Stojan Jennifer (Orcid ID: 0000-0002-7743-8014)

Title: Dealing with uncertainty in clinical reasoning: a threshold model and the roles of experience and task framing

Author Byline: Jennifer N Stojan MD, MHPE, Michelle Daniel MD, MHPE, Sarah Hartley MD, and Larry Gruppen, PhD

Author Biographies:
Jennifer N. Stojan MD, MHPE is an Associate Professor in the Department of Internal Medicine and the Department of Pediatrics, University of Michigan Medical School, Ann Arbor, Michigan.

Michelle Daniel MD, MHPE is Vice Dean for Medical Education and a Clinical Professor of Emergency Medicine at the University of California San Diego School of Medicine, La Jolla, California. orcid.org/0000-0001-8961-7119

Sarah Hartley, MD is an Associate Professor of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan

Larry D. Gruppen PhD, is Professor in the Department of Learning Health Sciences and Director of the University of Michigan Master of Health Professions Education (MHPE) program in Ann Arbor, Michigan. orcid.org/0000-0002-2107-0126

Manuscript Type: Original Research Article

Corresponding Author:
Jennifer N. Stojan, MD
Institution: University of Michigan
Phone: 586-770-6642
Address: 3119 Taubman Center, SPC5376
1500 East Medical Center Drive
Ann Arbor, MI 48109
E-mail: jstojan@umich.edu

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/medu.14673

Abstract

Introduction:

Uncertainty is integral to clinical practice and clinical reasoning but has proven difficult to study and model. Little is known about how clinicians manage uncertainty. According to evidence-based medicine theory, clinicians should utilize new information to reduce uncertainty until reaching action thresholds for further information gathering or treatment. We examined the impact of experience and task framing on uncertainty thresholds and the extent to which these thresholds guided clinical decisions. Finally, we sought to determine the impact of framing by having participants provide threshold responses as a range or as specific numbers.

Methods:

One-hundred-sixty-eight fourth year medical students, 93 residents, and 72 faculty were presented a case of viral pneumonia with a suspected superimposed bacterial infection. Participants identified their testing and treatment thresholds with either a specific number or an inter-threshold range of probabilities that would compel them to test further. Afterwards, they were told the patient had a 20% pretest probability of a superimposed infection and asked

whether they would treat the patient with antibiotics, order additional testing or neither. Responses were compared to their previously stated threshold values to assess decision-making consistency.

Results:

Testing thresholds were 15.8%, 20.6%, and 25.8%, treatment thresholds were 78.5%, 71.6% and 73.4% and threshold spans (difference between testing and treatment thresholds) were 62.7, 51, and 47.6 for students, residents, and faculty respectively. Sixty-four percent of respondents made judgments consistent with their thresholds, 28% escalated their decision (doing more than their thresholds predicted), and 7.6% de-escalated their decision (doing less than their thresholds predicted). Framing had an impact on both faculty and resident decisions and a larger impact on students.

Discussion:

These findings help us understand how clinical reasoning and threshold determinations vary with clinical experience. As uncertainty can lead to unnecessary testing and cognitive discomfort, examining decision thresholds helps us ascertain how diagnostic and treatment decisions are made.

Introduction

Uncertainty in reasoning and decision making is fundamental to real-world problems.[1] Dealing with uncertainty is a major component of the education and practice of physicians, posing a challenge for individuals and for the discipline of medicine as a whole.[2,3,4] Part of this challenge is that uncertainty comes from various sources and has different meanings.[5] There is a distinction between “corporate” uncertainty that is inherent to the current state of biomedical knowledge in general, and “individual” uncertainty that stems from an individual’s ignorance or inexperience.[6] Uncertainty is also a positive factor in promoting motivation for problem solving through curiosity and interest.[7,8,9] Reducing uncertainty reduces cognitive discomfort and clarifies the choices among alternative solutions.[10] The most common way to reduce uncertainty is to seek more information. In medicine, this often involves testing, sometimes to excess.[11]

Evidence-based medicine (EBM) is one approach to supporting physicians' efforts to quantify and manage uncertainty. EBM codifies methods for applying current best evidence to diagnostic and therapeutic decisions as well as criteria for judging the quality of evidence.[12,13] According to EBM theory, management of a patient should proceed via a Bayesian analytic process, in which the prior probabilities of disease(s) are updated as more information becomes available until a threshold for action is reached. This information may come from the research literature, or it may come from the patient history, physical examination, or laboratory studies.

The concept of thresholds for clinical decision making was first described by Pauker and Kassirer.[14] (Figure 1) If a physician's probability of a given disease is very low, she will not feel the need to order further testing prior to making the decision not to treat the disease. As the pre-test probability of a disease increases, at some point, a testing threshold will be reached wherein she will want additional information before feeling confident enough in her decision to treat or not treat a disease. As the probability of the disease continues to increase, a treatment threshold will be met where she will no longer feel the need to gain further information and will feel confident enough in her decision to just treat the disease.

However reasonable the concept, few "standard" thresholds for action are available. Instead, practitioners develop their own, idiosyncratic action thresholds based on their experiences with patients, outcomes, and their own tolerance of uncertainty.[15] The position of these two thresholds in an individual is dependent on several factors, including the characteristics of the disease, costs to the patient and to society, efficacy and toxicity of the treatment, physician experience and tolerance of uncertainty, and context.

These thresholds may be dynamic, changing with experience. Medical students have little experience to draw upon when forming thresholds for action, making them more likely to be uncomfortable with uncertainty and requiring additional information prior to making treatment decisions.[16] Residents have more experience to draw upon when estimating probabilities and forming thresholds while faculty have the most experience. However, we know little about probability estimates and action thresholds in general and even less about how experience influences them. In addition, we know little about how likely an individual would be to make decisions consistent with their stated thresholds in a clinical setting.

The best methodology for obtaining and studying probability thresholds and probability adjustments has not been established. Indeed, there is ample reason to be concerned that study methods may bias the results of studies of clinical reasoning.[17] The framing effect is one example of how apparently equivalent presentations of a clinical decision can lead to different solutions or decisions due to seemingly minor differences in how they are presented. For example, a decision question about pursuing surgery as a treatment option will produce different results when the situation is framed in terms of the probability of death versus the probability of survival. Early examples of this bias were reported by Tversky and Kahneman and framing has been shown to be active in medical decision making.[18,19] Therefore, we considered it important to examine not only the process of defining and using action thresholds, but also the impact of the study methodology on framing differences in this process.

In this study, we examined the testing and treatment thresholds of fourth-year medical students, and Internal Medicine and Emergency Medicine residents and faculty. We aimed to determine: (1) at what thresholds of uncertainty each of the three groups would decide to test, and when they would decide to treat, (2) if respondents make case-based decisions in ways consistent with their prior test and treatment threshold determinations, and (3) whether differences in framing of the task for identifying these two thresholds would have an impact on the resulting estimates. We hypothesized that increased clinical experience would be correlated with an increase in the threshold for testing and a decrease in the threshold for treatment, i.e., a narrower band of probabilities requiring testing. We anticipated that participants would not make case-based decisions consistent with their predetermined thresholds but would err on the side of more testing and treatment. We also expected that the framing of the problem would have an effect on the threshold estimates but were not confident about the form of this impact, and so, made no hypothesis.

Methods

Procedure

We built this study on our prior work examining action thresholds.[16] The stimulus case for this study presented a woman with fever, cough and myalgias for five days. A viral swab is positive for H1N1 influenza, and a chest radiograph shows focal consolidation. It is not entirely clear whether the patient has a viral pneumonia with atelectasis represented on chest radiograph or a viral pneumonia with a superimposed bacterial infection. A viral infection is treated supportively whereas antibiotics are required to treat a superimposed bacterial infection. Participants were informed that a procalcitonin test may help determine if this is a bacterial infection.

To evaluate framing effects, we developed two variants of the case, essentially asking for the same information framed in two different ways. The "range" version asked participants to list the range of disease pre-test probabilities for which they would feel compelled to order additional testing prior to making a treatment decision on prescribing antibiotics. The "endpoints" version asked participants under what disease pre-test probability they would feel confident not treating the patient with antibiotics without additional testing (testing threshold) as well as a separate estimate of what pre-test probability they would feel confident treating with antibiotics without additional testing (treatment threshold). Both the testing and treatment thresholds use a 0-100% probability scale, with arithmetic means and standard deviations reflecting this scale. These versions are logically identical, but one focuses on the range of values and the other on the boundaries to determine the extent to which the description or framing of the task would alter responses.

After providing their threshold values, participants were then told that, based on the patient's clinical presentation and chest radiograph, it has been estimated that the pretest probability of a superimposed bacterial infection was 20%. Given this new information, they were asked whether they would treat with antibiotics, not treat with antibiotics or order a procalcitonin test prior to making their decision. Providing a specific pre-test probability offered clarity as to the probability of a superimposed bacterial infection in this case that was not there in the initial presentation. According to the EBM model and Bayes' theorem, the participants should use the 20% probability in the context of their testing and treatment thresholds to decide on their next step (not treat, test, treat). The testing and treatment thresholds that each participant had determined prior to being given the 20% pre-test probability were then used to see if the participant was consistent with their prior thresholds when making decisions based on a 20% pre-test probability.

Participants

All 168 fourth year medical students participated in the study as part of a summative clinical competency assessment (CCA) that measured clinical skills deemed necessary of first-year postgraduate trainees, using both standardized patient and computer-based stations. The case used in this study was part of a computer-based evidenced based medicine (EBM) station in the 2018 administration of the CCA. Eighty-four students took the "range" version of the test and 84 took the "endpoints" version.

The same case was then given to internal medicine and emergency medicine residents in years 1-4 of their training and faculty to determine differences in tolerance of uncertainly based on level of training. Resident and faculty participation was solicited over email, with an equal number assigned to each group being sent the "range" and "endpoints" versions. The response rate for residents and faculty were 44% and 50% respectively. Sixty-four residents responded to the "range" version and 29 residents responded to the "endpoints" version, whereas 32 faculty responded to the "range" version and 40 faculty responded to the "endpoints" version.

This study received exemption status from the University of Michigan Institutional Review Board.

Data analysis

The threshold estimates provided by the participants defined three dependent variables: a testing threshold, describing the point at which they would switch their decision between doing nothing and testing for more information; a treatment threshold, describing the point at which they would switch their decision from testing for more information to treatment without additional testing; and a threshold span or the range of probabilities between the testing and the treatment thresholds describing the range of confidence values in which participants considered that testing would provide useful additional data. Participants in the three groups and the two case versions were compared using a 2-way ANOVA with eta-square as an effect size measure to determine the proportion of the total variance in the dependent variable that was accounted for by the two main effects (experience group and framing condition) and their interaction.

Chi-square statistics were calculated for comparing the judgments of individuals in the abstract and in response to a paper case. Statistical significance was defined as $p<0.05$. Analyses were performed using JMP Pro 14 (SAS Institute, Cary, NC, 2019).

Results

For all three dependent variables, there were statistically significant interactions between the framing version of the problem and the experience level of the participant. When examining testing thresholds, both faculty and residents included a mean probability range (the "range" version) that was lower than when asked to list a specific probability endpoint (the "endpoints" version). Students, on the other hand, utilized a lower threshold for the "endpoints" version than for the "range" version, eta^2 (version x group) = 0.06*, $p<0.05$ (Table 1). When the two versions are combined to examine the main effect of experience, students utilized a lower threshold value than residents and both groups utilized a lower mean threshold than faculty (a difference of about 5 percentage points for each contrast), eta^2 (group) = 0.04, $p<0.05$.

There were larger effects for the treatment threshold, which distinguished decisions to test before treating with decisions to treat without testing (Table 2). Students had higher mean estimates of their treatment thresholds (by 5-7 percentage points) than residents and faculty, who provided very similar means. This contrast was similar for both framing conditions. eta2 (version x group) = 0.11, p<.05. For all respondents, the mean treatment threshold for the “endpoints” framing version was 8-9 percentage points higher than the means for the “range” version (overall means 79.9% vs 71.2%, eta2 (version) = 0.07, p<.0.05).

Examining the difference between the testing and treatment thresholds, as the span of probabilities over which a respondent would order a procalcitonin test, there was again a significant interaction between respondent group and framing condition. Framing condition (eta^2 (version) = 0.04, $p<0.05$) made a small difference for faculty (the “range” version=45.1 vs the “endpoints” version=49.7 points) and residents (49.9 vs. 52.8 points) but the impact of framing version was considerably greater for students (56.1 vs. 69.3 points, eta^2 (version x group) = 0.13, $p<0.05$), Table 3) Overall, faculty had narrower spans (47.6 percentage points) than residents (51.0 percentage points) and students (62.7 percentage points, eta^2 (group) = 0.09, $p<0.05$). The ‘endpoints’ framing condition generated smaller spans than did the ‘range’ format (eta^2 (version) = 0.04, p<.05)

Comparison of the predicted decisions, based on prior thresholds provided by the respondents, with the actual decision made for the case vignette with the probability of a superimposed bacterial infection being 20% revealed that 64.4% of respondents made decisions in this case that were consistent with their thresholds, 28% escalated their decision (were more likely to test or treat than predicted by their thresholds), and 7.6% de-escalated their decision (were less likely to treat or test than predicted by their thresholds) when looked at overall ($chi^2(4, 342) = 44.60$, phi =0.4; (Table 4) When looking at only the "range" version, 61.1% made case-based decisions consistent with their thresholds, 31.1% escalated their decision, and 7.8% de-escalated their decision ($chi^2(4, 180) = 30.27$, phi = 0.4). When looking just at the "endpoints" version, 67.9% made case-based decisions consistent with their thresholds, 24.7% escalated their decision, and 7.4% de-escalated their decision ($chi^2(4, 162) = 11.70$, phi = 0.3). No statistically significant differences were found between experience levels, with all three subgroups (students, residents, and faculty) escalating their actions.

Discussion

These results help us explore how subjective uncertainty guides decision-making and the establishment of testing and treatment thresholds using evidence-based medicine theory. One could hypothesize that with increasing experience comes a higher tolerance for ambiguity and a decrease in cognitive discomfort. For example, if a physician has treated pneumonia for 10 years, she has seen enough patients to develop the confidence needed to make decisions related to the diagnosis and management of the disease process. One would therefore expect the experienced physician to utilize less testing, having a higher testing threshold, a lower treatment threshold, and a tighter threshold range for a common disease process such as pneumonia. However, those with less clinical exposure, may not have seen enough patients to have acquired that comfort level.

Our study demonstrated that faculty had the highest testing threshold, followed by residents and then students. This is what we would have predicted given those with more clinical experience should be more comfortable (lower probability threshold) making a decision to not treat without having to order another test to validate their decision. The treatment threshold however, demonstrated that residents had the lowest threshold followed by faculty and then students. We would have predicted that faculty would have had the lowest treatment threshold, however faculty and resident means were very similar and therefore may not be meaningfully different. Residents may have had enough clinical exposure at that point in their training to match faculty comfort in offering treatment without having to order additional testing. The larger discrepancy between residents and faculty with regards to the testing threshold may suggest that growing experience impacts the development of testing and treatment thresholds at different rates. The clinical context may also lead clinicians to treat a patient with antibiotics for possible bacterial pneumonia even if a subset of those patients would not have required treatment, rather than running the risk of not prescribing antibiotics to a patient who would benefit from treatment.

As predicted, faculty had the smallest threshold span, followed by residents and then students. Students overwhelmingly had the largest span, showing that inexperience tends to lead to more discomfort with uncertainty, a finding described in our previous work as well.[16] This greater intolerance for ambiguity is important to recognize, as it is associated with reduced psychological well-being and calls to support our learners as they acquire the necessary medical knowledge, clinical reasoning skills and experience, to grapple with uncertainty in the clinical environment.[10]

We examined consistency in decision-making by comparing predicted decisions, based on thresholds initially provided by the respondents, with actual decisions made when participants were given a 20% probability of a superimposed bacterial infection. Recall bias would assume that participants would make decisions based on the thresholds to which they had previously committed. However, roughly one third of respondents were inconsistent with their stated thresholds in their decision-making (a medium to large effect size). For those who were inconsistent, participants were more likely to err on the side of testing or treating than predicted by their initial thresholds. One reason for this could be that clinical reasoning is impacted by consideration for the individual patient being cared for, resulting in a more cautious decision rather than running the risk of missing a serious diagnosis with the potential for a bad outcome.

Antibiotic stewardship is likely a more abstract construct and acting for the good of the 'public' may not seem as important in the moment.

The framing effects reflected much of the related literature. Individuals are inapropriately distracted by the way decision tasks are framed and make different judgments when confronted with logically equivalent alternatives.[18,19] Our finding that the framing effect interacts with experience level may suggest that this bias is less influential for those with greater experience and expertise.

It might also be noted that the variability of these threshold estimates are highly variable among these subjects. We believe this reflects the combination of informational (individual) and intrinsic uncertainty. Intrinsic uncertainty in clinical care could conceivably be shared by other clinicians and have a relatively small amount of variability. Informational uncertainty, in contrast, is idosyncratic, reflecting lack of knowledge, inexperience, or biased experiences. This form of uncertainty will be highly situational and increase the variability of group estiamtes.[20]

A very useful extension of this study would be to examine the change in and impact of uncertainty as a result of experience by following subjects over time. A longitudinal study would help to identify how much variability in these threshold estimates might be attributed to the case and how much to the individual and individual characteristics, like experience and tolerance for ambiguity.

Our study has several limitations. Participants were from a single institution with its own culture and protocols. However, a majority of faculty and residents likely trained at other instituions making this less of an issue. We only asked participants for testing and treatment thresholds for a single case of a suspected superimposed bacterial pneumonia under "laboratory" conditions.. A case with greater complexity or a case with a less common presentation may have resulted in different findings. We had a lower response rate of faculty and residents when compared to students. Students were required to participate as part of an already established educational activity whereas faculty and residents were volunteers. Finally, we did not discriminate residents based on level of training which may have reduced the precision of our estimates of the impact of experience. Some residents may have been in the beginning of their training, making their level of clinical experience more like that of a student, whereas others may have been at the end of their training making their level of clinical experience more like that of faculty. Finally, this is a tightly constructed problem that avoids much of the complexity of an actual clinical case, which may alter the dynamics of uncertainty management.

Defining testing and treatment thresholds are just one part of the larger, and largely neglected area of management reasoning.[21] There are many more aspects of management reasoning that warrant attention. Even in the narrow implementation of the present paper, many questions remain, such as; what is the mechanism by which framing of the case causes differing threshold estimates? How do testing and treatment thresholds change with characteristics of the disease or the patient? Are individual clinicians systematically more or less comfortable with uncertainty or does this comfort level vary with the clinical context?

In addition, while our results demonstrate a relationship between testing and treatment thresholds and experience in medicine, there are larger economical implications of this work, including the impact of cost within medicine due to excessive testing, especially at academic medical centers. Each additional test ordered may help clinicians and particuarly learners deal with uncertainty and ambiguity, but there is a point in testing where the return on investment is limited and testing is only mitigating the psychologcal distress of the health care provider. This work is also a call to further explore if consistency in clinical decision making can be taught and whether there are experiences or tools such as best pratice alerts or decision rules ultilizing evidence-based medicine that can augment an individual's decisions. If successful, this may be a strategy that enhances both educational programs as well as patient outcomes.

Acknowledgements and Disclosures:

-- Acknowledgments: The authors wish to thank Laurie Whitman for help with the data gathering.

-- Funding/Support: None

-- Conflicts of Interest: None

-- Other disclosures: None

-- Ethical approval: This study received exemption status from the University of Michigan Institutional Review Board.

-- Disclaimers: None

-- Previous presentations: None

References

1. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1974;185:1124–1131.
2. Gerrity MS, Earp JAL, DeVellis RF, Light DW. Uncertainty and professional work: Perceptions of physicians in clinical practice. Am J Sociol. 1992;97(4):1022–1051.
3. Wray CM, Loo LK. The diagnosis, prognosis, and treatment of medical uncertainty. J Grad Med Educ. 2015;7:523–527.

4. Simpkin AL, Schwartzstein RM. Tolerating Uncertainty — The Next Medical Revolution? *N Engl J Med*. 2016;375(18):1713-1715.
5. Han PKJ, Klein WMP, Arora NK. Varieties of uncertainty in health care: A conceptual taxonomy. Med Decis Making. 2011;31(6):828–838.
6. Felder S, Mayrhofer T. Risk preferences: consequences for test and treatment thresholds and optimal cutoffs. *Med Decis Making*. 2014;34(1):33-41.
7. Berlyne DE. Uncertainty and epistemic curiosity. *Br J Psychol*. 1962;53(1):27-34. papers3://publication/uuid/F065C792-A2EA-4AB6-9D69-1CF022AD8240
8. Litman J a. Interest and deprivation factors of epistemic curiosity. *Pers Individ Dif*. 2008;44:1585-1595. doi:10.1016/j.paid.2008.01.014
9. Dyche L, Epstein RM. Curiosity and medical education. *Med Educ*. 2011;45(7):663-668. doi:10.1111/j.1365-2923.2011.03944.
10. Hancock J, Mattick K. Tolerance of ambiguity and psychological well-being in medical training: A systematic review. Med Educ. 2020;54(2):125–137.
11. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. N Engl J Med. 1989;320:1489–1491.
12. Sackett DL, Straus SE, Richardson WS, Rosenberg WM, Haynes B. Evidence Based Medicine: How to Practice and Teach EBM. 2nd ed. Toronto, Ontario, Canada: Churchill Livingstone; 2000.

13. Hawkins RC. The evidence based medicine approach to diagnostic testing: Practicalities and limitations. Clin Biochem Rev. 2005;26:7–18.

14. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. N Engl J Med. 1980;302:1109–1117.

15. Hillen MA, Gutheil CM, Strout TD, Smets EMA, Han PKJ. Tolerance of uncertainty: Conceptual analysis, integrative model, and implications for healthcare. Soc Sci Med. 2017;180:62-75.

16. Stojan JN, Daniel M, Morgan HK, Gruppen LD. A randomized cohort study of diagnostic and therapeutic thresholds in medical student clinical reasoning. academic medicine. 92(11S): S43-S47, 2017.

17. Daniel M, Rencic J, Durning SJ, et al. Clinical Reasoning Assessment Methods. Acad Med. 2019;94:902-912.

18. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science*.* 1981;211(4481):453-458.

19. Siminoff LA, Fetting JH. Effects of outcome framing on treatment decisions in the real world: impact of framing on adjuvant breast cancer decisions. Med Decis Making. 1989;9(4):262-71.

20. Wray CM, Loo LK. The Diagnosis, Prognosis, and Treatment of Medical Uncertainty. *J Grad Med Educ*. 2015;7(December):523-527.

21. Cook DA, Sherbino J, Durning SJ. Management Reasoning. JAMA. 2018;319(22):2267.

Tables and Figures

Figure 1: Clinical decision making threshold model. (Adapted from Sackett, et al.)[7]

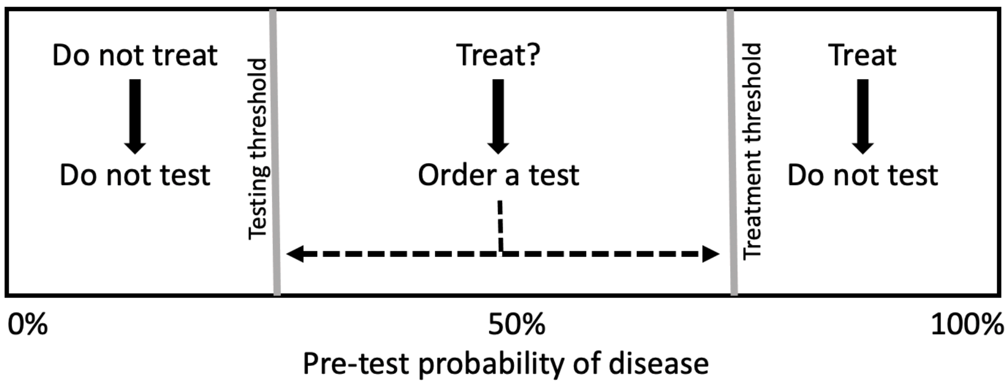

Table 1: Mean Testing Thresholds of Students, Residents and Faculty Based on Framing of the Clinical Scenario.

Version	Student Mean (s.d.)	Resident Mean (s.d.)	Faculty Mean (s.d.)	Version mean Mean (s.d.)
"Range"	18.4 (12.0)	18.4 (11.2)	23.7 (8.9)	19.4 (11.3)
"Endpoints"	13.3 (19.0)	24.2 (28.9)	27.5 (27.1)	19.4 (24.5)
Group mean	15.8 (16.1)	20.6 (20.0)	25.8 (20.8)	eta^2 (version x group) = 0.06* eta^2 (version) = 0.00 eta^2 (group) = 0.04*

* $p<0.05$

Table 2: Mean Treatment Thresholds of Students, Residents, and Faculty Based on Framing of the Clinical Scenario.

Version	Student Mean (s.d.)	Resident Mean (s.d.)	Faculty Mean (s.d.)	Version mean
"Range"	74.5 (14.6)	68.3 (19.1)	68.8 (14.4)	71.2 (16.5)
"Endpoints"	82.6 (8.7)	77.0 (18.3)	77.2 (16.7)	79.9 (13.9)
Group mean	78.5 (12.6)	71.6 (19.2)	73.4 (16.1)	eta^2 (version x group) = 0.11* eta^2 (version) = 0.07* eta^2 (group) = 0.04*

* $p<0.05$

Table 3: Mean Threshold Span from Testing Decision to Treatment Decision of Students, Residents and Faculty Based on Framing of the Clinical Scenario.

Version	Student Mean (s.d.)	Resident Mean (s.d.)	Faculty Mean (s.d.)	Version mean
"Range"	56.1 (19.2)	49.9 (20.0)	45.1 (13.8)	51.9 (19.0)
"Endpoints"	69.3 (19.7)	52.8 (29.1)	49.7 (26.2)	60.5 (25.4)
Group mean	62.7 (20.5)	51.0 (23.8)	47.6 (21.4)	eta^2 (version x group) = 0.13* eta^2 (version) = 0.04* eta^2 (group) = 0.09*

* $p<0.05$

Table 4: Comparison of predicted decisions, based on thresholds provided by the respondents, with the actual decision made for a case vignette with probability of a superimposed bacterial infection of 20%.

Overall:

Actual decision	Predicted decision		
	Not treat N (%)	Test N (%)	Treat N (%)
Overall (framing conditions combined), $chi^2(4, 342) = 44.60$, phi =0.4†*			
Not treat	42 (12.3)	25 (7.3)	0
Test	52 (15.2)	178 (52.1)	1 (0.3)
Treat	8 (2.3)	36 (10.5)	0
Framing with the "range" version, $chi^2(4, 180) = 30.27$, phi = 0.4*			
Not treat	32 (17.8)	13 (7.2)	0
Test	29 (16.1)	78 (43.3)	1 (0.6)
Treat	6 (3.3)	21 (11.7)	0
Framing with the "endpoints" version, $chi^2(4, 162) = 11.70$, phi = 0.3*			
Not treat	10 (6.2)	12 (7.4)	0
Test	23 (14.2)	100 (61.7)	0
Treat	2 (1.2)	15 (9.3)	0

*p<0.5

†phi effect size interpretation: 0.1 small, 0.3 medium, 0.5 large