



Published in final edited form as:

*Ann Thorac Surg.* 2014 July ; 98(1): 210–216. doi:10.1016/j.athoracsur.2014.04.141.

## The Impact of Perceived Frailty on Surgeons' Estimates of Surgical Risk

Mark K. Ferguson, MD<sup>1</sup>, Jeanne Farnan, MD, MHPE<sup>2</sup>, Josh A. Hemmerich, PhD<sup>2</sup>, Kris Slawinski, MA<sup>3</sup>, Julissa Acevedo, BS<sup>4</sup>, and Stephen Small, MD<sup>5</sup>

<sup>1</sup>Department of Surgery, The University of Chicago, Chicago, IL

<sup>2</sup>Department of Medicine, The University of Chicago, Chicago, IL

<sup>3</sup>Pritzker School of Medicine, The University of Chicago, Chicago, IL

<sup>4</sup>Center for Research Informatics, The University of Chicago, Chicago, IL

<sup>5</sup>Department of Anesthesiology and Critical Care, The University of Chicago, Chicago, IL

### Abstract

**Background**—Physicians are only moderately accurate in estimating surgical risk based on clinical vignettes. We assessed the impact of perceived frailty by measuring the influence of a short video of a standardized patient on surgical risk estimates.

**Methods**—Thoracic surgeons and cardiothoracic trainees estimated the risk of major complications for lobectomy based on clinical vignettes of varied risk categories (low, average, high). After each vignette, subjects viewed a randomly selected video of a standardized patient exhibiting either vigorous or frail behavior, then re-estimated risk. Subjects were asked to rate 5 vignettes paired with 5 different standardized patients.

**Results**—Seventy-one physicians participated. Initial risk estimates varied according to the vignette risk category: low,  $15.2 \pm 11.2\%$  risk; average,  $23.7 \pm 16.1\%$ ; high,  $37.3 \pm 18.9\%$ ;  $p < 0.001$  by ANOVA. Concordant information in vignettes and videos moderately altered estimates (high risk vignette, frail video:  $10.6 \pm 27.5\%$  increase in estimate,  $p = 0.006$ ; low risk vignette, vigorous video:  $14.5 \pm 45.0\%$  decrease,  $p = 0.009$ ). Discordant findings influenced risk estimates more substantially (high risk vignette, vigorous video:  $21.2 \pm 23.5\%$  decrease in second risk estimate,  $p < 0.001$ ; low risk vignette, frail video:  $151.9 \pm 209.8\%$  increase,  $p < 0.001$ ).

**Conclusions**—Surgeons differentiated relative risk of lobectomy based on clinical vignettes. The effect of viewing videos was small when vignettes and videos were concordant; the effect was more substantial when vignettes and videos were discordant. The information will be helpful in training future surgeons in frailty recognition and risk estimation.

### Keywords

geriatric; co-morbidity; education; lung cancer surgery

## Introduction

Judgment about the appropriateness of surgery or its anticipated outcomes is influenced by a variety of factors, some of which are surgeon-related, including training level, recent adverse patient outcomes, and prioritization of risk vs benefit [1–4]. Major complications after lung resection lead to increased length of hospitalization and increased costs [5,6]. Accurate risk estimation improves patient selection for surgery, obtaining informed consent, interventions for risk reduction, and allocation of appropriate perioperative resources.

Using vignettes of actual patients, practicing surgeons are more accurate than trainees in predicting risk, although neither group exhibits a high accuracy [7]. This caused us to query: how much of estimated risk is based on the surgeon's visual assessment of the patient (the "eyeball factor")? Seeing a patient has an important effect on how physicians view a patient's level of illness/risk in a clinical setting [8], but this effect has not been quantified in a surgical setting.

In this study we measured the degree to which the "eyeball factor" influences surgical risk estimation. We assessed the impact of a video of a standardized patient portraying vigorous or frail physical behavior on estimates of risk. Differences between practicing surgeons and trainees were also assessed. We hypothesized that trainees and practicing surgeons will differ in their risk estimates, that physicians' risk estimates will change significantly after they view a video of a standardized patient, and that the impact of videos on risk estimates will differ between these groups.

## Methods

A group of physicians and other geriatric specialists (MKF, JF, KS, SS), using an iterative process, developed a set of physical characteristics based on Fried's phenotypic criteria [9] that could be portrayed in a short, silent video, including weight loss, gait speed, strength, and fatigue (Appendix 1). Five standardized patients (SPs) portrayed differing degrees of these physical characteristics of frailty in videos ranging from "vigorous" to "frail" in a mock outpatient clinic setting. Five geriatrics specialists rated the level of frailty portrayed in each of 62 videos. Two videos were selected for each SP that had the highest inter-rater agreement for representing "somewhat vigorous" and "somewhat frail" behavior. Hereafter these are referred to as "vigorous" and "frail" videos.

Five short clinical vignettes were developed based on patients who underwent lobectomy for non-small cell lung cancer at our institution 2011–2012 and were aged 50 to 69. The Charlson score (for assessment of risk status related to comorbidities) and EVAD score (for physiologic risk status specifically related to lung resection) were calculated for each patient [10,11]; the possible range of Charlson scores is 0 to 37 and the possible range of EVAD scores is 0 to 12. In our dataset of 89 patients, Charlson scores ranged from 2 to 7 (mean  $3.0 \pm 1.2$ ; median 3), EVAD scores ranged from 0 to 10 (mean  $4.8 \pm 2.8$ ; median 4), and the range of combined scores was 2 to 15 (mean  $7.8 \pm 3.3$ ; median 7). Patients were selected for vignettes based on the mean combined score plus or minus one standard deviation. Two patient vignettes were classified as low risk with total score of 4 or 5, one was classified as

average risk with a total score of 8, and two patient vignettes were classified as high risk with a total score of 11. These scores were not revealed to the study subjects.

Thoracic surgeons serving as teaching physicians in traditional two or three year US-based cardiothoracic training programs (161) and cardiothoracic trainees in those programs (125) were solicited to participate in the study using e-mail addresses provided by the Thoracic Surgery Directors Association. Participants provided their age, sex, comfort level with performing a lobectomy, and the current year of training or number of years of experience since completing training.

Participants were asked to read a clinical vignette and score the patient's risk for major postoperative complications after lobectomy from 0% to 100% using an anchored Likert-like scale. They then rated the importance of each of 13 clinical variables in their risk assessment for that patient. Participants then viewed 1 of 2 videos of the SP paired with the clinical vignette, randomly selected between "vigorous" or "frail" behavior. Participants provided a second estimate of surgical risk based on the vignette and the video, rated the importance of the video in making the second risk estimate, and rated the importance of the physical behaviors portrayed in the video to their second risk estimate. This process was repeated for a total of five vignettes.

Study data were collected and managed using the REDCap tool hosted at the University of Chicago. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies [12]. Independent and paired t-tests, analysis of variance (ANOVA), and  $\chi^2$  analyses were used as appropriate. Analyses were performed using Minitab (Minitab Version 16, Minitab, Inc). The study was approved by our Institutional Review Board and the requirement for written consent was waived.

## Results

### Subjects

Fifty-three thoracic surgeons and 18 cardiothoracic trainees participated in the study. Practicing surgeons were significantly older than trainees. There was no difference in sex distribution between the groups; both were comprised predominantly of males. Practicing surgeons had considerable clinical experience, and this likely resulted in the high frequency with which they rated their ability to perform a lobectomy as "expert" (Table 1). In contrast, none of the trainees rated their ability as "expert" ( $p>0.001$ ). The 71 participants completed scoring for 294 vignette-video pairings; 53 (75%) completed all 5 vignette-video pairings. The distribution of frail and vigorous videos was similar between the trainees and surgeons ( $p=0.411$ ). All scores were used in the analyses.

### Initial risk estimates based only on clinical vignettes

Initial risk estimates, which were made without the influence of a paired video, are described in Table 2. When initial risk estimates for the five vignettes were combined into three risk categories, there was a significant difference among the risk scores. This illustrates good calibration of the participants' estimates relative to pre-assigned vignette risk category (Figure 1).

Some clinical factors in the vignettes (performance status, spirometry, diffusing capacity, smoking status, and possibly clinical impression) were instrumental in risk estimation, as determined by more than half of the sample rating them as “important,” whereas the majority of the elements were not considered “important” for most of the risk estimates (Table 3). The importance of clinical factors differed somewhat according to the level of risk evident in the vignettes. The impact of “smoking status,” “obesity,” and possibly “cardiac status” and “diffusing capacity” on risk estimates appeared to increase as vignette risk increased (Table 4).

### **Second risk estimates based on clinical vignettes and videos**

Overall, second risk estimates made after viewing videos of standardized patients were similar to initial estimates ( $25.1 \pm 16.2$  vs  $24.9 \pm 18.1$ ;  $p=0.863$ ). Initial and second risk estimates differed when assessed by vignette risk level. There was a significant increase in estimated risk for low risk vignettes ( $15.2 \pm 10.8$  vs  $19.4 \pm 14.1$ ;  $p<0.001$ ) and a significant decrease in estimated risk was associated with high risk vignettes ( $37.3 \pm 18.9$  vs  $33.5 \pm 16.9$ ;  $p=0.003$ ). There was no difference between initial and second risk estimates for average risk vignettes ( $23.7 \pm 16.1$  vs  $22.5 \pm 10.1$ ;  $p=0.393$ ). Second risk estimates increased by  $76.0 \pm 154.0\%$  for vignettes paired with frail videos, and decreased by  $18.0 \pm 34.9\%$  for vignettes paired with vigorous videos ( $p<0.001$ ).

The most striking differences between initial and subsequent scores were identified when the interaction between vignette risk level and video content was evaluated. Concordance between vignette and video content resulted in small but significant changes between initial and second risk estimates. Frail videos paired with high risk vignettes resulted in a  $2.3 \pm 8.7$  percentage point increase in the second risk estimate ( $p=0.050$ ), representing a  $10.6 \pm 27.5\%$  increase ( $p=0.006$ ). Vigorous videos paired with low risk vignettes resulted in a  $3.1 \pm 5.2$  percentage point decrease in the second risk estimate ( $p<0.001$ ), a decrease of  $14.5 \pm 45.0\%$  ( $p=0.009$ ). In contrast, discordance between video and vignette content was associated with substantially greater changes between the two risk estimates, particularly with regards to the effect of frail videos (Figure 2). Frail videos paired with low risk vignettes resulted in a  $12.5 \pm 10.1$  percentage point increase in the second risk estimate ( $p<0.001$ ), an increase of  $151.9 \pm 209.8\%$  ( $p<0.001$ ). Vigorous videos paired with high risk vignettes resulted in a  $10.6 \pm 13.3$  percentage point decrease in the second risk estimate ( $p<0.001$ ), representing a  $21.2 \pm 23.5\%$  decrease ( $p<0.001$ ).

The role of the videos in making the second risk estimates was rated as “important” for 117 (40%) of the 294 rated video-vignette pairings (Table 5). Because of the small numbers of videos rated as “somewhat unimportant” or “unimportant,” these two categories were combined for statistical analysis. There were no differences in the importance ratings of videos comparing video content or vignette risk (Table 5).

### **Trainees vs practicing surgeons**

Trainees initially estimated risk as greater than did practicing surgeons when all estimates were considered. This difference was more pronounced for low risk vignettes; there was no important difference between trainees and practicing surgeons for estimates for high risk

vignettes. Risk estimates for both groups corresponded to the pre-assigned risk levels for the vignettes, indicating good calibration of risk by each group (Table 2). In analyzing vignette content, when ratings were dichotomized as “important” versus “other,” the utility of “clinical impression” and “cancer stage” were rated higher by surgeons than by trainees in making risk estimates (Table 4).

As with initial estimates of risk, second risk estimates were higher for trainees than for practicing surgeons ( $28.9 \pm 14.9$  vs  $23.9 \pm 16.5$ ;  $p=0.018$ ). Second risk estimates did not differ from initial estimates for trainees ( $30.5 \pm 19.1$  vs  $28.9 \pm 14.9$ ;  $p=0.408$ ) or for practicing surgeons ( $23.1 \pm 17.5$  vs  $23.9 \pm 16.5$ ;  $p=0.252$ ). Trainees reduced their second risk estimates more in response to vigorous videos than practicing surgeons, whereas practicing surgeons increased their risk estimates more in response to frail videos; the differences were not statistically significant (Figure 3). The influence of frail and vigorous videos on trainees’ second risk estimates was similar for average and high risk vignettes. In contrast, attending surgeons produced graded changes in their second risk estimates that were better calibrated to vignette risk category.

Practicing surgeons rated the impact of videos on their risk estimates as “important” more often than did trainees (45% vs 25%;  $p=0.004$ ). Practicing surgeons also rated the impact of frail videos “important” more often than did trainees (44% vs 22%;  $p=0.025$ ); this difference was not evident for vigorous videos. Specific elements of physical frailty within the videos differed somewhat in their impact on risk estimates. Practicing surgeons rated “gait speed” and “strength” as more important than did trainees (Table 6).

## Comment

The ability to accurately estimate operative risk is a vital element in surgical decision making. Surgeons review a large amount of clinical data for estimating risk of more complex and potentially morbid operations, although it’s not clear how many of those data are actually used in risk estimation. A recent study determined that relying on clinical information alone resulted in poor accuracy of risk estimation for lung resection [7]. The ability to see a patient, which can be referred to as the “eyeball factor,” allows the surgeon to make judgments about physiologic age, fitness, cognition, and enthusiasm for a planned operation.

Frailty is a clinical syndrome characterized by decreasing physical functioning and impaired physiologic reserves. This syndrome can be suspected based on a brief in-person encounter with a patient, and its presence can be confirmed through fairly simple standardized testing. Despite the fact that frailty has been strongly associated with adverse outcomes in thoracic surgical patients [13–15], its characteristics, diagnosis, and impact on surgical outcomes are not widely understood, and these elements are not a standard part of the curriculum for cardiothoracic trainees. The present study was designed to measure the degree to which the “eyeball factor” influences surgical risk estimation. We focused on the specific impact of visually observable frailty in risk estimation, and compared characteristics of risk estimation between cardiothoracic trainees and practicing thoracic surgeons to help identify potential targets for educational opportunities.

In this study, participants accurately calibrated surgical risk estimates relative to clinical factors in the patient vignettes. In doing so, however, they appeared to use only a fraction of the available clinical factors in rendering their judgment. More factors were utilized as vignette risk increased, likely in response to increasing abnormalities in organ systems evident in the vignettes. The impact of the “eyeball factor” on risk assessment for lobectomy is complex, and depends in part on the clinical characteristics of the patient, the physical behavioral characteristics of the patient, and the physician involved in assessing risk. Participants altered risk estimates after viewing a patient video according to the degree of discordance between the perceived risk based on the clinical vignette and the physical behavior portrayed in the video.

Videos that depicted frail behavior had a more profound impact on risk estimates than did vigorous videos, especially for practicing surgeons. The degree of difference was larger than we had anticipated. For vigorous videos the most discordant findings resulted in a decrease in estimated operative risk of 21%. In comparison, the most discordant findings associated with frail videos resulted in an increase in estimated risk of over 150%. Clinical judgment based on physical behavior appeared to override risk estimates based on other clinical factors. The participants were substantially more concerned about outcomes when presented with a frail video paired with a low risk vignette. This may be reassuring, as it is likely more important to distinguish high risk patients from others than it is to distinguish low risk patients from others.

Trainees had substantially different levels of risk estimation than did practicing surgeons. This difference has not been well described, and indicates a more cautious approach to higher risk surgery among trainees. There was some indication that changes in trainee risk estimates in response to frail videos were more muted, while their responses to the vigorous videos were more pronounced, than those of the practicing surgeons. These differences were not statistically significant, but that may have been a result of the relatively low statistical power provided by small numbers among the subgroups that were analyzed. Finally, practicing surgeons and trainees appeared to use clinical factors depicted in the vignettes and physical characteristics evident in the videos somewhat differently in making their risk estimates. The combination of these findings suggests that even advanced level trainees possess clinical judgment that is quantitatively and qualitatively different than that of practicing surgeons.

There are potential limitations to this study. The patient vignette format does not replicate the true clinical setting well. Although our standardized patients performed their assigned tasks admirably, they represented “early to middle” middle age rather than the more appropriate “middle to late” middle age. This may have resulted in the vigorous videos appearing to represent a greater level of vigor than was intended. The disparity in numbers between practicing surgeons and trainees may have prevented identification of additional differences in the approaches of these two groups to risk estimation.

This is among the first studies to measure the impact of visually observed physical status on estimates of surgical risk. Understanding the elements that contribute to the “eyeball factor” and quantifying its impact are important first steps in developing more objectivity,

consistency, and reliability in risk estimation, and may provide information useful in developing risk estimation models. Surgeons appear to have some understanding of frailty and its impact on surgical risk. Teaching surgeons to explicitly consider frailty when evaluating a patient might result in a more nuanced assessment of risk. Referring potentially frail patients for a formal assessment of frailty might provide information that is very useful in shared decision making [16–19]. Finally, the differences between trainees and practicing surgeons in their use of objective and subjective information require further investigation. Such information may help in educating trainees on how to think like a master surgeon.

## Acknowledgments

The study was funded by the Donald J. Ferguson, MD, Surgical Research Fund.

The REDCap project at the University of Chicago is managed by the Center for Research Informatics and funded by the Biological Sciences Division and by the Institute for Translational Medicine, CTSA grant number UL1 RR024999 from the National Institutes of Health.

## References

1. Chatterjee S, Ng J, Kwan K, Matsumoto ED. Assessing the surgical decision making abilities of novice and proficient urologists. *J Urol*. 2009; 181:2251–2256. [PubMed: 19296971]
2. Dimick JB, Welch HG. The zero mortality paradox in surgery. *J Am Coll Surg*. 2008; 206(1):13–16. [PubMed: 18155563]
3. 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; 203:642–652. [PubMed: 17084325]
4. MacCormick A, Parry B. Judgment analysis of surgeons' prioritisation of patients for elective general surgery. *Med Decis Making*. 2006; 26:255–264. [PubMed: 16751324]
5. Wang J, Olak J, Ultmann RE, Ferguson MK. Assessment of pulmonary complications after lung resection. *Ann Thorac Surg*. 1999 May; 67(5):1444–1447. [PubMed: 10355428]
6. Birkmeyer JD, Gust C, Dimick JB, Birkmeyer NJ, Skinner JS. Hospital quality and the cost of inpatient surgery in the United States. *Ann Surg*. 2012; 255(1):1–5. [PubMed: 22156928]
7. Ferguson MK, Stromberg JD, Celauro AD. Estimating lung resection risk: a pilot study of trainee and practicing surgeons. *Ann Thorac Surg*. 2010; 89(4):1037–1043. [PubMed: 20338304]
8. Sündermann S, Dademasch A, Praetorius J, Kempfert J, Dewey T, Falk V, et al. Comprehensive assessment of frailty for elderly high-risk patients undergoing cardiac surgery. *Eur J Cardiothorac Surg*. 2011; 39:33–7. [PubMed: 20627611]
9. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Cardiovascular Health Study Collaborative Research Group. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001; 56(3):M146–56. [PubMed: 11253156]
10. Charlson ME, Sax FL, MacKenzie CR, Braham RL, Fields SD, Douglas RG Jr. Morbidity during hospitalization: can we predict it? *J Chronic Dis*. 1987; 40(7):705–12. [PubMed: 3110198]
11. Ferguson MK, Durkin AE. A comparison of three scoring systems for predicting complications after major lung resection. *Eur J Cardiothorac Surg*. 2003; 23(1):35–42. [PubMed: 12493501]
12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009; 42:377–81. [PubMed: 18929686]
13. Tsiouris A, Hammoud ZT, Velanovich V, Hodari A, Borgi J, Rubinfeld I. A modified frailty index to assess morbidity and mortality after lobectomy. *J Surg Res*. 2012; 183:40–46. [PubMed: 23273884]

14. Hodari A, Hammoud ZT, Borgi JF, Tsiouris A, Rubinfeld IS. Assessment of morbidity and mortality after esophagectomy using a modified frailty index. *Ann Thorac Surg.* 2013; 96(4): 1240–5. [PubMed: 23915593]
15. Velanovich V, Antoine H, Swartz A, Peters D, Rubinfeld I. Accumulating deficits model of frailty and postoperative mortality and morbidity: its application to a national database. *J Surg Res.* 2013; 183(1):104–10. [PubMed: 23415494]
16. Dalton AF, Bunton AJ, Cykert S, Corbie-Smith G, Dilworth-Anderson P, McGuire FR, et al. Patient characteristics associated with favorable perceptions of patient-provider communication in early- stage lung cancer treatment. *J Health Commun.* 2013 Dec 20. Epub ahead of print.
17. Knops AM, Legemate DA, Goossens A, Bossuyt PM, Ubbink DT. Decision aids for patients facing a surgical treatment decision: a systematic review and meta-analysis. *Ann Surg.* 2013; 257(5):860–6. [PubMed: 23470574]
18. Lim E. Patients' perspective in the surgical decision-making process. *Thorac Surg Clin.* 2012; 22(4):539–43. [PubMed: 23084617]
19. Ganai S, Ferguson MK. Quality of life in the high-risk candidate for lung resection. *Thorac Surg Clin.* 2012; 22(4):497–508. [PubMed: 23084613]

### Appendix 1. Categories for videos of standardized patients (from Ferguson, submitted for publication)

Clinical factor	Vigorous	Somewhat vigorous	Neither vigorous nor frail	Somewhat frail	Frail
<b>Weight loss</b>	None	None	Possible – either shirt or pants may appear somewhat loose	Possible – both shirt and pants may appear loose	Noticeable looseness of shirt and pants
<b>Gait</b>	Rapid, purposeful stride	Normal pace, medium stride	Normal pace and stride length	Slow, shortened stride	Very slow, shuffling, shortened stride
<b>Strength</b>	Normal, requires no aids to sit, stand, or climb	Normal, uses one hand for balance when sitting, standing, or climbing onto a step	Normal, uses two hands for balance and aid when sitting and standing; slightly slow to climb onto step and turn	Reduced, some difficulty sitting and standing despite use of two hands; some difficulty in climbing onto and turning on step in order to sit on table	Clearly weak, difficulty sitting and standing, uses upper body considerably in sitting and standing; clear difficulty in climbing onto and turning on step in order to sit on table
<b>Fatigue</b>	None – moves rapidly from chair to table	None – moves without effort from chair to table	None – moves somewhat slowly from chair to table	Mild – some rapid breathing with moderate effort, appears tired, moves slowly from chair to table	Definite – breathless with minimal effort; appears drawn; considerable



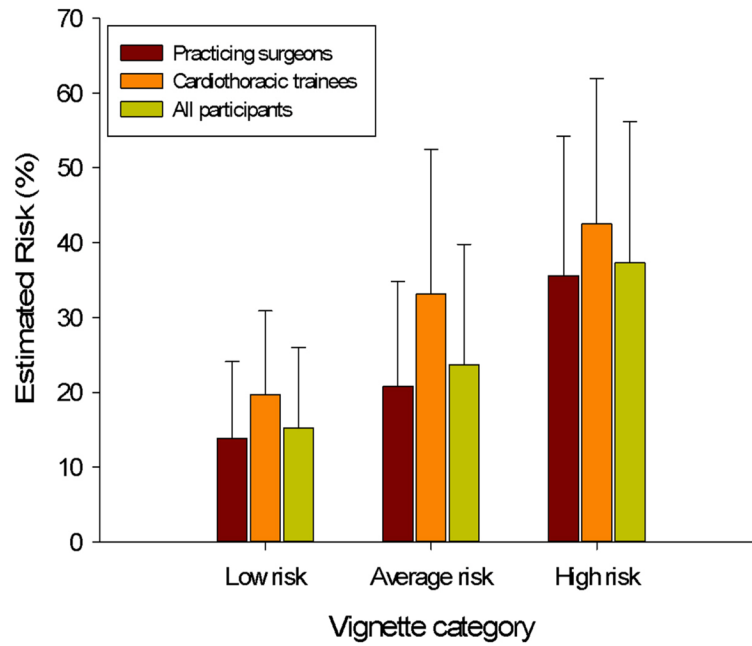
Clinical factor	Vigorous	Somewhat vigorous	Neither vigorous nor frail	Somewhat frail	Frail
					effort necessary to move from chair to table.

Author Manuscript

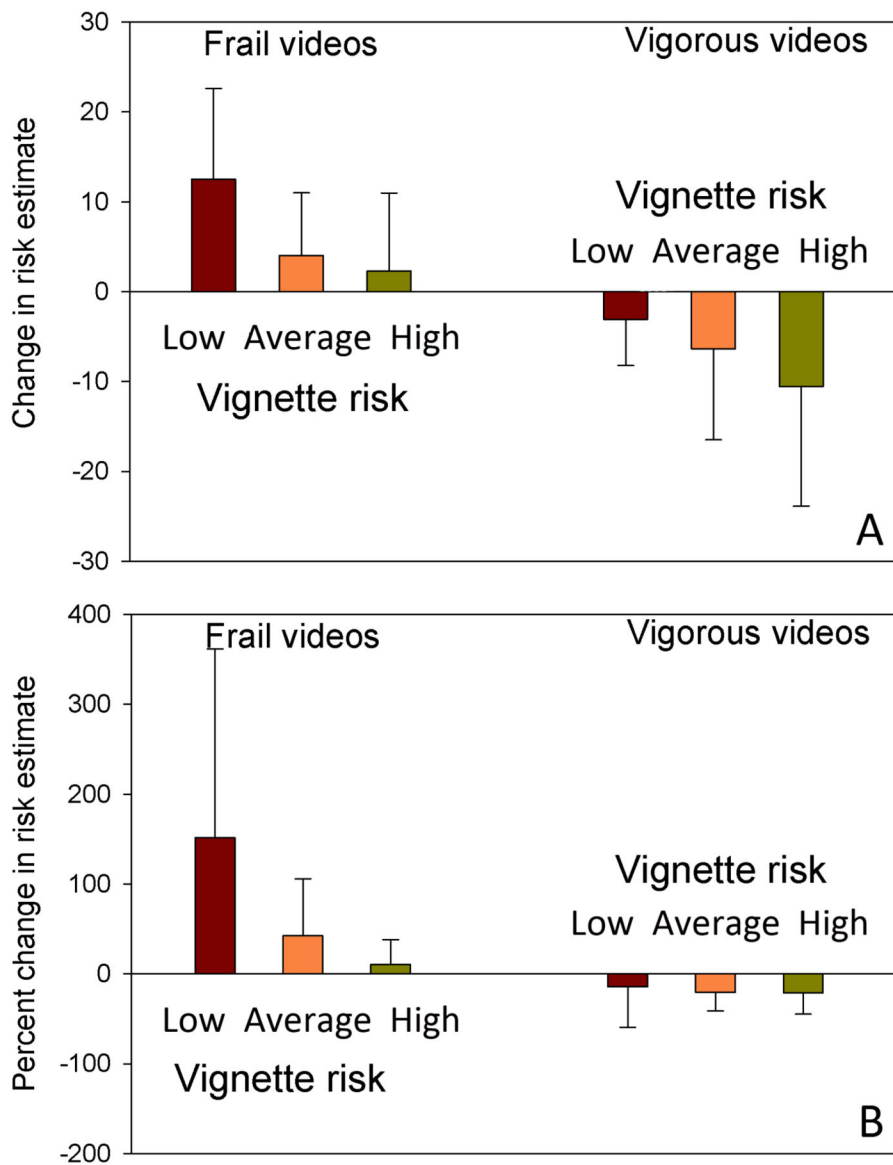
Author Manuscript

Author Manuscript

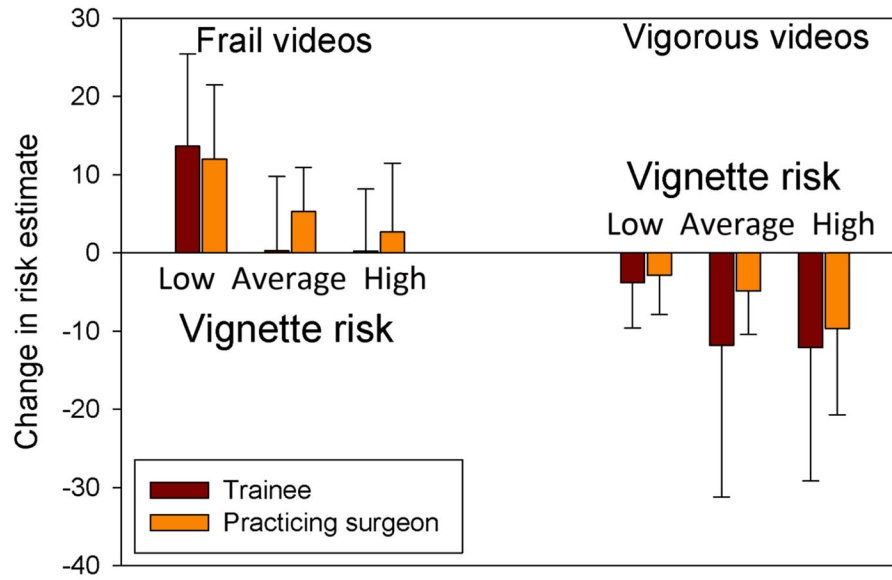
Author Manuscript



**Figure 1.** Initial risk estimates for vignette categories. Differences among the vignette categories were significant for all participants ( $p < 0.001$ ) and for practicing surgeons and trainees ( $p < 0.001$  for each). Differences between practicing surgeons and cardiothoracic trainees by risk category were: low risk,  $p = 0.012$ ; average risk,  $p = 0.051$ ; high risk,  $p = 0.120$ .



**Figure 2.** Changes in risk estimates grouped by video type and vignette risk category. Absolute change in percentage points is illustrated in Panel A, and percentage change is illustrated in panel B. The differences among values in each video category are statistically significant (ANOVA,  $p < 0.001$ ).



**Figure 3.** Changes in risk estimates comparing trainees to practicing surgeons, grouped by video type and vignette risk category. Change represents percentage point differences rather than percentage change. The differences between values for each vignette risk level in each video category were not statistically significant.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

Characteristics of participating surgeons and cardiothoracic trainees.

Characteristic	Trainees (18)	Practicing surgeons (53)	P value
Age	35.5 ± 3.4	48.7 ± 9.6	<0.001
Male	11 (85%)	42 (72%)	0.361
Years of cardiothoracic training or years in practice	2.6 ± 1.2	13.3 ± 10.6	n/a
Comfort level for performing lobectomy			<0.001
Naïve	0	0	
Learner	4 (22%)	1 (2%)	
Competent	14 (78%)	10 (19%)	
Expert	0	42 (79%)	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 2**

Initial risk estimates for vignettes by rater category.

	All vignettes	Vignette 1 Low risk	Vignette 2 Low risk	Vignette 3 Average risk	Vignette 4 High risk	Vignette 5 High risk
All participants	24.9 ± 18.1	14.9 ± 10.6	15.7 ± 11.1	23.7 ± 16.1	35.5 ± 18.9	39.1 ± 18.9
Trainees	30.5 ± 19.1	18.0 ± 8.9	21.8 ± 13.7	33.1 ± 19.4	40.8 ± 21.0	44.2 ± 18.5
Surgeons	23.1 ± 17.5	13.8 ± 11.0	13.9 ± 9.6	20.8 ± 14.0	33.8 ± 18.2	37.5 ± 19.0
p value	0.004	0.110	0.060	0.051	0.290	0.275

Table 3

The contribution of clinical factors in vignettes to initial risk estimates listed according to how often they were considered “important”.

Factor	Important	Somewhat important	Neither important nor unimportant	Somewhat unimportant	Unimportant
Performance status	212 (72)	70 (24)	8 (3)	3 (1)	1 (0)
DLCO	189 (64)	81 (28)	12 (4)	3 (1)	9 (3)
FEV1	156 (53)	118 (40)	5 (2)	6 (2)	9 (3)
Smoking status	147 (50)	122 (42)	15 (5)	5 (2)	5 (2)
Clinical impression	146 (50)	135 (46)	9 (3)	4 (1)	0 (0)
Cardiac status	123 (42)	134 (46)	25 (9)	9 (3)	3 (1)
Cancer stage	105 (36)	132 (45)	44 (15)	9 (3)	4 (1)
Renal function	35 (12)	161 (55)	71 (24)	15 (5)	12 (4)
Age	26 (9)	205 (70)	32 (11)	25 (9)	6 (2)
Obesity	21 (7)	171 (58)	64 (22)	21 (7)	17 (6)
Diabetes	20 (7)	144 (49)	79 (27)	25 (9)	26 (9)
Peripheral vascular disease	19 (6)	178 (61)	64 (22)	19 (6)	14 (5)
Hypertension	10 (3)	115 (39)	118 (40)	30 (10)	21 (7)

Numbers in parentheses represent percentages. DLCO: Diffusing capacity of the lung for carbon monoxide; FEV1: forced expiratory volume in the first second

**Table 4**

Percent of clinical factors rated as “important” in initial risk estimates according to participant category and vignette risk category.

Factor	Participants			Vignette Categories			
	Trainee (%)	Practicing surgeon (%)	p value	Low Risk (%)	Average Risk (%)	High Risk (%)	p value
Performance status	65	74	0.114	74	63	76	0.190
DLCO	62	65	0.640	62	55	72	0.076
FEV1	61	51	0.146	54	50	52	0.837
Smoking status	48	51	0.683	40	54	61	0.005
Clinical impression	38	53	0.024	45	50	55	0.271
Cardiac status	45	41	0.526	37	29	55	0.001
Cancer stage	25	39	0.036	37	30	37	0.635
Renal function	12	3	0.818	16	7	9	0.128
Age	12	8	0.192	11	11	6	0.329
Obesity	8	7	0.623	3	5	13	0.010
Diabetes	3	8	0.126	10	4	5	0.152
Peripheral vascular disease	4	7	0.379	8	4	7	0.578
Hypertension	1	4	0.287	2	2	7	0.082

DLCO: Diffusing capacity of the lung for carbon monoxide; FEV1: forced expiratory volume in the first second



**Table 5**

Distribution of ratings of the impact of videos on second estimates of risk.

Variable	Important	Somewhat important	Neither important nor unimportant	Somewhat unimportant	Unimportant	p value	p value important vs other
<b>Raters</b>							
Trainee	18 (25%)	43 (61%)	7 (10%)	2 (3%)	1 (1%)	0.019	0.004
Practicing surgeon	99 (45%)	107 (48%)	9 (4%)	5 (2%)	3 (1%)		
<b>Video type</b>							
Vigorous	62 (42%)	69 (46%)	9 (6%)	6 (4%)	3 (2%)	0.110	0.519
Frail	55 (38%)	81 (56%)	7 (5%)	1 (1%)	1 (1%)		
<b>Vignette risk category</b>							
Low	53 (40%)	64 (49%)	9 (7%)	3 (2%)	2 (2%)	0.104	0.590
Average	19 (34%)	36 (64%)	0	0	1 (2%)		
High	45 (42%)	50 (47%)	7 (7%)	4 (4%)	1 (1%)		

**Table 6**

Ratings of the importance of physical frailty elements in videos to surgical risk estimates, comparing trainees to practicing surgeons.

Physical frailty element	Important	Somewhat important	Neither important nor unimportant	Somewhat unimportant	Unimportant	p value combining "unimportant" values	p value important vs other
<b>Age</b>							
Trainee	7 (10%)	49 (69%)	14 (20%)	1 (1%)	0	0.079	0.820
Practicing surgeon	24 (11%)	157 (71%)	24 (11%)	14 (6%)	3 (1%)		
<b>Weight loss</b>							
Trainee	14 (20%)	33 (47%)	14 (20%)	9 (13%)	1 (1%)	0.944	0.793
Practicing surgeon	47 (21%)	94 (42%)	48 (22%)	21 (9%)	12 (5%)		
<b>Gait speed</b>							
Trainee	15 (21%)	45 (63%)	6 (8%)	4 (6%)	1 (1%)	0.055	0.011
Practicing surgeon	83 (37%)	115 (52%)	17 (8%)	7 (3%)	0		
<b>Strength</b>							
Trainee	19 (27%)	40 (56%)	6 (8%)	3 (4%)	3 (4%)	0.010	0.058
Practicing surgeon	87 (39%)	116 (52%)	16 (7%)	2 (1%)	1 (1%)		
<b>Fatigue</b>							
Trainee	27 (39%)	32 (45%)	10 (14%)	2 (1%)	0	0.580	0.242
Practicing surgeon	102 (46%)	92 (41%)	27 (8%)	10 (4%)	3 (1%)		