

A Randomized Trial of Ways To Describe Test Accuracy: The Effect on Physicians' Post-Test Probability Estimates

Milo A. Puhan, MD; Johann Steurer, MD, MME; Lucas M. Bachmann, MD, PhD; and Gerben ter Riet, MD, PhD

Background: Some people believe that likelihood ratios provide diagnostic information that is more useful than sensitivity and specificity estimates.

Objective: To assess how physicians' estimates about probability of illness are affected by the presentation of a diagnostic test's value as an estimate of sensitivity and specificity versus a likelihood ratio or an inexact numerical graphic.

Design: Random assignment of vignettes with different presentation formats of diagnostic test accuracy.

Setting: Auditorium at a continuing medical education conference.

Participants: 183 physicians.

Intervention: After estimating probabilities of 6 common illnesses described in patient vignettes, physicians reviewed pertinent test results presented in 1 of 3 formats.

Measurements: Physicians' probability estimates of illness before and after receiving test information, and post-test probability estimates based on the Bayes theorem.

Results: Absolute percentage point differences between the physicians' estimated and the Bayes-based post-test probabilities varied from -7 to 31 , from -7 to 28 , and from 1 to 29 for the sensitivity and specificity, likelihood ratio, and graphical groups, respectively. Mean differences of probability estimates between the sensitivity and specificity and the likelihood ratio groups were small for all vignettes (-2 to 3 percentage points; summary mean z value across the 6 vignettes, 0.04 [95% CI, -0.14 to 0.21]).

Limitations: The small pool of participants (who were potentially selected) and the limited number of vignettes prevented a more detailed analysis of relationships between the interpreted strength of diagnostic evidence and estimations of illness probability.

Conclusions: These findings suggest that presenting diagnostic test accuracy with likelihood ratios does not affect some physicians' estimates of illness probability compared with presenting diagnostic test results as sensitivity and specificity.

Ann Intern Med. 2005;143:184-189.

www.annals.org

For author affiliations, see end of text.

Diagnostic tests are used to reduce uncertainty about the presence or absence of an illness. The informativeness of a diagnostic test, usually expressed as its sensitivity and specificity or as a likelihood ratio, determines how much the illness probability changes from the pretest to the post-test situation. The change in illness probability can be calculated using the Bayes theorem, which combines the illness probability before testing with the information derived from the test (1).

Few physicians use this formal approach to weigh the informativeness of diagnostic test results (2) and, perhaps as a result, make mistakes when asked to use it (3, 4). Many investigators advocate the use of likelihood ratios (5–8). However, the effect of different presentation formats for diagnostic accuracy on post-test probability estimates has not been rigorously tested. We conducted a randomized study to see if physicians' estimates about illness probability more closely reflect estimates determined by the Bayes theorem when diagnostic information is presented as sensitivity and specificity estimates, a likelihood ratio, or an inexact numerical graphic.

METHODS

We randomly assigned experienced specialists in family and internal medicine attending a continuing medical education conference to complete 1 of 3 self-administered questionnaires that presented vignettes representative of

scenarios commonly encountered early in the diagnostic work-up. The vignettes differed only in how diagnostic test accuracy was presented. The questionnaires were sealed within 576 envelopes that had been randomized in 32 blocks of 18 to avoid order effects (6 different vignette orders for each of the 3 test accuracy presentation formats). We distributed the envelopes before participants entered the lecture hall so that we could not predict which participant received which envelope, thus concealing the randomization. The moderator of the lecture (which was unrelated to material presented in the vignettes) advised those present to complete the questionnaire during small breaks or at the end of the lecture. We collected 183 questionnaires 15 minutes after the lecture and excluded 5 questionnaires that were sent to us later.

The questionnaires asked participants to assess 6 clinical patient profiles and to estimate the probability of illness, specifying a percentage between 0% and 100%, for

See also:

Print

Editors' Notes 185

Web-Only

Appendix Table

Conversion of tables into slides

each scenario. With the exception of the chronic obstructive pulmonary disease vignette, all patient vignettes consisted of a description of presenting symptoms, additional clinical facts, and a risk profile (**Appendix Table**, available at www.annals.org). In the chronic obstructive pulmonary disease vignette, no presenting symptom but only the risk profile was given. For each vignette, we then gave the result of a diagnostic test commonly used in clinical practice and the test's accuracy derived from the literature, presented as sensitivities and specificities (group 1), as a positive or negative likelihood ratio (group 2), or as an inexact numerical graphical format (group 3). Using the test result and its diagnostic accuracy, participants were then invited to estimate the post-test probability. As shown in the **Appendix Table** (available at www.annals.org), we varied the informativeness of the diagnostic tests for both negative and positive results by presenting examples of weak (vignettes 1 and 4), intermediate (vignettes 2 and 5), and strong (vignettes 3 and 6) test accuracies. Sensitivities and specificities were given as percentages to group 1.

In group 2, the positive or negative likelihood ratios were presented as a numerical value and described in non-technical language. For example, the description of the D-dimer test's positive likelihood ratio of 1.7 in patients with high probability of pulmonary embolism (9) was additionally described as "in patients with pulmonary embolism, a positive test result is found almost twice as often as in patients without pulmonary embolism." For group 3, we depicted an inexact numerical presentation with a simple graphic of 5 circles (○○○○○). We darkened an increasing number of circles to correspond with increasing positive likelihood ratios or decreasing negative likelihood ratios (10). For example, positive likelihood ratios for 1 to 2.4 and for greater than 10 were presented as ●○○○○ and ●●●●●, respectively, while negative likelihood ratios for 1 to 0.41 and for less than 0.1 were presented as ●○○○○ and ●●●●●, respectively. Experienced clinicians designed and reviewed the vignettes.

Statistical Analysis

We calculated medians and interquartile ranges (25th to 75th percentile) for pretest and post-test probabilities. On the basis of each individual pretest probability estimate, we calculated a corresponding Bayes-based post-test probability by plugging the likelihood ratios corresponding to each clinical vignette into the Bayes theorem. We calculated the differences between Bayes-based post-test probabilities and the estimated post-test probabilities. Because these differences were distributed normally, we used 2-sample *t*-tests to compare the differences between groups 1 (sensitivity and specificity) and 2 (likelihood ratio), as well as between the groups with exact numerical information (combined groups 1 and 2) and the group with the inexact numerical graphical presentation. In addition, we used the summary measure procedure described by Schouten (11) to provide an overall estimate (across all 6

Context

There are several ways to present information about the accuracy of diagnostic tests.

Contribution

In this vignette-based study, 183 physicians read 6 scenarios that included patients' presenting symptoms, risk profiles, and test results. Physicians were randomly assigned to view information about test accuracy as sensitivity and specificity, a likelihood ratio, or an inexact graphic of the likelihood ratio. In general, the groups estimated similar post-test probabilities of disease regardless of how diagnostic accuracy was presented.

Implications

This preliminary work suggests some physicians may interpret sensitivity and specificity and likelihood ratio information in a similar manner.

—The Editors

vignettes) of the effect of the presentation format on differences between Bayes-based post-test probabilities and estimated post-test probabilities. These summary scores were compared between groups by using a 2-sample *t*-test.

To assess whether a presentation format was associated with illogical ways to derive post-test probability estimates, we performed 2 additional analyses. Physicians often multiply pretest probabilities by the likelihood ratios instead of using the correct method that requires the use of pretest odds (12). Therefore, we determined the proportion of physicians in each group who multiplied the pretest probabilities to derive post-test probabilities. Using the Fisher exact test, we tested the hypothesis that the proportion of physicians who made this mistake would be greatest in group 2 (likelihood ratios). Finally, we set out to learn whether the physicians in group 1 (sensitivity and specificity) were more likely to estimate post-test probabilities in the wrong direction. For example, a post-test probability estimate that is lower than the pretest probability despite a positive test result has the wrong direction. This type of error may occur if physicians were to associate low values of sensitivity and specificity with negative test results, thereby inadvertently lowering their illness probability estimates despite a positive test result. The reverse may occur with a negative test result and very high values for sensitivity and specificity.

All statistical analyses were performed with SPSS for Windows, version 12.0.1 (SPSS Inc., Chicago, Illinois).

Role of the Funding Sources

The funding sources had no role in the design, conduct, or analysis of the study and no influence on the decision to publish the results.

Table 1. Characteristics of Participants

Physician Characteristics	Group 1: Sensitivity/Specificity (n = 65)	Group 2: Likelihood Ratio (n = 60)	Group 3: Inexact Numerical Graphic (n = 58)
Median age (25th, 75th percentile), y	50 (42, 55)	45 (39, 52)	48 (40, 54)
Women, n (%)	11 (17)	19 (32)	7 (12)
Specialization, n (%)			
Family medicine	52 (79)	39 (66)	46 (79)
Internal medicine	10 (15)	14 (24)	10 (17)
Not specified	4 (6)	6 (10)	2 (3)
Median time since graduation (25th, 75th percentile), y	24 (17, 29)	18 (11, 24)	21 (13, 26)
Setting, n (%)			
Private practice	57 (86)	43 (73)	46 (79)
Hospital	9 (14)	14 (23)	11 (19)
Not specified	0	2 (4)	1 (2)
Attended continuing medical education in evidence-based medicine, n (%)	28 (42)	27 (46)	27 (47)

RESULTS

Of approximately 500 conference attendees, 183 (37%) participated in our study. Group 2 contained more women and hospital-based internists, whereas training in evidence-based medicine appeared balanced among the 3 groups (Table 1). Table 2 shows that physicians' pretest estimates varied widely for all vignettes, but the medians and interquartile ranges were similar across groups. Medians varied for the patient suspected of having a myocardial infarction, but interquartile ranges were similar across groups.

Table 2 also shows the differences between Bayes-based (that is, calculated) and estimated post-test probabilities. Absolute percentage point differences varied from -7 to 31 , from -7 to 28 , and from 1 to 29 for the sensitivity and specificity, likelihood ratio, and graphical groups, respectively. For the myocardial infarction, chronic obstructive pulmonary disease, temporal arteritis, and heart failure vignettes, physicians changed their illness probabilities to a lesser extent than was allowable under the Bayes theorem (positive values for differences). In groups 1 and 2, the diagnostic accuracy of the D-dimer test was overestimated; in the influenza vignette, however, estimated post-test probabilities were, on average, almost identical to Bayes-based post-test probabilities for all groups.

The 2 exact numerical presentation formats (groups 1 and 2) had similar effects. We found differences of -2 to 3 percentage points across the 6 vignettes and a difference of 0.04 (95% CI, -0.14 to 0.21) for the summary z score. Differences between the inexact numerical graphic group and the combined groups 1 and 2 were between -10 and 5 percentage points (summary z score difference, -0.21 [CI, -0.37 to -0.06]). Physicians' post-test probability estimates were closer to Bayes-based estimates when informed by the inexact numerical graphic for the pulmonary embolism (mean difference, 4 percentage points) and chronic obstructive pulmonary disease vignettes (mean difference, 10 percentage points) than when informed by other measures.

For 5 vignettes, we found 5 or fewer physicians per group who had post-test probabilities suggesting exact multiplication of the pretest probabilities by likelihood ratios instead of using pretest odds. For the myocardial infarction vignette, however, 20% of physicians in group 2 (likelihood ratios) exactly multiplied their pretest probability estimate by 0.15, compared with 14% of physicians in group 1 and 7% in group 3 ($P = 0.12$).

Overall, 6% (CI, 3.4% to 10.4%) of all post-test probability estimates were in the wrong direction. Nine percent (ranging from 3% to 19% across 6 vignettes) of all post-test probability estimates in group 1 (sensitivity and specificity) were in the wrong direction compared with 4% (range, 0% to 7%) in group 2 and 4% (range, 0% to 9%) in group 3 ($P = 0.005$). Of all post-test estimates that were in the wrong direction, 49% occurred in the vignettes with tests of low accuracy (pulmonary embolism and temporal arteritis).

DISCUSSION

This study showed that post-test probability estimates deviated to a small and similar extent from Bayes-based estimates in the groups informed by sensitivity and specificity or likelihood ratios. An inexact numerical graphic led physicians to come closer to Bayes-based estimates in the pulmonary embolism and chronic obstructive pulmonary disease vignettes. In addition, we found that some physicians estimated lower illness probabilities after a positive test result if it was accompanied by a low test accuracy value.

This study has several limitations. First, the small number of vignettes precluded a more detailed analysis of relationships between strength of the diagnostic evidence and the degree to which the inexact numerical method may lead to conservatism. Second, the analysis presented in Table 2 assumes constancy of likelihood ratios (or sensitivity-specificity combinations) across the range of pretest probabilities, although, in reality, they vary (13). Third, we

used published literature to estimate test accuracy characteristics, but we could not compare the physicians' post-test estimates with "correct" post-test probabilities, because "true" summary measures of test accuracy often were not

readily available (14). Fourth, the inexact numerical graphic may have oversimplified the test accuracy given in the chronic obstructive pulmonary disease vignette because it only informed physicians of a likelihood ratio of 10 or

Table 2. Probability Estimates and Deviation of Estimated from Bayes-Based Calculated Post-Test Probabilities

Target Illness	Group 1: Sensitivity and Specificity (25th, 75th Percentile)*	Group 2: Likelihood Ratio (25th, 75th Percentile)*	Group 1 vs. Group 2		Group 3: Inexact Numerical Graphic (25th, 75th Percentile)*	Groups 1 and 2 vs. Group 3†	
			2-Sample t-test (95% CI)	P Value		2-Sample t-test (95% CI)	P Value
Pulmonary embolism							
Pretest probability estimate, %	70 (50, 80)	70 (50, 80)			80 (55, 80)		
Post-test probability estimate, %	90 (75, 95)	88 (67, 95)			80 (60, 90)		
Mean Bayes-based minus estimated post-test probability, %	-7, SD 14	-7, SD 14	0 (-5 to 5)	0.95	4, SD 19	-10 (-15 to -5)	<0.001
Myocardial infarction							
Pretest probability estimate, %	40 (20, 60)	30 (20, 48)			50 (20, 60)		
Post-test probability estimate, %	10 (3, 21)	10 (3, 20)			20 (5, 40)		
Mean Bayes-based minus estimated post-test probability, %	7, SD 18	4, SD 13	3 (-3 to 8)	0.36	12, SD 16	-6 (-11 to -1)	0.02
Chronic obstructive pulmonary disease							
Pretest probability estimate, %	50 (30, 70)	50 (30, 70)			60 (40, 80)		
Post-test probability estimate, %	90 (70, 95)	90 (80, 95)			90 (80, 99)		
Mean Bayes-based minus estimated post-test probability, %	14, SD 18	16, SD 20	-2 (-9 to 4)	0.50	10, SD 14	5 (-1 to 10)	0.10
Temporal arteritis							
Pretest probability estimate, %	20 (10, 50)	20 (10, 50)			30 (10, 50)		
Post-test probability estimate, %	19 (10, 46)	20 (5, 50)			30 (10, 60)		
Mean Bayes-based minus estimated post-test probability, %	2, SD 22	0, SD 11	2 (-4 to 8)	0.52	6, SD 18	-5 (-10 to 1)	0.10
Influenza							
Pretest probability estimate, %	70 (50, 80)	65 (30, 80)			70 (33, 80)		
Post-test probability estimate, %	93 (88, 96)	90 (80, 95)			90 (80, 95)		
Mean Bayes-based minus estimated post-test probability, %	1, SD 20	-1, SD 16	2 (-5 to 8)	0.56	1, SD 20	-2 (-8 to 4)	0.57
Heart failure							
Pretest probability estimate, %	70 (50, 88)	80 (60, 90)			80 (70, 88)		
Post-test probability estimate, %	55 (38, 80)	55 (40, 90)			65 (30, 80)		
Mean Bayes-based minus estimated post-test probability, %	31, SD 21	28, SD 19	2 (-5 to 10)	0.51	29, SD 24	1 (-6 to 8)	0.79
Mean summary z score for all vignettes	0.64, SD 0.51	0.60, SD 0.39	0.04 (-0.14 to 0.21)	0.68	0.84, SD 0.50	-0.21 (-0.37 to -0.06)	0.007

* To facilitate interpretation, values for the 3 groups were calculated by using (post[Bayes-based]) minus (post[physician-based]) for vignettes with positive test results (pulmonary embolism, chronic obstructive pulmonary disease, and influenza) and (post[physician-based]) minus (post[Bayes-based]) for vignettes with negative test results (myocardial infarction, temporal arteritis, and heart failure). For all vignettes, positive values mean that physicians did not estimate post-test probabilities as high (or as low) as allowable under Bayes theorem (i.e., conservative estimates).

† Values in the last column were calculated by combining groups 1 and 2, then subtracting group 3 values from the combined groups. Thus, negative values mean that the presentation using inexact numerical graphics led to more conservative estimates than the exact numerical methods.

greater, whereas the positive likelihood ratio given to the numerical groups was 33.5. Fifth, the participation rate of approximately 37% led to a moderately sized and a potentially selected study sample. Some baseline imbalances existed with respect to sex and specialty, but we are not aware of any evidence that these characteristics are associated with skill in the application of the concepts of diagnostic test accuracy. Finally, the use of vignettes instead of real patients may have yielded greater variability in the pretest probability estimates, although we do not believe that this approach affected the application of the Bayes theorem.

There is the widespread assumption that likelihood ratios are more useful than sensitivity and specificity estimates because they are simpler to interpret (5) and can be applied easily by using the Fagan nomogram. Likelihood ratios also require the application of only 1 measure for test accuracy, compared with the 2 for sensitivity and specificity (5–8). However, our study challenges this belief and confirms earlier research that indicated that physicians arrive at similar post-test probabilities when informed by sensitivity and specificity estimates or by likelihood ratios (12).

For 5 vignettes featuring pictorial presentation, physicians changed their illness probability estimates to a lesser extent than was allowable under the Bayes theorem. This conservatism was more pronounced for vignettes with tests of higher diagnostic value. This observation may indicate that physicians intuitively adjust for the over-optimism inherent in results from single-test evaluation studies (15), on which the accuracy measures in our vignettes were based.

Our study explored different ways of communicating diagnostic evidence within the paradigm of the Bayes theorem. However, this concept ignores the context in which a test is performed (16, 17). In most reports about diagnostic studies, variability of test accuracy across different patient profiles is not taken into consideration despite repeated recommendations (18, 19). The meaning of evidence from single-test studies often remains unclear for patients with specific profiles like those in our vignettes. In the heart failure vignette, for example, physicians almost disregarded the high diagnostic value associated with the test result (20) and lowered their pretest probability estimates from 80% to around 60%. This finding is important because it shows that, in certain situations, physicians may not accept evidence that is not compatible with information available before testing.

Diagnostic studies with multivariable analyses that provide data on illness probabilities based on the patients' manifestational and risk profile are likely to provide a better knowledge base for clinicians (16). Results from multivariable analyses could inform physicians on illness probabilities given a certain patient profile and additional test results without the need for any calculations. Such data would also help physicians to choose a rational diagnostic work-up and to base decisions on further testing or treatments on valid illness probabilities.

In conclusion, we found that post-test probability es-

timates deviated to a similar extent from Bayes-based estimates, whether informed by sensitivity and specificity or by a likelihood ratio. This finding suggests that the likelihood ratio may not be more useful than sensitivity and specificity data to present test accuracy information.

From the University of Zurich, Zurich, Switzerland.

Acknowledgments: The authors thank Dr. Otto Brändli for the opportunity to conduct this study during the congress of the Zurich Lung League in Davos, Switzerland.

Grant Support: By the Helmut Horten Foundation. Dr. Bachmann's work was supported by the Swiss National Science Foundation (grants 3233B0-103182 and 3200B0-103183).

Potential Financial Conflicts of Interest: None disclosed.

Requests for Single Reprints: Milo A. Puhan, MD, Horten Centre, University Hospital of Zurich, Postfach Nord, 8091 Zurich, Switzerland; e-mail, milo.puhan@evimed.ch.

Current author addresses and author contributions are available at www.annals.org.

References

- Ledley RS, Lusted LB. Reasoning foundations of medical diagnosis; symbolic logic, probability, and value theory aid our understanding of how physicians reason. *Science*. 1959;130:9-21. [PMID: 13668531]
- Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med*. 1998;104:374-80. [PMID: 9576412]
- Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G. Communicating accuracy of tests to general practitioners: a controlled study. *BMJ*. 2002;324:824-6. [PMID: 11934776]
- Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Medicine. Communicating statistical information. *Science*. 2000;290:2261-2. [PMID: 11188724]
- Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA*. 1994;271:703-7. [PMID: 8309035]
- Sox HC, Blatt MA, Higgins MC, Marton KI. *Medical Decision Making*. Newton, MA: Butterworth-Heinemann; 1988:67-102.
- Hunink M, Glasziou P. *Decision Making in Health and Medicine: Integrating Evidence and Values*. New York: Cambridge Univ Pr; 2001:128-76.
- Weinstein MC, Fineberg HV. The use of diagnostic information to revise probabilities. In: *Clinical Decision Analysis*. Philadelphia: WB Saunders; 1980:75-130.
- Ginsberg JS, Wells PS, Kearon C, Anderson D, Crowther M, Weitz JI, et al. Sensitivity and specificity of a rapid whole-blood assay for D-dimer in the diagnosis of pulmonary embolism. *Ann Intern Med*. 1998;129:1006-11. [PMID: 9867754]
- Jaeschke R, Guyatt G, Lijmer JG. Diagnostic tests. In: Guyatt G, Rennie D, eds. *User's Guide to the Medical Literature*. Chicago: American Med Assoc; 2002:121-40.
- Schouten HJ. Combined evidence from multiple outcomes in a clinical trial. *J Clin Epidemiol*. 2000;53:1137-44. [PMID: 11106887]
- Bachmann LM, Steurer J, ter Riet G. Simple presentation of test accuracy may lead to inflated disease probabilities [Letter]. *BMJ*. 2003;326:393. [PMID: 12586679]
- Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med*. 2004;23:125-35. [PMID: 14695644]
- Puhan MA, Steurer J, Bachmann LM, ter Riet G. Variability in diagnostic probability estimates [Letter]. *Ann Intern Med*. 2004;141:578-9. [PMID: 15466782]

15. **Bachmann LM, ter Riet G, Clark TJ, Gupta JK, Khan KS.** Probability analysis for diagnosis of endometrial hyperplasia and cancer in postmenopausal bleeding: an approach for a rational diagnostic workup. *Acta Obstet Gynecol Scand.* 2003;82:564-9. [PMID: 12780428]
16. **Miettinen OS.** The modern scientific physician: 3. Scientific diagnosis. *CMAJ.* 2001;165:781-2. [PMID: 11584568]
17. **Moons KG, Biesheuvel CJ, Grobbee DE.** Test research versus diagnostic research. *Clin Chem.* 2004;50:473-6. [PMID: 14981027]
18. **Moons KG, van Es GA, Deckers JW, Habbema JD, Grobbee DE.** Limitations of sensitivity, specificity, likelihood ratio, and bayes' theorem in assessing diagnostic probabilities: a clinical example. *Epidemiology.* 1997;8:12-7. [PMID: 9116087]
19. **Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J.** Designing studies to ensure that estimates of test accuracy are transferable. *BMJ.* 2002;324:669-71. [PMID: 11895830]
20. **Davie AP, Francis CM, Love MP, Caruana L, Starkey IR, Shaw TR, et al.** Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction. *BMJ.* 1996;312:222. [PMID: 8563589]

Current Author Addresses: Drs. Puhan and Steurer: Horten Centre, University Hospital of Zurich, Postfach Nord, 8091 Zurich, Switzerland.

Dr. Bachmann: University of Berne, Department of Social and Preventive Medicine, Finkenhubelweg 11, 3012 Berne, Switzerland.

Dr. ter Riet: Academic Medical Center, Department of General Practice, Room J2-118, 1105 AZ Amsterdam, the Netherlands.

Author Contributions: Conception and design: M.A. Puhan, J. Steurer, G. ter Riet.

Analysis and interpretation of the data: M.A. Puhan, L.M. Bachmann, J. Steurer, G. ter Riet.

Drafting of the article: M.A. Puhan.

Critical revision of the article for important intellectual content: L.M. Bachmann, J. Steurer, G. ter Riet.

Final approval of the article: M.A. Puhan, L.M. Bachmann, J. Steurer, G. ter Riet.

Provision of study materials or patients: M.A. Puhan.

Statistical expertise: M.A. Puhan, G. ter Riet.

Collection and assembly of data: M.A. Puhan.

21. Innes G, Christenson J, Weaver WD, Liu T, Hoekstra J, Every N, et al. Diagnostic parameters of CK-MB and myoglobin related to chest pain duration. *Canadian Journal of Emergency Medicine*. 2002;4.

22. Badgett RG, Tanaka DJ, Hunt DK, Jelley MJ, Feinberg LE, Steiner JF, et al. Can moderate chronic obstructive pulmonary disease be diagnosed by historical and physical findings alone? *Am J Med*. 1993;94:188-96. [PMID: 8430714]

23. Salvarani C, Silingardi M, Ghirarduzzi A, Lo Scocco G, Macchioni P, Bajocchi G, et al. Is duplex ultrasonography useful for the diagnosis of giant-cell arteritis? *Ann Intern Med*. 2002;137:232-8. [PMID: 12186513]

24. Schultze D, Thomas Y, Wunderli W. Evaluation of an optical immunoassay for the rapid detection of influenza A and B viral antigens. *Eur J Clin Microbiol Infect Dis*. 2001;20:280-3. [PMID: 11399021]

Appendix Table. Clinical Patient Vignettes and Test Results

Target Illness	Manifestational Profile	Risk Profile	Test Result	Test Accuracy Presentation (Reference)
Vignette 1: pulmonary embolism	Pleuritic chest pain for 3 hours; heart rate, 96 beats/min; respiratory rate, 20 breaths/min	40-year-old woman, smokes 1 pack of cigarettes per day; uses oral contraceptives	D-Dimer test positive	Group 1: sensitivity, 93%; specificity, 45% Group 2: positive likelihood ratio, 1.7 Group 3: ●○○○○ (9)
Vignette 2: myocardial infarction	Emergency consultation for constant chest pain for 9 hours	42-year-old man, obese	Creatine kinase-MB and myoglobin both normal	Group 1: sensitivity, 87%; specificity, 88% Group 2: negative likelihood ratio, 0.15 Group 3: ●●●○○ (21)
Vignette 3: chronic obstructive pulmonary disease	No presenting symptoms	63-year-old man coming for a routine checkup, has smoked cigarettes for many years	90 pack-years and decreased breath sounds	Group 1: sensitivity, 67%; specificity, 98% Group 2: positive likelihood ratio, 33.5 Group 3: ●●●●● (22)
Vignette 4: temporal arteritis	Bilateral temporal headache for several weeks, no fever	67-year-old woman	Duplex ultrasonography of temporal arteries without any abnormal findings	Group 1: sensitivity, 40%; specificity, 79% Group 2: negative likelihood ratio, 0.8 Group 3: ●○○○○ (23)
Vignette 5: influenza	Cough and muscle pain for 2 days, fever (temperature, 38.2 °C) and chills for a few hours	34-year-old man	Influenza quick tests for influenza A and B viruses positive	Group 1: sensitivity, 51%; specificity, 93% Group 2: positive likelihood ratio, 7.0 Group 3: ●●●○○ (24)
Vignette 6: congestive heart failure	Dyspnea on exertion for several weeks, slightly bilaterally swollen ankles	72-year-old man	Electrocardiogram without any abnormal findings	Group 1: sensitivity, 94%; specificity, 61% Group 2: negative likelihood ratio, 0.1 Group 3: ●●●●● (20)