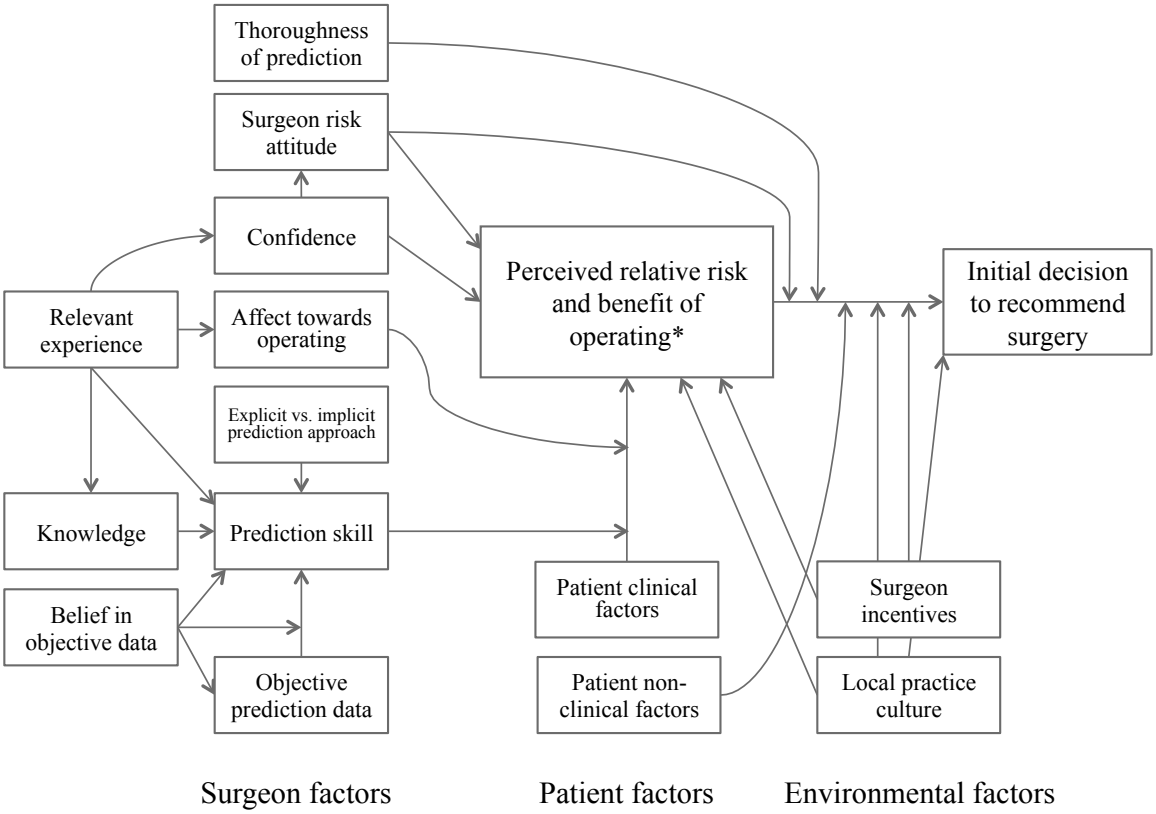



**Figure 3.1 Factors influencing the surgeon's initial decision to recommend an operation**
*Relative risk and benefit of operating in comparison to not operating. Perceived relative risk and perceived relative benefit of operating likely moderate each other's effect on the initial decision to recommend surgery.
Affect, in its psychological meaning, refers to a surgeon's involuntary, emotional response (dread vs. affection) to a clinical encounter. Responding to one's affect is also typically equivalent to "going with your gut."

## 3.2 Weighing the risks and benefits

The decision to operate, like many decisions in medicine, is made under varying degrees of uncertainty. According to normative decision theory, decisions made under uncertainty involve a tradeoff between the competing concepts of risks and benefits. Risks, as they pertains to medical decisions, are a function of the possibility of a negative outcome, such as death or a serious complication, and the probability with which that outcome may occur.[89] If there is no chance of a negative outcome occurring, the treatment is said to be risk free, a scenario that essentially never occurs in medicine. Benefits, on the other hand, refer to the possibility of a positive outcome and the probability of that outcome occurring. In surgery, the exact benefits

may differ from one patient to the next, as well as from one operation to the next, but usually include outcomes such as a technically successful operation, the alleviation of the patient's symptoms, and return of the patient to baseline health. It is worth noting that risks and benefits are not necessarily inversely related; it is possible for a negative outcome (e.g. a complication) to occur and for the patient to experience a full recovery from both the original illness and the operation performed to treat that illness. For instance, a patient may have an 80% chance of suffering a surgical site infection but also stand a 90% chance of experiencing full recovery within a defined time frame.

Risks and benefits of surgical treatments may include both short- and long-term outcomes, and may also range from physical outcomes (stroke, limb amputation, death) to emotional or psychological outcomes (stress, peace of mind, uncertainty). Finally, the risks and benefits of an operation may be borne by the patient alone, or also by the patient's family and community.

For many decisions in life, risks and benefits can be evaluated as attributes applied singularly to an action. Inaction, or choosing the status quo, is often free of risks and benefits. Choosing to bungee jump off a bridge requires accepting the risk of injury in order to gain the benefit of an adrenaline rush. Choosing not to jump is a risk-free and benefit-free decision. In medicine and particularly in surgery, however, there are risks and benefits associated with action, such as choosing to operate, as well as with inaction, such as choosing not to operate (if choosing not to operate were free of risk then there would rarely be a reason to consider an operation). Therefore, given the different possible outcomes for operating and not operating, a decision to operate needs to consider the *relative* risks and *relative* benefits of operating in comparison to not operating. An operation is therefore favored as the dominant choice when the risks of not

operating are greater than the risks of operating and when the benefits of operating are greater than the benefits of not operating. Of course, there are often more complicated scenarios in which there is no dominant choice, such as the case where relative to not operating, the risks of operating are greater, but so too are the benefits.

The effects of the relative risks of operating on the decision to operate likely depend directly on the relative benefits of operating, and vice versa. For example, given a high level of relative risks of operating, surgeons may be unlikely to recommend an operation even when the relative benefits of operating are favorable. Similarly, when the relative benefits of operating are high, surgeons may still be likely to recommend an operation despite high relative risks. Therefore, there may be a moderating effect between surgeons' perception of relative risk and relative benefit on their decision to recommend an operation.

While there have been tremendous advances in surgeons' ability to predict patient outcomes and to measure the risks and benefits of various treatments, the science of risk stratification is still in its infancy. Even when objective measures of risk and benefit are available and accurate, it is still left to the surgeon's discretion whether or not to trust these measures and whether or not to include them in the decision-making process. Therefore, the conceptual model of surgical decision-making described here focuses on the surgeon' *perceived* risks and benefits of operating and not operating instead of any direct and objective measure of these parameters. In fact, one of the goals of this dissertation is to measure how closely surgeons' perceptions of risk match the validated and objective measures of risk derived from large samples of comparable patients.

Examining the role of risk and benefit perception in surgical decision-making lends itself to several testable hypotheses that will subsequently be laid out in more detail. For example,

surgeons may differ in their treatment decisions because they vary in how they perceive the risks and benefits of the possible treatment options. Surgeons may be more likely to operate if they perceive the risks of operating to be less than the risks of not operating. Similarly, surgeons may also be more likely to operate if they perceive the benefits of operating to be greater than the benefits of not operating. In addition to examining these *relative* risks and benefits, one may make parallel conjectures about surgeons' perception of the *absolute* risks and *absolute* benefits of operating and not operating. Examining how these measures of risk and benefit are associated with surgeons' decisions allows for evaluation of how surgeons weigh each measure of risk and benefit relative to the others.

In order to paint a complete picture of the surgeon's perception of risk and benefit and describe how these factors influence the decision to recommend an operation, it is necessary to consider the myriad factors that might influence the level of risk and benefit perceived by the surgeon as well as those that more directly influence the decision to operate. These factors can be organized broadly into three categories: patient factors, surgeon factors, and environmental factors.

## 3.3 Patient factors

### 3.3.1 Clinical factors

Likely above all else, clinical factors inform the risks and benefits perceived by surgeons, regardless of the treatment pursued. The list of clinical factors that are relevant to a surgeon's decision to recommend an operation is extensive and far too detailed to cover here. In brief, these clinical factors include age, functional status, the presence of comorbid disease, the patient's presumed diagnosis and the associated uncertainty of that diagnosis, the acuity of the illness, the urgency with which treatment is needed, and the availability of other feasible treatment options.

The levels of risk and benefit for each patient vary both in terms of their absolute values as well as in terms of their values relative to one another. For example, a young, healthy patient with an inguinal hernia would be a low-risk patient regardless of treatment chosen. Conversely, an elderly, unhealthy patient with a ruptured abdominal aortic aneurysm would face exceedingly high risks both with or without an operation.

Even minor variations in clinical characteristics can alter surgeons' perceptions of risks and benefits. Such differences are one of the factors that make the study of clinical decision-making so difficult. It is nearly impossible to compare two surgeons' decisions since even two similar patients will always differ from one another in some measurable or immeasurable way. An example here would be two young, healthy patients with appendicitis, one with symptoms for 24 hours, the other with symptoms for 48 hours. It remains possible that the first patient actually has more advanced disease despite a shorter symptom course, making an operation more difficult and higher risk. Even under the most favorable circumstances, when careful measurement of all relevant clinical factors is possible, it will always be possible to point to possible reasons, reasonable or not, for how the two patients are not truly comparable. Nevertheless, to the extent that it can be accurately measured and recorded, a patient's clinical condition plays an important role in informing a surgeon's perception of risk and benefit.[27,90]

### 3.3.2 Non-clinical factors

While clinical factors primarily influence the decision by informing the level of perceived risks and benefits, there are also various non-clinical factors pertaining to a patient that may inform whether or not the surgeon responds to the perceived level of risk and benefit to recommend an operation. These non-clinical factors include a patient's social or cultural situation, family support, occupation, lifestyle choices, leisure activities, geographic location,

psychological characteristics, will to survive, prior advanced directives, current mental state, as well as any pressure placed by the patient on the surgeon in favor of one treatment over the other.[29]

As an illustration of these concepts, consider a simplified example of two patients for whom a surgeon perceives the relative risks and benefits of an operation to be identical. One patient has a terminal cancer, severe dementia, and a poor quality of life with limited social support. The other patient is also elderly and lives in a nursing home, but she enjoys a rich and fulfilling social life, is precise in her description of her life goals, and makes a coherent argument for the tremendous satisfaction she experiences from her biweekly visits with her granddaughter. It is easy to see how a surgeon, given the same risks and benefits, might preferentially choose to operate on the second patient, while possibly choosing to pursue palliative care for the first.

Other relevant non-clinical factors may include a patient's race, ethnicity, or sex. Similar to discriminative practices outside of medicine, there appear to be scenarios in which clinicians vary their clinical decisions depending on these demographic characteristics.[71] Therefore, given a level of perceived risk and benefit associated with a treatment, surgeons may choose different treatments depending on these patient factors. The goal of this dissertation, however, is to examine how surgeons make the decision to operate when the patient's clinical and non-clinical characteristics are held constant. Therefore, I will not examine any hypotheses relating to the effects of patient factors on surgical decision-making.

## 3.4 Surgeon factors

The decision to operate is a personal and carefully weighed decision that draws on, among other things, a surgeon's knowledge, experience, emotions, and confidence. These factors contribute to the surgeon-specific factors that must be considered in order to understand what

makes one surgeon choose to operate and another choose not to. According to the conceptual model outlined here, surgeon-specific factors inform the decision to operate primarily by influencing the surgeon's perception of risk and benefit. Some factors, including the surgeon's affect towards operating and the surgeon's prediction skill, inform the decision to operate by moderating the relationship between patient clinical factors and the perceived relative risks and benefits of operating. Other factors, such as a surgeon's risk attitude and a surgeon's consideration of all possible treatment options (thoroughness of prediction), moderate the relationship between the surgeon's perception of relative risk and benefit of operating and the decision to recommend an operation. The description of the surgeon factors in the conceptual model will roughly follow the organization of the conceptual model in Figure 3.1, moving from top to bottom.

### 3.4.1 Thoroughness of prediction

A critical part of the evaluation of risks and benefits is the consideration of all the possible treatment options for which risks and benefits should be weighed. A surgeon who predicts thoroughly takes into consideration the risks and benefits of all possible treatment options, which, in the case of the decision to recommend an operation, include most obviously the risks and benefits associated with operating as well as those associated with not operating. However, it is conceivable that this balanced assessment may not always take place. For example, a surgeon who is leaning towards operating may not fully consider the risks and benefits of the alternative—not operating. Without giving equal weight to both treatment options, the result is a biased assessment of the relative risks and benefits of operating. One potential manifestation of this theoretical framework, and a hypothesis that will be tested in this dissertation, is that when surgeons are only forced to consider the risks of operating (as opposed

to considering the risks of both operating and not operating), they may lend more weight to the risks of operating than to the risks of not operating when making treatment decisions. Conversely, surgeons may weigh non-operative risks more heavily than operative risks in their treatment decisions if they are forced to only explicitly consider outcomes associated with not operating.

### 3.4.2 Surgeon risk attitude

Predicting the relative risks and benefits of operating is nothing more than an intellectual exercise if no action is taken in response to these judgments. Therefore, once a surgeon has established the relative risks and benefits of each treatment option, he/she must decide whether or not the perceived level of risk and benefit calls for a decision of action, in the form of operating, or a decision of inaction, in the form of not operating. How a surgeon responds to a given level of risk is determined, at least in part, by the surgeon's overall risk attitude. Some surgeons may be willing to take on exceedingly high levels of risk and are inherently risk seeking, while others would avoid risky situations at almost any cost and can be said to be risk averse. While an individual's risk attitude appears to be an inherent personality trait, it is also well accepted by behavioral scientists that risk attitudes tend to be domain specific; an individual who gambles on the stock market is not necessarily the type of person who would BASE jump off the Eiffel Tower.

There is evidence to suggest that risk attitude influences decision-making in two important ways. First, given a level of risk, the surgeon's risk attitude affects the response to risk by determining an individual's risk taking behavior.[77] Secondly, the extent of risk aversion or risk seeking of a surgeon also directly informs the level of risk *perceived* and therefore influences the values of risk and benefit predicted. Risk-seeking individuals may be more willing

to take on higher levels of risk, but they may also simply perceive the absolute level of risk to be lower than would risk-averse individuals.[77] Therefore, two testable hypotheses arise. First is the hypothesis that risk-seeking surgeons perceive lower levels of relative (and absolute) risk of operating than risk-averse surgeons and second is the hypothesis that risk-seeking surgeons are more likely to recommend an operation for a given level of perceived operative risk.

### 3.4.3 Confidence

Risk attitude is also likely influenced by a surgeon's confidence, which has been shown to vary depending on an individual's demographic characteristics, such as age, race, gender, and surgical specialty.[27,53,77,78,91] Additionally, surgeon confidence also directly influences the perceived levels of risk and benefit associated with operating and not operating. Confidence in one's technical skill and operative ability likely results in a perception of lower risk and higher benefit associated with operating. Alternatively, surgeons may be confident in their clinical decision-making skills and therefore, even if they choose to not operate, they will remain confident in their decision and will perceive lower levels of risk and higher benefit associated with their chosen treatment to *not* operate.

Confidence is itself influenced by many other surgeon-specific factors. An abundance of relevant experience, as previously mentioned, can cause the surgeon to be more confident in her abilities to treat the patient successfully, an effect that would likely result in the perception of lower total risks and greater total benefits, regardless of which decision is made.

### 3.4.4 Affect towards operating

Feelings and emotions play an important role in all decision-making processes. Consistent with abundant evidence from the decision science literature, the emotional response provoked by a stimulus can have a pronounced affect on the decision-making process by

informing the levels of risks and benefits perceived. This emotional response—or affect—refers to the 'goodness' or 'badness' experienced by an individual in response to a stimulus. An individual that experiences a positive emotion in response to a stimulus tends to consider that stimulus to be low in risk and high in benefit.[68] On the other hand, an individual who responds negatively to the same stimulus is more likely to think the risks posed by that stimulus are higher and its benefits lower. In this way, our perceptions of risk and benefit are tied closely to the emotional valence that we attach to a specific event or activity.

In surgery, these effects, although not yet studied, likely play a role in shaping the decision to operate. Some surgeons enjoy performing certain types of operations and consulting on a patient that may need such an operation may evoke more positive emotions. For example, a vascular surgeon who particularly enjoys performing carotid endarterectomies will feel a more positive emotional response when seeing a patient with carotid stenosis (narrowing of the carotid artery) than a surgeon who does not particularly enjoy performing this operation. The *affect heuristic*, the psychological shortcut people use to make decisions as described by the psychologist Paul Slovic and others, predicts that the surgeon who enjoys the operation will judge lower risks and higher benefits for the operation than the surgeon who does not enjoy performing the operation. This psychological process is what is at play when people refer to decision-making based on "going with your gut." A surgeon's affective response towards a particular clinical encounter can thus weigh heavily in his/her decision to operate or not operate.

A surgeon's affective response to a clinical encounter can be shaped by many other, only some of which are included in this conceptual model. One such factor is a surgeon's technical skill for an operation. In the previous example, for instance, the surgeon who enjoys performing carotid endarterectomies may derive more joy because he/she is more skilled and adept at

performing the operation, as highlighted by the arrow connecting relevant experience and affect towards operating in Figure 3.1.

### 3.4.5 Prediction skill

Another surgeon factor that moderates the relationship between patient clinical factors and perceived relative risk and benefit of operating is a surgeon's skill at predicting clinical outcomes. As previously mentioned, for each treatment decision, it is necessary to predict the likelihood of various outcomes for each potential treatment strategy. Just as there are differences between surgeons in their technical skill and their ability to safely perform complex mechanical tasks, so too is there likely a difference in their ability to perform the cognitive task of judging the likelihood of various outcomes. Some surgeons may be better able than others to synthesize the complex and abundant clinical information before them, to consider all the possible positive and negative outcomes, and to assign each outcome an accurate probability.

One reason that a surgeon may be better equipped to accurately predict treatment outcomes is the extent of his/her knowledge base. Some surgeons have studied more extensively, read more widely, and remember more vividly the statistics from the published literature that pertain to a given patient. The ability to retrieve and apply this knowledge to patient care can serve to make treatment predictions more accurate.

In addition to the knowledge acquired from study, there is also the knowledge acquired from experience. In fact, relevant clinical experience may be more informative to a surgeon's knowledge base than even the most relevant journal article or textbook chapter. There are two ways that clinical experience may contribute to a surgeon's knowledge base. First, there is the conventional way that experience is thought to increase knowledge. By seeing many different types of patients over and over again, a surgeon is able to witness firsthand the variety of

different outcomes associated with each treatment option and is thus better equipped to judge their frequency. For example, surgeons with more experience treating patients with pancreatic cysts will have seen a great number of patients suffering through and recovering from the condition, and will therefore have more data to draw on to inform accurate predictions for various treatment outcomes. On the other hand, a surgeon that is seeing a patient with a liver abscess for the first time will have less experiential information on which to base a prediction.

Experience can also have another effect on a surgeon's prediction skill by accessing the less rational side of the mind. While it would be best to weigh all experiences equally by generating an outcome prediction based on an aggregate of all relevant experiences, there is much evidence to suggest that certain experiences are weighed more heavily by the mind than others. Drawing again on literature from decision science, experiences or thoughts that are more salient or more memorable tend to be weighed more heavily than routine, less significant experiences. The tendency to rely disproportionately on these experiences is referred to as the *availability heuristic*. Drawing an example from surgery, a situation may arise when a surgeon inadvertently causes a common bile duct injury during a laparoscopic cholecystectomy, a particularly devastating and dreaded complication. The theory suggests that despite having safely performed thousands of these operations, one extreme and negative experience will cause the surgeon to judge such a complication to be more common than it actually is. In fact, there is at least one study demonstrating that this effect may indeed play a role in influencing surgeon judgment.[66] Experience, which is typically considered beneficial to informing surgeons' knowledge, can thereby also be a source of prediction inaccuracy.

Also informing the skill and accuracy of a surgeon's predictions is the availability of objective data on the relevant outcomes. Objective data can arise from various sources, including

clinical trials, observational studies, or large databases that can be analyzed and used to develop risk calculators to derive patient-specific risk estimates. When these data are available, they can serve to improve the accuracy of a surgeon's predictions. However, this depends largely on whether or not the surgeon believes that the available objective data are relevant and apply to the patient in question. Although less frequently, many physicians still view evidence-based medicine and clinical guidelines with great skepticism and such surgeons are unlikely to incorporate such data into their own predictions.

Therefore, a surgeon's belief in objective data influences his/her prediction skill in two ways. First, it moderates the relationship between the availability of objective data and the surgeon's prediction skill or accuracy. Surgeons who believe in data will be more likely to incorporate those data into their predictions than those who tend to mistrust such data. Second, surgeons who believe in objective data may also routinely seek out such data—either in the published literature or by using surgical risk calculators—to inform their clinical predictions. These surgeons are therefore more rehearsed in making clinical predictions and may have, on average, better prediction skill. Two testable hypotheses arise here. First is the hypothesis that providing surgeons with objective clinical data improves their prediction accuracy. Second is the hypothesis that this improvement in prediction is greater for surgeons who actually believe in the value of objective data.

One final concept that contributes to the surgeon's prediction skill is how well the surgeon is able to incorporate risks and benefits into the decision to operate. The systematic approach, one that is consistent with expected utility theory, is to apply probabilities and utilities to each possible outcome and then to choose the treatment strategy that maximizes the overall outcome. However, it is also foreseeable that instead of only using probabilities associated with

risks and benefits to inform the decision to operate, a surgeon may choose to operate based on other, non-probability based constructs (such as an overall gestalt). The former, rational approach can be considered an explicit approach and the latter, less rational approach can be considered an implicit approach to outcome prediction.

## 3.5 Environmental factors

No clinical decision occurs in a vacuum. Each physician-patient encounter takes place in a particular clinical setting, where the surrounding people, rules, resources, and policies all play a role in determining whether or not a patient undergoes an operation. Some of these factors, such as the local practice culture of an institution, can influence the process before the surgeon makes a decision to recommend surgery. The vast majority of these systemic factors, however, do not directly influence the surgeon's decision. Instead, once a recommendation is made, these factors provide constraints, to which the patient and surgeon must attempt to adapt. In this context, the main environmental factors that are relevant to the decision to operate include local practice culture and the presence or absence of various surgeon incentives.

### 3.5.1 Local practice culture

The extensive literature on geographic variation in the surgical utilization rates suggests that surgeon practice differs independent of the forces of supply and demand, at least by the traditional measures of these concepts. Instead, there appear to be strong regional tendencies towards using surgery or not. There is no evidence to suggest exactly how local practice culture influences surgical practice patterns, however, this conceptual model theorizes that local practice culture has an effect on the decision to operate in two ways. First, local practice culture influences surgeon's perception of relative risks and benefits. And second, local practice culture predicts whether surgeons will choose to operate independent of the perceived risks and benefits.

By this latter mechanism, surgeons may choose their treatment strategy more on the prevailing local practice culture than on his or her risk/benefit assessment. For example, despite a belief in the value of non-operative management of appendicitis, a surgeon may choose to follow local convention and operate on all patients with appendicitis simply to avoid scrutiny or suspicion from his or her fellow surgeons.

Local practice culture is shaped by many factors, including the regional patterns of surgical practice as well as the social norms for each practice setting. Be they perceived or actual, these social norms can have profound effects on patient care, particularly when considering the differential use of surgery. For example, even within the same county, surgeons at one hospital may choose to surgically explore a patient with a bowel obstruction early on in the hospital course, while those at another hospital may be willing to wait several days before resorting to an operation. These forces are then further accentuated by constant peer feedback, which serve to help validate surgeons' decisions, but also perpetuate the local practice culture norms.[31,90]

### 3.5.2 Surgeon incentives

Aside from a patient's clinical presentation, there are certain contexts in which surgeons may tolerate a higher level of risk than they would in a different context. For example, a surgeon may find it financially lucrative to make one treatment decision over another.[39,92] Alternatively, the surgeon with a busy schedule the following day may choose to operate on a patient earlier than he otherwise would have to avoid having to work late the following day. Other examples of surgeon incentives that moderate the decision to operate include the type of surgical practice, teaching and time pressure, quality of the available assistant or equipment, desire to maintain favorable operative statistics, concern for personal image, as well as the surgeon's reputation

with referring physicians, colleagues, or nurses.[29,93-95] Similarly, fear of medical malpractice is frequently considered a culprit for excessive medical testing and treatment and it is likely that decisions to go to or stay away from the operating room are also affected by these concerns.[76]

### 3.5.3 External forces

As highlighted by the model, surgeons respond to their perceptions of relative risks and benefits to inform their decision to operate. However, once the surgeon has made a decision and a recommendation is communicated to a patient, there are various systemic forces that may determine whether or not the surgeon's recommendation is followed. These factors warrant mention but do not directly influence the outcome of interest—the surgeon's decision to recommend an operation—and are therefore not presented graphically in the conceptual model. For example, situational factors such as the available hospital resources, operating room availability, or other scheduling issues may act to prevent a recommended surgery from occurring and also to force an operation that was not recommended, particularly in the case where the initial recommendation was weak.

While clinical decisions are made on a patient-by-patient basis, the larger health care system, as sculpted by the various levels of health policy, can also alter the course of patient care by affecting whether or not a patient undergoes an operation. State or national policies on insurance coverage can have particularly significant effects, especially in the elective setting, sometimes preventing patients from receiving an operation recommended by a surgeon.[96]

In sum, the surgeon's decision to recommend an operation is complex and depends on a wide variety of factors ranging from a patient's clinical condition, to a surgeon's skill, experience, and cognitive processes, and even to environmental factors such as the local healthcare culture. At its core, however, the decision is informed by a careful assessment of the

risks and benefits associated with the two possible treatment options—to operate or not operate. Based on this conceptual model, I therefore aim to answer the following research questions by testing the associated hypotheses listed below. Each hypothesis also refers to the corresponding regression that will be further explained in the methods chapter.

## 3.6 Questions and hypotheses

**Question 1.** How does a surgeon's perception of the relative risks and benefits of operating influence the surgeon's decision to recommend an operation?

Not only is little known about how surgeons perceive the risks and benefits of different treatment options, it remains unclear how these perceptions might drive differences in treatment decisions. Specifically, I will explore how the perception of risk and benefit of surgery *relative* to those of medical management are associated with the decision outcome.

(Hypothesis 1.1) Controlling for patient characteristics and for any given level of perceived relative benefits of operating, surgeons are more likely to recommend an operation when the perceived relative risks of operating are lower (Regression 1).

(Hypothesis 1.2) Controlling for patient characteristics and for any given level of perceived relative risk of operating, surgeons are more likely to recommend an operation when the perceived relative benefits of operating are higher (Regression 1).

**Question 2.** How does a surgeon's risk attitude influence both risk perception and the decision to recommend an operation?

Considering clinical decisions are frequently made in the face of uncertainty in which risk is inherent, it is possible that a surgeon's risk attitude—propensity or aversion to risk—may

play a role in this decision. This role may relate to both how surgeons perceive the level of risk as well as how surgeons respond to a given level of risk.

(Hypothesis 2.1) Controlling for a surgeon's perceived relative risks and benefits of operating and for any given level of perceived relative benefits of operating, a surgeon's greater likelihood of operating in the face of lower perceived relative risks of operating is higher for risk-seeking surgeons than it is for risk-averse surgeons (Regression 1).

(Hypothesis 2.2) Risk-seeking surgeons perceive lower levels of relative risk than risk-averse surgeons (Regression 2).

(Hypothesis 2.3) Risk-seeking surgeons perceive lower levels of absolute operative risk than risk-averse surgeons (Regression 3).

**Question 3.** How does a surgeon's perception of the absolute risks and benefits of operating and not operating influence the surgeon's decision to recommend an operation?

In order to more carefully examine how risk and benefit perception influences the decision to operate, I will consider the risks and benefits of operating and the risks and benefits of not operating separately. This will allow for a more detailed examination of how these 4 concepts independently affect surgical decision-making.

(Hypothesis 3.1) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute operative risk are less likely to recommend an operation (Regression 4).

(Hypothesis 3.2) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute operative benefit are more likely to recommend an operation (Regression 4).

(Hypothesis 3.3) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute risks associated with not operating are more likely to recommend an operation (Regression 4).

(Hypothesis 3.4) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute benefit associated with not operating are less likely to recommend an operation (Regression 4).

**Question 4.** What are the surgeon characteristics that influence their perception of the risks of operating?

Surgeons differ in terms of their experience, their practice types, and the extent of their clinical training. I will examine how these characteristics influence their perception of risk and benefit.

(Hypothesis 4.1) Surgeons who have completed fellowships in trauma surgery or acute care surgery perceive lower relative risks of operating than those who did not complete a fellowship (Regression 2).

(Hypothesis 4.2) Surgeons perceive lower levels of absolute operative risk if they have more relevant experience, as measured by number of years in residency, greater number of years in practice, completion of a trauma/acute care surgery fellowship, and high operative volume (Regression 3).

(Hypothesis 4.3) Surgeons perceive higher levels of absolute operative benefit if they have more relevant experience, as measured by number of years in residency, greater number of years in practice, completion of a trauma/acute care surgery fellowship, and high operative volume (Regression 5).

**Question 5.** What is the effect of providing surgeons with objective data from a surgical risk calculator on their prediction of operative risk, their prediction skill, as measured by the difference between their prediction and the prediction of the risk calculator, and their decision to recommend an operation?

Risk calculators are intended to help surgeons judge the risks of operating on a given patient. However, their effects on actual surgeon perception in practice remain unclear. I will therefore evaluate the effects of providing surgeons with data from a risk calculator on their perception of operative risk. Ultimately, by helping surgeons improve their estimates of operative risk, risk calculators may help inform clinical decisions, yet as noted earlier, their effects on clinical decision-making are not known. Here I will examine whether providing surgeons with risk calculator data changes their clinical decisions.

(Hypothesis 5.1) Exposing surgeons to the risk calculator improves surgeons' prediction skill (Regression 6).

(Hypothesis 5.2) The improvement in prediction skill conferred by exposure to the risk calculator is greater for those surgeons who routinely use the calculator in their clinical practice than for those who do not (Regression 6).

(Hypothesis 5.3) If exposure to the risk calculator causes surgeons to perceive higher levels of absolute operative risk, then exposure to the risk calculator will also change their behavior and make surgeons less likely to recommend an operation (Regression 7).

However, if exposure to the risk calculator causes surgeons to predict lower average levels of operative risk, surgeons will change their behavior to be more likely to recommend an operation (Regression 7).

**Question 6.** What is the effect of asking surgeons to predict either the risks/benefits of operating or the risks/benefits of *not* operating on their decision to recommend an operation?

It is not known whether surgeons automatically incorporate the concepts of operative and non-operative risks and benefits into their treatment decisions. It is possible then, by prompting surgeons to think explicitly about only one of these alternatives (operative vs. non-operative) that their clinical decision might be biased accordingly.

(Hypothesis 6.1) The higher tendency for surgeons to operate when they perceive lower operative risks is more pronounced when surgeons are only asked to predict the risks and benefits of operating (and not the risks and benefits of *not* operating; regression 8).

(Hypothesis 6.2) The higher tendency for surgeons to operate when they perceive high levels of risks associated with *not* operating is more pronounced when surgeons are only asked to predict the risks and benefits of *not* operating (and not the risks and benefits of operating; regression 9).

# 4 Methods

## 4.1 Data source

Data for this dissertation were obtained from a prospective study that was conducted in partnership between researchers at the University of California, Los Angeles, the RAND Corporation, and the American College of Surgeons (ACS). The following sections explain how the data were obtained and provide details on the analytic plan for this dissertation.

### 4.1.1 Study sample

#### *4.1.1.1 Cohort*

The ACS is an organization comprised of roughly 80,000 surgeons from a wide variety of surgical subspecialties and represents the body of surgeons in matters relating to education, research, and politics. Due to the wide applicability of the study findings and because the study includes a test of the ACS-NSQIP risk calculator, the ACS agreed to endorse the study by recruiting surgeons to participate via an email invitation. In October 2014, the leadership at the ACS sent out an email invitation to all ACS members that were either currently enrolled in or had completed a residency in general surgery to participate in an online study. Both resident and attending surgeons were included in order to examine how decision-making patterns differ between junior and senior surgeons.

#### *4.1.1.2 Incentives*

For participation in the study, surgeons had the opportunity to participate in a Continued Medical Education activity after study completion and also had the option to enroll in a drawing to win one of four laptop computers.

#### *4.1.1.3 Response rate*

The email was sent to a total of 26,051 surgeons. Of these, a total of 10,197 surgeons opened the email and 2,204 surgeons clicked on the embedded link to participate in the study. Respondents were ineligible for participation if they had not completed and were not currently enrolled in a general surgery residency (n=36) or if they had retired (n=11). An additional 274 surgeons were excluded because they did not respond to any questions in the study, resulting in a final sample size of 1,880 participants. A response rate of 13.4% was calculated in accordance with the American Association for Public Opinion Research Standard Definitions Committee's guidelines for calculating response rate for internet-based studies.[97] To test for selection bias and assess the representativeness of the sample, surgeon demographic information was analyzed by comparing surgeons who responded to the first email, those who responded to the second email, and those from the general population of surgeons for whom data were obtained from the Association of American Medical Colleges (demographic data were not available from ACS).[98] These comparisons revealed that surgeons responding to the first and second email were demographically comparable and that although there were significant differences between the study sample and the overall surgeon population in terms of race and gender, these differences were moderate in magnitude (see Appendix 1).

### 4.1.2 Study design

#### *4.1.2.1 Study content*

After surgeons consented to participating in the study, they were asked to provide information on a range of demographic information, starting with whether they had completed a general surgery residency or if they were currently enrolled in one. Residency graduates were asked the year they graduated from residency, the type of clinical fellowship completed (if any),

type of practice (private practice, academic, county hospital), whether or not they see general surgery patients in their practice, and their annual operative volume. Residents were asked their last completed year of residency. Both current residents and residency graduates were asked their gender, race/ethnicity, type of hospital they work in (academic vs. community), and the zip code in which their hospital is located.

Collaborating with a panel of practicing general surgeons, I designed four clinical vignettes, each representing common general surgery scenarios: mesenteric ischemia, gastrointestinal bleed (GIB), small bowel obstruction (SBO), and appendicitis (Appendix 2). The vignettes were designed so that there was no obviously correct or incorrect treatment choice (either due to lack of supporting data or controversy surrounding the data) in order to ensure adequate variation in surgeon responses. The panel of surgeons revised the vignettes in an iterative fashion so that the detailed clinical descriptions provided surgeons with as much of the important clinical information needed as possible to make informed judgments about the patient's condition.

After the vignette, surgeons were asked to judge the likelihood of both positive and negative outcomes following either surgical or medical management. Specifically, for each of the two hypothetical scenarios (operative and non-operative management), they were asked to estimate the probability (on a scale from 0 to 100) that a patient would die, suffer a serious complication, or fully recover within 30 days of undergoing an operation. The probabilities of death or serious complications are considered "risks" of treatment and the probability of full recovery is considered "benefits" of treatment. Because the probability of death was so low for the small bowel obstruction and appendicitis case, these questions were not asked for these vignettes. Finally, surgeons were asked how likely they were to recommend an operation based

on the clinical information provided (1-5 Likert scale ranging from "very unlikely" to "very likely"). Both the content of the vignette and the subsequent questions were refined through multiple iterations of pilot testing using a separate, national sample of surgeons (n=26).

After completing the clinical vignettes, surgeons were asked whether or not they had previously been aware of the ACS-NSQIP risk calculator and if so, whether they use it in their clinical practice (never/occasionally/routinely). The full online questionnaire is provided in Appendix 2.

#### *4.1.2.2 Randomizations*

All surgeons were presented with clinical vignettes and then asked a series of questions for each vignette. However, upon initiating the study, surgeons were randomized to one of five study arms that differed by the type of information provided as well as the sequence of questions and the number of questions asked. A schematic of the study design is outlined in Figure 4.1. One group of surgeons, the "null group," was asked to judge the risks and benefits of operating, the risks and benefits of not operating and then to rate their likelihood of recommending an operation. A second group, the "risk calculator" group, was identical to the null group, but participants in that group were additionally provided objective data from the ACS-NSQIP risk calculator. For each case, the ACS-NSQIP risk calculator was used to estimate the probability of death (for the MI and GIB cases) and serious complications (for all 4 cases) using the specific demographic and clinical data provided in the vignette. The resulting calculator probability estimates, which only pertained to the risks associated with undergoing an operation, were provided to surgeons answering all 5 of the post-vignette questions. A third group, the "reverse sequence" group, was first asked to rate their likelihood of recommending an operation and then to judge the risks and benefits of operating and not operating. A fourth "operate only" group was

asked only to judge the risks and benefits of operating and to rate their likelihood of recommending an operation, while a fifth "non-operative only" group was only asked to judge the risks and benefits of not operating and to rate their likelihood of recommending an operation. The randomization occurred on a between-subjects level so that each surgeon was assigned to the same study arm for all 4 vignettes. This approach was selected over a within-subjects approach for two reasons. First, it might be disorienting for a surgeon to face different types of questions following each vignette. By remaining consistent, surgeons would be able to more seamlessly move through each vignette, answering the same questions throughout, without having to pause and reorient. Second, it is possible that varying the types of questions asked to surgeons might have inadvertently revealed some of the research hypotheses. For example, had a surgeon seen risk calculator data for the first vignette but then seen no such data for subsequent vignettes, the participant might be warned of one of the study's manipulations, possibly resulting in behavior change. The effectiveness of the randomization process was confirmed by comparing several surgeon demographic and clinical characteristics across the randomly assigned treatment arms. This comparison demonstrated no statistically significant differences in these variables between the groups (Appendix 3).

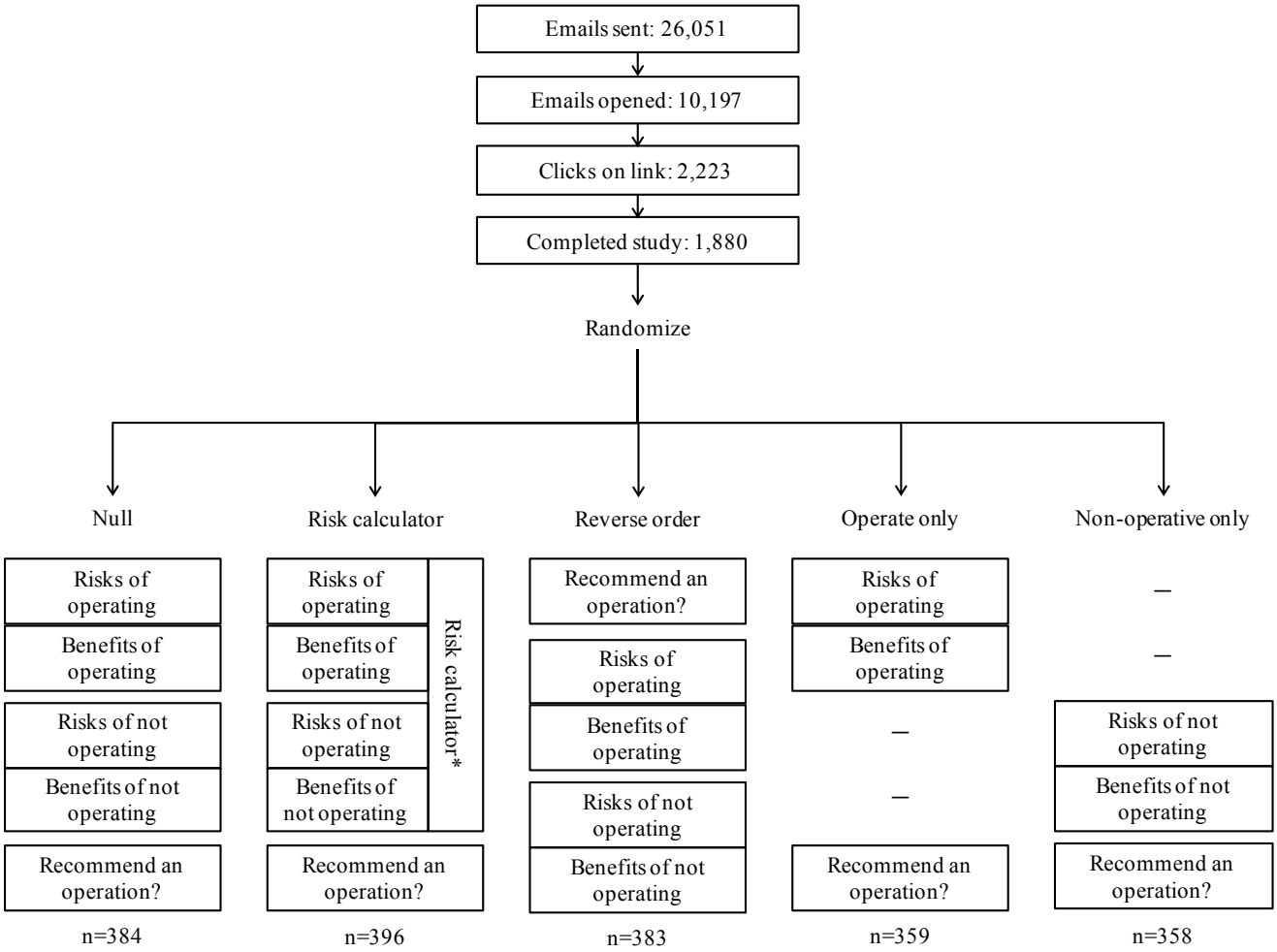



**Figure 4.1 Flow diagram of study participants**
*Risk calculator only provides data on the risks associated with operating

### *4.1.2.3 Risk attitude*

At the end of the study, surgeons were asked to complete a brief instrument that was adapted to measure their overall risk attitude. Due to its brevity and prior use in studying clinicians, I chose to use Pearson's adaptation of the Jackson Personality Index, which asks surgeons to respond on a 5-point Likert scale to 6 items that pertain to an individual's comfort with risk (Table 4.1).[99] Since risk attitude is well known to be domain specific[77] and since I was primarily interested in risk attitude in the surgical setting, I adapted the instrument by adding the words "in surgery" to each item.

**Table 4.1 Pearson's adapted Jackson Personality Index**

| | |
|---|---|
| 1. | People have told me that I enjoy taking chances in my surgical practice |
| 2. | In my surgical practice, I try to avoid situations that have uncertain outcomes |
| 3. | Taking risks in my surgical practice does not bother me if the potential for patient improvement is high |
| 4. | I enjoy taking risks in my surgical practice |
| 5. | I rarely, if ever, take risks in my surgical practice when there is another alternative |
| 6. | I consider security an important element in every aspect of my life |

All questions were asked on a six-point Likert scale from "strongly agree" to "strongly disagree". Risk attitude score=sum of responses x (6/[6-number of missing responses]). Items 2, 5, and 6 were reverse coded.

## 4.2 Measurement model/variable definitions

Following the framework laid out in the conceptual model chapter, the following section will describe the empirical proxies for each concept as well as details on where the relevant data were obtained and how they were measured. The discussion will follow Figure 4.2, starting at the concept of the decision to recommend an operation and moving counterclockwise from there.

### 4.2.1 Surgeon factors

#### *4.2.1.1 Decision to recommend surgery*

The decision to operate was determined from the question "how likely are you to recommend an operation?" which was asked to every surgeon after each clinical vignette. The response was recorded on a 5-point Likert scale (very unlikely, unlikely, neutral, likely, very likely). However, for regression analyses, this variable was dichotomized into those that would recommend an operation (very likely and likely) and those that would not recommend an

operation (very unlikely, unlikely, and neutral). Neutral responses were included as not operating since it was considered to be a passive choice, similar to not operating, as opposed to the active choice of operating. However, to avoid making this assumption, a sensitivity analysis was performed for hypothesis 1.1 and 1.2, whereby the neutral responses were dropped from the analysis. This analysis only resulted in losing 246 (7.5%) of the total surgeon-vignette observations.

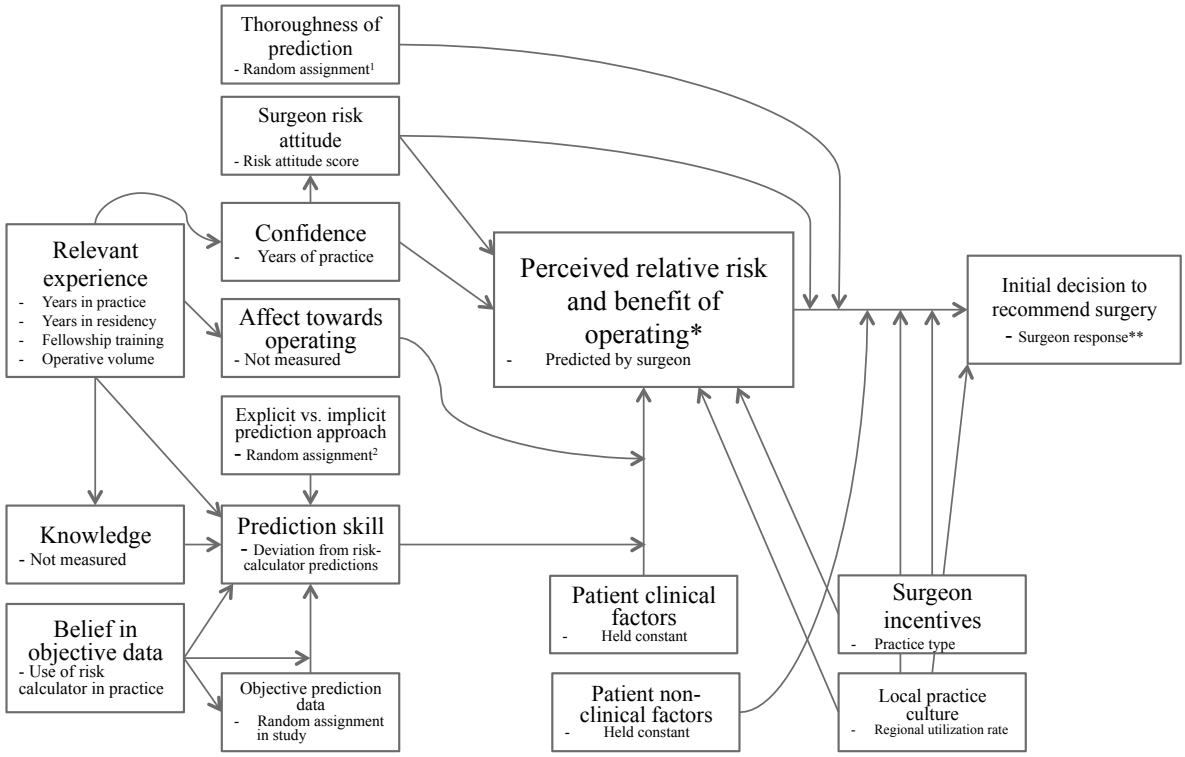



**Figure 4.2 Conceptual model with corresponding empirical measurement proxies**

*Relative risk and benefit of operating in comparison to not operating
** On a 5 point Likert scale in response to the question "how likely are you to recommend an operation for this patient?"
[1]Surgeons randomly assigned to either judge the risks and benefits of operating or the risks and benefits of not operating
[2] Surgeons randomly assigned to judge the risks and benefits of operating and not operating before or after they are asked if they would recommend an operation.
Affect, in its psychological meaning, refers to a surgeon's involuntary, emotional response (dread vs. affection) to a clinical encounter. Responding to one's affect is also typically equivalent to "going with your gut."

#### *4.2.1.2 Perception of relative risks and benefits*

For each vignette, surgeons were asked to estimate the risks and benefits of operating and not operating. To calculate the perceived relative risk of operating, for each vignette, the surgeon's estimate for the probability that a patient would have a serious complication after not operating was subtracted from the surgeon's estimate for the probability of the same outcome after operating.

Perceived Relative Risk of Operating=

Estimated probability of serious complication after operating –

Estimated probability of serious complication after not operating

The more negative the number, therefore, the greater the surgeon's perception that the risks favor operating. For the perception of relative benefit, I performed the same calculation using surgeon's estimated probability of full recovery instead of serious complication. For relative benefit, the more positive the number, the more the benefits favor operating.

#### *4.2.1.3 Thoroughness of prediction*

The thoroughness of prediction was determined based on exposure to one of the random assignments in the study. Those surgeons who were asked only to estimate the risks and benefits for operating or the risks and benefits for not operating were considered to not have thoroughly evaluated risk, since they were not forced to estimate the likelihood of outcomes for both treatment options. While surgeons may have estimated these outcomes themselves, either consciously or subconsciously without the prompt, I used this random assignment as a proxy since these surgeons were less likely to consider all risk and benefit parameters than those who were forced to do so. Assignment to either the "operative only" or the "non-operative" only

groups was coded as two separate dummy variables. For both of these dummy variables, the "null group" served as the reference category.

#### *4.2.1.4 Surgeon risk attitude*

Surgeon risk attitude was determined from responses to the Pearson Adapted Jackson Personality Index. This instrument yields a continuous measure of risk attitude with the higher number corresponding to a more risk-seeking attitude and a lower number corresponding to a more risk-averse attitude.

#### *4.2.1.5 Confidence*

While confidence was not measured directly, there is at least one empirical proxy that has, in prior studies, been shown to have high association with confidence: experience. For these purposes, however, it is not immediately clear what the relationship is between experience and confidence in surgery. On the one hand, surgeons who have been in practice for longer may be more confident due to accumulated experience, but on the other hand this relationship may also be reversed. It is possible that "ignorance is bliss" and that younger surgeons who have not yet experienced many bad outcomes retain a youthful confidence that eventually weathers with age.[27,53] [27,91]

#### *4.2.1.6 Affect towards operating*

There are no available empirical proxies for this concept.

#### *4.2.1.7 Explicit vs. implicit prediction approach*

Explicit prediction requires careful ex-ante consideration of treatment risks and benefits before deciding whether or not to recommend an operation. Implicit prediction, which may also represent the absence of risk prediction, occurs when surgeons decide to operate before such

careful consideration. In this study, surgeons were randomized to either the "null group" or the "reverse sequence" group, which was coded as a dummy variable to represent this concept. The null group was coded as the reference category.

#### *4.2.1.8 Relevant experience*

A surgeon's relevant experience was measured in three ways. First, by the surgeon's number of years since graduating from residency, which was coded as a continuous variable (current residents were coded as having zero years of experience) and by the number of years of residency completed (attending surgeons received the maximum value of 5 years for this variable). Second, since all cases represented emergency general surgery cases, surgeons who completed either a trauma or acute care surgery fellowship were considered to have more relevant experience. Additionally, surgeons that complete another type of fellowship may be less likely to treat basic general surgery patients and may therefore have less relevant experience. This variable was coded as a categorical variable (trauma or acute care surgery fellowship/different type of fellowship/no fellowship; the latter served as the reference category). Finally, surgeons who operate more frequently gather more relevant experience. I therefore included a surgeon's self-reported operative volume as a measure of experience as a categorical variable corresponding to how it was measured in the study (current resident, <50 (reference category), 50-100, 100-250, 250-500, >500; this variable was coded as a series of indicator variables).

#### *4.2.1.9 Knowledge*

There are no available empirical proxies for this concept.

#### *4.2.1.10 Prediction skill*

Prediction skill was measured as the absolute value of the difference between each surgeon's estimate of operative risk and the estimate derived from the risk calculator. A larger number for this variable represents a greater deviation from the calculator value, corresponding to worse prediction skill.

#### *4.2.1.11 Objective prediction data*

Surgeons assigned to the "risk calculator" group received objective prediction data from the risk calculator. Assignment to this group was coded as a dummy variable with the reference category representing the null group.

#### *4.2.1.12 Belief in objective data*

Surgeon belief in objective data was measured by their prior experience and use of the risk calculator in their clinical practice. The rationale behind this choice is that surgeons who do not believe in the use of data for these purposes in surgery may have possibly avoided the news surrounding the calculator's release and may be more likely to be unaware of the calculator prior to the study than a surgeon who believes strongly in the value of such data and seeks out this information. By extension, surgeons who use the calculator in their clinical practice were presumed to believe the data. This variable was coded as a categorical variable: (never heard of calculator prior to study (reference category)/heard of calculator but never use it in practice/use it in practice occasionally/use it routinely; coded as a series of dummy variables).

### 4.2.2 Patient factors

#### *4.2.2.1 Patient clinical factors*

These were held constant for each case through the use of vignettes.

#### *4.2.2.2 Patient non-clinical factors*

These were also held constant for each case through the use of vignettes.

### 4.2.3 Environmental factors

#### *4.2.3.1 Surgeon incentives*

Surgeons who get paid more for each additional operation, such as those who work in a fee-for-service system, may have additional incentive to operate for any given level of perceived relative risk and benefit. As a proxy, I created a dummy variable for surgeons who work in private practice or a community hospital as opposed to an academic center, a Veterans Affairs Hospital, or a county hospital (the latter group were combined as the reference category).

#### *4.2.3.2 Local practice culture*

Local practice culture is difficult to measure for at least two reasons. First, it is difficult to define what is meant by the term local since culture may vary between countries, states, counties, and even within a hospital. Second, culture can be an elusive term to define, let alone measure.[100] As already mentioned, much of the work on regional variations in care has been performed by the Dartmouth Atlas using either Hospital Referral Regions or Hospital Service Area (HSA) as the unit of regional measurement. Borrowing from the Dartmouth Atlas, I define local in the unit of HSA's. The culture of each HSA is defined by using publicly available data on the Dartmouth Atlas website for the average utilization rate of inpatient cholecystectomy. This operation was chosen because it represents a type of emergency general surgery operation, similar to those outlined in this study and is likely performed by the same types of surgeons who participated in the present study. The utilization rate is measured by number of operations per 1,000 Medicare enrollees.

## 4.3 Statistical analysis

For all study participants, univariate statistics are calculated to describe demographic information. Univariate statistics are also calculated for the main outcome variable (the decision to recommend an operation) as well as for the primary predictors (perceived risks and benefits of operating and not operating—both absolute and relative). In instances in which variables were assigned randomly to participants (risk calculator, reverse order, and operate only), surgeon characteristics are compared in each group to ensure that randomization was successful.

The variables included in each regression analysis as well as the direction of the hypothesized effect for all relevant variables are summarized in Table 4.2. The regression analyses combine all 4 clinical scenarios into a single regression model using a random intercept for the surgeon and fixed effects dummy variables for each of the 4 scenarios. For regressions using the decision to operate as a dichotomous outcome variable (regressions 1, 2, 7, 8, and 9), random effects logistic regression is used to calculate absolute risk differences based on the predicted probabilities calculated by the regression equation. Confidence intervals are calculated using the Delta Method (Taylor series approximation). For regressions using outcome variables that are continuous, such as the surgeon's perception of either the absolute or relative risk or operating (regressions 3, 4, 5, and 6), random effects linear regression is used to calculate beta coefficients and marginal effects. According to the conceptual model, there are several variables that moderate the relationship between surgeons' perception of risk and benefit and their decision to recommend an operation. However, because many of these moderating relationships are not relevant to the hypotheses tested here, these interaction terms are often suppressed, yielding results that are weighted averages across the omitted variables. For example, the relationship between perceived relative risk and benefit and the decision to operate is moderated

by a surgeons' financial incentives. Yet, the variable "private practice" is excluded from regressions 7, 8, and 9, since this variable is not directly relevant to the hypotheses being tested. The resulting regression coefficients for relative risk and benefit will therefore represent weighted average effects across surgeons who work in private practice and those who do not. Furthermore, the main effect of the "private practice" variable is also excluded from the regression since its effect on the decision to recommend an operation is mediated by the surgeons' perception of relative risks and benefits.

Approximately 10% of the study sample included members of the ACS that reported working primarily outside the United States (most commonly Canada). Since local practice patterns, as measured by the regional rates of inpatient cholecystectomy, were not available for these surgeons, they were excluded from the analyses. The remaining variables had relatively low rates of missing data and therefore, listwise deletion was used. However, a sensitivity analysis was performed for Regression 1 using multiple imputation (Appendix 4) and because the results were comparable, listwise deletion was used for all analyses.

As shown in the final row of Table 4.2, each regression equation includes different patient cohorts. This is done by necessity since some of the study participants did not provide data on all parameters included in the regression. For example, participants in the *operate only* arm were only asked to provide risk and benefit estimates for operative and not for non-operative management. Therefore, it is not possible to calculate these surgeons' perception of the *relative* risks or benefits of operating. Because surgeons were exposed to subtly different "treatments," all regression equations included dummy variables corresponding to the surgeon's randomly assigned experimental arm.

**Table 4.2 List of variables and relevant study cohorts included in each regression**

| Variable | Decision to operate; relative outcomes (1) | Perceived relative risks of operating (2) | Perceived relative benefits of operating (3) | Decision to operate; absolute outcomes (4) | Perceived absolute benefits of operating (5) | Prediction skill and objective data (6) | Decision to operate and objective data (7) | Decision to operate bias (8) | Decision to not operate bias (9) |
|---|---|---|---|---|---|---|---|---|---|
| Decision to operate | O | | | O | | | O | O | O |
| Relative risks and benefits | | | | | | | | | |
| Prediction of relative risks of operating | X° (-) | O | | | | | | | |
| Prediction of relative benefits of operating | X° (+) | | O | | | | | | |
| Absolute risks and benefits | | | | | | | | | |
| Prediction of absolute risks of operating | | | | X (-) | | | | X | |
| Prediction of absolute benefits of operating | | | | X (+) | O | | | X | |
| Prediction of absolute risks of not operating | | | | X (+) | | | | | X |
| Prediction of absolute benefits of not operating | | | | X (-) | | | | | X |
| Risk attitude score (lower=more risk-averse) | Xi* (+) | X (-) | X | Xi** (+) | X | | | | |
| Years in residency | | X | X | | X (+) | X | X | | |
| Years in practice | | X | X | | X (+) | X | X | | |
| Fellowship in acute care/trauma surgery | | X (-) | X | | X (+) | X | X | | |
| Operative volume | | X | X | | X (+) | X | X | | |
| Prediction skill§ | | | | | | O | | | |
| Random study assignments: | | | | | | | | | |
| Risk calculator | X | X | X | X | X | X (-) | X (+/-) | | |
| Reverse order | X | X | X | X | X | | | | |
| Operate only | | | | | X | | | Xi¶ (+) | Xi¶ (+) |
| Routine use of calculator in practice | | X | | | X | X, Xi (-) | X | | |
| Private practice | Xi* | X | X | Xi** | X | | | - | - |
| Regional surgical utilization rate | X | X | X | X | X | | | X | X |
| Study cohorts included | Null, risk calculator, reverse order | Null, risk calculator, reverse order | Null, risk calculator, reverse order | Null, risk calculator, reverse order | Null, risk calculator, reverse order, operate only | Null, risk calculator | Null, risk calculator | Null, operate only | Null, non-operative only |

O=outcome; X=covariate; Xi=interaction term

° Includes both main effect and interaction term between prediction of relative risks of operating and relative benefits of operating

*Interaction with prediction of relative risks and relative benefits of operating

**Interaction with prediction of absolute risks and absolute benefits of operating and not operating

§ Prediction skill measured by difference between surgeon's prediction of absolute risk of operating compared with same value from risk calculator

¶ Interaction with prediction of absolute risks/benefits of operating or absolute risks/benefits of not operating

- Reflects suppressed interaction term

The equations for each regression are notated using vector notation in Table 4.3 alongside the expected coefficient signs for each hypothesis. The table uses abbreviations to refer to the different variables included in the model. These abbreviations include the following: O Decision to operate; RR Relative risks of operating; RB Relative benefits of operating; RO Absolute risks of operating; BO Absolute benefits of operating; RN Absolute risks of not operating; BN Absolute benefits of not operating; RA Risk attitude score; AE Attending years of experience; RE Resident years of experience; FT Fellowship in acute care/trauma surgery; OV Operative volume; PS Prediction skill; RC Random assignment to risk calculator; CU Risk calculator use; CUU Unaware of risk calculator; CUN Never use risk calculator; CUO Occasionally use risk calculator; CUR Routinely use risk calculator; Op Random assignment to "operate only"; Nonop Random assignment to "non-operative only"; FI Financial incentive associated with private practice; FII Financial incentive interacted with RO, BO, RN, BN; Ex Vector for randomly assigned experimental group; LPC Local practice culture; Vg Dummy variables for clinical vignettes. The vectors refer to clinical decisions made by a surgeon *j* in a clinical scenario *i*. Using this notation, I lay out below each hypothesis accompanied by an explanation of the corresponding regression equation and the intended interpretation of the regression outputs.

(Hypothesis 1.1) Controlling for patient characteristics and for any given level of perceived relative benefit of operating, surgeons are more likely to recommend an operation when the perceived relative risks of operating are lower (Regression 1).

The regression model for the hypothesis 1.1 takes a form of observation of a clinical decision made by a surgeon *j* in a clinical scenario *i*, where *O* indicates whether a surgeon recommends an operation (1=yes, 0=no). The $pr(O_{ij}=1)$ is a (nonlinear) function *f* of the

following covariates: $RR_{ij}$ and $RB_{ij}$ correspond respectively to surgeon *j*'s perception of the relative risks and the relative benefit of operating for each clinical scenario *i*. RA refers to a surgeon's risk attitude score. FI refers to a surgeon who works in a private practice setting. Ex is a vector of dichotomous indicators for the randomly assigned experimental arm (null, risk calculator, or reverse order), LPC refers to the regional surgical utilization rate, and Vg is a vector of dichotomous indicators for each different clinical scenario. As noted above, RR is coded such that lower perceived relative risks of operating correspond to a (larger) negative value. Thus, support for hypothesis 1.1 will be found if $\beta_1$ is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 1.2) Controlling for patient characteristics and for any given level of perceived relative risk of operating, surgeons are more likely to recommend an operation when the perceived relative benefits of operating are higher (Regression 1).

Support for hypothesis 1.2 will be found if $\beta_2$ in equation (1) is positive and statistically significant at the p-value<0.05 level.

(Hypothesis 2.1) Controlling for a surgeon's perceived relative risks and benefits of operating and for any given level of perceived relative risk of operating, a surgeon's greater likelihood of operating in the face of lower perceived relative risks of operating is greater for risk-seeking surgeons than it is for risk averse surgeons (Regression 1)

RA is coded so that a higher value implies that the surgeon is more risk prone. Thus support for hypothesis 2.1 will be found if $\beta_4$ in equation (1) is positive and statistically significant at the p-value<0.05 level.

**Table 4.3 Summary of regression notation**

| Regression | Equation | Hypothesis | Coefficient of interest | Expected sign of coefficient |
|---|---|---|---|---|
| 1 | $pr(O_{ij}=1) = f(\beta_0 + \beta_1 RR_{ij} + \beta_2 RB_{ij} + \beta_3 RR_{ij} * RB_{ij} + \beta_4 RR_{ij} * RA_{ij} + \beta_5 RB_{ij} * RA_{ij} + \beta_6 RR_{ij} * FI_{ij} + \beta_7 RB_{ij} * FI_{ij} + \beta_8 Ex_{ij} + \beta_9 LPC_{ij} + \beta_{10} Vg_{ij})$ | 1.1<br>1.2<br>2.1 | $\beta_1$<br>$\beta_2$<br>$\beta_4$ | -<br>+<br>+ |
| 2 | $RO_{ij} = \beta_0 + \beta_1 RA_{ij} + \beta_2 RE_{ij} + \beta_3 AE_{ij} + \beta_4 FT_{ij} + \beta_5 OV_{ij} + \beta_6 RC_{ij} + \beta_7 Ex_{ij} + \beta_8 CU_{ij} + \beta_9 FI_{ij} + \beta_{10} LPC_{ij} + \beta_{11} Vg_{ij} + \epsilon_{ij}$ | 2.3<br>4.2 | $\beta_1$<br>$\beta_2, \beta_3, \beta_4, \beta_5$ | -<br>- |
| 3 | $pr(O_{ij}=1) = f(\beta_0 + \beta_1 RO_{ij} + \beta_2 BO_{ij} + \beta_3 RN_{ij} + \beta_4 BN_{ij} + \beta_5 RA_{ij} * RO_{ij} + \beta_6 RA_{ij} * BO_{ij} + \beta_7 RA_{ij} * RN_{ij} + \beta_8 RA_{ij} * BN_{ij} + \beta_9 FII_{ij} + \beta_{10} Ex_{ij} + \beta_{11} LPC_{ij} + \beta_{12} Vg_{ij})$ | 3.1<br>3.2<br>3.3<br>3.4 | $\beta_1$<br>$\beta_2$<br>$\beta_3$<br>$\beta_4$ | -<br>+<br>+<br>- |
| 4 | $RR_{ij} = \beta_0 + \beta_1 RA_{ij} + \beta_2 RE_{ij} + \beta_3 AE_{ij} + \beta_4 FT_{ij} + \beta_5 OV_{ij} + \beta_6 Ex_{ij} + \beta_7 CU_{ij} + \beta_8 FI_{ij} + \beta_9 LPC_{ij} + \beta_{10} Vg_{ij} + \epsilon_{ij}$ | 2.2<br>4.1 | $\beta_1$<br>$\beta_4$ | -<br>- |
| 5 | $BO_{ij} = \beta_0 + \beta_1 RA_{ij} + \beta_2 RE_{ij} + \beta_3 AE_{ij} + \beta_4 FT_{ij} + \beta_5 OV_{ij} + \beta_6 Ex_{ij} + \beta_7 CU_{ij} + \beta_8 FI_{ij} + \beta_9 LPC_{ij} + \beta_{10} Vg_{ij} + \epsilon_{ij}$ | 4.3 | $\beta_2, \beta_3, \beta_4, \beta_5$ | + |
| 6 | $PS_{ij} = \beta_0 + \beta_1 RC_{ij} + \beta_2 CU_{ij} + \beta_3 CUU_{ij} * RC_{ij} + \beta_4 CUN_{ij} * RC_{ij} + \beta_5 CUO_{ij} * RC_{ij} + \beta_6 CUR_{ij} * RC_{ij} + \beta_7 RE_{ij} + \beta_8 AE_{ij} + \beta_9 FT_{ij} + \beta_{10} OV_{ij} + \beta_{11} Vg_{ij} + \epsilon_{ij}$ | 5.1<br>5.2 | $\beta_1$<br>$\beta_6$ | -<br>- |
| 7 | $p(O_{ij}=1) = f(\beta_0 + \beta_1 RC_{ij} + \beta_2 CU_{ij} + \beta_3 RE_{ij} + \beta_4 AE_{ij} + \beta_5 FT_{ij} + \beta_6 OV_{ij} + \beta_7 Vg_{ij})$ | 5.3 | $\beta_1$ | +/-* |
| 8 | $p(O_{ij}=1) = f(\beta_0 + \beta_1 RO_{ij} + \beta_2 BO_{ij} + \beta_3 RO_{ij} * Op_{ij} + \beta_4 BO_{ij} * Op_{ij} + \beta_5 LPC_{ij} + \beta_6 Vg_{ij})$ | 6.1 | $\beta_3$ | - |
| 9 | $p(O_{ij}=1) = f(\beta_0 + \beta_1 RN_{ij} + \beta_2 BN_{ij} + \beta_3 RN_{ij} * Nonop_{ij} + \beta_4 BN_{ij} * Nonop_{ij} + \beta_5 LPC_{ij} + \beta_6 Vg_{ij})$ | 6.2 | $\beta_3$ | + |

Abbreviations: O Decision to operate; RR Relative risks of operating; RB Relative benefits of operating; RO Absolute risks of operating; BO Absolute benefits of operating; RN Absolute risks of not operating; BN Absolute benefits of not operating; RA Risk attitude score; AE Attending years of experience; RE Resident years of experience; FT Fellowship in acute care/trauma surgery; OV Operative volume; PS Prediction skill; RC Random assignment to risk calculator; CU Risk calculator use; CUU Unaware of risk calculator; CUN Never use risk calculator; CUO Occasionally use risk calculator; CUR Routinely use risk calculator; Op Random assignment to "operate only"; Nonop Random assignment to "non-operative only"; FI Financial incentive associated with private practice; FII Financial incentive interacted with RO, BO, RN, BN; Ex Vector for randomly assigned experimental group; LPC Local practice culture; Vg Dummy variables for clinical vignettes

*$\beta_1$ will be positive or negative depending on whether $\beta_6$ in regression 4 is negative or positive, respectively

(Hypothesis 2.2) Risk-seeking surgeons perceive lower levels of relative risk than risk-averse surgeons (Regression 2).

Support for hypothesis 2.2 will be found if $\beta_1$ in equation (2) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 2.3) Risk-seeking surgeons perceive lower levels of absolute operative risk than risk-averse surgeons (Regression 3).

Support for hypothesis 2.3 will be found if $\beta_1$ in equation (3) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 3.1) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute operative risk are less likely to recommend an operation (Regression 4).

Support for hypothesis 3.1 will be found if $\beta_1$ in equation (4) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 3.2) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute operative benefit are more likely to recommend an operation (Regression 4).

Support for hypothesis 3.2 will be found if $\beta_2$ in equation (4) is positive and statistically significant at the p-value<0.05 level.

(Hypothesis 3.3) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute risks associated with not operating are more likely to recommend an operation (Regression 4).

Support for hypothesis 3.3 will be found if $\beta_3$ in equation (4) is positive and statistically significant at the p-value<0.05 level.

(Hypothesis 3.4) Controlling for patient characteristics and other risk/benefit parameters, surgeons who perceive higher levels of absolute benefit associated with not operating are less likely to recommend an operation (Regression 4).

Support for hypothesis 3.4 will be found if $\beta_4$ in equation (4) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 4.1) Surgeons who have completed fellowships in trauma surgery or acute care surgery perceive lower relative risks of operating than those who did not complete a fellowship (Regression 2).

Support for hypothesis 4.1 will be found if $\beta_4$ in equation (2) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 4.2) Surgeons perceive lower levels of absolute operative risk if they have more relevant experience, as measured by their number of years in residency, greater number of years in practice, their completion of a trauma/acute care surgery fellowship, and their high operative volume (Regression 3).

Support for hypothesis 4.2 will be found if $\beta_2$, $\beta_3$, $\beta_4$, and $\beta_5$ in equation (3) are negative and statistically significant at the p-value<0.05 level.

(Hypothesis 4.3) Surgeons perceive higher levels of absolute operative benefit if they have more relevant experience, as measured by their number of years in residency, greater number of years in practice, their completion of a trauma/acute care surgery fellowship, and their high operative volume (Regression 5).

Support for hypothesis 4.3 will be found if $\beta_2$, $\beta_3$, $\beta_4$, and $\beta_5$ in equation (5) are positive and statistically significant at the p-value<0.05 level.

(Hypothesis 5.1) Exposing surgeons to the risk calculator improves surgeons' prediction skill (Regression 6).

Because prediction skill is calculated as the absolute difference between a surgeon's prediction of operative risk and the prediction made by the risk calculator, a lower value of prediction skill indicates better prediction. Therefore, support for hypothesis 5.1 will be found if $\beta_1$ in equation (6) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 5.2) The improvement in prediction skill conferred by exposure to the risk calculator is greater for those surgeons who routinely use the calculator in their clinical practice than for those who do not (Regression 6).

Support for hypothesis 5.2 will be found if $\beta_6$ in equation (6) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 5.3) If exposure to the risk calculator causes surgeons to perceive higher levels of absolute operative risk, then exposure to the risk calculator will also change their behavior and make surgeons less likely to recommend an operation (Regression 7).

However, if exposure to the risk calculator causes surgeons to predict lower average levels of operative risk, surgeons will change their behavior to be more likely to recommend an operation (Regression 7).

If $\beta_6$ in equation (4) is positive, then support for hypothesis 5.3 will be found if $\beta_1$ in equation (7) is negative and statistically significant at the p-value<0.05 level. However, if $\beta_6$ in equation (4) is negative, then support for the alternate hypothesis will be found if $\beta_1$ in equation (7) is positive and statistically significant at the p-value<0.05 level.

(Hypothesis 6.1) The higher tendency for surgeons to operate when they perceive lower operative risks is more pronounced when surgeons are only asked to predict the risks and benefits of operating (and not the risks and benefits of *not* operating; regression 8).

Support for hypothesis 6.1 will be found if $\beta_3$ in equation (8) is negative and statistically significant at the p-value<0.05 level.

(Hypothesis 6.2) The higher tendency for surgeons to operate when they perceive high levels of risks associated with *not* operating is more pronounced when surgeons are only asked to predict the risks and benefits of *not* operating (and not the risks and benefits of operating; regression 9).

Support for hypothesis 6.2 will be found if $\beta_3$ in equation (9) is positive and statistically significant at the p-value<0.05 level.

# 5 Results

## 5.1 Descriptive statistics

Across all 4 clinical vignettes, surgeons varied widely in their perception of the risks and benefits of operative and non-operative management (Figure 5.1). For example, in the appendicitis scenario, surgeons varied in their estimates of the likelihood of serious complication (mean 24%, range 0-100%) or recovery (mean 86%, range 1-100%) following operative management as well as the likelihood of serious complication (mean 31%, range 0-100%) or recovery (mean 68%, mean 0-100%) following non-operative management. For the mesenteric ischemia and GIB scenarios, the variation in surgeons' risk and benefit estimates for operative management were similar to those for non-operative management. However, for the SBO and appendicitis cases, although the range of estimates remained wide, surgeons mostly reported that the risks of operating were low and the benefits of operating were high. For these cases, there was substantially more variation in surgeons' predictions of non-operative risks and benefits.

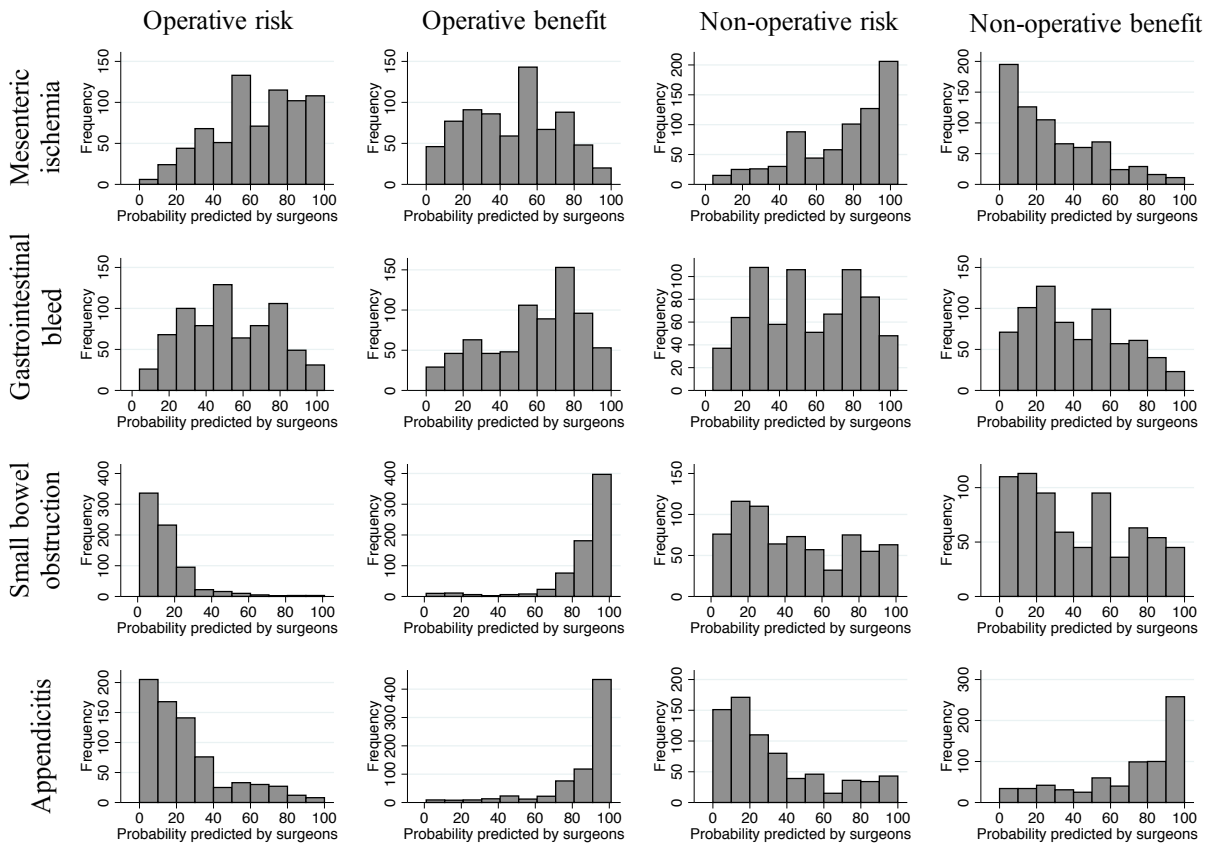



**Figure 5.1 Variation in surgeons' judgment of operative and non-operative risk and benefit**
Each histogram shows the number of surgeons (n=1,880 total) predicting each value of risk or benefit. The x-axis, which is divided into intervals of 10, refers to surgeons' judged risk (the likelihood of serious complication) or benefit (likelihood of patient recovery) following operative or non-operative management.

Surgeons also varied widely in their decision to recommend an operation (Figure 5.2), although the extent of that variation depended on the clinical scenario. For example, in the SBO case, most surgeons recommended an operation (84%), while there was more disagreement for the other cases on whether or not an operation should be performed (67% recommended an operation for mesenteric ischemia, 54% for GIB, and 49% for appendicitis).

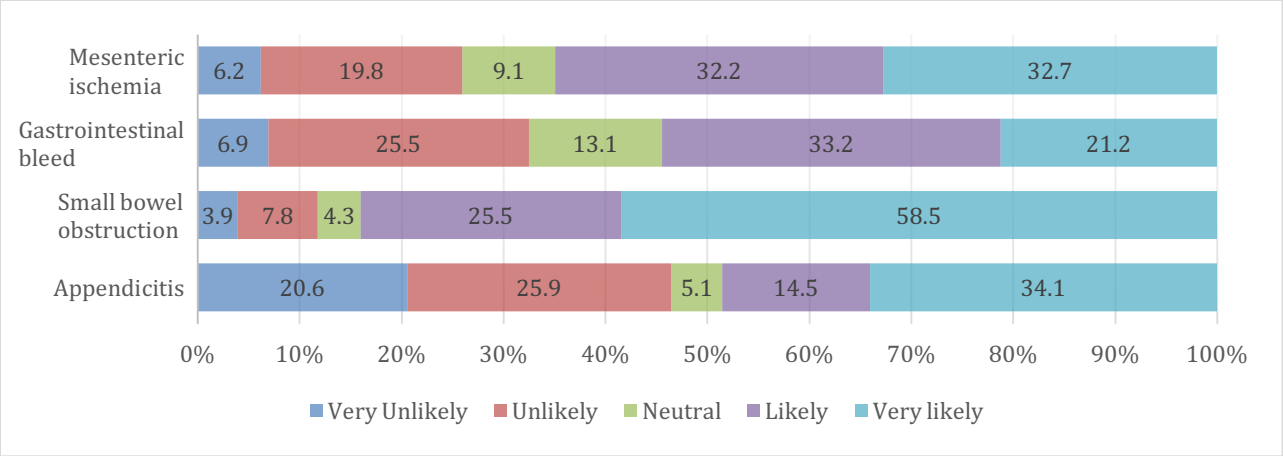



**Figure 5.2 Variation in surgeon response to "how likely are you to recommend an operation?"**

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 shows percentage in each row

The risk attitude scores, as measured by surgeons' responses to the Pearson Adapted Jackson Personality Index, ranged from 6 to 31 (mean 17.3, standard deviation 4.6; Figure 5.3).

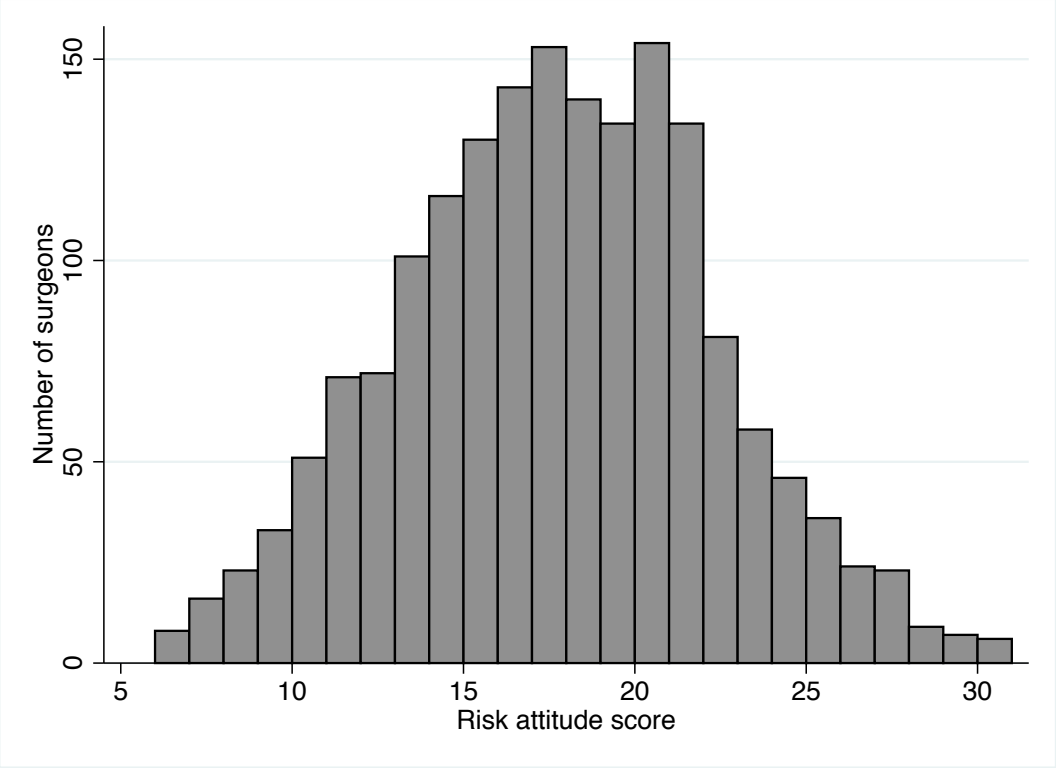



**Figure 5.3 Distribution of risk attitude scores for surgeons in study sample**

Risk attitude score derived from surgeon responses to the Pearson Adapted Jackson Personality Index instrument. Higher scores correspond to being more risk-seeking.

Overall, almost half of the surgeons (46.4%) were unaware of the ACS-NSQIP risk calculator prior to participating in this study. The remaining surgeons were divided almost equally between those who were previously aware of the tool but were not using it in their clinical practice (25.9%) and those who used it occasionally (24.5%). Only 3.2% of study participants reported using the calculator routinely in their practice.

## 5.2 Regression results

(Hypothesis 1.1) Controlling for patient characteristics and for any given level of perceived relative benefit of operating, surgeons are more likely to recommend an operation when the perceived relative risks of operating are lower (Regression 1).

Before reporting the results to this hypothesis, it should be noted that the effect of relative risk perception on the decision to operate is dependent on the level of perceived relative benefit as noted by the significant p-value associated with the interaction term between perceived relative risk and perceived relative benefit ($p<0.001$). However, at all levels of perceived relative benefit, surgeons who judged lower relative risk of operating were more likely to recommend an operation (Table 5.1). For example, holding relative benefit constant at 1 standard deviation (SD) below its mean (-8.4%), surgeons who judged relative risk at 1 SD below the mean would have an 84.5% predicted probability of recommending an operation, while those judging relative risk at 1 SD above the mean would have a 19.9% predicted probability of recommending an operation [absolute difference -64.7%, 95% confidence interval (CI) (-70.8, -58.5)]. When the relative benefits of operating are held constant at 1 SD above its mean (57.7%), the effect of relative risk perception on the decision to recommend an operation remains significant, albeit

smaller in magnitude [-15.3% absolute difference (95% CI -21.0, -9.5) between surgeons judging 1 SD above vs. below the mean]. In summary, these data show support for hypothesis 1.1.

**Table 5.1 Association between surgeon perception of relative risks and benefits of operating and the decision to recommend an operation**

| Risk/benefit parameter* | Probability that a surgeon at 1 standard deviation below mean recommends an operation | Probability that a surgeon at 1 standard deviation above mean recommends an operation | Absolute difference in probability that a surgeon recommends an operation* | 95% Confidence Interval |
|---|---|---|---|---|
| Relative risk (at relative benefit μ-sd) | 84.5 | 19.9 | -64.7 | (-70.8, -58.5) |
| Relative risk (at relative benefit μ+sd) | 94.3 | 79.0 | -15.3 | (-21.0, -9.5) |
| Relative benefit (at relative risk μ-sd) | 84.5 | 94.3 | 9.7 | (5.2, 14.2) |
| Relative benefit (at relative risk μ+sd) | 19.9 | 79.0 | 59.1 | (52.0, 66.3) |

* Because of significant interaction between relative risk and relative benefit, effect sizes are calculated twice. First the effect is calculated while holding the other parameter constant at 1 standard deviation below the mean and then while holding the other parameter constant at 1 standard deviation above the mean.

(Hypothesis 1.2) Controlling for patient characteristics and for any given level of perceived relative risk of operating, surgeons are more likely to recommend an operation when they perceived relative benefits of operating are higher (Regression 1).

The effect of perceived relative benefit on the decision recommend an operation also depended on the level of perceived relative risk. Taking this into account, at all levels of perceived relative risk, surgeons were more likely to recommend an operation if they perceived the relative benefits of operating to be higher (Table 5.1). For example, when holding perceived relative risk constant at 1 SD below its mean (-46.1%), there would be a 9.7 percentage point difference (95% CI 5.2, 14.2) in the predicted probability that surgeons would recommend an operation between those judging relative benefit at 1 SD below vs. above the mean (84.5% vs. 94.3%). This effect was consistent, albeit much larger in effect, when perceived relative risk was held constant at 1 SD above its mean (15.8%), where the same comparison noted a 59.1%

percentage point difference (19.9% vs. 79.0%). In summary, these data show support for hypothesis 1.2. Of note, controlling for the variables included in regression 1 explained a total of 37% of the observed variation in surgeons' decision to recommend an operation across all four clinical scenarios. Further, controlling only for surgeons' perception of relative risk and benefit of operating (and the interaction between the two variables) explained a total of 36% of the variation in decision-making. The results of hypothesis 1.1 and 1.2 changed only modestly in the sensitivity analysis whereby the observations in which surgeons reported a neutral likelihood of recommending an operation were dropped (Appendix 5).

(Hypothesis 2.1) Controlling for a surgeon's perceived relative risks and benefits of operating, a surgeon's greater likelihood of operating in the face of lower perceived relative risks of operating is lower for risk-averse surgeons than it is for risk seeking surgeons (Regression 1).

There was no significant interaction between surgeon perception of relative risk and their level of risk aversion on the decision to recommend an operation (p-value for interaction term=0.119). As the perceived relative risk of operating increased, the lower likelihood of recommending an operation was similar for all levels of risk attitude. For example, when risk attitude is held constant at 1 SD below the mean (12.5), surgeons were 36.6% (95% CI -42.3%, -30.9%) less likely to recommend an operation as relative risk perception increased from 1 SD below the mean to 1 SD above the mean. Similarly, when risk attitude is held constant at 1 SD above the mean (22.0), the same absolute difference was -44.7% (95% CI -51.1%, -38.2%), which translates into an 8.1% difference between the effect sizes for risk-seeking and risk-averse surgeons (95% CI -0.3%, 16.4%). This finding is further displayed in Figure 5.4, which

compares the relationship between perceived relative risk and the decision to operate at different risk attitude scores. In summary, I found no evidence to support hypothesis 2.1.

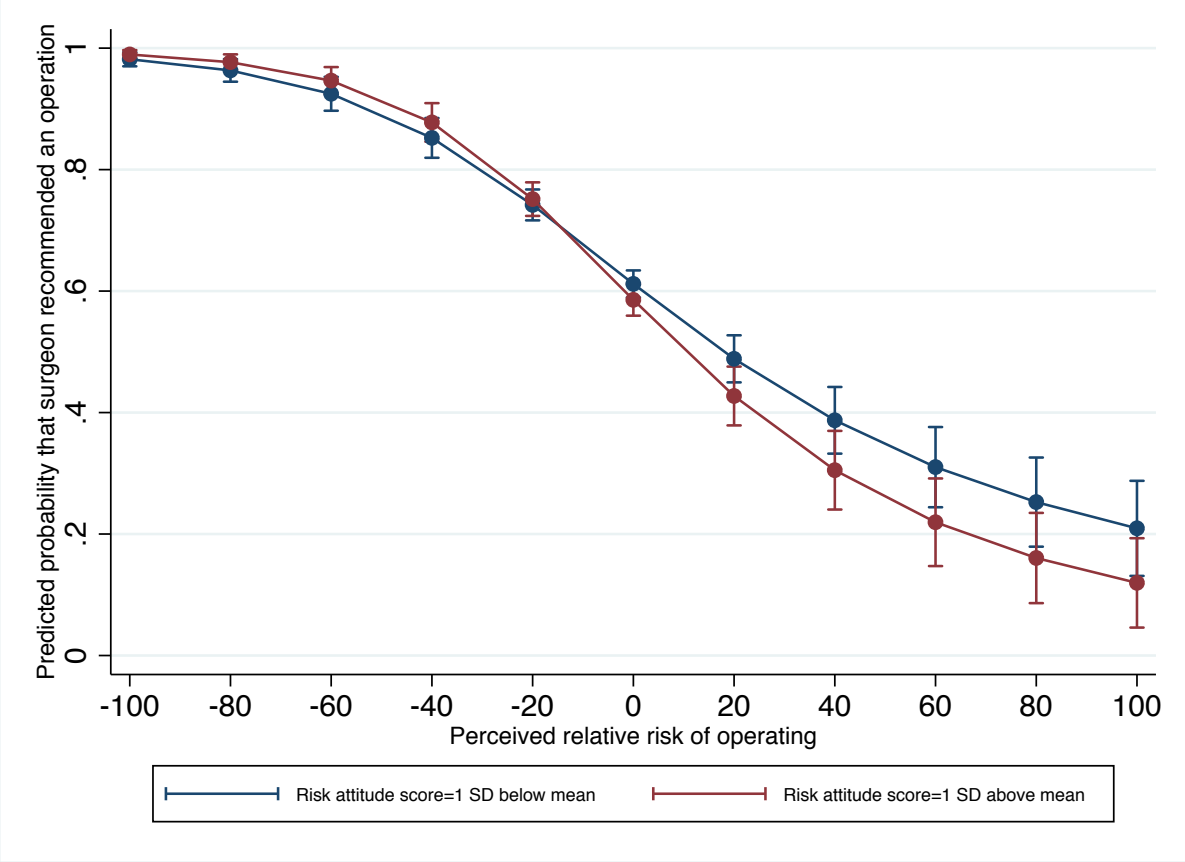



**Figure 5.4 Association between risk attitude and decision to recommend an operation at different values of perceived relative risk**

Of note (not specifically hypothesized), there was a statistically significant interaction effect between surgeons' risk attitude and their perceived relative benefit on the decision to recommend an operation (Figure 5.5). For example, when risk attitude is held constant at 1 SD below the mean, surgeons were 38.9% (95% CI -33.5%, -44.2%) more likely to recommend an operation as relative risk perception increased from 1 SD below the mean to 1 SD above the mean. However, when risk attitude is held constant at 1 SD above the mean, the same absolute difference decreases to 31.0% (95% CI 24.8%, 37.3%; p-value for interaction term=0.038). The

difference between this effect for risk-seeking and risk-averse surgeons was 7.8% (95% CI 0.8%, 14.9%).

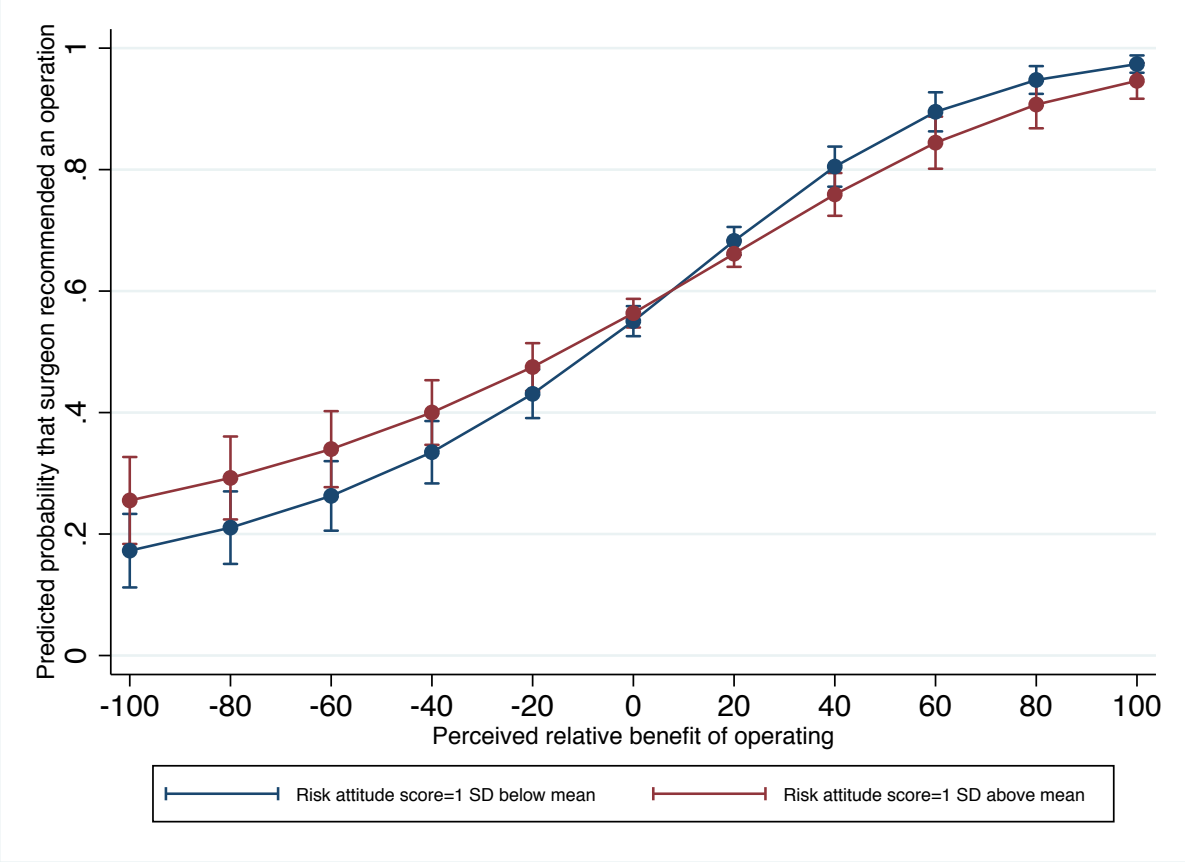



**Figure 5.5 Association between risk attitude and decision to recommend an operation at different values of perceived relative benefit**

(Hypothesis 2.2) Risk-averse surgeons perceive higher levels of relative risk than risk-seeking surgeons (Regression 3).

All else held constant, there was no significant association between a surgeon's risk attitude score and their perception of relative risks of an operation (beta=-0.16, p=0.20). Therefore, I found no evidence to support hypothesis 2.2.

(Hypothesis 2.3) Risk-seeking surgeons perceive lower levels of absolute operative risk (Regression 4).

All else held constant, there was a significant association between a surgeon's risk attitude score and their perception of absolute operative risk (beta=-0.2, p=0.027; Figure 5.6). Surgeons that are more risk-seeking perceived lower levels of absolute operative risk. For example, compared to surgeons with risk attitude scores 1 SD below the mean, those with risk attitude scores 1 SD above the mean predicted 1.8 percentage points (95% CI -3.5, -0.2) lower levels of absolute operative risk (34.8% vs. 36.6%).

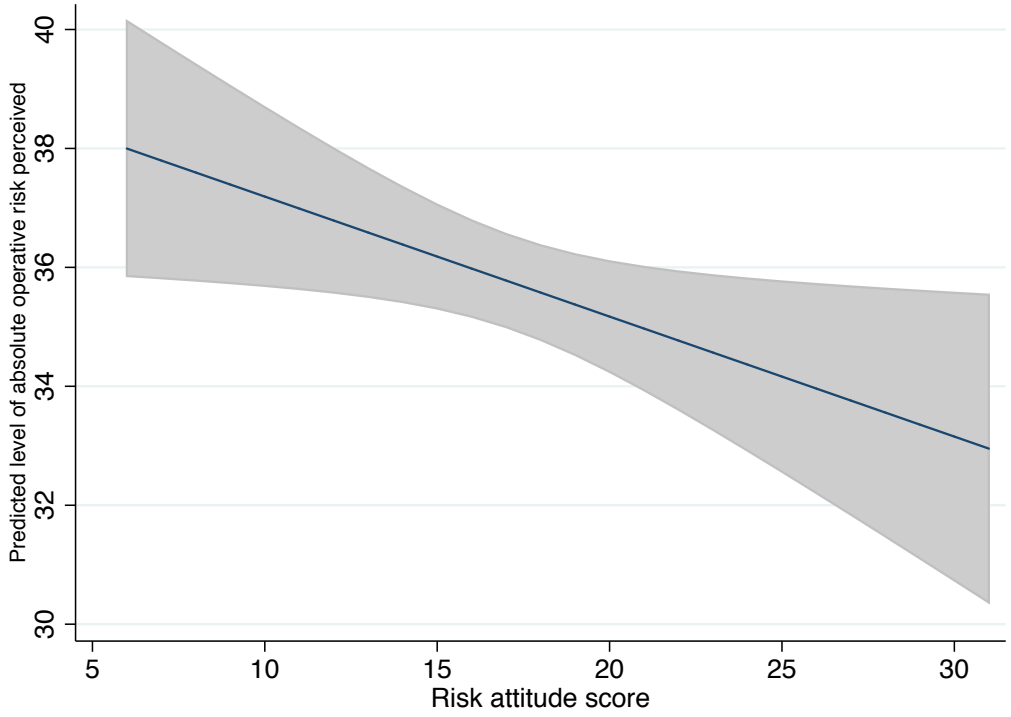



**Figure 5.6 Association between surgeon risk attitude and perceived level of absolute operative risk**

Regression controls for surgeon experience (years in training or since residency graduation, type of fellowship completed, operative volume), experience using a surgical risk calculator, and local surgical utilization rates. P-value for trend=0.027. Shaded area includes 95% confidence intervals.

(Hypothesis 3.1) Controlling for patient characteristics, surgeons who perceive higher levels of absolute operative risk are less likely to recommend an operation (Regression 2).

There was a significant negative association between surgeons' perception of absolute operative risk and their decision to recommend an operation. For example, surgeons judging the risks of operating at 1 SD below the mean would have a predicted probability of operating to be 77.5% (95% CI 75.5, 79.5), while those judging the risks of operating at 1 SD above the mean would have a predicted probability of 48.8% (95% CI 46.2, 51.4), translating into a 28.7 percentage point difference (95% CI -33.5, -24.8). Therefore, there is strong evidence in support of hypothesis 3.1. The results for hypotheses 3.1 through 3.4 are summarized in Table 5.2.

**Table 5.2 Association between surgeons' decision to operate and their perception of treatment risks and benefits**

| | Predicted probability at 1 SD below mean | Predicted probability at 1 SD above mean | Risk difference | 95% Confidence interval |
|---|---|---|---|---|
| Perceived absolute risks of operating | 77.5 | 48.8 | -28.7 | (-32.5, -24.8) |
| Perceived absolute benefits of operating | 54.0 | 73.6 | 19.6 | (16.0, 23.2) |
| Perceived absolute risks of not operating | 47.3 | 82.5 | 35.1 | (31.3, 39.0) |
| Perceived absolute benefits of not operating | 80.8 | 50.3 | -30.5 | (-34.5, -26.5) |

Predicted probabilities calculated using hierarchical logistic regression while controlling for surgeons' perception of all 4 risk/benefit parameters. Model included random intercept for the surgeon and fixed effect dummy variables for each clinical scenario and for the randomly assigned study arm.
Abbreviations: SD standard deviation

(Hypothesis 3.2) Controlling for patient characteristics, surgeons who perceive higher levels of absolute operative benefit are more likely to recommend an operation (Regression 2).

As also evident in Table 5.2, there was a significant association between surgeon perception of operative benefit and the decision to recommend an operation. Surgeons judging operative benefit at 1 SD above the mean would be 19.6 percentage points (95% CI 16.0, 23.2) more likely to recommend an operation than surgeons judging operative benefit at 1 SD below

the mean [73.6% (95% CI 71.5, 75.8) vs. 54.0% (95% CI 51.7, 56.4)]. This finding provides strong support for hypothesis 3.2.

(Hypothesis 3.3) Controlling for patient characteristics, surgeons who perceive higher levels of absolute risks associated with not operating are more likely to recommend an operation (Regression 2).

There was a significant positive association between surgeon perception of non-operative risk and the decision to recommend an operation. Surgeons judging non-operative benefit at 1 SD below its mean would have a 47.3% predicted probability of recommending an operation (95% CI 44.7, 49.9), while those judging 1 SD above the mean would have a predicted probability of 82.5% (95% CI 80.3, 84.6). Therefore, surgeons judging higher non-operative risk are 35.1 percentage points (95% CI 31.3, 39.0) more likely to recommend an operation than those judging lower non-operative risks. This finding supports hypothesis 3.3.

(Hypothesis 3.4) Controlling for patient characteristics, surgeons who perceive higher levels of absolute benefit associated with not operating are less likely to recommend an operation (Regression 2).

Finally, as surgeons' perception of non-operative benefits increased from 1 SD below the mean to 1 SD above the mean, they were 30.5 percentage points (95% CI -34.5, -26.5) less likely to recommend an operation [80.8% (95% CI 78.5, 83.1) vs. 50.3% (95% CI 47.8, 52.9); Table 5.2]. This finding provides evidence in support of hypothesis 3.4.

(Hypothesis 4.1) Surgeons who have completed fellowships in trauma surgery or acute care surgery perceive lower relative risks of operating than those who did not

complete a trauma/acute care surgery/burn surgery fellowship or those who completed another surgical fellowship (Regression 3).

Holding all else equal, there was no significant difference in the level of perceived relative risk of operating for surgeons who completed a trauma/acute care surgery/burn surgery fellowship and those who completed no fellowship [-14.5% vs. -12.0%, absolute difference -2.5% (95% CI -5.9, 1.0)]. However, surgeons who completed another type of fellowship did perceive significantly lower relative risks of operating compared to those who completed no fellowship [-18.1% vs. -12.0%, absolute difference -6.1% (95% CI -9.1, -3.1)]. Therefore, although there is no evidence to support hypothesis 4.1, there does appear to be an association between the type of fellowship training completed and the perception of relative risks of operating.

(Hypothesis 4.2) Surgeons perceive lower levels of absolute operative risk if they have more relevant experience, as measured by their number of years in residency, greater number of years in practice, their completion of a trauma/acute care surgery fellowship, and their high operative volume (Regression 4).

For each year of residency training completed, there is a 1.7 percentage point increase in the level of perceived absolute operative risk (p=0.041; Table 5.3). For example, residents in their first year of training judged the absolute risk of operating to be 27.8% averaged across all scenarios (95% CI 20.1, 35.5) and those in their final year of training (4 years completed) judged the level of risk to be 34.7% (95% CI 33.4, 36.0), translating into an absolute difference of 6.9% (95% CI 0.3, 13.6)].

**Table 5.3 Association between surgeon experience and the level of perceived absolute operative risk**

| Measure of experience | Marginal effect* | p-value | 95 percent confidence interval |
|---|---|---|---|
| Number of years in residency | 1.7 | 0.041 | (0.1, 3.4) |
| Years in practice | -0.1 | 0.009 | (-0.2, 0) |
| Fellowship completed | | | |
| No fellowship | Reference | | |
| Other fellowship | 0.4 | 0.75 | (-1.8, 2.5) |
| Trauma/ACS/critical care/burns | -0.2 | 0.84 | (-2.7, 2.2) |
| Operative volume, cases per year | | | |
| <50 | Reference | | |
| Resident | 2.4 | 0.52 | (-4.8, 9.6) |
| 51-100 | -0.1 | 0.96 | (-5.6, 5.3) |
| 101-250 | -1.9 | 0.44 | (-6.6, 2.8) |
| 251-500 | -2.3 | 0.33 | (-7, 2.3) |
| >500 | -2.9 | 0.26 | (-8, 2.1) |

Abbreviations: ACS acute care surgery

*Reflects the marginal change in absolute operative risk predicted (measured in percentage points) for each unit change in the independent variable (for continuous variables) or in comparison to the reference group (for categorical variables)

For each year of experience in clinical practice, there is a 0.1 percentage point decrease in the level of perceived absolute operative risk (p=0.009; Figure 5.7). For example, surgeons who graduated residency 3 years ago (1 SD below the mean) would judge the level of absolute operative risk to be 37.0% (95% CI 35.8, 38.2), while those who graduated 28 years ago (1 SD above the mean) would judge the risk to be 34.2% (95% CI 32.9, 35.6), translating into an absolute difference of -2.8 percentage points (95% CI -4.8, -0.7).

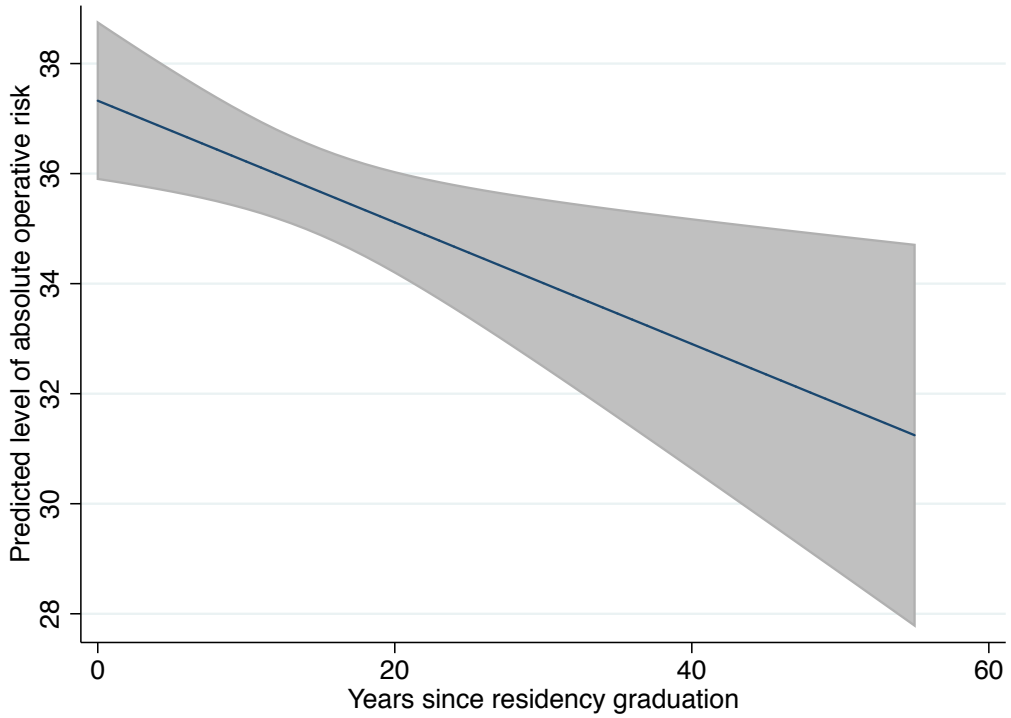



**Figure 5.7 Association between surgeon experience and perception of absolute operative risk**

There was no significant association between the type of fellowship training completed and the level of absolute operative risk perceived. Compared to surgeons who completed no fellowship, those who completed a fellowship in either trauma, acute care surgery, critical care, or burns judged similar levels of operative risk (p=0.81) as did those who completed another type of fellowship (p=0.54).

There was also no significant association between a surgeon's annual operative volume and the level of absolute operative risk perceived (p=0.34 for Wald test of overall significance).

(Hypothesis 4.3) Surgeons perceive higher levels of absolute operative benefit if they have more relevant experience, as measured by the number of years in residency, greater number of years in practice, their completion of a trauma/acute care surgery fellowship, and their high operative volume (Regression 5).

For each year of residency training completed, there is a 0.89 percentage point decrease in the level of perceived absolute operative benefit, although this finding did not reach statistical significance (p=0.35; Table 5.4).

**Table 5.4 Association between surgeon experience and the level of perceived absolute operative benefit**

| Measure of experience | Marginal effect* | p-value | 95 percent confidence interval |
|---|---|---|---|
| Number of years in residency | -0.9 | 0.35 | (-2.7, 1) |
| Years in practice | 0.1 | 0.012 | (0, 0.2) |
| Fellowship completed | | | |
| No fellowship | Reference | | |
| Other fellowship | -0.4 | 0.72 | (-2.8, 1.9) |
| Trauma/ACS/critical care/burns | -3.4 | 0.014 | (-6.2, -0.7) |
| Operative volume, cases per year | | | |
| <50 | Reference | | |
| Resident | 5.0 | 0.22 | (-3, 13.1) |
| 51-100 | 6.6 | 0.035 | (0.5, 12.7) |
| 101-250 | 7.5 | 0.005 | (2.3, 12.8) |
| 251-500 | 7.3 | 0.006 | (2.1, 12.5) |
| >500 | 7.9 | 0.006 | (2.2, 13.6) |

Abbreviations: ACS acute care surgery
*Reflects the marginal change in absolute operative benefit predicted (measured in percentage points) for each unit change in the independent variable (for continuous variables) or in comparison to the reference group (for categorical variables)

For each year of experience in clinical practice, there is a 0.1 percentage point increase in the level of perceived absolute operative benefit (p=0.012; Figure 5.8). For example, surgeons who graduated residency 3 years ago (1 SD below the mean) would judge the level of absolute operative benefit to be 68.0% (95% CI 66.7, 69.4), while those who graduated 28 years ago (1 SD above the mean) would judge the risk to be 71.0% (95% CI 69.5, 72.6), translating into an absolute difference of 3.0 percentage points (95% CI 0.7, 5.3).

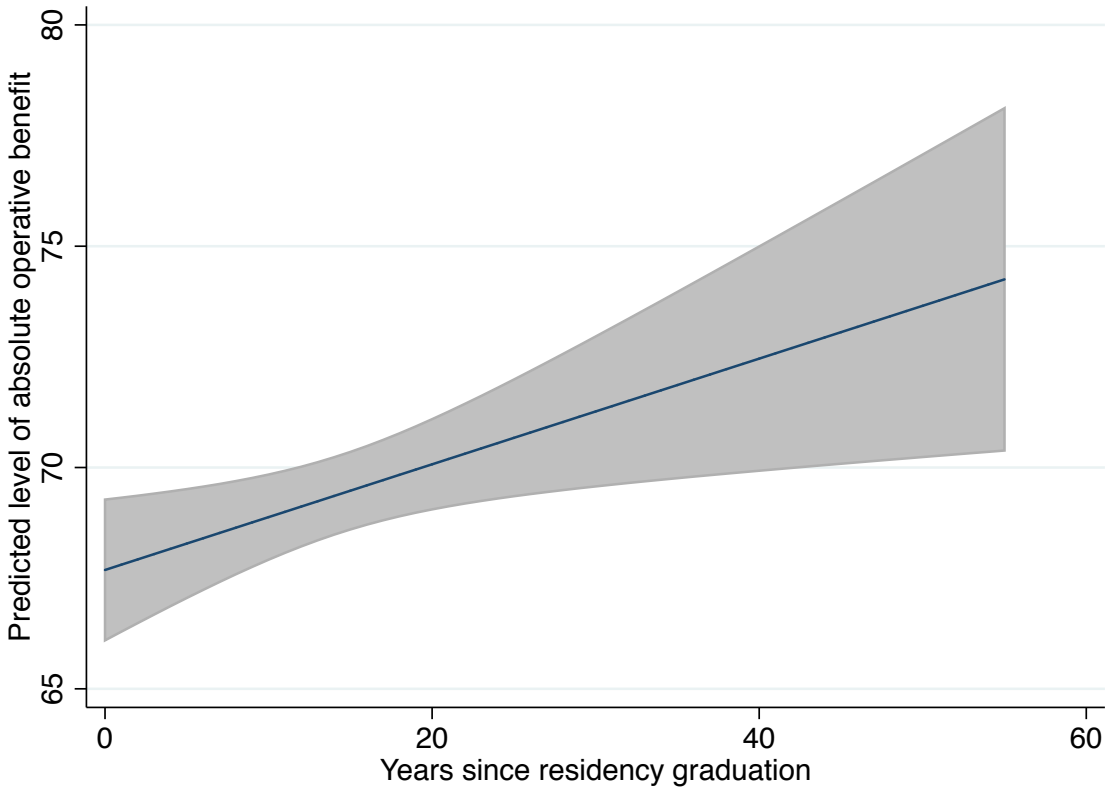



**Figure 5.8 Association between surgeon experience and perception of absolute operative benefit**

Surgeons who completed fellowship training perceived significantly lower levels of absolute operative benefit compared to those who completed no fellowship training [66.7% vs. 70.1%, absolute difference -3.4% (95% CI -6.2, -0.7)].

There was also a significant association between a surgeon's annual operative volume and the level of absolute operative benefit perceived (Figure 5.9). Compared to surgeons who perform fewer than 50 cases per year, those who perform greater than 50 cases perceive significantly higher levels of absolute operative benefit (Table 5.4). For example, surgeons who perform fewer than 50 cases per year perceive the average level of absolute operative benefit to be 62.6% (95% CI 57.5, 67.6), while those who perform greater than 500 operations per year

perceive the average level of operative benefit to be 70.5% (95% CI 67.7, 73.2), translating to an absolute increase of 7.9% (95% CI 2.2, 13.6).

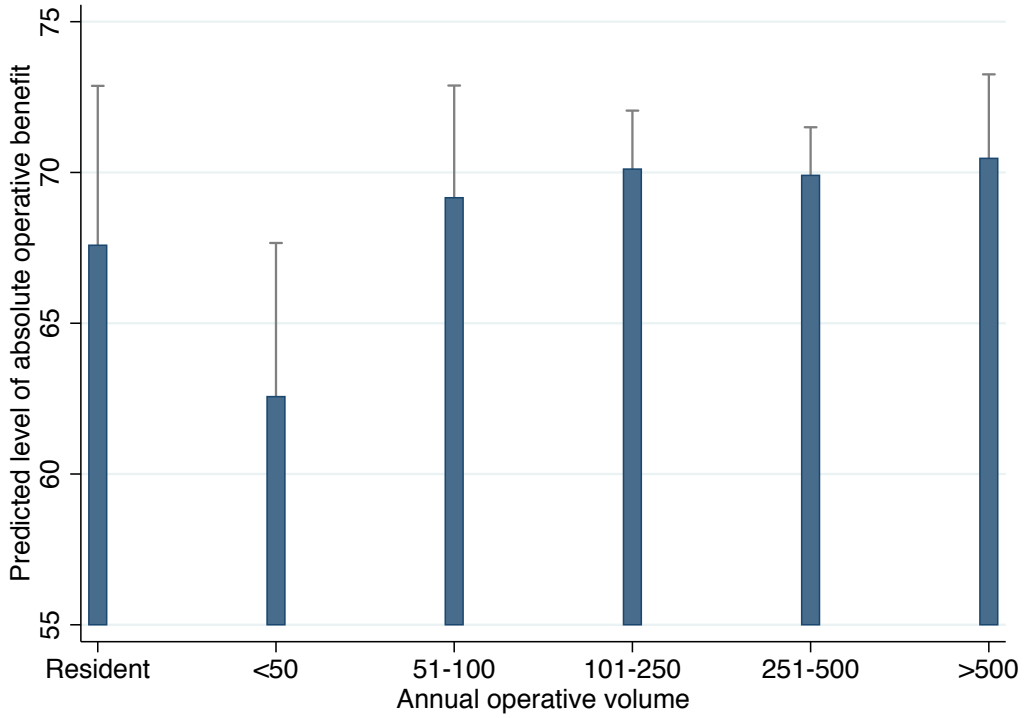



**Figure 5.9 Association between annual operative volume and the perceived level of absolute operative benefit**

(Hypothesis 5.1) Exposing surgeons to the risk calculator improves surgeons' prediction skill (Regression 6).

Prediction skill was measured as the difference between a surgeon's estimates and those calculated by the risk calculator. Defined this way, the risk calculator conferred an average improvement in prediction skill of 9.3 percentage points (95% CI 7.7%, 10.8%). Surgeons in the control group deviated from the risk calculator estimates by an average of 26.4 percentage points, while those in the risk calculator group deviated by an average of 17.2 percentage points.

Additional descriptive analyses are further indicative of the risk calculator effects and shed light on subsequent analyses. For example, unadjusted comparison of the surgeons' judgments of operative risk in the risk calculator and the null group reveals both how surgeons' naive judgments match those of the risk calculator and also how risk calculator exposure influences surgeons' judgments. In comparison to level of risk estimates by the risk calculator, surgeons in the control group overestimated the risks of operating for all 4 scenarios, while those in the risk calculator group estimated significantly lower risks of operating that more closely approximated the risk calculator values (Table 5.5). For example, in the mesenteric ischemia vignette, surgeons in the risk calculator group estimated the risk of death following an operation to be 23.2% while those in the control group estimated the risk of death to be 42.5% ($p<0.001$; actual risk calculator value 12%). Thus, in all 4 scenarios, surgeons in the risk calculator group clustered their predictions around the given risk calculator value (Figure 5.10). For 2 predictions (risk of death in the gastrointestinal bleed vignette and risk of serious complication in the small bowel obstruction vignette) the average estimates for surgeons in both groups closely reflected that of the risk calculator.

**Table 5.5 Surgeon judgment of risks and benefits for operative and non-operative management with and without risk calculator**

| Judgment of operative risk | Risk calculator value | Control group (n=384) | Risk calculator group (n=395) | p-value* | Control group (n=384) | Risk calculator group (n=395) | p-value* |
|---|---|---|---|---|---|---|---|
| | Percent | Mean, % | | | Standard deviation, % | | |
| Mesenteric ischemia | | | | | | | |
| 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 |
| Gastrointestinal bleed | | | | | | | |
| 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 |
| Small bowel obstruction | | | | | | | |
| Serious complication | 14 | 17.5 | 13.6 | <0.001 | 14.9 | 10.6 | <0.001 |
| Appendicitis | | | | | | | |
| Serious complication | 5 | 24.4 | 13.4 | <0.001 | 21.8 | 15.2 | <0.001 |

* 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.

There was also significantly less variation in operative risk judgments among surgeons in the risk calculator group compared to those in the control group (Table 5.5). For example, in the gastrointestinal bleed vignette, the standard deviation for the judged likelihood of operative death was lower in the risk calculator group than in the control group (12.8% vs. 20.2%, p<0.001). The findings of lower means and lower standard deviations for these risk estimates were consistent across all vignettes.

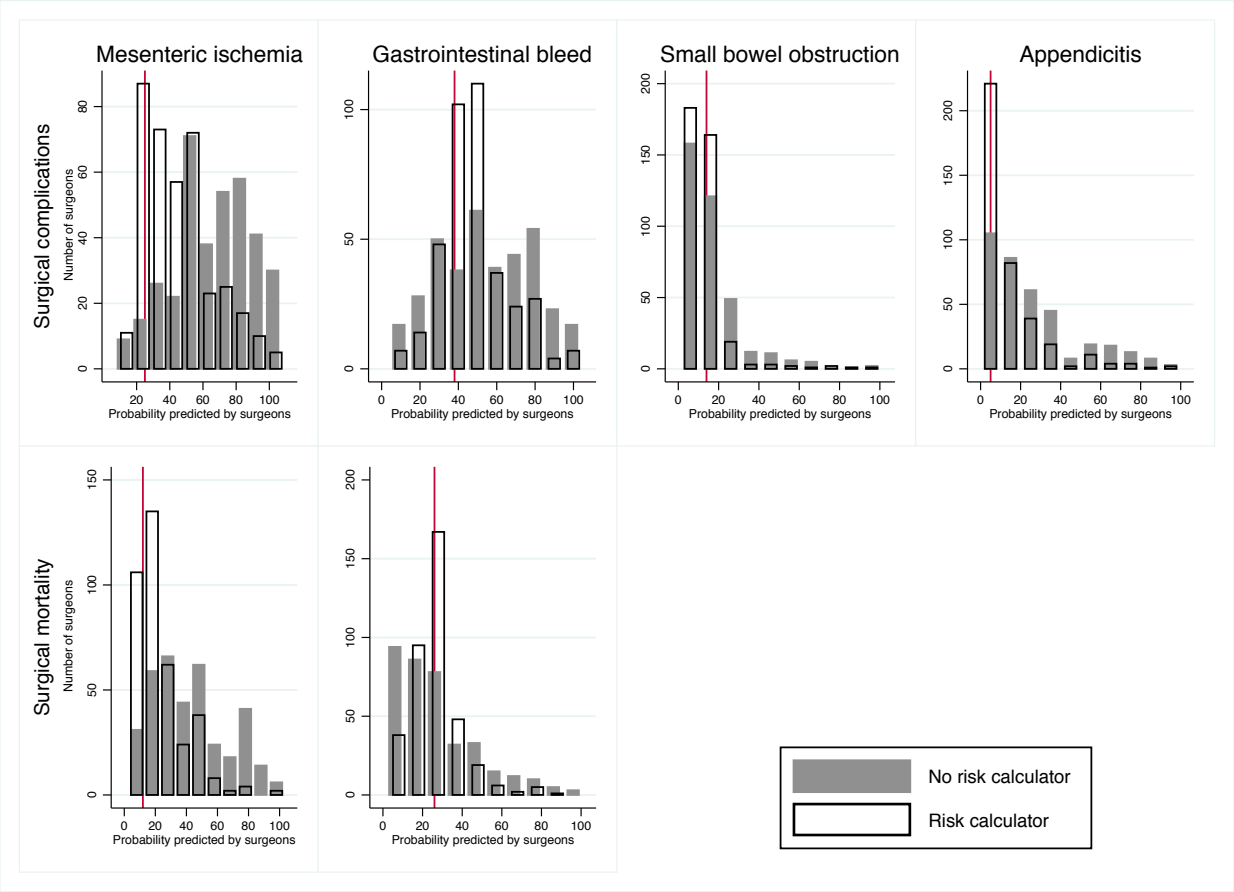



**Figure 5.10 Surgeon's judgment of operative risk with and without risk calculator data**
Red line signifies value from risk calculator
Risk of mortality was exceedingly low for the small bowel obstruction and appendicitis cases and therefore surgeons were not asked to make these predictions.

Finally, averaged across all 4 vignettes, surgeons exposed to the risk calculator predicted lower risks of operating (-10.4 percentage points, p<0.001; Table 5.6) compared to surgeons in the control group. Yet, even though the calculator only provides data on the risks of operating,

surgeons exposed to these data also judged higher benefits of operating (+2.2 percentage points, p=0.059), 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 to surgeons in the control group.

**Table 5.6 Adjusted effect of risk calculator on surgeon judgments and clinical decisions across all four clinical vignettes**

| Surgeon estimates | Control group (n=384) | Risk calculator group (n=395) | Absolute difference between groups | Relative difference between groups | p-value* |
|---|---|---|---|---|---|
| | | | 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 |

*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”

(Hypothesis 5.2) The improvement in prediction skill conferred by exposure to the risk calculator is greater for those surgeons who routinely use the calculator in their clinical practice than for those who do not (Regression 6).

Before testing this hypothesis directly, it is informative to first examine the different levels of prediction skill displayed by surgeons based on their prior experience with the risk calculator. Among surgeons in the control group (not exposed to the risk calculator), those who were previously unaware of the ACS-NSQIP risk calculator had the lowest level of prediction skill, on average deviating from the risk calculator number by 28.4 percentage points. This deviation was significantly different from the 24.5 percentage point deviation by those surgeons who were previously aware of the calculator but did not use it in practice (AD=-2.1%, 95% CI -

3.9, -0.2; Table 5.7). Surgeons who used the calculator occasionally in their clinical practice also predicted on average 2.0 percentage points closer to the risk calculator's prediction (95% CI -4.0, -0.1) than surgeons who were previously unaware of the calculator. However, there was no difference between surgeons who were previously unaware of the calculator and those who use the calculator routinely in their practice (AD=-0.5, 95% CI -5.0, 4.0). Nevertheless, this absent association may be due to the fact that only 56 surgeons (3.2%) used the calculator routinely in their practice.

On average, the improvement in prediction skill conferred by exposure to the ACS-NSQIP risk calculator was not greater for surgeons who routinely use the calculator. In fact, contrary to the hypothesis, the surgeons who demonstrated the greatest improvement in prediction skill were those who were previously unaware of the calculator (marginal effect -10.9, 95% CI -13.2, -8.6). However, this effect was not statistically significantly different from the 7.3% improvement demonstrated by surgeons who were aware of but did not use the calculator (AD=3.6%, 95% CI -0.1%, 7.4%), the 9.1% improvement for the surgeons who occasionally used the calculator (AD=1.8%, 95% CI -2.0%, 5.6%), or the 3.3% improvement for surgeons who used the calculator routinely (AD=7.7%, 95% CI -1.4%, 16.7%).

**Table 5.7 Effects of routine use of risk calculator on prediction skill**

| | Prediction skill among surgeons in control group | Prediction skill among surgeons in risk calculator group | Marginal effect of the risk calculator[1] | 95 percent confidence interval |
|---|---|---|---|---|
| Prior experience with risk calculator** | | | | |
| Unaware | 28.4 | 17.5 | -10.9 | (-13.2, -8.6) |
| Aware but don't use | 24.5 | 17.2 | -7.3 | (-10.2, -4.3) |
| Use occasionally | 25.5 | 16.3 | -9.1 | (-12.2, -6.1) |
| Use routinely | 24.1 | 20.8 | -3.3 | (-12.0, 5.5) |

[1]Reflects the marginal change in prediction skill, which was measured by comparing surgeons' predictions of absolute operative risk with the same predictions made by the ACS-NSQIP risk calculator and taking the absolute value of the difference; therefore, the smaller the value, the greater the prediction skill.
**Surgeons' self reported experience using the ACS-NSQIP risk calculator

(Hypothesis 5.3) If hypothesis 5.1 shows that exposure to the risk calculator causes surgeons to perceive higher levels of absolute operative risk, then exposure to the risk calculator will also change their behavior and make surgeons less likely to recommend an operation (Regression 7).

However, if exposure to the risk calculator causes surgeons to predict lower average levels of operative risk, surgeons will change their behavior to be more likely to recommend an operation (Regression 7).

Reviewing the coefficient for the risk calculator dummy variable in regression 4 reveals that the risk calculator caused surgeons to judge absolute operative risk to be an average of 10.4 percentage point lower (Table 5.6) than surgeons in the control group. Given this finding, the alternate version of hypothesis 5.3 will be formally tested here. In unadjusted analysis, the average response to the question "how likely are you to recommend an operation as opposed to pursuing medical management?" was similar between surgeons in the risk calculator and control

group for all 4 vignettes (Figure 5.11). For example, for the mesenteric ischemia vignette, the mean response was nearly identical between the 2 groups (3.7 vs. 3.8; p=0.59).

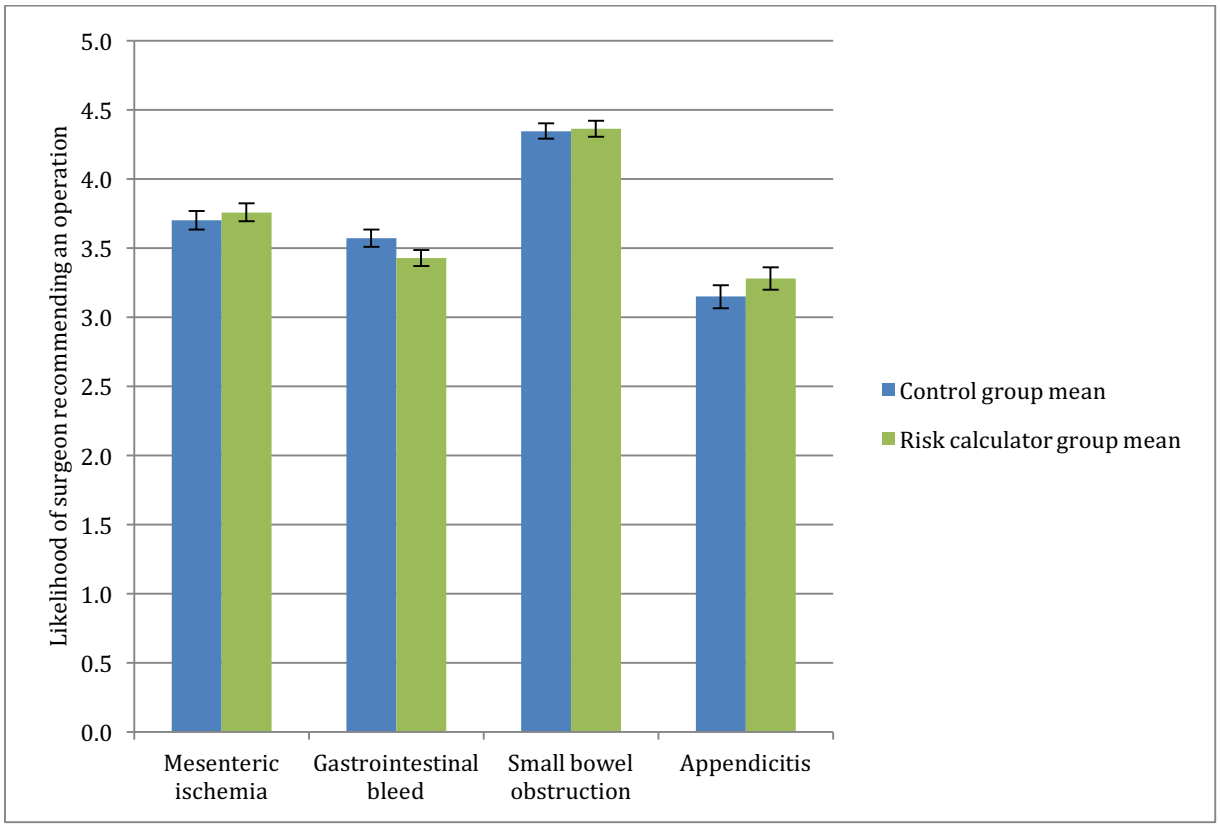



**Figure 5.11 Surgeon decision to recommend an operation with and without risk calculator**
Likelihood of surgeon recommending an operation measured by surgeon response to the question "how likely are you to recommend an operation at this time?" as answered on a 5-point Likert scale 1 “very unlikely”, 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)

In adjusted analysis of the average effect across all clinical scenarios, there was still no difference between the 2 groups in the likelihood that the surgeon recommended an operation (65.3% vs. 65.8%, p=0.78; Table 5.8). In summary, these results support the null hypothesis and provide no evidence in favor of the alternate hypothesis 5.3.

**Table 5.8 Effects of risk calculator on surgeons' decision to recommend an operation**

| | Control group | Risk calculator group | Absolute difference | 95% confidence interval |
|---|---|---|---|---|
| Predicted probability that surgeon recommended an operation | 65.3 | 65.8 | 0.5 | (-3.1, 4.1) |

(Hypothesis 6.1) The higher tendency for surgeons to operate when they perceive lower operative risks is more pronounced when surgeons are only asked to predict the risks and benefits of operating (and not the risks and benefits of *not* operating; regression 8).

Averaged across all levels of perceived operative risk and benefit, surgeons who were, by random assignment, only asked to assess the risks and benefits of operating (operate only group) and not the risks and benefits of not operating were significantly less likely to recommend an operation than those surgeons (control group) who were asked to assess both (60.0% vs. 64.5%, AD=-4.5%, 95% CI -8.4%, -0.7%). However, the higher tendency to operate when they perceived lower operative risk was equivalent between the 2 randomly assigned groups. For example, comparing surgeons who perceived operative risk at 1 SD below the mean to those 1 SD above the mean, surgeons in the operate only group were 23.4 percentage points (95% CI -30.0, -16.8) less likely to recommend an operation, while those in the null group were 20.4 percentage points (95% CI -26.3%, -14.5%) less likely to recommend an operation (AD=3.0, 95% CI -2.4%, 8.3%; Figure 5.12). Therefore, these findings do not provide support for hypothesis 6.1.

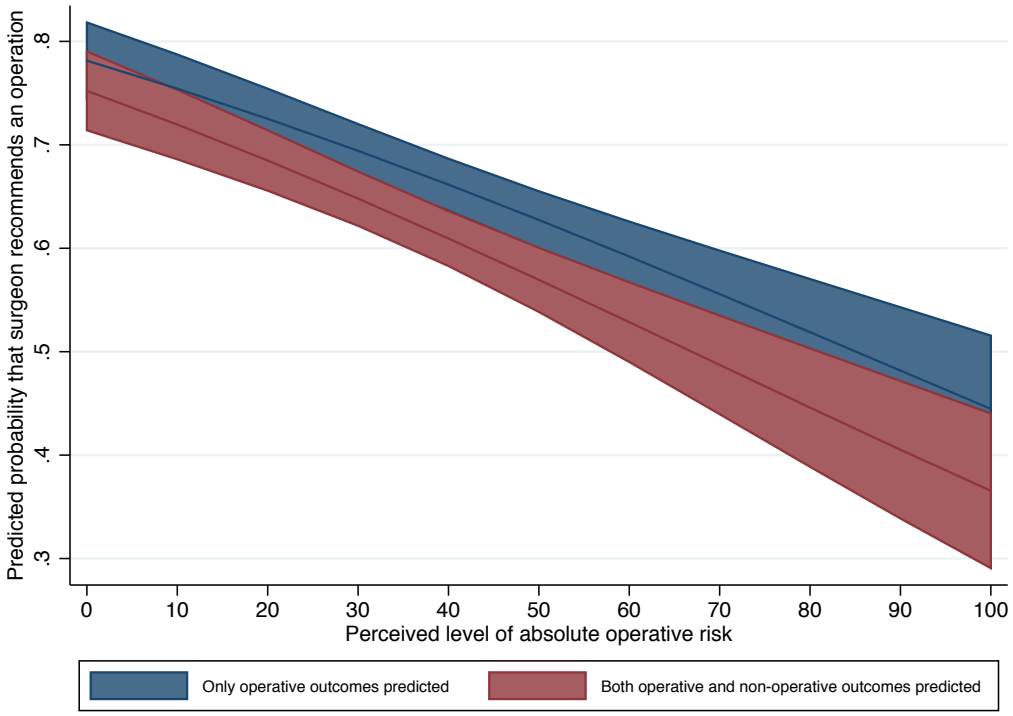



**Figure 5.12 Effect of only asking surgeons to assess operative outcomes on the decision to recommend an operation by perceived level of absolute operative risk**

(Hypothesis 6.2) The higher tendency for surgeons to operate when they perceive high levels of risks associated with *not* operating is more pronounced when surgeons are only asked to predict the risks and benefits of *not* operating (and not the risks and benefits of operating; regression 9).

Asking surgeons to only assess the risks and benefits of non-operative management tended to make them weigh those outcomes more heavily in their decision to recommend an operation, although this finding did not reach statistical significance (p=0.16 for interaction term between random experimental assignment and perceived non-operative risk). For example, in the control group, there would be a 17.7 percentage point increase (95% CI 11.3%, 24.0%) in the likelihood of recommending an operation as perceived non-operative risk increased (from 1 SD below the mean to 1 SD above the mean). Among surgeons who were only asked to assess the

risks and benefits of non-operative management, that same increase was 20.6 percentage points (95% CI 14.3%, 26.9%). This translates into an absolute difference in effect size of -3.0% (95% CI -0.7%, 1.3%). Therefore, these findings provide some evidence for hypothesis 6.2, though not at statistically significant levels (Figure 5.13).

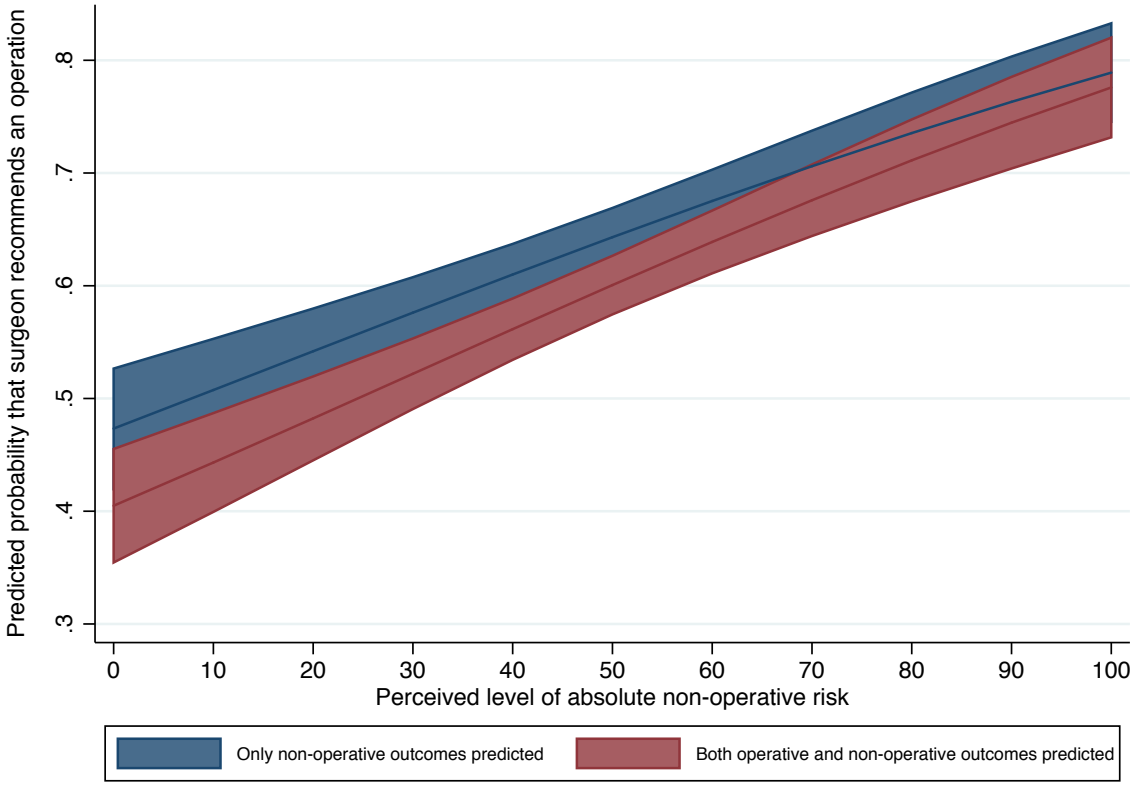



**Figure 5.13 Effect of only asking surgeons to assess non-operative outcomes on the decision to recommend an operation by perceived level of absolute non-operative risk**

Although there was no significant difference in how surgeons weighed non-operative risks in their decision to recommend an operation, there was a difference in how surgeons weighed non-operative benefits in this decision if they were only asked to assess non-operative outcomes (Figure 5.14). For example, in the control group, there would be a 31.2 percentage point decrease (95% CI -37.7%, -24.8%) in the likelihood or recommending an operation as perceived non-operative risk increased (from 1 SD below the mean to 1 SD above the mean).

Yet, among surgeons who were only asked to assess the risks and benefits of non-operative management, that same decrease was 42.2 percentage points (95% CI -48.9%, -35.6%). This translated into an absolute difference in effect size of 11.0% (95% CI 5.5% 16.6%).

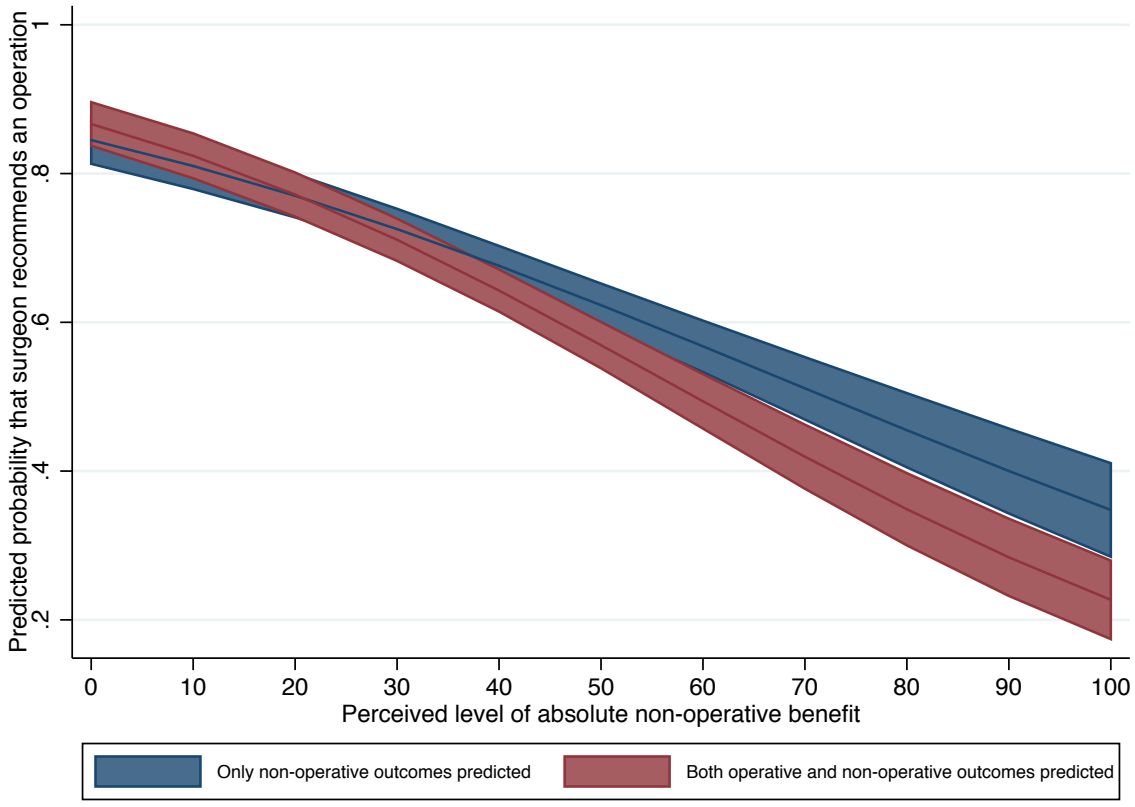



**Figure 5.14 Effect of only asking surgeons to assess non-operative outcomes on the decision to recommend an operation by perceived level of absolute non-operative risk**

## 5.3 Summary

In summary, the findings for each of the tested hypotheses were mixed, with empirical support found for only about half of the 18 hypotheses tested (Table 5.9). The overall findings show support for the fact that surgeons vary widely in how they perceive the risks and benefits of operative and non-operative management and that these perceptions are strongly predictive of their treatment decisions. In fact, accounting for only surgeon perception of the relative risks and benefits of operating explained a total of 36% of the observed variation in the decision to

recommend an operation. There also appears to be some association between surgeon characteristics, such as their experience and overall risk attitude, and their perception of these risks and benefits. Furthermore, surgeons are responsive to the provision of external and objective risk data. Providing surgeons with data from the ACS-NSQIP risk calculator resulted in less varied and more accurate risk predictions. However, these changes did not affect their ultimate clinical decision-making. Finally, the findings also suggest that surgeons' decision-making is somewhat susceptible to subtle changes in how the risks and benefits are framed. By only asking surgeons to predict either the operative or non-operative risks and benefits, I observed some significant differences in how surgeons weighed these risks and benefits in their clinical decisions.

**Table 5.9 Summary of evidence for hypotheses**

| Hypothesis | Evidence supporting hypothesis[†] | Comments |
|---|---|---|
| 1.1 | + | Surgeons between 15.3 and 64.7 percentage points less likely to recommend an operation as level of perceived relative risk of operating increased |
| 1.2 | + | Surgeons between 9.7 and 59.1 percentage points more likely to recommend an operation as level of perceived relative risk of operating increased |
| 2.1 | - | No interaction between surgeons' risk attitude and perception of relative risk on decision to recommend an operation |
| 2.2 | - | No significant association between a surgeon's risk attitude score and their perception of relative risks of operation (beta=-0.16, p=0.20) |
| 2.3 | + | Significant association between a surgeon's risk attitude score and their perception of absolute operative risk (beta=-0.2, p=0.027) |
| 3.1 | + | Surgeons less likely to recommend an operation as perceived absolute operative risk increases |
| 3.2 | + | Surgeons more likely to recommend an operation as perceived absolute benefit risk increases |

| | | |
|---|---|---|
| 3.3 | + | Surgeons more likely to recommend an operation as perceived absolute non-operative risk increases |
| 3.4 | + | Surgeons less likely to recommend an operation as perceived absolute non-operative benefit increases |
| 4.1 | - | No difference in level of perceived relative risk of operating for surgeons who completed a trauma/acute care surgery/burn surgery fellowship and those who completed no fellowship; however, surgeons who completed another fellowship perceived lowest levels of relative risk |
| 4.2 | +/- | Higher level of operative risk perceived as experience in residency increases, lower level of operative risk perceived as years in clinical practice increased; no effects for fellowship completed or operative volume |
| 4.3 | +/- | Higher level of operative benefit as years in clinical practice increased, lower, not higher, level of operative benefit for surgeons completing trauma/acute care surgery fellowship, and higher perceived operative benefit among high volume surgeons |
| 5.1 | + | Risk calculator caused surgeons to demonstrate better prediction skill |
| 5.2 | -* | Surgeons with prior experience with risk calculator had better predictive skill on average. Improvement in prediction skill was greatest for surgeons with the least prior experience with the risk calculator. |
| 5.3 | - | No effect of risk calculator on the decision to recommend an operation |
| 6.1 | - | Asking surgeons to only assess the risks and benefits of operating did not result in greater weighting of these parameters in the decision to recommend an operation. |
| 6.2 | - | Asking surgeons to only assess the risks and benefits of not operating did not result in greater weighting of these parameters in the decision to recommend an operation. However, surgeons in the experimental group weighed non-operative benefits significantly more in their decision. |

† + implies evidence was found in support of hypothesis, - implies no evidence was found, and +/- implies some findings supported the hypothesis, while other findings did not support the hypothesis (in cases where multiple coefficients were tested); asterisk implies that findings were significant but not in the direction hypothesized.

# 6 Discussion

Deciding whether or not to operate on a patient is an enormously consequential decision—for the patient, for the surgeon, and for the entire healthcare system. Yet what has been known for decades, is that whether or not a patient undergoes an operation depends not only on the patient's clinical condition, but also on which surgeon evaluates the patient and makes the resulting clinical decision. Previously documented by the abundant literature on regional variations in care, this dissertation further confirms this phenomenon using detailed clinical vignettes, in which there was no clearly dominant treatment option. Under such uncertainty, even when confronted with identical patient scenarios, different surgeons reach very different and often opposing decisions on whether or not to use surgical intervention.

Despite the extensive research on variations in the use of surgery, prior studies have largely fallen short of fully answering the question of what makes one surgeon choose to operate on a patient and another surgeon choose not to operate. This dissertation attempts to answer this question by first creating a conceptual model that outlines how surgeons might make this challenging clinical decision. The conceptual model recognizes that the decision to operate on a patient, like many clinical decisions, is often made under conditions of uncertainty. Borrowing from traditional economic theory as well as more recent lessons from the field of behavioral economics, this model centers on how surgeons perceive the risks and benefits of each available treatment options, most relevantly the option to operate on a patient or manage the patient medically without an operation. In this way, the model serves as both a descriptive and a prescriptive model for surgical decision-making.

The model is prescriptive in that it takes a normative stance on how the decision to operate *should* be made, given an individual patient. According to the model, a surgeon should

recommend an operation if the risks of operating are low relative to the risks of not operating and if the benefits of operating are high relative to the benefits of not operating. What is notable about this model is that it does not dictate what decision *should* be made but rather *how* the decision should be made. For each level of perceived relative risk and benefit, two surgeons could easily and justifiably reach opposing treatment decisions, a discrepancy that may be explained in part by characteristics of the surgeon, such as how comfortable the surgeon is with a given risk/benefit tradeoff, as well as by characteristics of the clinical setting, such as the local culture or local incentive structure. Nevertheless, the important operative component of the conceptual model is the careful and deliberate consideration of the treatment risks and benefits and then using those parameters to inform the clinical decision.

In addition to being prescriptive, the model also serves in a descriptive capacity in that several parts of the model were validated using data from a large, national sample of surgeons responding to a series of standardized clinical vignettes. In so doing, the results offer a novel explanation for the otherwise unexplained variation in the use of surgery. The findings presented here suggest that surgeons differ in their use of surgery, at least in part, because they have very different ideas on what the risks and benefits are for both operative and non-operative management. In fact, for many of these estimates, surgeons varied by as much as 0-100%. These differences in risk/benefit estimates were strongly associated with surgeons' clinical decisions. Accounting for only the differences in perceived relative risks and benefits, I was able to explain 36% of the observed variation in the decision to recommend an operation.

It is important to note that the analyses and results presented here pass no judgment on what constitutes the right or wrong treatment decision. In fact, the vignettes used to collect these data were deliberately designed so that no best treatment option was obvious. Absent a clearly

dominant treatment option, there may in fact not be a strictly "right" or "wrong" decision. Rather, decisions like these may best be made on a patient-by-patient basis, where 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.

## 6.1 Interpretation of findings

### 6.1.1 Perceived risks and benefits

Normative decision theory stipulates that under conditions of uncertainty, decisions should be made based on a consideration of the risks and benefits associated with each decision option. In order to perform effectively, this model of decision-making supposes that the probabilities associated with these risks and benefits are either known or easily calculated. Where these probabilities are known, decisions may be relatively straightforward. However, in many situations, particularly in the clinical realm, the relevant probabilities needed to make an informed decision are either unknown, or at the very least, uncertain. As a result, the findings presented in this dissertation suggest that in the case of 4 uncertain clinical scenarios, surgeons vary widely in how they perceive the risks and benefits relevant to making an informed clinical decision to operate or not operate on a patient.

Of the 4 predictions analyzed in this dissertation (risks and benefits of operating and not operating), perhaps the most surprising is the variation in surgeons' judgment of the risks of operating. Assessment of patients' operative risk has in fact been a hallmark of surgical practice for almost half a century[5] and is now a feature of hospital benchmarking algorithms and pay-for-performance programs.[101,102] Yet, the fact that surgeons vary so widely in their operative risk

assessment suggests that improvements in risk prediction have not yet influenced clinical practice on a broad scale.

The finding that surgeons vary in how they perceive the benefits of operating, on the other hand, is slightly less surprising given the perhaps unexpected fact that there is currently no well-accepted definition for the concept of operative benefit and, as a result, it is usually not used as a probabilistic construct in clinical decision-making. Although the definition used in this study—the probability that a patient recovers within 30 days of an operation—may not be equally applicable across all clinical scenarios and patients, the results of this dissertation suggest that surgeons do incorporate their judgment of this outcome into their clinical decisions, indicating the need to establish a practical measure for this concept. In addition to guiding individual clinical decisions, focusing explicitly on the intended benefits of care may also be instrumental in the development of clinical guidelines, as shown by one recent simulation study on the use of antihypertensive medications, which demonstrated the potential for treatment guidelines tailored directly to the intended benefits to not only save hundreds of thousands of lives but do so at a fraction of the cost of more contemporary guideline strategies.[103] However it is derived, any useful definition for the concept of operative benefit will likely differ from one clinical condition to the next but it should, if intended for wide use, have sufficient breadth to encompass the postoperative outcomes that are most important to patients[104] and also the flexibility to be adapted to different clinical scenarios.

Least surprising of all is the finding that surgeons vary in their judgment of outcomes for non-operative management. This is largely due to the fact that there remains a paucity of data on outcomes for patients who undergo medical management of potentially surgical conditions. One explanation for this unavailable data is that absent any prospective clinical trials, data on these

patients would likely need to be collected from other data sources, such as administrative billing data. Yet, the ability to accurately identify patients who are managed non-operatively and to meaningfully compare them to patients who were managed surgically, is limited. Further, existing surgical registries track outcomes only for patients who undergo surgery and do not keep track of patients who are managed non-operatively.[105] It will therefore be important moving forward to develop reliable ways to measure these outcomes so that such data can be incorporated into the decision-making process.

By one interpretation, the findings presented in this dissertation suggest, encouragingly, that surgeons make different treatment decisions simply because they are rationally responding to their own careful analysis of treatment risks and benefits. The finding then that surgeons choose to operate when the risks of operating are low and the benefits of operating are high borders on tautology. In fact, to even speculate otherwise would seem both ungenerous and indicting. Yet, the revelation of these findings is not that surgeons operate when the risks are low and the benefits are high but rather that they operate when they *think* the risks are low and the benefits are high. That surgeons vary so widely in what they think these risks and benefits are and that such variation in perception then drives their clinical decision-making is far less tautological and does, in fact, have important implications that will later be elaborated on. But it is important to note that by this interpretation of the data, surgeons appear to be using a model of decision-making that carefully takes into account the risks and benefits of both operative and non-operative management. They are, however, simply populating that model with very different inputs on what those risks and benefits are.

Also encouraging is the finding that surgeons were not particularly susceptible to subtle changes in how the questions about risk and benefit were framed. There are several examples

from the literature that show that physicians fall prey to irrational thinking when problems are framed in different ways. For example, physicians are less likely to choose an operation for themselves when they are informed on the risks of dying than if they are informed on the risks of surviving, even when the two risks are mathematically equal.[74] One might imagine, then, that surgeons may demonstrate similar framing bias depending on whether they are asked to assess either the risks and benefits of operating or the risks and benefits of not operating. Being asked only to assess the risks and benefits of operating may cause the surgeon to preferentially focus on that treatment option and weigh those probabilities more heavily than if they had been forced to consider the risks and benefits of not operating. The lack of evidence to support this hypothesis noted in this dissertation may be attributable to the fact that surgeons are already considering the risks and benefits of both treatments when making decisions and forcing them to only explicitly consider one treatment does not negate the implicit process already underway of considering both treatments simultaneously. Although I found no evidence to support such preferential weighting in the analyses for hypothesis 7.1 and 7.2, there was some evidence that surgeons tended to weigh more heavily the benefits of not operating in their decision to operate when they were only asked to assess non-operative outcomes. This finding points towards the possibility that the lack of an identified framing effect may be due more to nebulous problems surrounding the clinical vignettes, the attached questions, or other study design particulars than to the absence of these effects in surgical decision-making.

This dissertation unpacks the relationship between surgical decision-making and the perception of risks and benefits in two different ways: first, by analyzing the perceived risks and benefits of operating *relative* to those of not operating; and second, by analyzing the *absolute* level of perceived risk and benefit for each treatment option. Although the conceptual model

focuses primarily on the *relative* risk and benefit approach, the *absolute* risk and benefit approach is implicitly included in the model by virtue of the fact that relative risk is measured simply as the difference between two absolute risks. This distinction is important because moving forward, it will be important to understand which model more accurately describes surgeons' decisions so that it can be used to plan future studies and analysis plans. A comparison of regressions 1 (which uses relative risk and benefit) and 2 (which uses absolute risk and benefit) demonstrates that both approaches are comparable, although the relative risk and benefit approach is slightly superior in explaining the observed variation in the outcome (R-squared 0.366 vs. 0.342; these were calculated using models identical in specification to regression 1 and 2 except using a linear instead of a logistic regression to predict the decision to recommend an operation on its original 5 point scale).

### 6.1.2 Risk attitude

Despite the vast literature demonstrating an association between individuals' attitude towards risk and their decision-making, many of the hypotheses regarding individual risk attitude in this dissertation yielded null results. For example, there was no difference between risk-seeking and risk-averse surgeons in the magnitude of association between perceived relative risk and the decision to operate. Nor was there a direct association between risk attitude and perceived relative risk. On the other hand, there was evidence to support the hypothesis that risk-seeking surgeons perceive the absolute risks of operating to be lower than risk-averse surgeons. There are at least 3 possible explanations for these apparently conflicting findings. One, discussed in more detail in the limitation section, is that the risk attitude instrument used in the study did not accurately identify aspects of an individual's personality that are relevant to risk tolerance in a clinical setting. This may explain, in part, the combination of positive and negative

findings using this instrument. Second, it is possible that in addition to perceiving lower absolute risks of operating, risk-seeking surgeons may also perceive lower absolute risks of not operating, thereby canceling out the effect of risk attitude on perceived *relative* risk when the two absolute risks are subtracted from one another. Finally, it is also possible that there is only an association between risk attitude and perceived absolute risk of operating and not absolute risk of not operating, resulting in the null effects of risk attitude on perceived non-operative risk diluting out the effects of risk attitude on perceived operative risk. In fact, the data support this latter explanation rather than the second explanation. Using a regression that is identical in specification to regression 4 but with the outcome of interest changed from perceived operative risk to perceived non-operative risk, surgeon risk attitude was not a significant predictor of this outcome ($\beta$=-0.00, p=0.99).

### 6.1.3 Experience

According to the conceptual model outlined in this dissertation, the effect of experience on surgeons' clinical decisions is fully mediated by their perception of treatment risk and benefit. Validation of this theoretical model produced mixed results. There was, for example, an association between surgeons' fellowship training and their estimates of treatment outcomes in that surgeons who completed fellowships in trauma/acute care surgery perceived the benefits of operating to be lower than those who completed no fellowship. This finding may be, in part, due to the fact that all 4 of the vignettes analyzed fell under the domain of trauma/acute care surgery and surgeons who completed fellowships in this field may be more aware of the somewhat recent literature in this specialty demonstrating that diseases that were once traditionally thought to be operative in nature, are, in fact, often better managed without an operation. Therefore, this literature supports less relative benefit of operative intervention.

The number of years experienced, either in training or in clinical practice, seems to also be an important predictor of how surgeons perceive risk and benefit. Interestingly, as surgeons gained experience in practice, they were more likely to perceive the risks of operating to be lower and the benefits of operating to be higher. A simple explanation for this finding may be that the risks and benefits of operating are, in fact, more favorable when performed by a surgeon with more years under the belt and a greater number of such operations in their case logs. However, it is also possible that by doing more operations, one becomes convinced by the relative benefits of surgery, thereby driving a more optimistic view on the low risks and high benefits of the intervention. This may also explain why higher volume surgeons also perceived higher benefits of operating. A final explanation again relates to possibly lower awareness among senior surgeons of the more modern literature supporting the relative benefits of non-operative management of these clinical conditions.

### 6.1.4 Risk calculator

The surgical specialty has long been ahead of the curve in collecting data to rigorously scrutinize its quality of care. In the realm of risk prediction, surgeons have again taken the lead in making such estimates available at the point of care. Not only does the ACS-NSQIP risk calculator provide risk-adjusted estimates for an individual patient's risk of various complications following surgery, but it does so using high quality clinical data that are now available from a national registry comprised of over one million patients from around the country. As such, the calculator has been shown to have very good prediction capability (c-statistic 0.944 for mortality and 0.816 for morbidity).[7] Well-validated as the calculator is in its prediction ability, the data presented in this dissertation are the first to suggest how surgeons

might respond to these data and how they might incorporate the calculator's estimates in to their clinical practice.

Effect on risk prediction

The analyses on the effects of the risk calculator on surgeon risk perception yielded promising results, suggesting that surgeons do in fact incorporate the calculator's risk estimates into their own clinical judgment. For each prediction where risk calculator data were available, surgeons demonstrated greater accuracy and less variation in their judgments. Interestingly, relying only on their intuition (without the risk calculator data), surgeons tended to overestimate the risk of adverse outcomes following surgery relative to the risk calculator value, sometimes by as much as 20 percentage points. This finding contrasts to some extent with the findings of prior studies comparing physician estimates of risk parameters with those derived from validated risk assessment tools. These prior studies have drawn contradictory conclusions, with some suggesting high concordance between surgeons' probability estimates of adverse outcomes and those derived from risk assessment tools,[58-60] and others revealing systematic over- or under-estimation by physicians compared to the output of these risk assessment tools.[47,106,107] These discrepancies may be partially explained by the fact that the systems used to create these prediction tools vary with respect to the variables included, the extent to which data definitions were standardized, the types of outcomes predicted, the statistical techniques used, and the validation methodology employed.[9] Of all these risk assessment tools, the ACS-NSQIP risk calculator uses the highest quality data abstracted directly from patients' clinical charts, the most advanced statistical modeling strategies, and the largest number of patient observations on which to base its estimates. If it is assumed that these modeling improvements result in more accurate predictions, then it appears that surgeons without knowledge of these estimates do in fact tend to

overestimate the risks of operating, at least for the four clinical scenarios examined in this dissertation. This is particularly interesting since that suggests a close resemblance between surgeons and patients, who also tend to overestimate risks associated with medical intervention.[108]

In addition to judging lower likelihood of adverse operative outcomes, surgeons in the risk calculator group also predicted slightly higher benefits of operating than surgeons in the control group when averaged across all scenarios, despite the fact that no data on operative benefit are available from the risk calculator. This finding may be explained in part by what behavioral scientists call the affect heuristic,[62] whereby decisions and judgments (including those of risk and benefit) are driven in part by people's emotional response to an activity.[67,109-111] The affect heuristic is a plausible explanation here because if surgeons base their decision to operate on their affective response to operating (i.e., 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.[67] But because affect was not measured directly in this study, its role in surgical decision-making warrants further investigation.

Effect on decision to operate

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, 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 two groups, while 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. These explanations are unlikely, however, considering the strong empiric association between surgeons' decisions to recommend an operation and their judgments of the risks and benefits of operative and non-operative management. Second, it appears that when risk calculator data suggested that the risks of operating are lower than surgeons expected, this led to not only lower judgments of operative risk (and higher judgments of benefit) but also lower judgments of non-operative 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 irrelevant non-operative risk assessments.[63,112,113] This second explanation is likely as it is supported by empirical evidence from this study population using the following tests. If the risk calculator's null effect on the decision to recommend an operation was due to the simultaneous effect on judgments for all four risk/benefit parameters, then it would be expected that using these judgments to predict surgeons' decisions (instead of looking directly at the decisions themselves), surgeons in the risk calculator group would have a similar predicted probability of recommending an operation as surgeons in the control group (null group). Using the regression coefficients from regression 4 (which models the decision to recommend an operation from the perceived absolute risks and benefits of operative and non-operative management), surgeons in the null group would have a 74.0% (95% CI 70.0, 78.2) predicted probability of recommending an operation across the four vignettes. If the calculator had only influenced surgeons' judgment of operative risk, then the predicted probability that a surgeon would recommend an operation would be 82.0% (95% CI

78.1, 86.0)—a slightly higher likelihood of recommending an operation. However, because the risk calculator influenced all four risk/benefit judgments, the actual predicted probability for the surgeons in the risk calculator group was 78.1% (95% CI 74.0, 82.1), similar to the 74.0% of the control group (with overlapping confidence intervals).

One final explanation for the null effect of the risk calculator on the decision to recommend an operation is the potential for reverse causality. That is, it could be 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. It is thus possible that surgeons (perhaps non-consciously) modified their judgments of risks and benefits for both operative and non-operative management to be consistent with the intuitive decision that they had made based on the vignette alone before seeing the 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 non-operative risks and benefits, thereby leaving the relative favorability of operating unchanged. However, somewhat diminishing the concern for reverse causality is the finding that correlations between judgments of risks and benefits of operating versus not operating are relatively low (median Spearman correlation for operative versus non-operative risks = .34; median Spearman correlation for operative versus non-operative benefits = .20). These correlations appear to mitigate some, though not all, of the concern for reverse causality.

Further exploring the role of affect in surgical decision-making, it is worth noting that the overall effects of the risk calculator might have been different if surgeons had first been asked to consider their own estimates of risks before they were given the risk calculator estimates. In one study, patients were randomized to either estimate an average woman's lifetime risk of getting

breast cancer before being told the risk is 13% or they were simply told the risk estimate without making their own prediction.[114] Because participants in the former group estimated the risk to be, on average, 46%, they tended to feel more relieved, and perceived the true risk of 13% to be lower than the participants who had not come up with their own risk estimate. These findings are applicable here because had the surgeons examined in this dissertation first been asked to make risk estimates before learning the risk calculator value, it is possible that the risk calculator value would have been perceived as even lower, possibly making them feel a greater sense of relief and, as a result, be more likely to recommend an operation.

Effect by prior use of calculator

According to the conceptual model, any improvement in the accuracy of surgeons' predictions conferred by use of the risk calculator would be moderated by how much they believe the reliability and accuracy of the data provided. As a measure of this concept, it might be assumed that those surgeons who were already aware of the calculator and those who already used it to some degree in their clinical practice would also be the surgeons who most readily believe in the value of the tool and therefore would also be more responsive to its estimates. In fact, however, evidence for the opposite effect was found. The risk calculator conferred the greatest prediction skill to those surgeons who were previously unaware of the tool, while it conferred the least improvement in prediction skill (no statistically significant improvement at all) to the surgeons who were already using the calculator in their practice. There are several explanations for this finding. First, only a small number of surgeons in the study sample used the calculator routinely. The small sample size may have prevented detection of statistically significant improvement for these surgeons. Second, the previous exposure and prior use of the calculator may have had a durable effect, leaving surgeons with a lasting ability to more

accurately estimate the surgical risks for the patients described in the clinical vignettes. By this reasoning, these surgeons would have been able to draw on their more developed knowledge of operative risk without actually having access to the calculator data. On the other hand, the surgeons who were previously unaware of the tool were essentially blank slates. Without the calculator data, their prediction skill was the weakest and the calculator therefore produced the greatest improvement. One final explanation is that the surgeons who routinely used the calculator in their practice may have, without prompting, used the risk calculator while completing the study. It would be feasible while completing the online study to simultaneously enter the relevant clinical information into the ACS-NSQIP risk calculator—which is also an online tool—and report numbers aligned with the risk calculator estimates.

## 6.2 Limitations

There are several issues that should be acknowledged that may limit the interpretation of the findings in this dissertation. Broadly speaking, these threats can be organized by their threats to either the internal or external validity of the findings.

### 6.2.1 Threats to internal validity

One of the primary purposes of this dissertation is to examine the association between how surgeons perceive risk and benefit and how they make clinical decisions. Yet, inherent in this examined relationship is the threat of endogeneity caused by possible reverse causality in the relationship between perceived risk and benefit and the decision to operate. Both the conceptual model and the analysis plan used here suggest a unidirectional relationship whereby perceived risk and benefit influence the decision. However, it is also plausible that to the extent that causality exists, it may run in the reverse direction whereby surgeons make their decisions using other cues, including the way they were trained to do things, their experience, and even their gut

instinct, and then use their decision to inform their estimates of risk and benefit. In fact, some of the findings from the risk calculator analyses suggest that such reverse causality may actually be present, since exposure to data from the risk calculator did not alter surgeons' decisions but it did affect how they perceived the risks and benefits of *not* operating. However, as previously mentioned, the relatively low correlations between judged operative and non-operative risks, as well as operative and non-operative benefits, suggest that each of these parameters are relatively independent of one another, which diminishes some, though not all, of the concern for reverse causality.

Another limitation is the use of variables that may imprecisely approximate the intended measures. For example, an individual's risk attitude is known to be an important predictor for both how that individual perceives risk, as well as for how he/she responds to a given level of risk. However, I was only able to demonstrate an association between surgeons' risk attitudes and how they perceive the absolute level of operative risk, not their resulting clinical decision. These findings may be attributable to the crudeness of the instrument used to assess risk attitude. Although the same measure has been used to demonstrate an association between risk attitude and other types of physician decisions,[115] it is plausible, if not even likely, that the measure does not fully capture the dimensions of risk attitude relevant to surgical decision-making.

It should also be noted that there were no available data on the financial incentives faced by surgeons in their practice, making it necessary to use the imperfect proxy of surgeon practice type. A legitimate criticism of this measure stems from the marked heterogeneity of physician payment policies across practice type. For example, even when surgeons in academic centers are paid on salary, there is still often a component of their pay that is based on clinical productivity. Furthermore, aside from payment policies, there are several other surgeon characteristics that

may be associated both with their choice of practice type and their clinical decision-making. As just one example, academic surgeons tend to be more oriented to research and teaching, while surgeons in private practice tend to be exclusively committed to clinical responsibilities.

The variables relating to surgeon experience used in this dissertation may also be inadequate to measure the complex dimensions of experience that influence risk perception and decision-making. For example, surgeons who had previously experienced particularly negative adverse events associated with either a particular disease process or a specific type of operation may incorrectly perceive the risks for a comparable patient or operation to be much higher than they otherwise would have thought. These negative experiences are emotionally charged and therefore come to mind with greater ease than memories of less consequential events, even if the less dramatic events are far more common. This process, known as the *availability heuristic*, causes an individual to overweigh certain extreme experiences when estimating the likelihood of an event.[116] Unfortunately, data on such experiences were not available for analysis but it is unclear how the absence of those data would bias the findings presented here.

According to expected utility theory, decisions are made by multiplying the likelihood of each resulting event by the desirability of that outcome. There are therefore two parts to each outcome: the probability with which it occurs and the value attached to that outcome. However, this dissertation remains agnostic to the value assigned by each surgeon in their decision-making process and focuses only on the probabilities assigned. Much like the variation observed in both the decision to operate and in the perception of risk and benefit, it is likely that similar variation would also be found in how surgeons assign value or utility to each desired or undesired health state. However, the omission of outcome utility in these analyses is unlikely to bias the outcomes presented here.

Finally, a basic assumption of the analyses for the risk calculator is that the tool produces the most accurate and reliable estimates of a patient's operative risk. Surgeons who deviated most from the calculator estimate were considered to be most inaccurate in their predictions. However, it is also possible that the calculator produced inaccurate estimates for these particular clinical scenarios. Risk estimates, after all, are entirely dependent on what variables are included in the prediction model.[117] Therefore, inclusion of any other variables that are relevant to the particular patient's risk would only make the estimates more accurate.

### 6.2.2 Threats to external validity

Perhaps the most notable limitation is that the results presented here were based entirely on hypothetical, vignette-based scenarios, as opposed to actual clinical encounters where risk prediction and decision-making tend to occur in a more dynamic environment, in which surgeons must consider the competing clinical, social, financial, and legal implications of their decisions. Vignettes were chosen to study decision-making primarily to ensure that all patient characteristics were held constant. Studying decision-making in clinical practice is always limited by the inability to adequately measure all features of the patient that may influence a surgeons' decision or affect how they perceive the risks and benefits of different treatments. But given the evidence that vignette-based studies often reflect actual patterns of clinical care,[118,119] it is possible that the results presented here are also a reflection of surgical decision-making in practice. However, it is important to note that there are other studies that have also shown inconsistencies between clinician responses to vignettes and behavior in actual patient encounters.[120,121]

Second, because data were only available for 4 clinical vignettes, all of which represented urgent general surgery cases, these results may not generalize to elective operations, to

operations for other surgical specialties, or to operations for which there is a more obviously "correct" or appropriate treatment. Lastly, the results may not generalize to surgeons who did not participate in this study. However, given that online studies typically have low response rates, particularly among surgeons,[122] that these data included a large and mostly representative sample, and that there was little difference between the first and second wave of respondents, the results presented here are likely to be generalizable to other surgeons.[123]

## 6.3 Implications and future work

Because the decision to operate on a patient is so fundamental to the practice of surgery, the results presented in this dissertation have far-reaching implications, ranging from how clinical care is delivered, to how resources are allocated, and how patients interact with their providers when deciding what treatments to pursue.

Variation in perceived risk/benefit

The variation in how surgeons judge the risks and benefits for both operating and non-operative management (ranging from 0-100%) is substantial and is an important finding in itself. For one, the variation reveals that in the clinical scenarios studied here, surgeons do not know or at least disagree on what the actual risks and benefits of treatment are. One possible consequence of this variation is that each surgeon likely uses very different risk and benefit estimates when counseling patients during the decision-making process. Not only might this variation misinform patients, but it may also result in patients agreeing to certain treatments that they otherwise would not have if they had known the actual risks and benefits involved. Even under ideal circumstances, patient understanding of risk is often very limited,[124] thus restricting their ability to use that information to make informed clinical decisions. Patients and surgeons alike should therefore be aware of differences in surgeon judgment and should, whenever possible, seek out

objective data to inform their estimates.[7] These estimates, when communicated carefully, will enable patients to meaningfully participate in their own care and to make decisions that are congruent with their values and preferences, regardless of which surgeon is seen.[125-127] Such changes hold the potential to advance surgery into a field where treatment decisions are data driven and whenever possible, evidence-based.

Nevertheless, the variation in risk and benefit perception is not necessarily an indictment of surgeons themselves. Rather, it points to the fact that thinking about risks and benefits explicitly, as surgeons were forced to do in this study, is not yet a formalized component of routine surgical care, at least not to the point where a consensus has developed on what the true risks and benefits are. For that reason, surgeons are not yet accustomed to assigning specific values to these parameters. In fact, the ability to accurately calculate and assign such values is only a recent phenomenon, one that has become far more reliable given the availability of high-quality clinical data from ACS-NSQIP and the resulting ACS-NSQIP risk calculator.

The variation in perception of risk and benefit is given even greater importance by its significant association with surgeons' decisions to operate. The reported estimates of risk and benefit are therefore not simply random probabilities given to answer the questions posed by researchers. Rather, these estimates appear to form the surgeons' underlying impressions of how useful a treatment is and, accordingly, how likely they are to use that treatment. The results therefore provide empirical evidence for a conceptual framework that surgeons appear to use—be it explicitly or implicitly—in making treatment decisions. Surgeons choose to operate when they think operating is low risk and high benefit and choose not to operate when they think that treatment strategy is low risk and high benefit. This explanatory model may open new avenues for further evaluating surgical decision-making. Potential applications of this model range from

training surgeons to navigate uncertainty to counseling patients on treatment options. In both applications, surgeons and patients should both be encouraged to consider explicitly the risks and benefits of the available treatment options before making a clinical decision.

Risk attitude and experience

To varying degrees, both the risk attitude and the experience of a surgeon appears to play a role in how surgeons perceive treatment risks and benefits and whether they choose one particular treatment over another. One of the main implications of the related findings presented in this dissertation is that in order to develop a more comprehensive understanding of how these surgeon characteristics influence clinical care, further research is needed. For risk attitude, one of the main constraints on further research is the lack of reliable tools with which to measure this trait. While several such tools exist from the psychology literature, only few have been developed specifically for physicians, and even fewer have been adapted for surgeons.[128] Assuming a surgeon's risk attitude can reliably be measured and its role in surgical decision-making clearly defined, it is conceivable that such tools may one day be used in clinical practice. Surgeons may use these instruments to improve their understanding of their own risk tolerance or they may even be required to inform their patients about their general attitude towards risk. Just as patients are increasingly empowered to select their surgeon based on experience or complication rates, so too might patients one day seek out surgeons whose risk attitude is aligned with their own.

Risk calculator

The risk calculator's effects on surgical judgment and decision-making have several important theoretical and practical implications for its use in surgery, which may also translate to the use of risk calculators in other medical specialties. Encouragingly, the results presented here

suggest that the ACS-NSQIP risk calculator does in fact achieve its intended goal of helping surgeons judge the risks of an operation with more accuracy and less variation. This finding alone should encourage surgeons to use the risk calculator any time they are attempting to quantify a patient's operative risk.

However, the results of this dissertation also suggest that the risk calculator may not change a surgeon's clinical decision, even though it influences their perception of operative risk and differences in risk perception are significantly associated with recommending an operation. The risk calculator provides high-quality estimates for the risks of an operation but no complementary tools exist to assist surgeon judgment of other important variables, such as the benefits of operating or the risks and benefits of not operating. Lacking these data, surgeons are susceptible to rely on various cognitive heuristics, like the anchoring and affect heuristics, to inform those estimates, resulting in a clinical decision that is no different than if no risk calculator were available in the first place.

Naturally, these findings point to the need for decision-support tools that estimate all relevant parameters, not just the risks of an operation. As already mentioned, it will be challenging to construct measures for concepts such as operative benefit and perhaps even more challenging to collect reliable data to help predict non-operative outcomes. Nevertheless, overcoming these challenges will be a critical step towards ensuring that when confronted with marked clinical uncertainty, surgeons and patients both have access to reliable estimates for the risks and benefits of all available treatment options so that those estimates can inform their clinical decisions.

One further consideration regarding the risk calculator is that, in its current form, the tool only provides point estimates on the probability of certain outcomes, but it does not provide the

uncertainty surrounding those estimates (i.e. confidence intervals). Just as surgeons differ in their decision-making depending on the level of risk involved, so too might surgeons differ depending on the amount of uncertainty in outcome estimates. One study demonstrated that members of the lay public placed high value on higher levels of precision for estimates of risk and benefit when considering a hypothetical medical decision.[129] Future studies may further categorize how surgeons differ in their estimation of uncertainty and how such differences might influence their clinical decision-making. Comfort with uncertainty, much like attitude towards risk, may be as much a feature of a specific clinical encounter as it is a reflection of a deeply ingrained personality trait.[130]

Given its high predictive performance, the risk calculator is a valuable tool for counseling patients before surgery. All too often, the informed consent process in surgery can be cursory and on occasion, may not even include an explicit discussion of the relevant risks and benefits.[131,132] And since sharing such information with patients can have a significant influence on their treatment decisions,[125,126] it is critical to ensure that risk/benefit estimates are accurate and communicated in a meaningful way to patients. Routine use of the risk calculator may therefore be a useful means of facilitating accurate communication, not only by providing the most accurate risk estimates, but also by providing a formalized structure for the conversation between patient and surgeon, which will encourage such conversations to take place more frequently.

The tension between medical paternalism and patient-centered care

The variation identified in this dissertation—both in clinical decisions and the perception of risk and benefit—raises several important questions about variations in health care utilization that have implications far beyond the specialty of surgery. Is variation in care a distinct problem

that needs to be addressed? In that case, would it be desirable for all similar patients to receive identical care? If not the case, then what is an acceptable amount of variation? And finally, to the extent that care differs from one patient to the next, how and by whom should these decisions be made? Should agency belong to physicians, with their knowledge and expertise in medicine, with the patient for whom these decisions are being made, or by some third party, such as a government or insurance agency? These questions have no simple answers. But to at least frame the debate, it is helpful to consider them in the context of the evolving tension between two schools of thought in the medical community. That is the debate between medical paternalism and patient-centered care. At the root of this debate is the recognition that both the physician and the patient face an asymmetry of information, where one party has access to vital information that the other lacks. Physicians bring to the encounter an in-depth knowledge and experience in treating disease, while patients bring to the decision their own preferences and values, even if they are not immediately aware of how these preferences relate to the clinical decisions they confront.

Paternalism has deep roots in the practice of medicine, dating back to Hippocrates. Until somewhat recently, physicians themselves were the unilateral drivers of care decisions. They decided which diagnostic tests to order and which treatments to administer. On many occasions, these decisions were made without even informing patients on the details of their illness or how the treatments might affect their recovery and wellbeing. However, beginning in the 1960's, the medical community began to evolve from within in its thinking on how best to make clinical decisions, resulting in a movement, still underway, that seeks to make medical decision-making more patient-centered.

This development established a natural tension between opposing ideas, each with their own assumptions and limitations, on how treatment decisions should be made. On the one end of the spectrum, medical paternalism suggests that clinical decisions should be made based on the clinical expertise of physicians. This approach maximizes the discretion of individual physicians and allows each physician to choose a management strategy that they think will be best for the patient. Inherently, though, paternalism directly limits patient autonomy by risking the possibility of imposing a physician's values and preferences onto a patient, even if those values are not shared by the patient. On the other end of the spectrum, patient-centered care suggests that because the patient is primarily shouldering the potential risks and benefits of care, he/she should be able to directly determine what type of care is provided. However, by granting most power to patients, this approach limits (or at least redefines) the role of clinical expertise in clinical decision-making, inherently limiting the autonomy of individual providers and possibly forcing them to provide treatments that are less beneficial and possibly even harmful simply because they are desired by the patient. The debate between medical paternalism and patient-centered care thus raises the question of how to protect patient autonomy, while respecting the wealth of knowledge, experience, and expertise of individual clinicians.

These questions raised by this debate represent deep philosophical concerns that, unfortunately, cannot be answered by empirical research. Nor, for that matter, can these questions be answered in a dissertation. Rather, the answers, which have huge implications for the way medicine is practiced, will have to be wrestled with, debated, and perpetually reconsidered on both a societal and professional level. Community groups, professional organizations, governmental agencies, and health care stakeholders from across the system will

all have a vested interest in voicing their opinion on these issues and shaping the paradigm for how clinical decisions are made.

Implications for quality

Over the past three decades, there has been a strong movement in medicine—surgery in particular—to improve the quality of care delivered. The most widely used framework for understanding quality in healthcare is the Donabedian model, which divides quality into three unique domains: structure, process, and outcomes.[133] All three of these domains have been applied to the field of surgery in an attempt to improve care quality. For structure, hospitals have been encouraged to maintain a minimum volume of high-risk surgical procedures[134] and to meet certain staffing requirements to be designated as a trauma center.[135] For process, hospitals are graded based on whether patients receive the appropriate antibiotics before surgery and whether they receive appropriate venous thromboembolism prophylaxis after surgery.[136] And for outcomes, both surgeons and hospitals are increasingly held accountable for their rates of postoperative complications.[137]

As these examples demonstrate, quality measures in surgery have focused primarily on what takes place during the operation (process) or after the operation (outcomes) and have largely neglected the care that takes place leading up to the operation. Most importantly, what has thus far escaped scrutiny as a quality measure is the decision to even go to the operating room in the first place.[12] Given the substantial consequences of this decision, it is intriguing to conceive of ways in which this moment in a patient's care pathway can be evaluated and graded so that surgeons and patients alike can be encouraged to make the best possible treatment decisions. However, before the surgical decision-making process can reasonably be evaluated as

a quality measure, it will first be necessary to distinguish between a high-quality and a low-quality decision.

Typically, when variation is identified in healthcare that is not attributable to differences in clinical variables, it is considered a reflection of inappropriate care. Variation exists, by this reasoning, because of overuse at the high end of the spectrum, underuse at the low end of the spectrum, and possible misuse across the entire spectrum. The challenge in translating these concerns into practical interventions is determining what exactly constitutes appropriate care. Once that is determined, each health care provider or each hospital can be judged on their use of health care—be it a diagnostic test or a medical procedure—against an objective and established benchmark. When actual use matches the predetermined criteria for appropriateness, that care is deemed high quality. Conversely, when there is a mismatch between actual use and appropriate use, that care is deemed low quality.

However, applying these concepts to the variation in decision-making identified in this dissertation is not possible because for all four of the studied clinical vignettes, no consensus exists on what constitutes appropriate care. That is likely because it is very difficult to find such consensus among experts when there is such great uncertainty on the potential outcomes for different treatment options. Even more discouraging is that it is quite possible that scenarios with high levels of uncertainty—the discretionary gray areas of medicine—represent the rule rather than the exception.[1] Of all the possible clinical scenarios in which surgery is a possible treatment, surgeons may disagree with each other on the appropriate use of surgery far more than they agree. If that is the case, it is probably unreasonable to expect robust appropriateness criteria to be developed for every possible clinical scenario and without such criteria, it will not be possible to judge whether or not a decision was high or low quality.

However, there may be another way to judge the quality of a decision, one that depends less on the actual decision made and more on *how* the decision was made. Much like Donabedian distinguishes between the process and the outcome of healthcare, it is possible to distinguish between the process and the outcome of a surgical decision. As already established, it may not be possible to judge the quality of a decision based on its outcome (the actual decision made) because it will never be possible to know what might have happened had another decision been made (the counterfactual). However, the results of this dissertation suggest that a surgical decision can in fact be broken up into its corresponding parts and perhaps by analyzing each of these parts individually, it will be possible to judge one decision-making process as being better or worse than another.

There are several possible ways in which the process of a decision can be judged, many of which are already used in decision-making research, but none of which have been adopted as a quality measure. The most commonly used measures of quality for the process of decision-making include patients' self-reports of their satisfaction with decision-making, the nature of the interaction between the provider and the patient, and the patient's state, as measured by factors such as their level of understanding, their knowledge of the decision content, or their eventual comfort with the decision made.[138] Yet another way to measure the quality of the decision-making process is to evaluate its patient-centeredness, as in how well the decision-making reflects the needs, values, and expressed preferences of a well-informed patient.[139]

The content of this dissertation points to another possible way to measure the process of decision-making, one that relies not just on patients' and physicians' subjective assessments but also on an objective measure of the accuracy of the information used to inform the decision. When it comes to predicting outcomes, some estimates are better than others. It is simply not

reasonable to claim that two surgeons who predict the likelihood of a surgical complication to be 80% and 20% are equally correct. One prediction is inevitably more accurate than the other. This point is not always intuitively obvious because there are many experts across a wide range of fields who doubt the possibility of predicting discrete and rare events. However, the work of Phillip Tetlock and his colleagues has done much to disprove these doubters and in doing so, has also revealed a possible way to measure surgical decision-making quality.[140]

Tetlock has shown that some people are inherently more skilled than others at predicting the likelihood of future events. By assembling teams of so-called "superforecasters", these researchers were able to predict with remarkable accuracy the likelihood of events as varied as North Korea detonating a nuclear weapon in the next 3 years, the number of refugees fleeing Syria, and the speed of growth of China's economy.[140] The ability for some individuals to forecast events better than others should come as no surprise. Already in surgery, we grade individuals based on their knowledge and technical performance. Much like one surgeon can be better than another at doing a laparoscopic gastric bypass operation,[141] so too can one surgeon be better at predicting the likelihood of risks and benefits associated with different treatments.

In this dissertation, I demonstrated a range of prediction skill among surgeons for four clinical scenarios. Assuming the accuracy of the risk calculator, there were some surgeons who predicted events with near-perfect accuracy, while others were off the mark by an order of magnitude. This variation in prediction skill may offer a useful way to judge the quality of the decision-making process. Of course, this measure of quality depends on a gold standard in prediction but tools like the ACS-NSQIP risk calculator and its eventual iterations offer what is likely the best possible estimates available. Implementing such a quality measure will only further encourage surgeons to use these risk prediction tools in their practice. In fact, use of the

tool could itself be considered a supplemental measure of the quality of the decision-making process.

There are then several different possibilities for how to measure the quality of a surgical decision. The best of these approaches will have to be determined by subsequent research and refined by further testing. But however it is measured, it will be important to recognize that the decision to recommend an operation is in and of itself a critical event in the care of a surgical patient and for that reason, that decision, perhaps above all others, deserves intense scrutiny to ensure that surgical decisions, whatever their outcome, are made in accordance with each patient's preferences and values.

# 7 Appendix

**Appendix 1 Comparison of first-wave respondents, second-wave respondents, and the general population of surgeons**

| | 1st wave* (n=1,187) | 2nd wave* (n=693) | p-value | Full study sample (1,880) | General population° (n=48,635) | p-value§ |
|---|---|---|---|---|---|---|
| Gender | | | 0.26 | | | <0.01 |
| Male | 72.5 | 74.9 | | 73.4 | 81.5 | |
| Female | 27.5 | 25.1 | | 26.6 | 18.5 | |
| Race | | | 0.49 | | | <0.01 |
| White | 70.6 | 71.9 | | 71.1 | 52.4 | |
| Asian | 13.0 | 13.6 | | 13.2 | 11.4 | |
| Hispanic/Latino | 8.4 | 9.3 | | 8.7 | 4.8 | |
| Black | 2.4 | 1.5 | | 2.1 | 4.4 | |
| Hawaiian/PI | 0.4 | 0.2 | | 0.3 | – | |
| American Indian/Alaskan Native | 0.2 | 0.2 | | 0.2 | 0.4 | |
| Other | 5.0 | 3.5 | | 4.4 | 0.4 | |
| Missing | – | – | | – | 26.3 | |
| Graduation year | | | 0.47 | | | |
| 2010 or later | 17.3 | 15.0 | | 16.4 | – | |
| 2000-2010 | 23.5 | 22.8 | | 23.2 | – | |
| Before 2000 | 44.7 | 48.0 | | 45.9 | – | |
| Resident | 14.6 | 14.2 | | 14.4 | – | |
| Practice type | | | <0.01 | | | |
| Private | 31.8 | 26.8 | | 29.9 | – | |
| Community | 25.0 | 27.9 | | 26.1 | – | |
| Academic | 36.5 | 33.6 | | 35.4 | – | |
| VA | 2.4 | 1.9 | | 2.2 | – | |
| County | 2.1 | 2.0 | | 2.1 | – | |
| Military | 1.2 | 0.9 | | 3.3 | – | |
| Other | 1.1 | 7.0 | | 1.1 | – | |

Abbreviations: VA-Veterans Affairs; PI-Pacific Islander

° Data on general population of surgeons obtained from the Association for American Medical Colleges.

* 1st wave refers to surgeons who responded to the first email; 2nd wave refers to surgeons who responded to the reminder email.

° Data on general population of surgeons obtained from the Association for American Medical Colleges.

§ p-value for comparison of all study respondent (1st and 2nd wave) to general population.

**Appendix 2. Full questionnaire from online study**

# Let's get started!

Thank you for participating in our study, which aims to describe surgeons' preferences for operative or non-operative management in various clinical conditions. The study is a collaboration between the UCLA Robert Wood Johnson Clinical Scholars program, the UCLA Center for Surgical Outcomes and Quality (CSOQ), and the American College of Surgeons.

Please note:

This survey will take approximately **10-15 minutes** to complete.
Participation is **voluntary**. Clicking on the next button below implies that you have read these instructions and voluntarily agree to participate.
The survey is ***confidential***.
You will have the option to provide your email address for the **CHANCE TO WIN** 1 of 4 MacBook Air computers.
You will also have the option, at the end of the study, to complete brief self-assessment questions to **EARN CME CREDIT** (up to 1.25 *AMA PRA Category 1 Credits™*).
Aggregate results that do not permit identification of any individual respondent may subsequently be published or presented.

If you have any questions or concerns about your rights as a research subject, please contact:

The Human Subjects Protection Committee at RAND, 1776 Main Street, PO Box 2138, Santa Monica, CA 90407, 310-393-0411 x6369.

Background questions

Would you like to be enrolled in a drawing to **win 1 of 4 MacBook Air** computers?
- Yes
- No

Have you **completed** a general surgery residency?
- Yes
- No

Which of the following best describes your surgical practice (select one)?
- Private practice
- Community Hospital
- Academic Medical Center
- Veterans Affairs hospital
- County hospital
- Other

Have you completed or are you currently enrolled in any of the following clinical fellowships (select one)?

- No fellowship
- Acute care surgery
- Bariatrics/minimally invasive surgery
- Breast
- Cardiac
- Colorectal
- Endocrine
- Hepatobiliary
- Pediatric surgery
- Surgical Oncology
- Thoracic
- Transplant
- Trauma
- Vascular
- Other

Please enter the zip code for the hospital you work at most often (if in USA). If not in USA, please enter name of country.

In your clinical practice, do you take care of patients with emergency general surgery conditions (small bowel obstruction, appendicitis, mesenteric ischemia, etc.)?
- Routinely
- Occasionally
- Rarely

How many operations do you typically perform each year?
- <50
- 50-100
- 101-250
- 251-500
- <500

Gender
- Male
- Female

Race/ethnicity (select all that apply)
- American Indian or Alaska Native
- Asian
- Black or African American
- Hispanic or Latino
- Native Hawaiian or Pacific Islander
- White
- Other

Are you currently **enrolled** in a general surgery residency?
- Yes
- No

# Instructions (please read)

The following **FOUR clinical vignettes** represent surgical scenarios. Each case is followed by **questions** in which you will be asked to predict the likelihood of positive and negative outcomes. The scenarios will be followed by a few brief questions. Please **consider** yourself as the operating surgeon and primary decision-maker. Unless otherwise specified, **assume** that the patient and his/her family defer to your recommendations. The following **30 day**-outcome definitions are available within the survey by **hovering** the mouse over the blue words. Give it a try now... Serious complication (30-day) Death (30-day) Recovery (30-day)

*****Keep in mind that a patient can have a serious complication (30-day) AND meet criteria for recovery (30-day). Therefore, when asked to predict the probability of these events, the numbers are not required to add up to 100%*****

[Serious complication (30-day): At least one of the following within 30 days of making 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 if operation performed: return to the operating room, deep incisional or organ space surgical site infection, or wound disruption (consistent with American College of Surgeons National Surgical Quality Improvement Project).

Death (30-day): Death within 30 days of making the decision to operate or not operate.

Recovery (30-day): The patient is free of the immediate threats of the surgical disease process in question and is back to a reasonable level of baseline health (within 30 days of making the decision to operate or not operate)]

**Vignettes (risk calculator version shown)**

# NEW CASE: Question 1 of 5

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

**[The following questions were asked similarly for all vignettes, with minor alterations to suit clinical scenario]**

## RISK CALCULATOR DATA

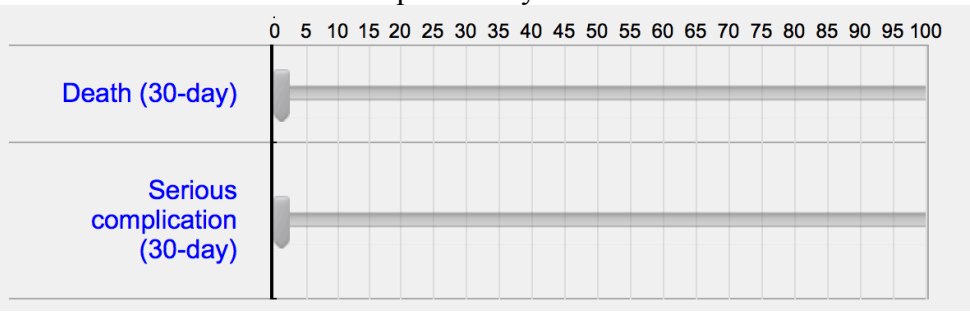
The American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) has developed a risk calculator that estimates the chance of an unfavorable outcome (such as a complication or death) within 30 days after surgery. The estimates are calculated using data from a large number of 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%

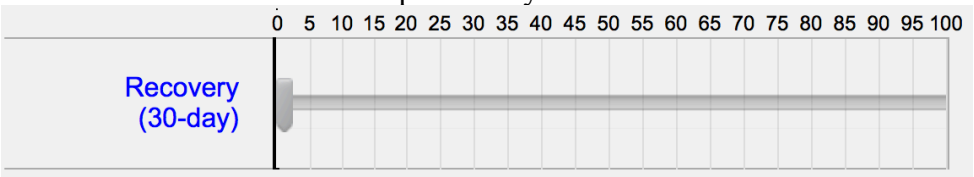
**OPERATIVE SCENARIO**

Given the ACS-NSQIP risk calculator data as well as your own clinical judgment, if you **were** to operate, what is the probability of the following outcomes?
Slide the bar below to match a probability from 0% to 100% for each



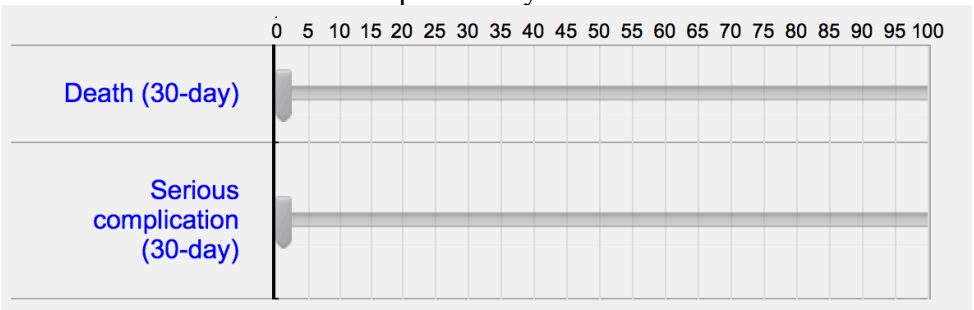

**OPERATIVE SCENARIO**

If you **were** to operate, what is the probability of the following outcome?
Slide the bar below to match a probability from 0% to 100%



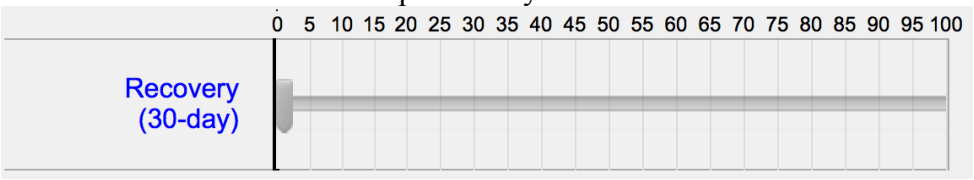

**NON-OPERATIVE SCENARIO**

If you decide to **NOT** operate and instead pursue medical management (including fluid resuscitation, antibiotics, etc.), what is the probability of the following outcomes?
Slide the bar below to match a probability from 0% to 100% for each




**NON-OPERATIVE SCENARIO**

If you decide to **NOT** operate and instead pursue medical management (including fluid resuscitation, antibiotics, etc.), what is the probability of the following outcome?
Slide the bar below to match a probability from 0% to 100%




How likely are you to **recommend an exploratory surgery** (laparoscopic or open) at this time, as opposed to pursuing medical management, including fluid resuscitation and antibiotics?

- Very unlikely
- Unlikely
- Neutral
- Likely
- Very likely

# **NEW CASE**: Question 1 of 5

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** 7.0g/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

**[Subsequent questions similar to first case]**

# **NEW CASE**: Question 1 of 5

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.

**[Subsequent questions similar to first case]**

**NEW CASE**: Question 1 of 5 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.

**[Subsequent questions similar to first case]**

Before taking this survey, were you aware of the ACS-NSQIP risk calculator?

- No
- Yes

Do you use the ACS-NSQIP risk calculator in your surgical practice?

- Never
- Use it occasionally
- Use it routinely

Please rate these statements on the scale very unlikely, unlikely, neutral, likely, very likely.

- People have told me that I seem to enjoy taking chances in my surgical practice.
- In my surgical practice, I try to avoid situations that have uncertain outcomes.
- Taking risks in my surgical practice does not bother me if the potential for patient improvement is high.
- I enjoy taking risks in my surgical practice.
- I rarely, if ever, take risks in my surgical practice when there is another alternative.
- I consider security an important element in every aspect of my life.

**Appendix 3. Comparison of demographic and clinical characteristics across randomly assigned study arms**

| | Null group | Risk calculator | Reverse order | Operate only | Non-operative only | p-value |
|---|---|---|---|---|---|---|
| Attendings, n | 327 | 346 | 319 | 311 | 301 | 0.56 |
| Residents, n | 57 | 49 | 61 | 47 | 55 | |
| Gender, % | | | | | | |
| Female | 19.3 | 22.7 | 20.6 | 20.1 | 17.2 | 0.81 |
| Male | 20.9 | 21.0 | 19.7 | 19.2 | 19.2 | |
| Race, % | | | | | | |
| American Indian/Alaskan Native | 28.6 | 0.0 | 14.3 | 42.9 | 14.3 | 0.41 |
| Asian | 22.1 | 17.8 | 16.7 | 21.0 | 22.5 | 0.14 |
| Black | 15.9 | 31.8 | 15.9 | 15.9 | 20.5 | 0.46 |
| Hawaiian/PI | 20.0 | 0.0 | 10.0 | 30.0 | 40.0 | 0.23 |
| Other | 12.2 | 22.0 | 14.6 | 24.4 | 26.8 | 0.08 |
| White | 20.4 | 21.7 | 21.2 | 18.6 | 18.2 | 0.30 |
| Hispanic/Latino | 23.5 | 22.2 | 21.6 | 16.7 | 16.1 | 0.66 |
| Practice general surgery, % | | | | | | 0.54 |
| Routine | 21.3 | 20.3 | 19.6 | 18.9 | 19.9 | |
| Occasional | 18.2 | 22.9 | 22.9 | 19.1 | 17.0 | |
| Rarely | 19.4 | 23.4 | 20.3 | 21.6 | 15.4 | |
| Cases per year, % | | | | | | 0.73 |
| <50 | 21.7 | 23.3 | 13.3 | 26.7 | 15.0 | |
| 51-100 | 23.0 | 19.1 | 19.8 | 21.4 | 16.7 | |
| 101-250 | 18.4 | 22.1 | 22.3 | 19.4 | 17.8 | |
| 251-500 | 22.6 | 20.8 | 19.0 | 18.3 | 19.4 | |
| >500 | 16.5 | 22.6 | 18.9 | 20.8 | 21.2 | |
| Aware of calculator, % | | | | | | 0.27 |
| No | 20.3 | 19.3 | 20.7 | 21.1 | 18.6 | |
| Yes | 20.0 | 22.7 | 18.8 | 18.4 | 20.1 | |
| Use calculator, % | | | | | | 0.97 |
| Never | 21.1 | 22.2 | 17.9 | 17.7 | 21.1 | |
| Occasional | 18.7 | 23.3 | 19.2 | 19.6 | 19.2 | |
| Routinely | 21.4 | 21.4 | 21.4 | 16.1 | 19.6 | |

p-value calculated using chi-square test

**Appendix 4 Comparison of listwise deletion and multiple imputation for regression 1**

| | | Listwise deletion | Multiple imputation |
|---|---|---|---|
| Relative risk | b | -0.01584038 | -0.01509076 |
| | se | 0.01325353 | 0.01224219 |
| | p | 0.232 | 0.2177 |
| Relative benefit | b | 0.05647114 | 0.03666889 |
| | se | 0.01174957 | 0.01091939 |
| | p | 0 | 0.0008 |
| Relative risk # Relative benefit | b | 0.00042914 | 0.00023295 |
| | se | 0.00007726 | 0.00006847 |
| | p | 0 | 0.0007 |
| Risk attitude | b | 0.01754101 | 0.00305176 |
| | se | 0.01347542 | 0.01243758 |
| | p | 0.193 | 0.8062 |
| Risk attitude # Relative risk | b | -0.00083337 | -0.00079186 |
| | se | 0.00054398 | 0.00046503 |
| | p | 0.1255 | 0.0886 |
| Risk attitude # Relative benefit | b | -0.00114986 | -0.00048235 |
| | se | 0.00046924 | 0.00042662 |
| | p | 0.0143 | 0.2582 |
| Private practice # Relative risk | b | 0.00046052 | 0.00169032 |
| | se | 0.00573896 | 0.00482515 |
| | p | 0.936 | 0.7261 |
| Private practice # Relative benefit | b | -0.00570266 | -0.00613671 |
| | se | 0.00446855 | 0.00386796 |
| | p | 0.2019 | 0.1126 |
| Local practice culture | b | 0.06246919 | 0.02501949 |
| | se | 0.08689279 | 0.0857633 |
| | p | 0.4722 | 0.7705 |
| Local practice culture # Relative risk | b | -0.00705581 | -0.0046516 |
| | se | 0.00353566 | 0.00333905 |
| | p | 0.046 | 0.1636 |
| Local practice culture # Relative benefit | b | 0.00024775 | 0.00149019 |
| | se | 0.00315376 | 0.00302172 |
| | p | 0.9374 | 0.6219 |

Abbreviations: # interaction between 2 variables, b log odds, se standard error, p p-value.
Estimates calculated using hierarchical logistic regression models. Multiple imputation performed using 50 imputations.

**Appendix 5. Association between surgeon perception of relative risks and benefits of operating and the decision to recommend an operation**

| Risk/benefit parameter* | 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 |
|---|---|---|---|---|
| Relative risk (at relative benefit μ-sd) | 89.1 | 24.0 | -65.1 | (-71.1, -59.1) |
| Relative risk (at relative benefit μ+sd) | 95.7 | 85.4 | -10.2 | (-15, -5.4) |
| Relative benefit (at relative risk μ-sd) | 89.1 | 95.7 | 6.5 | (2.9, 10.2) |
| Relative benefit (at relative risk μ+sd) | 24.0 | 85.4 | 61.4 | (54.7, 68.2) |

* Because of significant interaction between relative risk and relative benefit, effect sizes are calculated twice. First the effect is calculated while holding the other parameter constant at 1 standard deviation below the mean and then while holding the other parameter constant at 1 standard deviation above the mean.

# 8 References


1. Birkmeyer JD, Reames BN, McCulloch P, Carr AJ, Campbell WB, Wennberg JE. Understanding of regional variation in the use of surgery. *Lancet*. 2013;382(9898):1121–1129.

2. The Dartmouth Institute of Health Policy, Practice C. *The Dartmouth Atlas of Healthcare*. Hanover, NH: Dartmouth Medical School; 2013.

3. National Center for Health Statistics. Centers for Disease Control and Prevention. [CDC web site]. Available at: http://www.cdc.gov/nchs/fastats/inpatient-surgery.htm#. Accessed December 12, 2015.

4. Leape LL, Park RE, Solomon DH, Chassin MR, Kosecoff J, Brook RH. Does inappropriate use explain small-area variations in the use of health care services? *JAMA*. 1990;263(5):669–672.

5. Neuman MD, Bosk CL. What We Talk about When We Talk about Risk: Refining Surgery's Hazards in Medical Thought. *Milbank Quarterly*. 2012;90(1):135–159.

6. Moyer CA. The Assessment of Operative Risk. In: Moyer CA, Rhoads JE, Allen JG, Harkins HN, eds. *Surgery: Principles and Practice*. 4 ed. Philadelphia; 1970.

7. 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(5):833–42.e1–3.

8. Ramanan B, Gupta PK, Gupta H, Fang X, Forse RA. Development and validation of a bariatric surgery mortality risk calculator. *J Am Coll Surg*. 2012;214(6):892–900.

9. Cohen ME, Bilimoria KY, Ko CY, Hall BL. 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(6):1009–1016.

10. 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(1):39–52.

11. 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(25 Suppl 2):S49–73.

12. Krumholz HM. Variations in health care, patient preferences, and high-quality decision making. *JAMA: The Journal of the American Medical Association*. 2013;310(2):151–152.

13. Institute of Medicine. *Crossing the Quality Chasm: a New Health System for the Twenty-First Century*. Washington: National Academy Press; 2001.

14. Rutkow IM. The surgical decision-making process: determinants of surgical rates. *Health Services Research*. 1982;17(4):379–385.

15. Chand M, Armstrong T, Britton G, Nash GF. How and why do we measure surgical risk? *Journal of the Royal Society of Medicine*. 2007;100(11):508–512.

16. Kahneman D. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux; 2011.

17. Keeney RL, Raiffa H. *Decisions with Multiple Objectives*. Cambridge University Press; 1993.

18. Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science*. 1959;130(3366):9–21.

19. Henschke UK, Flehinger BJ. Decision theory in cancer therapy. *Cancer*. 1967;20(11):1819–1826.

20. Services USDOHAH, Agency for Healthcare Research and Quality. *Decision and Simulation Modeling in Systematic Reviews*. Createspace Independent Pub; 2013.

21. Cronenwett JL, Birkmeyer JD, Nackman GB, et al. Cost-effectiveness of carotid endarterectomy in asymptomatic patients. *J Vasc Surg*. 1997;25(2):298–309– discussion 310–1.

22. Stroupe KT, Morrison DA, Hlatky MA, et al. Cost-effectiveness of coronary artery bypass grafts versus percutaneous coronary intervention for revascularization of high-risk patients. *Circulation*. 2006;114(12):1251–1257.

23. Pauly MV. What Is Unnecessary Surgery? *The Milbank Memorial Fund Quarterly Health and Society*. 1979;57(1):95–117.

24. Eddy DM. Clinical decision making: from theory to practice. Anatomy of a decision. *JAMA*. 1990;263(3):441–443.

25. Loewenstein GF, Weber EU, Hsee CK, Welch N. Risk as feelings. *Psychol Bull*. 2001;127(2):267–286.

26. Yule S, Flin R, Paterson-Brown S, Maran N. Non-technical skills for surgeons in the operating room: a review of the literature. *Surgery*. 2006;139(2):140–149.

27. 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(2):298–304.

28. Waterworth PD, Soon SY, Govindraj R, Sivaprakasam R, Jackson M, Grayson AD. Factors which influence the cardiac surgeon's decision not to operate on patients referred for consideration of surgery. *J Cardiothorac Surg*. 2008;3(1):9.

29. Orri M, Farges O, Clavien P-A, Barkun J, Revah-Lévy A. Being a surgeon--the myth and the reality: a meta-synthesis of surgeons' perspectives about factors affecting their practice and well-being. *Ann Surg*. 2014;260(5):721–8– discussion 728–9.

30. Walton NA, Martin DK, Peter EH, Pringle DM, Singer PA. Priority setting and cardiac surgery: a qualitative case study. *Health Policy*. 2007;80(3):444–458.

31. Gagliardi AR, Wright FC, Anderson MAB, Davis D. The role of collegial interaction in continuing professional development. *J Contin Educ Health Prof*. 2007;27(4):214–219.

32. Leaper DJ, Horrocks JC, Staniland JR, de Dombal FT. Computer-assisted diagnosis of abdominal pain using "estimates" provided by clinicians. *British Medical Journal*. 1972;4(5836):350–354.

33. Bakwin H. Pseudodoxia Pediatrica. *N Engl J Med*. 1945;232(24):691–697.

34. Weinstein JN, Bronner KK, Morgan TS, Wennberg JE. Trends and geographic variations in major surgery for degenerative diseases of the hip, knee, and spine. *Health Aff (Millwood)*. 2004;Suppl Variation:VAR81–9.

35. McPherson K, Wennberg JE, Hovind OB, Clifford P. Small-area variations in the use of common surgical procedures: an international comparison of New England, England, and Norway. *N Engl J Med*. 1982;307(21):1310–1314.

36. Chassin MR, Kosecoff J, Park RE, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *JAMA*. 1987;258(18):2533–2537.

37. Miller DC, Saigal CS, Banerjee M, Hanley J, Litwin MS, Urologic Diseases in America Project. Diffusion of surgical innovation among patients with kidney cancer. *Cancer*. 2008;112(8):1708–1717.

38. Glover JA. The Incidence of Tonsillectomy in School children. *International Journal of Epidemiology*. 2008;37(1):9–19.

39. Matlock DD, Groeneveld PW, Sidney S, et al. Geographic variation in cardiovascular procedure use among Medicare fee-for-service vs Medicare Advantage beneficiaries. *JAMA: The Journal of the American Medical Association*. 2013;310(2):155–162.

40. Rutkow IM, Starfield BH. Surgical Decision Making and Operative Rates. 1984;119(8):899–905.

41. Weeks WB, Paraponaris A, Ventelou B. Geographic variation in rates of common surgical procedures in France in 2008–2010, and comparison to the US and Britain. *Health Policy*. September 2014.

42. Keyhani S, Falk R, Bishop T, Howell E, Korenstein D. The relationship between geographic variations and overuse of healthcare services: a systematic review. *Med Care*.

2012;50(3):257–261.

43. Roos N, LL R Jr, Henteleff P. Elective surgical rates--do high rates mean lower standards? Tonsillectomy and adenoidectomy in Manitoba. *N Engl J Med*. 1977;297(7):360–365.

44. Shao Y-H, Albertsen PC, Shih W, Roberts CB, Lu-Yao GL. The impact of PSA testing frequency on prostate cancer incidence and treatment in older men. *Prostate Cancer Prostatic Dis*. 2011;14(4):332–339.

45. Lucas FL, Siewers AE, Malenka DJ, Wennberg DE. Diagnostic-therapeutic cascade revisited: coronary angiography, coronary artery bypass graft surgery, and percutaneous coronary intervention in the modern era. *Circulation*. 2008;118(25):2797–2802.

46. Hawker GA, Wright JG, Coyte PC, et al. Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences. *Med Care*. 2001;39(3):206–216.

47. McClellan M, Brook RH. Appropriateness of care. A comparison of global and outcome methods to set standards. *Med Care*. 1992;30(7):565–586.

48. Winslow CM, Solomon DH, Chassin MR, Kosecoff J, Merrick NJ, Brook RH. The appropriateness of carotid endarterectomy. *N Engl J Med*. 1988;318(12):721–727.

49. Pilpel D, Fraser GM, Kosecoff J, Weitzman S, Brook RH. Regional differences in appropriateness of cholecystectomy in a prepaid health insurance system. *Public Health Rev*. 1992;20(1-2):61–74.

50. Brook RH. Assessing the appropriateness of care--its time has come. *JAMA: The Journal of the American Medical Association*. 2009;302(9):997–998.

51. Rutkow IM, Gittelsohn AM, Zuidema GD. Surgical decision making. The reliability of clinical judgment. *Ann Surg*. 1979;190(3):409–419.

52. Rutkow IM. Surgical decision making: the reproducibility of clinical judgment. *Archives of Surgery*. 1982.

53. Wilson NP, Wilson FP, Neuman M, et al. Determinants of surgical decision making: a national survey. *Am J Surg*. 2013;206(6):970–7– discussion 977–8.

54. Eddy DM. The Challenge. *JAMA*. 1990;263(2):287.

55. Oddone EZ, Samsa G, Matchar DB. Global Judgments versus Decision-model-facilitated Judgments Are Experts Internally Consistent? *Med Decis Making*. 1994;14(1):19–26.

56. MacCormick AD, Parry BR. Judgment analysis of surgeons' prioritization of patients for elective general surgery. *Med Decis Making*. 2006;26(3):255–264.

57. Eddy DM, Billings J. The quality of medical evidence: implications for quality of care.

*Health Aff (Millwood)*. 1988;7(1):19–32.

58. Hobson SA, Sutton CD, Garcea G, Thomas WM. Prospective comparison of POSSUM and P-POSSUM with clinical assessment of mortality following emergency surgery. *Acta Anaesthesiol Scand*. 2007;51(1):94–100.

59. Hartley MN, Sagar PM. The surgeon's 'gut feeling' as a predictor of post-operative outcome. *Annals of The Royal College of Surgeons of England*. 1994;76(6 Suppl):277–278.

60. Markus PM, Martell J, Leister I, Horstmann O, Brinker J, Becker H. Predicting postoperative morbidity by clinical assessment. *Br J Surg*. 2005;92(1):101–106.

61. Pettigrew RA, Hill GL. Indicators of surgical risk and clinical judgement. *British Journal of Surgery*. 1986;73(1):47–51.

62. Slovic P, Finucane ML, Peters E, MacGregor DG. The affect heuristic. *European Journal of Operational Research*. 2007;177(3):1333–1352.

63. Strack F, Mussweiler T. Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *J Pers Soc Psychol*. 1997;73(3):437–446.

64. Croskerry P. The importance of cognitive errors in diagnosis and strategies to minimize them. *Academic Medicine*. 2003;78(8):775–780.

65. Jacklin R, Sevdalis N, Darzi A, Vincent CA. Efficacy of cognitive feedback in improving operative risk estimation. *Am J Surg*. 2009;197(1):76–81.

66. Dale W, Hemmerich J, Ghini EA, Schwarze ML. Can induced anxiety from a negative earlier experience influence vascular surgeons' statistical decision-making? A randomized field experiment with an abdominal aortic aneurysm analog. *J Am Coll Surg*. 2006;203(5):642–652.

67. Finucane ML, Alhakami A, Slovic P, Johnson SM. The affect heuristic in judgments of risks and benefits. *J Behav Decis Making*. 2000;13:1–17.

68. Dreman D, Johnson S, MacGregor D, Slovic P. A Report on the March 2001 Investor Sentiment Survey. *Journal of Psychology and Financial Markets*. 2001;2(3):126–134.

69. Ganzach Y. Judging Risk and Return of Financial Assets. *Organizational Behavior and Human Decision Processes*. 2000;83(2):353–370.

70. Reyna VF. A theory of medical decision making and health: fuzzy trace theory. *Med Decis Making*. 2008;28(6):850–865.

71. 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(8):618–626.

72. Caverly TJ, Prochazka AV, Combs BP, et al. Doctors and numbers: an assessment of the

critical risk interpretation test. *Medical Decision Making*. 2015;35(4):512–524.

73. Redelmeier DA, Rozin P, Kahneman D. Understanding patients' decisions. Cognitive and emotional perspectives. *JAMA*. 1993;270(1):72–76.

74. McNeil BJ, Pauker SG, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. *N Engl J Med*. 1982;306(21):1259–1262.

75. Kahneman D, Tversky A. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*. 1979:263–291.

76. Lin GA, Dudley RA, Redberg RF. Cardiologists' use of percutaneous coronary interventions for stable coronary artery disease. *Arch Intern Med*. 2007;167(15):1604–1609.

77. Weber EU, Blais AR, Betz NE. A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *J Behav Decis Making*. 2002;15(4):263–290.

78. Nakata Y, Okuno-Fujiwara M, Goto T, Morita S. Risk attitudes of anesthesiologists and surgeons in clinical decision making with expected years of life. *J Clin Anesth*. 2000;12(2):146–150.

79. Nightingale SD. Risk preference and laboratory test selection. *J Gen Intern Med*. 1987;2(1):25–28.

80. MacDermid E, Young CJ, Young J, Solomon M. Decision-making in rectal surgery. *Colorectal Dis*. 2014;16(3):203–208.

81. Grol R, Whitfield M, De Maeseneer J, Mokkink H. Attitudes to risk taking in medical decision making among British, Dutch and Belgian general practitioners. *Br J Gen Pract*. 1990;40(333):134–136.

82. Baldwin RL, Green JW, Shaw JL, et al. Physician Risk Attitudes and Hospitalization of Infants with Bronchiolitis. *Acad Emergency Med*. 2005;12(2):142–146.

83. Arkes HR, Wortmann RL, Saville PD, Harkness AR. Hindsight bias among physicians weighing the likelihood of diagnoses. *J Appl Psychol*. 1981;66(2):252–254.

84. Redelmeier DA, Shafir E. Medical decision making in situations that offer multiple alternatives. *JAMA*. 1995;273(4):302–305.

85. Berner ES, Graber ML. Overconfidence as a cause of diagnostic error in medicine. *Am J Med*. 2008;121(5 Suppl):S2–23.

86. Friedman CP, Gatti GG, Franz TM, et al. Do physicians know when their diagnoses are correct? Implications for decision support and error reduction. *J Gen Intern Med*. 2005;20(4):334–339.

87. Barber BM, Odean T. Trading is hazardous to your wealth: The common stock investment

performance of individual investors. *The Journal of Finance*. 2000;55(2):773–806.

88. Jin CJ, Martimianakis MA, Kitto S, Moulton C-AE. Pressures to "Measure Up" in Surgery. *Ann Surg*. 2012;256(6):989–993.

89. Bogardus ST, Holmboe E, Jekel JF. Perils, pitfalls, and possibilities in talking about medical risk. *JAMA*. 1999;281(11):1037–1041.

90. Larsson G, Weibull H, Larsson BW. Analysis of the decision-making process leading to appendectomy: a grounded theory study. *Scand J Psychol*. 2004;45(5):449–454.

91. Irwin ZN, Hilibrand A, Gustavel M, et al. Variation in surgical decision making for degenerative spinal disorders. Part II: cervical spine. *Spine*. 2005;30(19):2214–2219.

92. LoGerfo JP, Efird RA, Diehr PK, Richardson WC. Rates of surgical care in prepaid group practices and the independent setting: what are the reasons for the differences? *Med Care*. 1979;17(1):1–10.

93. Dhupar R, Evankovich J, Klune JR, Vargas LG, Hughes SJ. Delayed operating room availability significantly impacts the total hospital costs of an urgent surgical procedure. *Surgery*. 2011;150(2):299–305.

94. Bauer AS, Blazar PE, Earp BE, Louie DL, Pallin DJ. Characteristics of emergency department transfers for hand surgery consultation. *Hand (N Y)*. 2013;8(1):12–16.

95. Worni M, Schudel IM, Østbye T, et al. Worse Outcomes in Patients Undergoing Urgent Surgery for Left-Sided Diverticulitis Admitted on Weekends vs Weekdays: A Population-Based Study of 31 832 Patients. 2012;147(7):649–655.

96. Hadley J, Steinberg EP, Feder J. Comparison of uninsured and privately insured hospital patients. Condition on admission, resource use, and outcome. *JAMA*. 1991;265(3):374–379.

97. *The American Association for Public Opinion Research. 2011. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 7th Edition. AAPOR.*

98. Association of American Medical Colleges. Diversity in the Physician Workforce: Facts & Figures 2014. Available at: *aamcdiversityfactsandfigures.org*. Accessed February 1, 2016.

99. Pearson SD, Goldman L, Orav EJ, et al. Triage decisions for emergency department patients with chest pain: do physicians' risk attitudes make the difference? *J Gen Intern Med*. 1995;10(10):557–564.

100. Sacks GD, Shannon EM, Dawes AJ, et al. Teamwork, communication and safety climate: a systematic review of interventions to improve surgical culture. *BMJ Qual Saf*. 2015;24(7):458–467.

101. Dimick JB, Osborne NH, Hall BL, Ko CY, Birkmeyer JD. Risk-Adjustment for Comparing Hospital Quality with Surgery: How Many Variables Are Needed? *J Am Coll Surg*.

2010;210(4):503–508.

102. Kansagara D, Englander H, Salanitro A, et al. Risk Prediction Models for Hospital Readmission: A Systematic Review. *JAMA*. 2011;306(15):1688–1698.

103. Sussman J, Vijan S, Hayward R. Using benefit-based tailored treatment to improve the use of antihypertensive medications. *Circulation*. 2013;128(21):2309–2317.

104. Schwarze ML, Brasel KJ, Mosenthal AC. Beyond 30-day mortality: aligning surgical quality with outcomes that patients value. *JAMA Surgery*. 2014;149(7):631–632.

105. 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(5):777–782.

106. Farges O, Vibert E, Cosse C, et al. "Surgeons' Intuition" Versus 'Prognostic Models'. *Ann Surg*. 2014;260(5):923–930.

107. Morse SB, Haywood JL, Goldenberg RL, Bronstein J, Nelson KG, Carlo WA. Estimation of Neonatal Outcome and Perinatal Therapy Use. *Pediatrics*. 2000;105(5):1046–1050.

108. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Risky feelings: Why a 6% risk of cancer does not always feel like 6%. *Patient Education and Counseling*. 2010;81:S87–S93.

109. Zajonc RB. Feeling and thinking: Preferences need no inferences. *American Psychologist*. 1980;35(2):151–175.

110. Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sci*. 1978;9(2):127–152.

111. Slovic P, Kraus N, Lappe H, Major M. Risk perception of prescription drugs: report on a survey in Canada. *Can J Public Health*. 1991;82(3):S15–20– S38–44.

112. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974;185(4157):1124–1131.

113. Chapman G, Johnson E. Anchoring, Activation, and the Construction of Values. *Organizational Behavior and Human Decision Processes*. 1999;79(2):115–153.

114. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. How making a risk estimate can change the feel of that risk: shifting attitudes toward breast cancer risk in a general public survey. *Patient Education and Counseling*. 2005;57(3):294–299.

115. Harrison JD, Young JM, Butow P, Salkeld G, Solomon MJ. Is it worth the risk? A systematic review of instruments that measure risk propensity for use in the health setting. *Social Science & Medicine*. 2005;60(6):1385–1396.

116. Tversky A, Kahneman D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*. 1973;5(2):207–232.

117. Sniderman AD, D'Agostino RB Sr, Pencina MJ. The Role of Physicians in the Era of Predictive Analytics. *JAMA*. 2015;314(1):25–26.

118. Peabody JW, Luck J, Glassman P, Dresselhaus TR, Lee M. Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of 3 Methods for Measuring Quality. *JAMA*. 2000;283(13):1715–1722.

119. Cutler D, Skinner J, Stern AD, Wennberg D. *Physician Beliefs and Patient Preferences: a New Look at Regional Variation in Health Care Spending*. Cambridge, MA: National Bureau of Economic Research; 2013.

120. Mohan D, Fischhoff B, Farris C, et al. Validating a vignette-based instrument to study physician decision making in trauma triage. *Medical Decision Making*. 2014;34(2):242–252.

121. Jones TV, Gerrity MS, Earp J. Written case simulations: do they predict physicians' behavior? *J Clin Epidemiol*. 1990;43(8):805–815.

122. Thoma A, Cornacchi SD, Farrokhyar F, Bhandari M, Goldsmith CH, Evidence-Based Surgery Working Group. How to assess a survey in surgery. *Can J Surg*. 2011;54(6):394–402.

123. Livingston EH, Wislar JS. Minimum response rates for survey research. *Arch Surg*. 2012;147(2):110.

124. Lloyd A, Hayes P, Bell PR, Naylor AR. The role of risk and benefit perception in informed consent for surgery. *Med Decis Making*. 2001;21(2):141–149.

125. Nabozny MJ, Kruser JM, Steffens NM, et al. Constructing High-stakes Surgical Decisions: It's Better to Die Trying. *Ann Surg*. 2016;263(1):64–70.

126. Pecanac KE, Kehler JM, Brasel KJ, et al. It's Big Surgery. *Ann Surg*. 2014;259(3):458–463.

127. Minneci PC, Mahida JB, Lodwick DL, et al. Effectiveness of Patient Choice in Nonoperative vs Surgical Management of Pediatric Uncomplicated Acute Appendicitis. *JAMA Surgery*. December 2015:1–8.

128. Tubbs EP, Elrod JAB, Flum DR. Risk taking and tolerance of uncertainty: implications for surgeons. *J Surg Res*. 2006;131(1):1–6.

129. Bansback N, Harrison M, Marra C. Does Introducing Imprecision around Probabilities for Benefit and Harm Influence the Way People Value Treatments? *Medical Decision Making*. 2016;36(4):490–502.

130. Wakker PP. Uncertainty aversion: a discussion of critical issues in health economics. *Health Econ*. 2000;9(3):261–263.

131. Kalet A, Roberts JC, Fletcher R. How do physicians talk with their patients about risks? *J Gen Intern Med*. 1994;9(7):402–404.

132. Braddock CH, Edwards KA, Hasenberg NM, Laidley TL, Levinson W. Informed decision making in outpatient practice: time to get back to basics. *JAMA*. 1999;282(24):2313–2320.

133. Donabedian A. *Explorations in Quality Assessment and Monitoring Volume I*. Vol Volume I. Ann Arbor, Michigan: Health Administration Press; 1980.

134. JAMA Forum: Back to the Future: Volume as a Quality Metric. June 2015.

135. Demetriades D, Martin M, Salim A, et al. Relationship between American College of Surgeons trauma center designation and mortality in patients with severe trauma (injury severity score > 15). *J Am Coll Surg*. 2006;202(2):212–5– quiz A45.

136. Stulberg JJ. Adherence to Surgical Care Improvement Project Measures and the Association With Postoperative Infections. *JAMA*. 2010;303(24):2479–2485.

137. Maggard-Gibbons M. The use of report cards and outcome measurements to improve the safety of surgical care: the American College of Surgeons National Surgical Quality Improvement Program. *BMJ Qual Saf*. 2014;23(7):589–599.

138. Sepucha KR, Fowler FJ, Mulley AG. Policy support for patient-centered care: the need for measurable improvements in decision quality. *Health Aff (Millwood)*. 2004;Suppl Variation:VAR54–62.

139. Llewellyn-Thomas HA. Patients' health-care decision making: a framework for descriptive and experimental investigations. *Med Decis Making*. 1995;15(2):101–106.

140. Mellers B, Stone E, Murray T, et al. Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions. *Perspectives on Psychological Science*. 2015;10(3):267–281.

141. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical Skill and Complication Rates after Bariatric Surgery. *N Engl J Med*. 2013;369(15):1434–1442.