

Accuracy of Practitioner Estimates of Probability of Diagnosis Before and After Testing

Daniel J. Morgan, MD, MS; Lisa Pineles, MA; Jill Owczarzak, PhD; Larry Magder, PhD; Laura Scherer, PhD; Jessica P. Brown, PhD; Chris Pfeiffer, MD, MHS; Chris Terndrup, MD; Luci Leykum, MD, MBA; David Feldstein, MD; Andrew Foy, MD; Deborah Stevens, LCSW-C, MPH; Christina Koch, MD; Max Masnick, PhD; Scott Weisenberg, MD; Deborah Korenstein, MD

IMPORTANCE Accurate diagnosis is essential to proper patient care.

OBJECTIVE To explore practitioner understanding of diagnostic reasoning.

DESIGN, SETTING, AND PARTICIPANTS In this survey study, 723 practitioners at outpatient clinics in 8 US states were asked to estimate the probability of disease for 4 scenarios common in primary care (pneumonia, cardiac ischemia, breast cancer screening, and urinary tract infection) and the association of positive and negative test results with disease probability from June 1, 2018, to November 26, 2019. Of these practitioners, 585 responded to the survey, and 553 answered all of the questions. An expert panel developed the survey and determined correct responses based on literature review.

RESULTS A total of 553 (290 resident physicians, 202 attending physicians, and 61 nurse practitioners and physician assistants) of 723 practitioners (76.5%) fully completed the survey (median age, 32 years; interquartile range, 29-44 years; 293 female [53.0%]; 296 [53.5%] White). Pretest probability was overestimated in all scenarios. Probabilities of disease after positive results were overestimated as follows: pneumonia after positive radiology results, 95% (evidence range, 46%-65%; comparison $P < .001$); breast cancer after positive mammography results, 50% (evidence range, 3%-9%; $P < .001$); cardiac ischemia after positive stress test result, 70% (evidence range, 2%-11%; $P < .001$); and urinary tract infection after positive urine culture result, 80% (evidence range, 0%-8.3%; $P < .001$). Overestimates of probability of disease with negative results were also observed as follows: pneumonia after negative radiography results, 50% (evidence range, 10%-19%; $P < .001$); breast cancer after negative mammography results, 5% (evidence range, <0.05%; $P < .001$); cardiac ischemia after negative stress test result, 5% (evidence range, 0.43%-2.5%; $P < .001$); and urinary tract infection after negative urine culture result, 5% (evidence range, 0%-0.11%; $P < .001$). Probability adjustments in response to test results varied from accurate to overestimates of risk by type of test (imputed median positive and negative likelihood ratios [LRs] for practitioners for chest radiography for pneumonia: positive LR, 4.8; evidence, 2.6; negative LR, 0.3; evidence, 0.3; mammography for breast cancer: positive LR, 44.3; evidence range, 13.0-33.0; negative LR, 1.0; evidence range, 0.05-0.24; exercise stress test for cardiac ischemia: positive LR, 21.0; evidence range, 2.0-2.7; negative LR, 0.6; evidence range, 0.5-0.6; urine culture for urinary tract infection: positive LR, 9.0; evidence, 9.0; negative LR, 0.1; evidence, 0.1).

CONCLUSIONS AND RELEVANCE This survey study suggests that for common diseases and tests, practitioners overestimate the probability of disease before and after testing. Pretest probability was overestimated in all scenarios, whereas adjustment in probability after a positive or negative result varied by test. Widespread overestimates of the probability of disease likely contribute to overdiagnosis and overuse.

JAMA Intern Med. 2021;181(6):747-755. doi:10.1001/jamainternmed.2021.0269
Published online April 5, 2021.

← Invited Commentary page 756

+ Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Daniel J. Morgan, MD, MS, Department of Epidemiology and Public Health, University of Maryland School of Medicine, 10 S Pine St, Medical Student Teaching Facility Room 334, Baltimore, MD 21201 (dmorgan@som.umaryland.edu).

Diagnosis of disease is complex and taught using estimated probabilities based on the patient's history, physical examination findings, and diagnostic test results.¹⁻³ Correct ordering and interpretation of tests are increasingly important given the increase in the number and complexity of tests, with more than 14 billion tests performed yearly in the US alone.⁴ Although practitioners are taught to estimate pretest probability and to apply the sensitivity and specificity of a test to interpret a positive or negative result, data suggest that historically most practitioners perform poorly on assessments of these skills and do not use these approaches in day-to-day practice.⁵⁻¹¹

Test ordering and interpretation are taught briefly in medical schools,¹² with curricular evaluation often limited to self-assessment of skills.¹³ The impact of such education on clinical practice is unclear. Estimating the probability of disease and deciding to test may be influenced by training, experience, and personality.^{8,14} Medical decisions, like other human decisions, may not be rational and are prone to errors associated with poor knowledge of the base rate of disease or other errors associated with probability.¹⁴ Test performance and interpretation have increasingly become a point of discussion in medicine and for the general public during the COVID-19 pandemic.¹⁵ Erroneous estimates of disease probability likely impact practitioner treatment decisions.^{3,16} Lack of accurate diagnostic reasoning may lead to overdiagnosis and overtreatment.¹⁷

Few studies have systematically examined how practitioners interpret diagnostic test results within the context of actual clinical scenarios. We performed a multicenter survey of practitioners in primary care practice to explore practitioner understanding of the probability of disease before and after test results for common clinical scenarios.

Methods

Survey

We developed a survey to assess practitioner test understanding and the process of making a diagnosis using probability as well as actions taken by practitioners in similar scenarios in their practice. The survey also included items regarding basic demographic characteristics, educational background, and practice setting. Institutional review board approval was obtained at each of the 3 coordinating sites (Baltimore, Maryland; San Antonio, Texas; and Portland, Oregon). Verbal informed consent with a waiver of documentation was approved at all sites. The study followed the American Association for Public Opinion Research (AAPOR) reporting guideline.

A draft survey was developed by primary investigators (D.J.M., L.L., D.K., D.F., L.S., J.P.B., A.F., S.W., C.P., J.O., and L.P.) based in part on previous surveys of risk understanding.^{5,8-11,18} This survey was reviewed by an expert panel of practitioners with different areas of expertise, practicing in community and academic settings (D.J.M., L.L., D.F., A.F., S.W., and D.K.), a qualitative research expert (J.O.), an epidemiologist (J.P.B.) and a psychologist (L.S.) with expertise in survey design, and a senior biostatistician (L.M.). The survey

Key Points

Question Do practitioners understand the probability of common clinical diagnoses?

Findings In this survey study of 553 practitioners performing primary care, respondents overestimated the probability of diagnosis before and after testing. This posttest overestimation was associated with consistent overestimates of pretest probability and overestimates of disease after specific diagnostic test results.

Meaning These findings suggest that many practitioners are unaccustomed to using probability in diagnosis and clinical practice. Widespread overestimates of the probability of disease likely contribute to overdiagnosis and overuse.

was further revised by the expert panel during an in-person meeting and 2 conference calls. A pilot test of the survey was conducted with 10 practitioners for comprehension and interpretation of questions, and minor language adjustments were made.

Practitioner Risk Understanding

The survey assessed risk understanding for common testing clinical decisions encountered by primary care practitioners in routine scenarios similar to previous small surveys.^{8-11,18} Individual testing questions pertained to mammograms for breast cancer, stress testing for cardiac ischemia, chest radiography for pneumonia, and urine cultures for urinary tract infection (UTI) (eAppendix 1 in the [Supplement](#)).

Practitioners were presented with a clinical scenario and asked to estimate pretest probability of disease and posttest probabilities after both positive and negative test results. Each scenario was created for a general situation but included essential details to calculate true risk for patients (eg, age and absence of any risk factors for breast cancer in mammogram screening questions). The primary outcome of testing questions was to accurately identify the probability that a patient had disease after positive or negative results. Questions were designed to assess whether errors in test interpretation associated with poor pretest estimates or inaccurate updating of probability after testing. Additional questions provided sensitivity and specificity of a theoretical test and asked participants to calculate positive and negative predictive value at particular levels of disease prevalence.

To assess the accuracy of participant responses, we used a hierarchical method to identify the scientific evidence for pretest probability, sensitivity, and specificity from the literature, which was completed after survey finalization. We first reviewed high-quality recent systematic reviews and meta-analyses. If only older systematic reviews and meta-analyses were available, with newer high-impact studies after publication, we considered data from both (attempting to understand the most accurate numbers for current technology and practice). If no systematic reviews or meta-analyses were available, we used data from studies commonly cited in recent guidelines, creating weighted means by consensus. The expert panel of physicians overseeing the study was presented

Table 1. Survey Responses

Variable	No. (%) of practitioners			
	Maryland and Middle Atlantic states	Oregon and Washington	Texas	All sites
Invited to participate	390	150	183	723
No response	41 (11)	0	16 (9)	57 (8)
Refusals ^a	10 (3)	3 (2)	3 (2)	16 (2)
Not interested	2	0	2	4
Too busy or bad timing	6	2	2	10
Too difficult	2	1	0	3
Other	3	0	0	3
Agreed to participate (of all invited)	339 (87)	147 (98)	164 (90)	650 (90)
Agreed but did not complete survey	27 (7)	23 (15)	15 (8)	65
Total surveys received	312 (80)	124 (83)	149 (81)	585 (81)
Failed to complete ≥ 1 questions required for final analysis	7 (2)	12 (8)	13 (7)	32 (4)

^a May list more than 1 reason for refusing to complete the survey.

with the best evidence identified, had a comment and question period, and determined consensus evidence-based answers presented in the Results section (eAppendix 2 in the Supplement).

Enrollment Procedure

People in leadership positions for group practices or residency programs were contacted and informed of the study. Investigators sought permission to give a short presentation or email introduction that described the study during a group practice meeting. Individual practitioners were then approached by a coordinator and/or physician investigator to request participation. The survey was offered to 723 primary care physicians, nurse practitioners, and physician assistants practicing in Delaware, Maryland, Oregon, Pennsylvania, Texas, Virginia, Washington, and the District of Columbia (Table 1). The survey was administered in paper format. The coordinator generally remained at the clinic, office, or meeting location until the practitioner had completed the survey. If practitioners requested to complete the survey at a later date, they were provided with an addressed, stamped envelope and could return the survey by mail, email, or clinic drop-off. Respondents were provided with a US \$50 gift card for completion, if permitted by their employer.

Practitioners who initially agreed to participate but did not return the survey within 2 weeks were contacted by study staff via email and/or in person up to 5 times during 3 months. Practitioners who did not complete the survey after these subsequent contacts were considered nonparticipants. Practitioners who declined to participate at initial enrollment or after reminders were asked to provide a reason for not participating from a standardized list to assess for selection bias. Of the contacted practitioners, 585 responded to the survey, and 553 answered all the questions.

Imputed Likelihood Ratios

To understand the adjustment in probability of disease after a positive or negative test result, we calculated an imputed likelihood ratio.¹⁹ By comparing estimated probability of disease before and after testing, we could impute the likelihood ratio

that was consciously or unconsciously applied to modify probabilities. The imputed likelihood ratio was calculated by dividing posttest odds by pretest odds, where odds were calculated as probability divided by 1 minus probability.¹⁹ Responses of 0% or 100% were modified to 0.1% and 99.9% to allow for calculation of a likelihood ratio. Likelihood ratios were estimated from the literature as described above by the expert panel of physicians (eAppendix 2 in the Supplement).

Statistical Analysis

Survey responses were entered into a REDCap (Research Electronic Data Capture) database with double data entry. A sample size of 500 was planned based on desire for generalizable results across enrollment sites. The target sample was surpassed while we collected outstanding surveys. Data were analyzed with R software (R Foundation for Statistical Computing) for creation of density plots. SAS statistical software, version 9.4 (SAS Institute Inc) was used for calculation of descriptive statistics and all other statistical analyses. Comparison of those who completed all key survey questions with those who did not was performed with the χ^2 test. To assess the statistical significance of differences between respondent estimates of diagnostic probabilities and the probabilities determined from scientific evidence, we used Wilcoxon signed-rank tests. To display the range of results for estimates of probability, we used density plots. These were created using R software (GGPlot2). A 2-sided $P < .05$ was considered to be statistically significant.

Results

Participant Demographics

A total of 553 of 723 practitioners (76.5%) fully completed the survey (median age, 32 years; interquartile range, 29-44 years; 293 female [53.0%]; 296 [53.5%] White) from June 1, 2018, to November 26, 2019 (Table 2). A total of 492 of the 553 respondents (89.0%) had MD or DO degrees, and 290 (52.4% were in residency). The survey required a median of 20 minutes to complete (interquartile range [IQR], 15-25 minutes).

Table 2. Variables Associated With Practice Among Enrolled Practitioners

Variable	No. (%) of respondents
Practitioner type	
MD or equivalent	492 (89.0)
NP	48 (8.7)
PA	13 (2.4)
Race/ethnicity	
White	296 (53.5)
Black	37 (6.7)
Asian	142 (25.7)
Hispanic/Latino	45 (8.1)
>1 Race/ethnicity	19 (3.4)
Other or missing	14 (2.5)
Female sex	293 (53.0)
Age, median (IQR), y	32 (29.0-44.0)
Medical, nursing, or PA school	
International	113 (20.4)
Osteopathy school	21 (3.8)
Current resident	290 (52.4)
Type of residency	
Internal medicine	336 (60.8)
Family medicine	142 (25.7)
Other or NA	75 (13.6)
Type of practice (n = 633; may be >1 type)	
Academic	343 (54.2)
Rural	7 (1.1)
Suburban	60 (9.5)
Urban	83 (13.1)
VA	141 (22.3)
Ever sued for malpractice	31 (5.6)
Other graduate degree	115 (20.8)
Time in practice, median (IQR), y	
Resident	1.5 (1.0-3.0)
Nonresident	11 (5.0-22.0)

Abbreviation: IQR, interquartile range; MD, physician; NA, not applicable; NP, nurse practitioner; PA, physician assistant; VA, Veterans Affairs.

We compared the 32 respondents who did not complete all necessary questions with the final cohort of 553 practitioners with complete responses. We found that those not completing the survey were more likely to be female (26 [81.3%] noncompleters vs 293 [53.0%] final cohort, $P < .001$), to have been in practice more than 10 years (15 [46.9%] noncompleters vs 145 [26.2%] final cohort, $P = .01$), to be nonresidents (27 [84.4%] noncompleters vs 263 [47.6%] final cohort, $P < .001$), or to be nurse practitioners or physicians assistants (13 [40.6%] noncompleters vs 61 [11.0%] final cohort, $P < .001$).

Estimates of Disease Probability

Estimates of probability of disease were consistently higher than scientific evidence (Figure). We also broke down answers by type of practitioner (resident physician, attending physician, and nurse practitioner or physician assistant) (Table 3). All types of practitioners overestimated probability of disease before and after testing.

For pneumonia, the median clinical scenario-based estimate of pretest probability by participants was 80% (IQR, 75%-

90%; evidence range, 25%-42%; $P < .001$). Median estimated probability of pneumonia was 95% (IQR, 90%-100%; evidence range, 46%-65%; $P < .001$) after a positive radiology result and 50% (IQR, 30%-80%; evidence range, 10%-19%; $P < .001$) after a negative radiology result. After a positive radiology result, 551 practitioners (99.6%) would treat with antibiotics, whereas 401 (72.5%) would treat with antibiotics after a negative radiology result.

For breast cancer, the clinical scenario-based estimate of pretest probability by participants was 5% (IQR, 1%-10%; evidence range, 0.2%-0.3%; $P < .001$). Median estimated probability of breast cancer was 50% (IQR, 30%-80%; evidence range, 3%-9%; $P < .001$) after a positive mammography result and 5% (IQR, 1%-10%; evidence range, <0.05%; $P < .001$) after a negative mammography result.

For cardiac ischemia, the median clinical scenario-based estimate of pretest probability by participants was 10% (IQR, 5%-20%; evidence range, 1%-4.4%; $P < .001$). The median estimated probability of cardiac ischemia was 70% (IQR, 50%-90%; evidence range, 2%-11%; $P < .001$) after a positive exercise stress test result and 5% (IQR, 1%-10%; evidence range, 0.43%-2.5%; $P < .001$) after a negative exercise stress test result. After a positive test result, 432 (78.1%) would treat for cardiac ischemia.

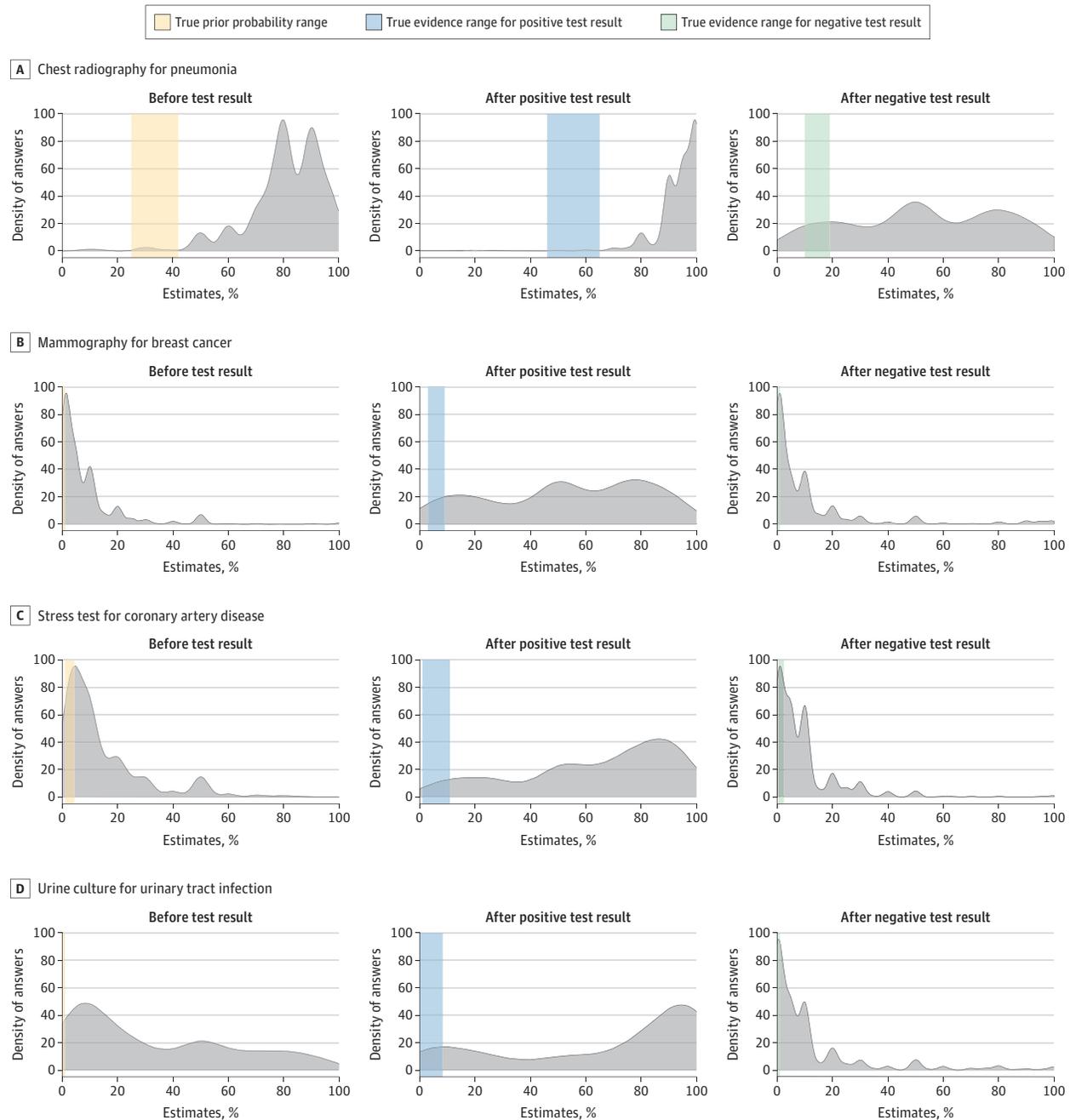
For UTI, the description was of asymptomatic bacteriuria. The median clinical scenario-based estimate of pretest probability by participants was 20% (IQR, 10%-50%; evidence range, 0%-1%; $P < .001$). The median estimated probability of a UTI was 80% (IQR, 30%-95%; evidence range, 0%-8.3%; $P < .001$) after a positive urine culture result and 5% (IQR, 0%-10%; evidence range, 0%-0.11%; $P < .001$) after a negative urine culture result. After a positive test result, 393 (71.1%) would treat with antibiotics. After a negative test result, 43 practitioners (7.8%) would treat with antibiotics.

Scenarios requesting identical test interpretation based on hypothetical numbers revealed similar tendencies. For the question, "A test to detect a disease for which prevalence is 1 out of 1000 has a sensitivity of 100% and specificity of 95%. What is the chance that a person found to have a *positive* result actually has the disease?" the median answer was 95% (IQR, 95%-100%), whereas the correct answer was 2%. For the related question, "What is the chance that a person found to have a *negative* result actually has the disease?" the median answer was 5% (IQR, 0%-5%), whereas the correct answer was 0%.

Imputed Likelihood Ratios

Imputed likelihood ratios were of variable accuracy across clinical scenarios. The most accurate were those for the impact of chest radiography for the diagnosis of pneumonia and urine culture for the diagnosis of UTI; the least accurate were those for negative mammography results for breast cancer and positive exercise stress test results for cardiac ischemia (imputed median positive and negative likelihood ratios for practitioners for chest radiography for pneumonia: positive likelihood ratio, 4.8; evidence, 2.6; negative likelihood ratio, 0.3; evidence, 0.3; those for mammography for breast cancer: positive likelihood ratio, 44.3; evidence range, 13.0-33.0; negative likelihood ratio, 1.0;

Figure. Distribution of Practitioner Assessments of Probability of Disease Before Testing and After Positive or Negative Test Results for 4 Testing Questions Representing Scenarios Commonly Encountered in Primary Care



A, Scenario: a previously healthy 35-year-old woman who smokes tobacco presents with 5 days of fatigue, productive cough, worsening shortness of breath, temperatures to 38.9°C, and decreased breath sounds in the lower right field. She has a heart rate of 105 beats/min, but vital signs are otherwise normal. B, Scenario: a 45-year-old woman comes in for an annual visit. She has no specific risk factors or symptoms for breast cancer. C, Scenario: a 43-year-old

premenopausal woman presents with atypical chest pain and normal ECG results. She has no risk factors and has normal vital signs and examination findings. D, Scenario: a 65-year-old man is seen for osteoarthritis. He has noted foul-smelling urine and no pain or difficulty with urination. A urine dipstick shows trace blood. ECG indicates electrocardiography.

evidence range, 0.05-0.24; those for exercise stress test for cardiac ischemia: positive likelihood ratio, 21.0; evidence range, 2.0-2.7; negative likelihood ratio, 0.6; evidence range, 0.5-0.6; those for urine culture for urinary tract infection: positive like-

likelihood ratio, 9.0; evidence, 9.0: negative likelihood ratio, 0.1; evidence, 0.1.) (Table 4). Estimates of probability and imputed likelihood ratios were similar between residents and primary care practitioners (Table 4).

Table 3. Estimates of Probability of Disease Before Testing and After Positive or Negative Test Results for 5 Testing Questions

Clinical scenario	Scientific evidence range, %	Median (IQR)		
		Resident physician estimate, % (n = 290)	Attending physician estimate, % (n = 202)	Nurse practitioner or physician assistant estimate, % (n = 61)
Pneumonia				
Pretest probability	25-42	80 (75-90)	85 (80-90)	80 (70-90)
After positive test result	46-65	95 (90-99)	95 (95-100)	95 (90-100)
After negative test result	10-19	60 (40-80)	50 (20-80)	50 (20-50)
Breast cancer				
Pretest probability	0.2-0.3	5 (1-10)	2 (1-10)	10 (5-20)
After positive test result	3-9	60 (35-75)	50 (20-80)	60 (50-80)
After negative test result	<0.05	5 (1-10)	1 (1-10)	10 (2-20)
Cardiac ischemia				
Pretest probability	1-4.4	10 (5-20)	5 (3-10)	15 (6.25-30)
After positive test result	2-11	75 (50-90)	60 (25-80)	90 (60-95)
After negative test result	0.43-2.5	5 (1-10)	5 (1-10)	10 (5-20)
Urinary tract infection				
Pretest probability	0-1	25 (10-60)	20 (5-50)	30 (10-50)
After positive test result	0-8.3	77.5 (25-95)	90 (40-98)	90 (75-100)
After negative test result	0-0.11	5 (0.1-20)	5 (0-10)	5 (0-10)
Hypothetical testing situation				
After positive test result	2	95 (95-95)	95 (80-100)	95 (95-100)
After negative test result	0	2 (0-10)	5 (0-5)	5 (5-75)

Abbreviation: IQR, interquartile range.

Table 4. Imputed Positive and Negative Likelihood Ratios Calculated for Each Practitioner Based on Their Pretest and Posttest Positive or Negative Responses

Test and LR result	Scientific evidence of likelihood ratio	Imputed LR, median (IQR)		
		Resident physician	Attending physician	Nurse practitioner or physician assistant
Chest radiography for pneumonia				
Positive	2.57	4.75 (2.25-11.00)	5.21 (2.25-52.58)	9.00 (2.15-111.00)
Negative	0.33	0.35 (0.12-0.67)	0.21 (0.06-0.44)	0.25 (0.06-0.42)
Mammography for breast cancer				
Positive	13.00-33.00	36.00 (9.00-196.00)	54.79 (13.22-428.14)	19.00 (4.00-49.00)
Negative	0.05-0.24	1.00 (0.47-1.26)	1.00 (0.44-1.00)	1.00 (0.33-1.56)
Exercise stress test for cardiac ischemia				
Positive	2.03-2.65	22.67 (8.12-81.00)	19.00 (5.44-73.50)	51.00 (9.00-218.50)
Negative	0.51-0.56	0.50 (0.17-1.00)	0.58 (0.11-1.00)	0.59 (0.26-1.00)
Urine culture for UTI				
Positive	9.00	5.10 (1.89-44.33)	16.86 (3.00-147.00)	27.00 (8.14-333.00)
Negative	0.11	0.16 (0.02-0.58)	0.11 (0.01-0.47)	0.12 (0.01-0.63)

Abbreviations: IQR, interquartile range; LR, likelihood ratio; UTI, urinary tract infection.

Discussion

In this survey study, in scenarios commonly encountered in primary care practice, practitioners overestimated the probability of disease by 2 to 10 times compared with the scientific evidence, both before and after testing. This result was mostly associated with overestimates of pretest probability,

which were observed across all scenarios. Adjustments to probability in response to test results varied from accurate to overestimates of risk by type of test. There was variation in accuracy between type of practitioner that was small compared with the magnitude of difference between practitioners and the scientific evidence. Many practitioners reported that they would treat patients for disease for which likelihood had been overestimated.

The most striking finding from this study was that practitioners consistently and significantly overestimate the likelihood of disease. Small studies with limited generalizability have had similar findings, often asking practitioners to perform one isolated aspect of diagnosis, such as interpreting a test result. However, past studies⁸⁻¹¹ have not explored a range of questions or clarified estimates at different steps in the diagnostic pathway. The reason for inaccurate estimates of probability are not clear, although anecdotes reported during the current study imply that practitioners often do not think in terms of probability. One participant stated that estimating probability of disease “isn’t how you do medicine.” This attitude is consistent with a previous study²⁰ of diagnostic strategies that describe an initial pattern recognition phase of care with only 10% of practitioners engaging in a secondary phase of probabilistic reasoning.

This study found that probability estimates were consistently biased toward overestimation, as has been seen in other contexts, such as expectations of high stock returns among investors.²¹ This overestimation is consistent with cognitive biases, including base rate neglect, anchoring bias, and confirmation bias.¹⁴ These biases drive overestimation because true base rates are usually lower than expected and anchoring tends to reflect experiences that represent improbable events or those in which a diagnosis was missed. Such cognitive biases have been associated with diagnostic errors that may occur from errors in estimating risk.^{5,22,23} Notably, practitioners in this survey were often residents or academic physicians who often practice with populations with higher prevalence of disease. This experience may have also contributed to higher estimates of disease.

Pretest probabilities were consistently overestimated for all questions, but overestimates were particularly apparent for the pneumonia and UTI scenarios. Estimates of pretest probability generally reflect clinical knowledge. Reasons for overestimates for these infectious diseases may relate to the fact that antibiotics are often appropriately given even when the likelihood of infection is moderate. In the UTI scenario, estimates of high pretest probability may reflect the evolution of the definition of asymptomatic bacteriuria as a separate entity from UTI.²⁴

In contrast to past literature,^{8-10,19} practitioners accurately adjusted estimates of disease based on the results of some tests, as demonstrated by the imputed likelihood ratios. This adjustment could be artifactual because of inability to adjust probability for tests that had high pretest estimates (ie, pneumonia and UTI). In other cases, practitioners markedly overestimated the probability of disease after testing, specifically after a positive or negative mammography result or a positive exercise stress test result. Practitioners are known to overestimate chance of disease when completing a theoretical estimate of likelihood of disease after a positive test result when pretest probability was 1 in 1000 tests.^{9,10} The current study included the identical question with an identical response, with participants estimating the likelihood of disease at 95% when the correct answer was 2%.^{5,8-10,19} The findings regarding real-life examples are consistent with evidence from limited past studies,⁸⁻¹¹ for example, physician interpre-

tation of a positive mammography result in a typical woman as conveying 81% probability of breast cancer.⁸

The assessment of test results in this study was simplified to positive or negative. This dichotomization reflects the literature on the sensitivity and specificity of testing.^{5,6} However, in clinical medicine, these tests often present a range of descriptions for a positive result from mild positives, such as well-circumscribed density on a mammogram, to a strongly positive result, such as inducible ischemia on a stress test or spiculated mass on a mammogram. A more strongly positive or abnormal result would be less sensitive but more specific for disease. This study did not evaluate interpretation of more complex test results.

There are important implications of the finding of a gap between practitioner estimates and scientific estimates of the probability of disease. Practitioners who overestimate the probability of disease would be expected to use that overestimation when deciding whether to initiate therapy, which could lead to overuse of medications and procedures with associated patient harms. Practitioners in the study reported that they would initiate treatment based on estimates of disease, including 78.2% who would treat cardiac ischemia and 71.0% who would treat a UTI when a positive test result would place their patient at 11% or less chance of disease. These errors would similarly corrupt shared decision-making with patients, which relies on practitioner understanding and communication of the likelihood of various outcomes.²⁵⁻²⁷ Training in shared decision-making has focused on communication skills,²⁸ not on understanding the probability of disease,²⁹ but the findings suggest another important educational target.

More focus on diagnostic reasoning in medical education is important. The finding of a primary problem with pretest probability estimates may be more amenable to intervention than the more commonly discussed bayesian adjustment to probability from test results.³⁰ Pretest probability is commonly discussed in medical education, but a standard method for estimating pretest probability has not been described.³⁰ Ideally, such estimates incorporate knowledge of disease prevalence and the predictive value of components of the history and physical examination, but for many conditions this information is difficult to find. The fact that estimates are so far from scientific evidence identifies a pressing need for improvement. There are a limited number of well-characterized diseases with pretest probability calculators, notably cardiac ischemia.^{31,32} Despite the fact that respondents in this study had no access to external aids while completing the survey, pretest estimates of cardiac ischemia were more accurate than for other clinical scenarios, implying that access to these calculators may improve knowledge and impact clinical reasoning. There is also a need to improve bayesian adjustment in probability from test results, which requires readily accessible references for clinical sensitivity and specificity. Computer visual decision aids that guide estimates of probability may also have a role.^{5,33} Alternative approaches, such as natural frequencies and naturalistic decision-making or use of heuristics, may improve decisions.³⁴

Limitations

This study has limitations. One is that the small fraction of respondents who did not complete the survey were more likely to be female, nurse practitioners, or physician assistants or to have been in practice for more than 10 years. However, the overall response rate was high. The format of survey questions required participants to estimate pretest probability before giving interpretation of positive or negative test results, which may not reflect their natural practice. Finally, although validity was extensively assessed via a multidisciplinary expert panel, reliability of our novel survey was not assessed.

Conclusions

In this study, large overestimates of the probability of disease before and after diagnostic testing were observed. Probability adjustments in response to test results varied from accurate to overestimates of risk by type of test. This significant overestimation of disease likely limits the ability of practitioners to engage in precise and evidence-based medical practice or shared decision-making.

ARTICLE INFORMATION

Accepted for Publication: November 21, 2020.

Published Online: April 5, 2021.

doi:10.1001/jamainternmed.2021.0269

Open Access: This is an open access article distributed under the terms of the [CC-BY License](#). © 2021 Morgan DJ et al. *JAMA Internal Medicine*.

Author Affiliations: Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore (Morgan, Pineles, Magder, Brown, Stevens); Veterans Affairs (VA) Maryland Healthcare System, Baltimore (Morgan); Department of Health, Behavior, and Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland (Owczarzak); Adult and Child Consortium of Health Outcomes Research and Delivery Science, University of Colorado School of Medicine, Aurora (Scherer); Division of Cardiology, University of Colorado School of Medicine, Aurora (Scherer); Center of Innovation for Veteran-Centered and Value-Driven Care, VA Denver, Denver, Colorado (Scherer); Division of General Internal Medicine & Geriatrics, Department of Medicine, Oregon Health & Science University, Portland (Pfeiffer, Terndrup); Department of Medicine, Dell Medical School, the University of Texas at Austin, Austin (Leykum); Department of Medicine, South Texas Veterans Health Care System, San Antonio (Leykum); Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison (Feldstein); Department of Medicine, Penn State College of Medicine, Hershey, Pennsylvania (Foy); Department of Public Health Sciences, Penn State College of Medicine, Hershey, Pennsylvania (Foy); Department of Medicine, University of Maryland School of Medicine, Baltimore (Koch); Department of Informatics, Genomic Medicine Institute, Geisinger, Danville, Pennsylvania (Masnick); Division of Infectious Diseases, New York University Grossman School of Medicine, New York (Weisenberg); Division of General Internal Medicine, Memorial Sloan Kettering Cancer Center, New York, New York (Korenstein).

Author Contributions: Dr Morgan had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Morgan, Pineles, Owczarzak, Magder, Scherer, Brown, Terndrup, Feldstein, Foy, Stevens, Koch, Weisenberg, Korenstein.

Acquisition, analysis, or interpretation of data: Morgan, Pineles, Magder, Scherer, Pfeiffer, Leykum, Stevens, Koch, Masnick.

Drafting of the manuscript: Morgan, Pineles, Magder, Stevens.

Critical revision of the manuscript for important intellectual content: Morgan, Pineles, Owczarzak, Magder, Scherer, Brown, Pfeiffer, Terndrup, Leykum, Feldstein, Foy, Koch, Masnick, Weisenberg, Korenstein.

Statistical analysis: Morgan, Magder.

Obtained funding: Morgan, Pineles.

Administrative, technical, or material support: Morgan, Pineles, Owczarzak, Scherer, Brown, Pfeiffer, Terndrup, Leykum, Stevens.

Supervision: Morgan, Terndrup.

Conflict of Interest Disclosures: Dr Morgan reported receiving grants from the National Institutes of Health (NIH) during the conduct of the study and grants from the US Department of Veterans Affairs, the Agency for Healthcare Research and Quality, and the Centers for Disease Control and Prevention outside the submitted work. Ms Pineles reported receiving grants from the NIH to the University of Maryland School of Medicine during the conduct of the study. Dr Scherer reported receiving grants from the NIH during the conduct of the study. Dr Brown reported receiving grants from the NIH during the conduct of the study. Dr Pfeiffer reported receiving grants from Pfizer to serve as site investigator for a *Clostridium difficile* vaccine trial (protocol B5091007) since July 2020 under a Cooperative Research and Development Agreement with VA Portland outside the submitted work. Dr Korenstein reported receiving grants from the NIH and grants from the National Cancer Institute to Memorial Sloan Kettering Cancer Center during the conduct of the study and that her spouse serves on the scientific advisory board and as a consultant for Vedanta Biosciences, serves as a consultant for Takeda, serves on the scientific advisory board and as a consultant for Opentrons. No other disclosures were reported.

Funding/Support: This project was funded by grant NLM DP2LM012890 (New Innovator Award) from the NIH (Dr Morgan, principal investigator). Dr Korenstein's work on this project was supported in part by Cancer Center Support Grant P30 CA008748 from the National Cancer Institute to Memorial Sloan Kettering Cancer Center.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

- Hunter DJ. Uncertainty in the era of precision medicine. *N Engl J Med*. 2016;375(8):711-713. doi:10.1056/NEJMpt1608282
- Schünemann HJ, Mustafa RA, Brozek J, et al; GRADE Working Group. GRADE guidelines: 22. the GRADE approach for tests and strategies—from test accuracy to patient-important outcomes and recommendations. *J Clin Epidemiol*. 2019;111:69-82. doi:10.1016/j.jclinepi.2019.02.003
- Armstrong KA, Metlay JP. Annals clinical decision making: using a diagnostic test. *Ann Intern Med*. 2020;172(9):604-609. doi:10.7326/M19-1940
- Centers for Disease Control and Prevention. About DLS. Published March 26, 2020. Accessed May 20, 2020. <https://www.cdc.gov/csels/dls/about-us.html>
- Whiting PF, Davenport C, Jameson C, et al. How well do health professionals interpret diagnostic information? a systematic review. *BMJ Open*. 2015;5(7):e008155. doi:10.1136/bmjopen-2015-008155
- Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med*. 1998;104(4):374-380. doi:10.1016/S0002-9343(98)00054-0
- Berwick DM, Fineberg HV, Weinstein MC. When doctors meet numbers. *Am J Med*. 1981;71(6):991-998. doi:10.1016/0002-9343(81)90325-9
- Gigerenzer G. *Reckoning With Risk: Learning to Live With Uncertainty*. Penguin Books; 2003.
- Manrai AK, Bhatia G, Strymish J, Kohane IS, Jain SH. Medicine's uncomfortable relationship with math: calculating positive predictive value. *JAMA Intern Med*. 2014;174(6):991-993. doi:10.1001/jamainternmed.2014.1059
- Casscells W, Schoenberger A, Graboyes TB. Interpretation by physicians of clinical laboratory results. *N Engl J Med*. 1978;299(18):999-1001. doi:10.1056/NEJM197811022991808
- Krouss M, Croft L, Morgan DJ. Physician understanding and ability to communicate harms and benefits of common medical treatments. *JAMA Intern Med*. 2016;176(10):1565-1567. doi:10.1001/jamainternmed.2016.5027
- Korenstein D. Charting the route to high value care: the role of medical education. *JAMA*. 2015;314(22):2359-2361. doi:10.1001/jama.2015.15406
- Gagliardi JP, Stinnett SS, Schardt C. Innovation in evidence-based medicine education and assessment: an interactive class for third- and fourth-year medical students. *J Med Libr Assoc*.

- 2012;100(4):306-309. doi:10.3163/1536-5050.100.4.014
14. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*. 1974;185(4157):1124-1131. doi:10.1126/science.185.4157.1124
 15. CNNPolitics. Top doctor says White House Coronavirus Task Force still missing 50% of testing data. Accessed May 6, 2020. <https://www.cnn.com/2020/04/02/politics/birx-task-force-coronavirus-testing/index.html>
 16. Morgan DJ, Scherer LD, Korenstein D. Improving physician communication about treatment decisions: reconsideration of "risks vs benefits". *JAMA*. 2020;324(10):937-938. Published online March 9, 2020. doi:10.1001/jama.2020.0354
 17. Morgan DJ, Brownlee S, Leppin AL, et al. Setting a research agenda for medical overuse. *BMJ*. 2015;351:h4534. doi:10.1136/bmj.h4534
 18. Fagerlin A, Sepucha KR, Couper MP, Levin CA, Singer E, Zikmund-Fisher BJ. Patients' knowledge about 9 common health conditions: the DECISIONS survey. *Med Decis Making*. 2010;30(5)(suppl):355-525. doi:10.1177/0272989X10378700
 19. Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ*. 2002;324(7341):824-826. doi:10.1136/bmj.324.7341.824
 20. Heneghan C, Glasziou P, Thompson M, et al. Diagnostic strategies used in primary care. *BMJ*. 2009;338:b946. doi:10.1136/bmj.b946
 21. Bernstein PL. *Against the Gods: The Remarkable Story of Risk*. John Wiley & Sons; 1996.
 22. Saposnik G, Redelmeier D, Ruff CC, Tobler PN. Cognitive biases associated with medical decisions: a systematic review. *BMC Med Inform Decis Mak*. 2016;16(1):138. doi:10.1186/s12911-016-0377-1
 23. Crowley RS, Legowski E, Medvedeva O, et al. Automated detection of heuristics and biases among pathologists in a computer-based system. *Adv Health Sci Educ Theory Pract*. 2013;18(3):343-363. doi:10.1007/s10459-012-9374-z
 24. Nicolle LE, Gupta K, Bradley SF, et al. Clinical practice guideline for the management of asymptomatic bacteriuria: 2019 update by the Infectious Diseases Society of America. *Clin Infect Dis*. 2019;68(10):1611-1615. doi:10.1093/cid/ciz021
 25. Wells' Criteria for Pulmonary Embolism. MDCalc. Accessed April 27, 2020. <https://www.mdcalc.com/wells-criteria-pulmonary-embolism>
 26. Sheridan SL, Donahue KE, Brenner AT. Beginning with high value care in mind: a scoping review and toolkit to support the content, delivery, measurement, and sustainment of high value care. *Patient Educ Couns*. 2019;102(2):238-252. doi:10.1016/j.pec.2018.05.014
 27. Morgan DJ, Leppin AL, Smith CD, Korenstein D. A practical framework for understanding and reducing medical overuse: conceptualizing overuse through the patient-clinician interaction. *J Hosp Med*. 2017;12(5):346-351. doi:10.12788/jhm.2738
 28. Back AL, Fromme EK, Meier DE. Training clinicians with communication skills needed to match medical treatments to patient values. *J Am Geriatr Soc*. 2019;67(5):S435-S441. doi:10.1111/jgs.15709
 29. Armstrong KA, Metlay JP. *Annals* clinical decision making: communicating risk and engaging patients in shared decision making. *Ann Intern Med*. 2020;172(10):688-692. doi:10.7326/M19-3495
 30. Richardson WS. Five uneasy pieces about pre-test probability. *J Gen Intern Med*. 2002;17(11):882-883. doi:10.1046/j.1525-1497.2002.20916.x
 31. Genders TSS, Steyerberg EW, Hunink MGM, et al. Prediction model to estimate presence of coronary artery disease: retrospective pooled analysis of existing cohorts. *BMJ*. 2012;344:e3485. doi:10.1136/bmj.e3485
 32. Fihn SD, Gardin JM, Abrams J, et al; American College of Cardiology Foundation; American Heart Association Task Force on Practice Guidelines; American College of Physicians; American Association for Thoracic Surgery; Preventive Cardiovascular Nurses Association; Society for Cardiovascular Angiography and Interventions; Society of Thoracic Surgeons. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS Guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. *J Am Coll Cardiol*. 2012;60(24):e44-e164. doi:10.1016/j.jacc.2012.07.013
 33. Testing Wisely. Accessed February 25, 2021. <https://calculator.testingwisely.com/>
 34. Klein G. A naturalistic decision making perspective on studying intuitive decision making. *J Appl Res Memory Cognit*. 2015;4(3):164-168. doi:10.1016/j.jarmac.2015.07.001