Is It Time to Raise the Bar? Age-Adjusted D-dimer Cutoff Levels for Excluding Pulmonary Embolism

Answers to the July 2014 Journal Club Questions

Jonathan Kirschner, MD; Jeffrey Kline, MD

0196-0644/$-see front matter

http://dx.doi.org/10.1016/j.annemergmed.2014.07.450

Editor's Note: *You are reading the 40th installment of* Annals of Emergency Medicine *Journal Club. This Journal Club refers to the Righini et al*[1] *article published in* JAMA *on March 19, 2014. Information about journal club can be found at http://www.annemergmed.com/content/journalclub. Readers should recognize that these are suggested answers. We hope they are accurate; we know that they are not comprehensive. There are many other points that could be made about these questions or about the article in general. Questions are rated "novice," (NOV) "intermediate," (INT) and "advanced (ADV) so that individuals planning a journal club can assign the right question to the right student. The "novice" rating does not imply that a novice should be able to spontaneously answer the question. "Novice" means we expect that someone with little background should be able to do a bit of reading, formulate an answer, and teach the material to others. Intermediate and advanced questions also will likely require some reading and research, and that reading will be sufficiently difficult that some background in clinical epidemiology will be helpful in understanding the reading and concepts. We are interested in receiving feedback about this feature. Please e-mail journalclub@acep.org with your comments.*

DISCUSSION POINTS

1. The "Age-adjusted D-dimer Cutoff Levels to Rule Out Pulmonary Embolism" trial studied the diagnostic strategy of using clinical prediction tools to identify patients with nonhigh pretest probability for pulmonary embolism (PE) and then applied an age-adjusted D-dimer cutoff level (age×10 for patients aged 50 years or older). The reported outcome was failure rate of the diagnostic strategy, defined as adjudicated thromboembolic events at 3 months.

NOV A. In diagnostic test accuracy studies, the prevalence of disease in the study population will directly affect the derived positive and negative predictive values—so-called posttest probabilities. How does the reported primary outcome of "failure rate of the diagnostic test strategy" relate to negative predictive value? Calculate the negative predictive value of the age-adjusted D-dimer cutoff level. Because predictive values are, in part, based on disease prevalence, likelihood ratios are preferred measures of diagnostic test accuracy. Discuss why likelihood ratios inform clinical practice better than sensitivity and specificity. What is the range of possible negative likelihood ratios associated with the age-adjusted D-dimer cutoff level as reported in this study?

NOV B. "Consecutive patients" were enrolled. What does that description imply about enrollment methodology?

NOV C. Patients had "acute onset or worsening shortness of breath or chest pain without another obvious etiology." What questions do these inclusion criteria raise about patient eligibility?

INT D. The prevalence of PE in this study was 19%. How does that number compare with the prevalence of PE in studies performed in other countries?

INT E. How does prevalence of disease in the study population affect the sensitivity and specificity of the diagnostic strategy?

ADV F. The median age of this European study population was 63 years, which is older than most American populations tested for PE. What effect might older age of the population studied have on the diagnostic accuracy of the D-dimer assay? What effect might older age have on the overall findings of this study?

2. Two different clinical probability assessment tools and 6 different D-dimer assays were used in this multicenter trial.

INT A. The patients were risk-stratified with either the revised Geneva score[2] or 2-level Wells score.[3] Describe the performance of these risk-stratification aids. How can the use of subjective criteria in a risk-stratification aid affect the validity of a diagnostic accuracy study result?

ADV B. Although all of the D-dimer assays used in this study had the same 500 ng/mL cutoff for an abnormal value, many other quantitative D-dimer assays have different cutoffs for abnormal. What is the basis of the differences in cutoffs? Can the results of this study be translated to D-dimers with different cutoffs?

3. The reference standard used in this study was adjudicated thromboembolic events as determined by further testing or clinical follow-up including autopsy.

INT A. Who should perform adjudicated outcomes, and what data should they use?

INT B. The authors considered only segmental or larger pulmonary arterial filling defects observed on imaging as evidence of PE. What would you predict the effect of excluding subsegmental PEs as a positive diagnosis has on diagnostic accuracy of the age-adjusted D-dimer?

INT C. How can the use of different reference standards for a diagnostic test introduce bias? In this case, can you

predict the effect of using clinical follow-up as a reference standard? Would it strengthen or weaken the study conclusion?

(INT) D. Name 3 advantages to the patient of obviating contrast-enhanced pulmonary angiography in an individual who does not have PE.

4. Clinicians often rely on test performance (sensitivity, specificity, etc) as the ultimate determinant of a strategy's usefulness in improving health outcomes. Indeed, the study authors believe that the very low failure rate in this prospective validation study is evidence that an age-adjusted cutoff should be used clinically. Reliance on test performance to support widespread use, however, can have undesirable effects.[4,5]

(NOV) A. What non–disease-specific outcomes in this study are important to adoption of the age-adjusted D-dimer cutoff level?

(NOV) B. How might one design a study to determine whether a diagnostic strategy is safe and effective at improving patient-centered outcomes?

(INT) C. When is demonstrating good test performance sufficient to recommend a diagnostic test? When should further impact analysis be performed?

(INT) 5. If you were to design a diagnostic algorithm at your hospital, would you include an age-adjusted D-dimer cutoff level for non–high-risk patients?

ANSWER 1

Q1. The "Age-adjusted D-dimer Cutoff Levels to Rule Out Pulmonary Embolism" (ADJUST-PE) trial studied the diagnostic strategy of using clinical prediction tools to identify patients with nonhigh pretest probability for pulmonary embolism (PE) and then applied an age-adjusted D-dimer cutoff level (age×10 for patients aged 50 years or older). The reported outcome was failure rate of the diagnostic strategy, defined as adjudicated thromboembolic events at 3 months.

Q1.a In diagnostic test accuracy studies, the prevalence of disease in the study population will directly affect the derived positive and negative predictive values—so-called posttest probabilities. How does the reported primary outcome of "failure rate of the diagnostic test strategy" relate to negative predictive value? Calculate the negative predictive value of the age-adjusted D-dimer cutoff level. Because predictive values are, in part, based on disease prevalence, likelihood ratios are preferred measures of diagnostic test accuracy. Discuss why likelihood ratios inform clinical practice better than sensitivity and specificity. What is the range of possible negative likelihood ratios associated with the age-adjusted D-dimer cutoff level as reported in this study?

The diagnostic test strategy used by the authors included pretest probability assessment for venous thromboembolism (VTE), with subsequent D-dimer testing using an age-adjusted cutoff for patients with nonhigh probability. In patients with a D-dimer level below the age-adjusted threshold, no further testing was performed and these patients were followed for 3 months. Failure of the diagnostic test strategy was defined as patients with a negative D-dimer result who had an adjudicated VTE event (either deep venous thrombosis [DVT] or PE) during the follow-up period. The rate of these false-negative studies is inversely related to the true negative rate, another name for the negative predictive value.[1] In this case, the failure rate of the diagnostic strategy would could be calculated as 1–negative predictive value, or (false negatives)/(true negatives+false negatives). There were only 2 false-negative results. Thus, the negative predictive value for nonhigh pretest probability patients with a negative adjusted D-dimer test result would be 1,139/1,141, or 99.8%, and a failure rate or false negative rate of 0.2%.

Negative predictive value may be misleading for several reasons. The posttest probability given from the negative predictive value depends on the prevalence of disease in the study population.[1] If the prevalence of disease in the study population is low, the negative predictive value will be inflated. Sensitivity and specificity are frequently preferred by clinicians because the calculation may not be directly affected by disease prevalence; however, these test performance characteristics each focus on patients with a specific disease state: either disease present (sensitivity) or disease absent (specificity). The problem with using this characteristic clinically lies in the reason for performing the test. A clinician deciding to order a test is unaware of the disease state of the patient, so statistics specific to a disease state cannot be readily applied. Sensitivity and specificity are characteristics of the test and do not directly transform a pretest probability to a posttest probability,[2] thus limiting their utility to inform patient care decisions. A likelihood ratio is a characteristic of the diagnostic test result that does connect pretest probability to posttest probability,[3] similarly to predictive values; however, it is not directly affected by prevalence (see question 1E for further discussion of spectrum effects). Although the authors do not provide information on the true- and false-positive rates, a range of possible likelihood ratios may be calculated. The number of false positives may range from 1,048 to 1,474, though the upper limit is highly improbable. Positive likelihood ratio is the true-positive rate relative to the false-positive rate or (true positive rate)/(false positive rate) (Table). The range of possible positive likelihood ratios is 1.8 to 2.1. Conversely, the negative likelihood ratio is the false-negative rate relative to the true-negative rate. The range of possible negative likelihood ratios is 0.0072 to 0.018. A Web-based diagnostic test calculator can be found at http://araw.mede.uic.edu/cgi-bin/testcalc.pl.

Table. Results of age-adjusted D-dimer cutoff.*

		Adjudicated VTE	
		+	–
D-dimer	+	212–638	1,048–1,474
	–	2	1,139

*The range of possible values for a positive test result based on published data is listed. Likelihood ratio (+)=1.8–2.1; likelihood ratio (–)=0.01–0.02.

Q1.b "Consecutive patients" were enrolled. What does that description imply about enrollment methodology?

By its strictest definition, *consecutive* means one after the other without interruption, which is a remarkably difficult feat to accomplish in clinical research, especially in the emergency department (ED) with symptom-based inclusion criteria. When determining the answer to a research question, investigators will select a sample of patients who will be representative of a larger population of similar patients. Sampling methodology is a potential source of bias in clinical studies, and efforts should be made to limit the bias introduced when a nonrepresentative sample is enrolled. A random sampling of patients from an eligible pool would decrease the risk of bias; however, this is prohibitively costly and onerous in clinical trials. More frequently, a convenience sample of patients is enrolled, as is the case in the ADJUST-PE study. Patients are enrolled when circumstances make enrollment feasible. When patients are selected nonsystematically or at the whim of study personnel, the risk of bias is high. If selection is performed for consecutive patients with clearly defined inclusion and exclusion criteria, the risk of sampling bias will be decreased compared with less methodologically rigorous enrollment.[4] It is likely that because of the clear criteria for inclusion and the consecutive enrollment of patients, the study population represents the larger population it attempts to approximate. This conclusion would have been strengthened if the authors had reported the number of potential study participants who were not screened for inclusion during the study period.

Q1.c Patients had "acute onset or worsening shortness of breath or chest pain without another obvious etiology." What questions do these inclusion criteria raise about patient eligibility?

Patients with dyspnea or pain without clear cause are a unique subset of patients presenting to the ED with chest pain or dyspnea as a primary complaint. Whether another cause is obvious, however, is left up to the treating clinician. Determining an obvious cause will vary from one provider to another. The potential for poor interrater reliability of this inclusion criterion may introduce bias.[5] This narrowly defined inclusion criterion systematically limits the generalizability of the trial results and translation into clinical use, which raises uncertainty about the effect of the intervention on the more diverse or complex population treated in clinical practice.[6]

Q1.d The prevalence of PE in this study was 19%. How does that number compare with the prevalence of PE in studies performed in other countries?

In studies of clinical prediction rules, the prevalence of PE ranged from 4% to 44%, with a mean of 16%, lower than that reported in the ADJUST-PE trial. The reported 19% prevalence is approximately 3 times the prevalence of PE in studies conducted in US EDs, but is significantly lower than that of previous studies conducted in Europe. Additionally, sensitivity was found to vary with prevalence, with low prevalence studies reporting sensitivities of approximately 0.4 to 0.7 when the prevalence was less than 10%. The sensitivity increased to approximately 0.6 to 0.9 when the prevalence was greater than 20%.[7]

Q1.e How does prevalence of disease in the study population affect the sensitivity and specificity of the diagnostic strategy?

The mathematic calculation of sensitivity and specificity is not affected by disease prevalence and thus is presumed to be more stable across populations[2]; however, test accuracy has been found to vary with prevalence. A variety of mechanisms can account for this variability in test performance because of both true and perceived variation in the study population. Verification bias and sampling bias, for example, may produce a perceived increase in prevalence and sensitivity, whereas referral filters and spectrum effects will produce a skewed study population.[8]

Ideally, the study population should include patients with a degree of uncertainty about the diagnosis. The sensitivity and specificity of a test when used to identify individuals who clearly do or do not have the disease in question will likely overestimate the test's performance compared with its use when there is diagnostic uncertainty. Because the prevalence of VTE in this study population was higher than that of many previous studies, it is possible that the study overestimates sensitivity.[9]

Q1.f The median age of this European study population was 63 years, which is older than most American populations tested for PE. What effect might older age of the population studied have on the diagnostic accuracy of the D-dimer assay? What effect might older age have on the overall findings of this study?

The risk of VTE markedly increases with age. In one study, the rate among patients aged 85 years and older was 13-fold greater than that in those aged 45 to 55 years, with an absolute rate of 7 per 1,000 annually.[10,11] Predictably, elderly nursing home residents are at even greater risk.[12] As the age of the study population increases, so too does the spectrum of disease. Elderly patients are more likely to present with more severe clinical features of PE, to develop PE, and to have comorbid conditions that may mimic VTE.[13] As discussed in the previous answer, spectrum effects can have significant consequences on reported diagnostic test performance. If the study population is skewed toward greater prevalence and more severe disease spectrum, the resultant sensitivity will be higher than that for the general population.

ANSWER 2

Q2. Two different clinical probability assessment tools and 6 different D-dimer assays were used in this multicenter trial.

Q2.a The patients were risk-stratified with either the revised Geneva score[2] or 2-level Wells score.[3] Describe the performance of these risk-stratification aids. How can the use of subjective criteria in a risk-stratification aid affect the validity of a diagnostic accuracy study result?

A 2011 meta-analysis reported the performance risk-stratification aids in the diagnosis of VTE.[7] The revised Geneva rule[14] had a pooled sensitivity of 0.91 and specificity of 0.37 (4 studies), and the Wells criteria with cutoff less than 4.5 had a sensitivity of 0.60 and specificity of 0.80 (11 studies). Clinical gestalt was found to have a sensitivity of 0.85 and specificity of 0.51 (15 studies).[7] The 2 risk-stratification aids did perform similarly in direct comparison.[15]

These clinical prediction rules may perform worse in elderly populations. With an increasing prevalence, the sensitivities remain high; however, the specificities are markedly lower. Although safe for identifying low-risk patients, prediction rules need to be adapted to be more useful in elderly patients.[16]

The inclusion of subjective criteria in a clinical prediction rule can introduce potential bias. In regard to the specific question in Wells criteria about "the most likely diagnosis," the interobserver reliability had a κ of 0.57.[17] Both the revised Geneva score and Wells score perform well; however, whether the rules can be used interchangeably remains unclear.[18] Likewise, interobserver variability is good to excellent (κ=0.6 for low probability) for gestalt assessment of pretest probability for PE when the estimate is trifurcated into low, moderate, and high risk.[19] Standardization of the pretest probability assessment in the ADJUST-PE trial would limit variation in results; however, the potential for bias is difficult to quantify.

Q2.b Although all of the D-dimer assays used in this study had the same 500 ng/mL cutoff for an abnormal value, many other quantitative D-dimer assays have different cutoffs for abnormal. What is the basis of the differences in cutoffs? Can the results of this study be translated to D-dimers with different cutoffs?

A variety of D-dimer assays exist, including enzyme-linked immunoadsorbent assay, latex agglutination, and whole blood assays. The performance of different assays varies, and there is no standard against which all assays can be calibrated.[20] Significant differences may be noted in clinically important thresholds across assays.[21] Additionally, quantitative results may be reported in fibrinogen equivalent units or D-dimer units, with various units of measure, adding to the confusion surrounding D-dimer testing. As stand-alone tests, the assays vary significantly in regard to cutoff value, negative predictive value (93% to 100%), and specificity (0% to 67%).[22] Another study reported the sensitivities for rapid D-dimer methods: SimpliRED 82.3%; VIDAS D-Dimer 91.4%; MiniQuant D-Dimer 96.3%; and Advanced D-Dimer 97.1% (95% confidence interval 96.3% to 97.9%). These sensitivities notably improved in the inpatient setting.[23] There are a variety of explanations for the heterogeneity in performance characteristics: monoclonal antibodies with different specificities for fibrin and its derivatives, as well as differences in calibrators, reference intervals, cutoff values, and patient populations evaluated.[24] Pooling results from a variety of assays may increase variability within a study and prevent meaningful application of the study results. For example, if only a few subjects receive a lower-sensitivity assay, a higher-sensitivity assay also used in the study will artificially increase the sensitivity of the former assay. In the ADJUST-PE trial, it is unlikely that use of various assays introduced significant bias; however, standardization of these assays would strengthen the conclusions.

ANSWER 3

Q3. The reference standard used in this study was adjudicated thromboembolic events as determined by further testing or clinical follow-up including autopsy.

Q3.a Who should perform adjudicated outcomes, and what data should they use?

Adjudication committees are used in two thirds of randomized controlled trials of VTE[25] and one third of randomized controlled trials overall.[26] Adjudicators should be independent, should be unbiased and blinded to the study intervention, and should form a central committee conducting a systematic and standardized evaluation of outcomes according to prespecified definitions. Members of the committee should be experts in the outcome of interest and have experience in adjudication.[27,28] Data for adjudication are frequently transmitted electronically, stripped of identifiers and the index test result. In the ADJUST-PE trial, 3 adjudicators were used. There were 15 suspected VTE events in patients with D-dimer below the age-adjusted cutoff; only 2 were deemed by the adjudication committee to have had an event.

Q3.b The authors considered only segmental or larger pulmonary arterial filling defects observed on imaging as evidence of PE. What would you predict the effect of excluding subsegmental PEs as a positive diagnosis has on diagnostic accuracy of the age-adjusted D-dimer?

In all probability, had the criterion standard included subsegmental PE, the sensitivity of the adjusted D-dimer would have decreased because patients with small and distal clots are more likely to have a normal D-dimer concentration.[29] There is debate over the clinical value of identifying subsegmental PEs. As many as 40% of low-pretest-probability patients with PE diagnosed on multi-detector CT had false-positive results, and nonhigh-probability patients are more likely to have subsegmental PEs.[30,31] There are no randomized controlled trials of anticoagulation for the treatment of subsegmental PEs, and some have questioned the need for anticoagulation.[30,32]

Q3.c How can the use of different reference standards for a diagnostic test introduce bias? In this case, can you predict the effect of using clinical follow-up as a reference standard? Would it strengthen or weaken the study conclusion?

Differential verification bias is a common problem in studies of diagnostic tests because often subjects with a positive index test result receive one reference standard and those with negative index test result will have an alternative verification of disease state, commonly clinical follow-up. If clinical follow-up is as likely to detect the disease state as the preferred reference standard, little bias will be introduced.[33-35] However, if disease state can change—for example, symptoms of PE spontaneously resolve before follow-up or a PE develops after index testing and is identified at follow-up—there is potential for bias. The former example, which misclassifies a false negative as a true negative and seems more plausible than the latter example, would artificially inflate the sensitivity of the D-dimer test, slightly weakening the study conclusion.

Q3.d Name 3 advantages to the patient of obviating contrast enhanced pulmonary angiography in an individual who does not have PE.

Patients who escape exposure to contrast-enhanced pulmonary angiography avoid the approximately 2% probability of a false-positive diagnosis and its associated risk of unnecessary

anticoagulation, the cost of the study, the radiation exposure, a 1% risk of immediate complications (such as allergy), and the 15% probability of developing contrast-induced nephropathy.[36-40]

ANSWER 4

Q4. Clinicians often rely on test performance (sensitivity, specificity, etc) as the ultimate determinant of a strategy's usefulness in improving health outcomes. Indeed, the study authors believe that the very low failure rate in this prospective validation study is evidence that an age adjusted cutoff should be used clinically. Reliance on test performance to support widespread use, however, can have undesirable effects.[4,5]

Q4.a What non–disease-specific outcomes in this study are important to adoption of the age-adjusted D-dimer cutoff level?

Diagnostic accuracy studies focus on diagnosing a target condition. However, more accurate diagnosis does not always lead to better health outcomes. For example, a test with many false positives, resulting in treatment and further downstream testing that might worsen health outcomes, can increase rate of diagnosis but ultimately lead to harm when applied widely. This study reported VTE events at 3 months as the only outcome. More important than accuracy of the diagnostic strategy might be death, major bleeding events, or cost. To adequately assess these outcomes, comparison of this diagnostic strategy with usual care could better inform decisions to implement this strategy.

Q4.b How might one design a study to determine whether a diagnostic strategy is safe and effective at improving patient-centered outcomes?

Rather than focus on diagnosis of the condition of interest, stronger evidence for implementation of a diagnostic strategy comes from impact analysis. More than improving diagnostic accuracy, the new strategy should change diagnostic thinking, change management, and improve patient-centered outcomes.[41] A randomized controlled trial of the new diagnostic strategy compared with a standard strategy for improving patient-important outcomes, including nondisease-specific outcomes, provides better evidence for changing practice. A patient-centered study might also measure the patient's perception of safety, and understanding of the physician's communication about the clinical condition.

Q4.c When is demonstrating good test performance sufficient to recommend a diagnostic test? When should further impact analysis be performed?

In this study, using an age-adjusted cutoff improved specificity without decreasing sensitivity significantly in patients being evaluated for PE. Compared with the standard cutoff, the age-adjusted cutoff likely would decrease further testing without increasing missed PE. In this situation, further randomized controlled trials will be unnecessary to recommend use of the new strategy because the potential for harm will be quite low.[42] Conversely, if a new test has increased sensitivity, this test's benefit relates to the treatment response of the additional true-positive diagnoses as determined from randomized controlled trials of therapy. For example, are the additional patients with PE found by multidetector computed tomography equally likely to benefit from anticoagulation compared with the more severe cases detected by older imaging studies?[41] Often, this is difficult to determine from existing studies, and in that scenario, a direct comparison of the diagnostic strategies should be performed in a randomized trial. Because PE represents a spectrum of clinical severity and therapy for PE is largely unproven, focusing diagnosis on low-pretest-probability patients and those with non–life-threatening PE may be misguided.[43] Again, a clinical trial examining patient-important outcomes would be useful in recommending a different approach to diagnosing PE.

ANSWER 5

Q5. If you were to design a diagnostic algorithm at your hospital, would you include an age-adjusted D-dimer cutoff level for non–high-risk patients?

Despite some of the limitations discussed here, this was a relatively well-done study with a low risk of bias. The authors demonstrate an acceptably high sensitivity of adjusted D-dimer cutoffs to exclude PE in non–high-probability patients. There is concern about overtesting, given the notably poor specificity of this strategy in elderly patients. In patients who require exclusion of PE, however, it is likely that age-adjusted cutoffs are acceptably sensitive. Because there is no impact analysis on all patient-important outcomes, it is unclear how the use of a diagnostic strategy similar to that used in this study would affect health outcomes. A key limitation may be that many D-dimer assays use other measurement units and thresholds for abnormal than the 500 ng/mL value that was used for all D-dimer assays in this study. Transporting and changing the formula for age adjustment may need independent validation with D-dimer assays that do not use 500 ng/mL as their cutoff for abnormal. Acceptable posterior probability of VTE is unique to each clinical encounter, and diagnostic decisions should be made incorporating the values and preferences of the patient. Significant areas of inquiry remain; however, it appears that an age-adjusted D-dimer cutoff remains sensitive, with improved specificity.

Section editors: Tyler W. Barrett, MD, MSCI; David L. Schriger, MD, MPH

Author affiliations: From the Department of Emergency Medicine, Indiana University School of Medicine, Indianapolis, IN.

REFERENCES

1. Newman TB, Kohn MA. *Evidence-based Diagnosis*. Dallas, TX: Cambridge University Press; 2009.
2. Barrett TW, Schriger DL. *Annals of Emergency Medicine* Journal Club. Practical considerations in HIV testing in the emergency department, characteristics of diagnostic tests, and the role of sensitivity analysis in observational studies. Answers to March 2008 Journal Club questions. *Ann Emerg Med*. 2008;52:170-181.
3. Guyatt G. *Users' Guides to the Medical Literature: Essentials of Evidence-Based Clinical Practice. 2nd ed.* Dallas, TX: McGraw-Hill Medical; 2008.

4. Bjorn M, Brendstrup C, Karlsen S, et al. Consecutive screening and enrollment on clinical trials. *J Card Fail*. 1999;5:283-284.
5. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg*. 2010;126:619-625.
6. Donkin L, Hickie IB, Christensen H, et al. Sampling bias in an internet treatment trial for depression. *Transl Psychiatry*. 2012;2:e174.
7. Lucassen W, Geersing GJ, Erkens PM, et al. Clinical decision rules for excluding pulmonary embolism: a meta-analysis. *Ann Intern Med*. 2011;155:448-460.
8. Leeflang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol*. 2009;62:5-12.
9. Montori VM, Wyer P, Newman TB, et al. Tips for learners of evidence-based medicine: 5. The effect of spectrum of disease on the performance of diagnostic tests. *CMAJ*. 2005;173:385-390.
10. Deitelzweig SB, Johnson BH, Lin J, et al. Prevalence of clinical venous thromboembolism in the USA: current trends and future projections. *Am J Hematol*. 2011;86:217-220.
11. Tsai AW, Cushman M, Rosamond WD, et al. Cardiovascular risk factors and venous thromboembolism incidence: the longitudinal investigation of thromboembolism etiology. *Arch Intern Med*. 2002;162:1182-1189.
12. Reardon G, Pandya N, Nutescu EA, et al. Incidence of venous thromboembolism in nursing home residents. *J Am Med Dir Assoc*. 2013;14:578-584.
13. Robert-Ebadi H, Righini M. Diagnosis and management of pulmonary embolism in the elderly. *Eur J Intern Med*. 2014;25:343-349.
14. Klok FA, Mos IC, Nijkeuter M, et al. Simplification of the revised Geneva score for assessing clinical probability of pulmonary embolism. *Arch Intern Med*. 2008;168:2131-2136.
15. Klok FA, Kruisman E, Spaan J, et al. Comparison of the revised Geneva score with the Wells rule for assessing clinical probability of pulmonary embolism. *J Thromb Haemost*. 2008;6:40-44.
16. Siccama RN, Janssen KJ, Verheijden NA, et al. Systematic review: diagnostic accuracy of clinical decision rules for venous thromboembolism in elderly. *Ageing Res Rev*. 2011;10:304-313.
17. Rodger MA, Maser E, Stiell I, et al. The interobserver reliability of pretest probability assessment in patients with suspected pulmonary embolism. *Thromb Res*. 2005;116:101-107.
18. Kline JA. ACP Journal Club. Simplification of the revised Geneva score did not decrease accuracy for diagnosis of pulmonary embolism. *Ann Intern Med*. 2009;150:JC3-JC15.
19. Runyon MS, Webb WB, Jones AE, et al. Comparison of the unstructured clinician estimate of pretest probability for pulmonary embolism to the Canadian score and the Charlotte rule: a prospective observational study. *Acad Emerg Med*. 2005;12:587-593.
20. van Beek EJ, van den Ende B, Berckmans RJ, et al. A comparative analysis of D-dimer assays in patients with clinically suspected pulmonary embolism. *Thromb Haemost*. 1993;70:408-413.
21. Zucker M. D-dimer for exclusion of venous thromboembolism. *Lab Med*. 2011;42:503-504.
22. Gardiner C, Pennaneac'h C, Walford C, et al. An evaluation of rapid D-dimer assays for the exclusion of deep vein thrombosis. *Br J Haematol*. 2005;128:842-848.
23. Gosselin RC, Owings JT, Kehoe J, et al. Comparison of six D-dimer methods in patients suspected of deep vein thrombosis. *Blood Coagul Fibrinolysis*. 2003;14:545-550.
24. Raby A. D-dimer assay issues and standardization: QMP-LS studies. Presented at: Mayo/NASCOLA Coagulation Testing Quality Conference; Minneapolis, MN. April 2009.
25. Stuck AK, Fuhrer E, Limacher A, et al. Adjudication-related processes are underreported and lack standardization in clinical trials of venous thromboembolism: a systematic review. *J Clin Epidemiol*. 2014;67:278-284.
26. Dechartres A, Boutron I, Roy C, et al. Inadequate planning and reporting of adjudication committees in clinical trials: recommendation proposal. *J Clin Epidemiol*. 2009;62:695-702.
27. Kradjiam SGJ, Baratelle AM, Einstein S, et al. Development of a charter for an endpoint assessment and adjudication committee. *Drug Inform J*. 2005;39:53-61.
28. Vannabouathong C, Saccone M, Sprague S, et al. Adjudicating outcomes: fundamentals. *J Bone Joint Surg Am*. 2012;94(suppl 1):70-74.
29. Kline JA, Hogg MM, Courtney DM, et al. D-dimer threshold increase with pretest probability unlikely for pulmonary embolism to decrease unnecessary computerized tomographic pulmonary angiography. *J Thromb Haemost*. 2012;10:572-581.
30. Eyer BA, Goodman LR, Washington L. Clinicians' response to radiologists' reports of isolated subsegmental pulmonary embolism or inconclusive interpretation of pulmonary embolism using MDCT. *AJR Am J Roentgenol*. 2005;184:623-628.
31. Perrier A, Roy PM, Sanchez O, et al. Multidetector-row computed tomography in suspected pulmonary embolism. *N Engl J Med*. 2005;352:1760-1768.
32. Yoo HH, Queluz TH, El Dib R. Anticoagulant treatment for subsegmental pulmonary embolism. *Cochrane Database Syst Rev*. 2014;(4):CD010222.
33. Kohn MA, Carpenter CR, Newman TB. Understanding the direction of bias in studies of diagnostic test accuracy. *Acad Emerg Med*. 2013;20:1194-1206.
34. de Groot JA, Dendukuri N, Janssen KJ, et al. Adjusting for differential-verification bias in diagnostic-accuracy studies: a Bayesian approach. *Epidemiology*. 2011;22:234-241.
35. Reitsma JB, Rutjes AW, Khan KS, et al. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol*. 2009;62:797-806.
36. Mitchell AM, Jones AE, Tumlin JA, et al. Prospective study of the incidence of contrast-induced nephropathy among patients evaluated for pulmonary embolism by contrast-enhanced computed tomography. *Acad Emerg Med*. 2012;19:618-625.
37. Courtney DM, Miller C, Smithline H, et al. Prospective multicenter assessment of interobserver agreement for radiologist interpretation of multidetector computerized tomographic angiography for pulmonary embolism. *J Thromb Haemost*. 2010;8:533-539.
38. Mitchell AM, Jones AE, Tumlin JA, et al. Incidence of contrast-induced nephropathy after contrast-enhanced computed tomography in the outpatient setting. *Clin J Am Soc Nephrol*. 2010;5:4-9.
39. Mitchell AM, Jones AE, Tumlin JA, et al. Immediate complications of intravenous contrast for computed tomography imaging in the outpatient setting are rare. *Acad Emerg Med*. 2011;18:1005-1009.
40. Mitchell AM, Kline JA. Contrast nephropathy following computed tomography angiography of the chest for pulmonary embolism in the emergency department. *J Thromb Haemost*. 2007;5:50-54.
41. Pines JM, Everett WW. *Evidence-based Emergency Care: Diagnostic Testing and Clinical Decision Rules*. Dallas, TX: Blackwell/BMJ Books; 2008.
42. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? *Ann Intern Med*. 2006;144:850-855.
43. Newman DH, Schriger DL. Rethinking testing for pulmonary embolism: less is more. *Ann Emerg Med*. 2011;57:622-627.e623.