Every Peddler Praises His Own Needle: Have Clinical Rules in the Diagnosis of Subarachnoid Hemorrhage Supplanted Lumbar Punctures Yet?

Answers to the July 2013 Journal Club Questions

Malkeet Gupta, MD, MS; Tyler W. Barrett, MD, MSCI; David L. Schriger, MD, MPH

0196-0644/$-see front matter

http://dx.doi.org/10.1016/j.annemergmed.2013.08.009

Editor's Note: *You are reading the 34th installment of* Annals of Emergency Medicine *Journal Club. This Journal Club refers to the Mark et al*[1] *article published in the July 2013 edition. 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 authors state, "Research conducted among ED [emergency department] patients with possible acute coronary syndrome suggests that patients often have much higher risk thresholds for themselves than do the treating physicians.[1]"

(INT) A. Assuming both patients and physicians are making rational decisions, list some reasons why they may have different risk thresholds. Do you think that emergency physicians have higher or lower risk thresholds than other physicians (eg, internists, pediatricians, neurosurgeons)? Why might this be so? Should our health care system try to align patient and physician acceptable risk thresholds? If so, how might this be achieved?

(INT) B. *Editor's Note: This question was incorrectly worded in the original July 2013 publication. The question has been corrected in these Answers and on the* Annals *Website. The editors apologize for the error.* Assume that 2% of all patients presenting to the ED have a chief complaint of headache, that the percentage of these ED headache patients who have subarachnoid bleeding matches the percentage suggested in the Marks et al paper. Also assume that the 97% sensitivity reported for Mark et al's decision rule has been externally validated, and that a typical emergency physician treats 3500 patients per year. On average, how many years would an emergency physician have to apply this clinical rule to miss 1 subarachnoid hemorrhage? Do these sound like reasonable numbers to you? Why or why not?

(INT) C. Noncontrast computed tomography (CT) of the brain and lumbar punctures have associated complications. Summarize the reported frequency of complications associated with CT and lumbar puncture. Using the same data listed in question 1B, calculate how often these complications might occur. Does it matter that lumbar puncture-associated complications typically occur acutely, whereas noncontrast CT radiation exposure is associated with late-occurring morbidity and mortality? How might patient age factor into the likelihood of complications associated with CT and lumbar puncture?

(NOV) D. If this clinical rule were widely adopted, in 5 years would you expect that more or fewer patients would be investigated for subarachnoid hemorrhage (the use of similar decision aids in pulmonary embolism may serve as a useful example)?

2. The authors performed a matched control study.

(INT) A. Under what circumstances are case-control studies ideal? Consider the time course of exposure and disease, the nature of the exposure and outcomes, and the frequency of the disease.

(INT) B. Discuss the importance of matching in case-control studies. Why does one match? What should one match on? What should one not match on?

(NOV) C. Why did the authors choose 3 controls for each case? Why not 1? Why not 6?

(INT) D. What are the biases of case-control studies in general? In this study specifically?

3. The authors state, "Approximately 80% of subarachnoid hemorrhage cases are due to ruptured cerebral aneurysms.[1]"

(NOV) A. What are other common causes of nontraumatic subarachnoid hemorrhage?

(NOV) B. In general terms, describe the sensitivity of a noncontrast head CT in detecting these other etiologies. When choosing a screening test, is it preferable to be highly sensitive or highly specific? What test characteristic (ie, sensitivity, specificity) is desired for a confirmatory study?

(INT) C. After a patient complaint about a post-lumbar puncture headache, a hospital administrator proposes that the ED replace lumbar puncture with CT angiography when attempting to exclude subarachnoid hemorrhage in patients with a negative initial head CT. What do you think about this, assuming the combination of CT and CT angiography is 99% sensitive for diagnosing a subarachnoid hemorrhage, as recently reported[2]?

4. The authors performed a sensitivity analysis to assess the rule's stability using a 1,000-iteration bootstrap analysis.

(INT) A. What is a sensitivity analysis? Why are sensitivity analyses often performed in observational studies?

(ADV) B. What is a bootstrap analysis? Why are bootstrap analyses especially important when evaluating clinical decision rule performance? What assumptions are invoked by bootstrap analyses?

ANSWER 1

1. The authors state, "Research conducted among ED [emergency department] patients with possible acute coronary syndrome suggests that patients often have much higher risk thresholds for themselves than do the treating physicians.[1]"

Q1.a Assuming both patients and physicians are making rational decisions, list some reasons why they may have different risk thresholds. Do you think that emergency physicians have higher or lower risk thresholds than other physicians (eg, internists, pediatricians, neurosurgeons)? Why might this be so? Should our health care system try to align patient and physician acceptable risk thresholds? If so, how might this be achieved?

Increasingly, physicians practice in a time-constrained environment that further limits their already diminished ability to form longitudinal relationships with patients. Given that emergency physicians often make decisions for strangers with little time to engage in shared decisionmaking, they tend to be overly cautious in risk assessment. Physician risk aversion may be further consolidated and informed by their personal risk tolerance, fear of malpractice, patients' demands (perceived and real), and their own, as well as society's, increasing disquiet with uncertainty.[3-13] Increased risk tolerance may correlate with the chance of a missed diagnosis or bad outcome. Because these events are often incorrectly attributed to physician error (rather than systems errors) and subject physicians to personal, professional, and legal liabilities, it is not surprising that physicians often practice medicine in a manner that minimizes their personal risk.[14]

Additionally, physicians and patients may have different risk thresholds. Physicians may be more risk averse than patients because patients have not seen the rare but devastating outcomes that physicians have seen. Furthermore, physicians, by nature of their socioeconomic status, training, or personality, may be more risk averse than patients. Fear of malpractice and the fact that their cumulative risk for a bad outcome is much higher than each patient's cumulative risk (because they have multiple exposures and the patient has only 1) are other reasons that physicians may be more risk averse than patients.

Alternatively, some patients may have high levels of health anxiety and may be swayed by popular media to levels of caution that go beyond even the physician's estimate of risk. In EDs miles away from the site of outbreaks, physicians have been asked to test for hantavirus or Severe Acute Respiratory Syndrome or to order a head CT when intracranial pathology was the last thing on their mind.

Little research has been conducted on interspecialty comparisons of risk tolerance. However, some data exist with respect to specific clinical scenarios. For example, trauma surgeons have been shown to be more risk averse than emergency physicians in evaluating patients with even minor blunt trauma and have a consequent increased desire to obtain whole-body CT in these patients.[15] On the other hand, emergency physicians regularly engage in battles with admitting internists about whether to admit borderline patients with chest pain.

Aligning risk thresholds between patients and physicians seems desirable for numerous practical, financial, and ethical reasons. However, doing so would require multiple steps. First and foremost, we as a society would have to decide what is an acceptable level of risk for particular condition. Is a 2% miss rate for acute coronary syndromes acceptable? What about for grade 1 splenic lacerations in trauma patients? Clearly these questions are complex and involve multiple competing financial, philosophical, and practical interests. Concurrent with this discussion, physicians themselves would have to define guidelines on how to characterize the risk of a particular clinical scenario, a difficult task, given the amount of both disagreement and lack of accurate risk information that currently exists.[3,15-19] Additionally, if they adhered to such guidelines, physicians would have to feel legally protected so they could begin to manage population rather than patient or personal risk.[20,21]

Complicating this entire process is the fact that individuals do not always make normatively rational decisions. Our decisions have repeatedly been shown to be subject to biases, including priming bias, framing bias, anchoring bias, hindsight bias, and a strong preference for loss aversion, all of which will influence clinical decisions in ways that will complicate any formulation of societal or professional risk guidelines.[22-24]

Q1.b Editor's Note: This question was incorrectly worded in the original July 2013 publication. The question has been corrected in these Answers and on the Annals *Web site. The editors apologize for the error. Assume that 2% of all patients presenting to the ED have a chief complaint of headache, that the percentage of these ED headache patients who have subarachnoid bleeding matches the percentage suggested in the Marks et al paper. Also assume that the 97% sensitivity reported for Mark et al's decision rule has been externally validated, and that a typical emergency physician treats 3500 patients per year. On average, how many years would an emergency physician have to apply this clinical rule to miss 1 subarachnoid hemorrhage? Do these sound like reasonable numbers to you? Why or why not?*

We first calculate the number of subarachnoid hemorrhages that would be seen by the above emergency physician per year by multiplying the number of patients seen annually (3,500) by the percentage with acute headaches (2%) by the percentage of acute headaches that are ultimately subarachnoid hemorrhage (2.5%, the average of the 1% to 4% range given by the authors). This yields 1.75 (3,500×0.02×0.025) patients with subarachnoid hemorrhage seen per year. This figure is then multiplied by the proportion of subarachnoid hemorrhage missed by CT (1–0.92=0.08) and then multiplied by the miss rate of the decision rule (1–0.97=0.03) to yield 0.0042 (1.75×0.08×0.03) subarachnoid hemorrhage missed per year.[25] The inverse of this yields 238 (1/0.0042) years to miss 1 patient with subarachnoid hemorrhage using the clinical decision rule.

From a population perspective, this decision rule could decrease a substantial number of lumbar punctures and associated complications while rarely missing a case of subarachnoid hemorrhage and therefore could be seen as reasonable. The 97% sensitivity is also in line with the accepted miss rate of acute coronary syndromes.[26] The average emergency physician would never miss a case in his or her career. On the flip side, lumbar punctures are generally low cost, are facilely performed, and can decrease the risk of a missed subarachnoid hemorrhage to near zero. Given the relatively favorable risk and cost profile of the lumbar puncture, it seems reasonable to engage appropriate patients in shared decisionmaking about their specific risk tolerance.

Q1.c Noncontrast computed tomography (CT) of the brain and lumbar punctures have associated complications. Summarize the reported frequency of complications associated with CT and lumbar puncture. Using the same data listed in question 1B, calculate how often these complications might occur. Does it matter that lumbar puncture-associated complications typically occur acutely, whereas noncontrast CT radiation exposure is associated with late-occurring morbidity and mortality? How might patient age factor into the likelihood of complications associated with CT and lumbar puncture?

The major complication associated with noncontrast CT is the risk of death from cancer because of CT's ionizing radiation. Estimates of this risk are controversial and have been largely extrapolated from the atomic bomb survivors of Hiroshima and Nagasaki. A recent review estimates the risk of death from a single head CT to be just under 1 in 1,000 for neonates.[27] This risk rapidly decreases to approximately 1 in 10,000 for 15-year-olds and then even further subsequently. The pediatric risk figure has been reproduced in a recent retrospective cohort study.[28] The most common complication associate with lumbar puncture is a postdural headache, which can occur in up to 40% of patients and lasts less than 5 days in 80% of people.[29] The risk is highest in the 18- to 30-year-old age group and is much less in children and patients older than 60 years. Low back pain has also been described in up to 35% of patients. Other described complications are much rarer, including transient cranial neuropathies, local infection, meningitis, subdural hematoma, subarachnoid bleeding, and spinal hematoma. The recent deaths associated with contaminated medications remind us that any invasive procedure and routine medication administrations are not without potential severe morbidity and mortality.[30-32]

Estimating the incidence of radiation-induced mortality from the same emergency physician described in question 1B is difficult because of the age variance of patients who present to the ED with acute atraumatic headaches (14% younger than 18 years, 65% aged 18 to 49 years, and 21% older than 50 years).[33] We can grossly estimate the risk of these age groups to be 1/10,000, greater than 1/20,000, and near zero, respectively. For simplicity, taking the average as 1/20,000 and assuming that 30% of patients with acute atraumatic headaches receive a CT scan,[33] we estimate the risk of radiation-induced mortality to be as follows:

3,500 (number of patients seen)×0.02 (percentage with headache)×0.3 (percentage that get CT)×0.00005 (percentage who will die from induced cancer)=0.00105. Taking the inverse of this, we arrive at 952, meaning that a single emergency physician would have to practice 952 years to cause a single death from radiation of the brain CT performed for patients with acute nontraumatic headache. The risk of a postdural headache is much higher: 3,500 (number of patients seen)×0.02 (percentage with headache)×0.02 (percentage who undergo lumbar puncture[34])×0.4 (risk of post dural headache)=0.56. Taking the inverse, we obtain 1.78; thus, a single emergency physician causes a postdural headache more than every other year when he or she performs lumbar punctures on patients with acute nontraumatic headaches.

Because they are easier to imagine and conceptualize, it is not surprising that the near-term risks of a lumbar puncture often take primacy to the long-term risks of radiation-induced mortality from CTs. Yet, on a population basis, the risks from CTs appear to be more substantial compared with the largely transient complications associated with lumbar puncture. Between 1998 and 2008, the percentage of patients undergoing a head CT in the ED for a nontraumatic headache has increased from 12% to 30%, a trend observed in CTs of other body parts, as well as across specialties.[33] This cognitive ease bias, combined with the difficulty for people to account for small risks in decisionmaking, is just one example of the seemingly "irrational" but predictable ways in which discussions of risk among physicians and patients can be complicated.

The effect of age on complication rates is discussed above. The average age for patients who receive a diagnosis of a subarachnoid hemorrhage approaches the age at which CT and lumbar puncture–associated complications decrease.

Q1.d If this clinical rule were widely adopted, in 5 years would you expect that more or fewer patients would be investigated for subarachnoid hemorrhage (the use of similar decision aids in pulmonary embolism may serve as a useful example)?

The medical literature abounds with clinical decision rules to aid ED medical decisionmaking.[35] These clinical decision aids, by helping identify patients who require testing, putatively aim to decrease risk, frequency of testing, and the adverse effects associated with the diagnostic studies. Despite their proliferation in the literature and their use in the ED, studies on the effect of

clinical decision rules have had mixed results. The Ottawa Knee Rule and the Ottawa Ankle Rule have demonstrated either a reduction in imaging use or mixed results.[36,37] Decision rules for imaging in patients with low back pain, minor head trauma, and pulmonary embolism have unintentionally been associated with increased use of radiographic imaging.[35] During the past 20 years, there has been a vast increase in the use of CT angiography for testing ED patients for pulmonary embolism, without any real mortality benefit.[38-40] In fact, testing itself may lead to more harm than benefit as a consequence of the adverse effects of overdiagnosis and overtesting.[39] It is entirely possible that the validated clinical decision rule in this article will also lead to additional testing because it will be indiscriminately applied to all patients with a headache, not simply those with a suspected subarachnoid hemorrhage.

ANSWER 2

2. The authors performed a matched control study.

Q2.a Under what circumstances are case-control studies ideal? Consider the time course of exposure and disease, the nature of the exposure and outcomes, and the frequency of the disease.

When a disease is rare or occurs long after the exposure, randomized trials and cohort studies are infeasible. Consider, for example, the expense and logistic challenges of following cohorts of 16- to 20-year-old smokers and nonsmokers to determine how often they get lung cancer. First, it will take many, many years for the lung cancer to develop; the investigator will likely have retired before outcomes data are collected. Second, only a small fraction of smokers (15% at most) ever get lung cancer, meaning that one would need to follow a large cohort to ensure adequate power. Third, keeping contact with persons in the cohort will be difficult and expensive during the 50-year study period; accurately tracking their smoking status during this time may be even more challenging. In such situations, it may be more practical to identify a group of patients with lung cancer and a similar group without lung cancer and query both on their smoking history. This case-control study can be more easily accomplished than a prospective cohort study.

Q2.b Discuss the importance of matching in case-control studies. Why does one match? What should one match on? What should one not match on?

There is much confusion about the role of matching in case-control studies and we encourage interested readers to read a full treatment of this topic.[41] Matching in cohort studies is designed to make the exposed and unexposed groups as similar as possible with respect to known confounders. For example, if men are 5 times more likely to get a disease than women, it would make sense to have the ratio of men to women the same in exposed and unexposed groups to eliminate this factor when considering the relationship of exposure to outcome.

Matching in case-control studies is conducted for an entirely different reason. We match to make the study more efficient, to get the maximum information for a given sample size. Matching does not eliminate confounding and may create it when it was not there to begin with. In case-control studies, we match to make things more efficient and then control for the matched variable in the analysis to eliminate confounding by the matching variable.

Matching's effect in case-control studies differs from its effect in cohort studies and trials because matching is between groups with different outcomes, not groups with different exposures. Consider the example of cigarette smoking and lung cancer. Let us imagine we were concerned that radon exposure was a confounder because radon causes lung cancer (the confounder is associated with the outcome) and radon exposure may be associated with smoking status (the confounder is associated with the exposure). In a cohort study, we would attempt to match so that that the percentage of persons with radon exposure would be the same in smokers and nonsmokers, thereby reducing the possibility that radon could confound the relationship between smoking and lung cancer. In a case-control study on this topic, we would ensure that there was an equal percentage of persons with radon exposure in the lung cancer and no lung cancer groups. This does not mean that the percentage of persons who were exposed to radon will be the same in smokers and nonsmokers. In fact, we can be assured that the percentage will differ from whatever values those relationships would have had in an unmatched study. Because radon does cause lung cancer, we would expect that the percentage of subjects exposed to radon would be higher in the lung cancer group. By forcing more radon-exposed persons into the non–lung cancer group than normally would have been there, we alter the composition of that group. If there is a positive association between radon exposure and smoking status, then matching will ensure that more people in the control group will smoke, thereby decreasing the crude association of smoking and lung cancer. By stratifying the analysis on radon status, however, we can obtain estimates of the smoking–lung cancer association that are not confounded by radon status. Matching helps because it ensures that there is sufficient balance in the tables to ensure statistical efficiency. Without matching, certain cells may have too few individuals and confidence intervals of the estimates of the odds ratio may be needlessly wide.

Finally, one cannot match on the exposure of interest. To do so is to guarantee that the frequency of exposure in the cases is identical to that in the controls, and therefore the odds ratio will be 1. Although this seems self-evident, the literature is replete with examples of this error.

Q2.c Why did the authors choose 3 controls for each case? Why not 1? Why not 6?

Case-control studies are typically carried out for rare diseases. This makes finding cases difficult because there are not that many of them. Controls, on the other hand, are typically plentiful. The mathematics of statistical power is such that adding additional controls beyond the number of cases makes the estimate of the exposure rate in the controls more precise than it would have been if the number of controls were the same as the number of cases (ie, 1:1 case:control ratio). This increased precision on the control side will result in increased precision of the measure of the difference between case and control rates. Thus, adding

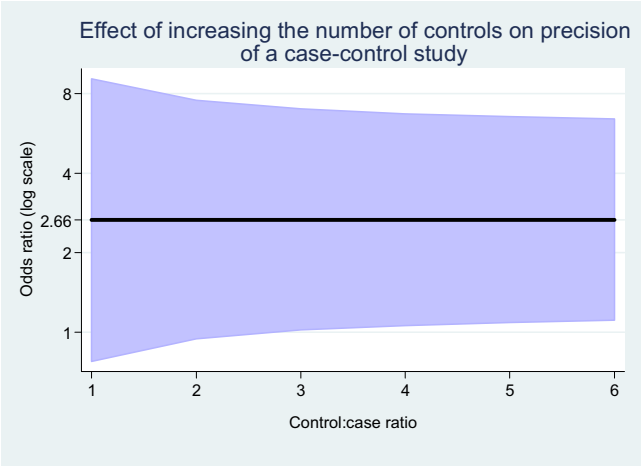

Figure. The 95% confidence interval (shaded area) narrows as the control:case ratio goes from 1 to 3, however confidence intervals for ratios greater than 3 are not sufficiently narrower to justify the cost and effort of recruiting additional controls.

controls is a way to increase the precision of the estimate without finding more cases. However, at some point making further improvements in the precision of the control estimate adds little to overall precision because the imprecision of the case estimate becomes the limiting factor. As shown in the Figure, although adding a second and third control for each case narrows the confidence interval considerably, additional controls provide marginal improvements in precision that are typically not worth the effort. Hence, it is unusual to see a case-control study with more than 3 controls per case.

Q2.d What are the biases of case-control studies in general? In this study specifically?

Just as in cohort studies, the main threat to bias relates to measurement of the exposures and measurement of the outcomes. In case-control studies, the outcomes are often the easier thing to measure because they are the organizing principle of the study and are the basis for study entry. Exposures, however, have often occurred many years ago and are subject to recall bias. For example, in a study questioning parents of children who had (did not have) a disease about a variety of exposures, the parents of the diseased children have a huge motivation for remembering every detail of the child's past, whereas control parents have no such motivation. This can lead to differential recall, and this bias can confound study results.

One of the potential biases of this study is the use of imputation to create values for missing data. Although the authors use state-of-the-art techniques to do so, such techniques are not immune to bias. All such techniques use information about patterns in other variables to infer the most likely value of the missing data. There is always the risk that the imputation algorithm is making incorrect assumptions about the nature of these patterns. As an illustration, consider variables A, B, and C, all coded as yes or no. In patients with no missing data, those with A and B=no always have C=no. An algorithm might see this and decide that in patients for whom A and B are no and C is missing C is likely no. Although that might be true, it is also possible that whatever made C be missing also makes those patients different in a way that makes it possible that C is yes. A more complete consideration of imputation can be found in answer 1c of the March 2010 Journal Club.[42]

ANSWER 3

Q3. The authors state, "Approximately 80% of subarachnoid hemorrhage cases are due to ruptured cerebral aneurysms.[1]"

Q3.a What are other common causes of nontraumatic subarachnoid hemorrhage?

When the standard subarachnoid hemorrhage evaluation fails to identify an aneurysm, the most common "other cause" remains an occult aneurysm. Occult aneurysms can occur in up to 49% of patients with an initial negative CT angiography result.[25] Reasons for the false-negative scan results include technical errors, small aneurysmal size, and obscuration of the aneurysm by thrombus or vasospasm. Other causes of nontraumatic subarachnoid hemorrhage include perimesencephalic nonaneurysmal subarachnoid hemorrhage, other vascular malformations (including cavernous malformations, venous angiomas, telangiectasias, and arterial dissections), cerebral venous thrombosis, sickle cell disease, bleeding disorders, amyloid angiopathy, and cocaine abuse–induced vasculopathy. Rare causes also include tumors, spinal aneurysms, moyamoya disease, vasculitis, and reversible vasoconstrictive syndromes.

Q3.b In general terms, describe the sensitivity of a noncontrast head CT in detecting these other etiologies. When choosing a screening test, is it preferable to be highly sensitive or highly specific? What test characteristic (ie, sensitivity, specificity) is desired for a confirmatory study?

In general, the cause of the subarachnoid hemorrhage will not be determined on the initial noncontrast head CT. Rather the definitive cause of the subarachnoid hemorrhage is often determined as a result of stepwise testing aimed first at evaluating patients for an aneurysm with either digital subtraction angiography or CT angiography.[25] These modalities will reveal most vascular lesions responsible for the subarachnoid hemorrhage such as an arteriovenous malformation or arterial dissection. Patients with subarachnoid hemorrhage and negative evaluation results based on digital subtraction angiography or CTA subsequently undergo a gadolinium-enhanced magnetic resonance imaging of the brain and spinal cord. One exception to the above may be perimesencephalic nonaneurysmal subarachnoid hemorrhage, which is often suspected according to the initial location of subarachnoid hemorrhage bleeding. However, even in these cases, aneurysmal subarachnoid hemorrhage must be excluded with additional testing before a definitive diagnosis can be made.

If one were to design a perfect test for a disease, it would capture all the people with the disease while not identifying healthy people as having it. Screening tests should be designed such that they capture all patients with the disease (otherwise

defined as true positives [TPs]) and minimize the patients who have the disease but are missed on the screening test (otherwise defined as false negatives [FNs]). The proportion of these variables is mathematically defined as the sensitivity of a test (TP/TP+FN) and can be described as the proportion of actual positives that are correctly identified as such. Conversely, confirmatory tests should be focused on ensuring that healthy people are identified as disease free (true negatives [TNs]) and minimize the number of healthy people incorrectly identified as having disease (false positives [FPs]). The proportion of these can be mathematically defined as the specificity of a test (TN/TN+FP).

In actuality, for every test there is a tradeoff between capturing all patients with a disease and falsely identifying healthy persons as diseased. This tradeoff can be articulated with the test characteristics of the test (sensitivity and specificity) and the loss function, the explicit description of how one values an FN compared with an FP. For example, if, for a certain condition, an FP test result can easily be identified as such by a second, inexpensive, harmless test, then an FP is not so bad, and if missing the disease is a bad thing, we would want to avoid FNs much more than FPs. In that case, we would be willing to gain a higher sensitivity at the expense of lower specificity because that combination meets our goal of minimizing FN even if we incur additional FPs. In a different circumstance, however, we might be willing to accept FNs but cannot tolerate FPs and we would choose a cut off point for the test that had a high specificity even if the sensitivity were lower. To recapitulate, there is no way to optimize a test without specifying how one values an FN versus an FP. Articles that claim optimization of sensitivity and specificity without explicitly stating the tradeoff are making an incomplete statement, and those that maximize the sum of sensitivity and specificity are essentially stating that FPs and FNs are equally distasteful, a condition that is seldom true.

Q3.c After a patient complaint about a post-lumbar puncture headache, a hospital administrator proposes that the ED replace lumbar puncture with CT angiography when attempting to exclude subarachnoid hemorrhage in patients with a negative initial head CT. What do you think about this, assuming the combination of CT and CT angiography is 99% sensitive for diagnosing a subarachnoid hemorrhage, as recently reported[2]?

Approximately 5% of the population has at least 1 saccular intracranial aneurysm, according to autopsy and radiography series.[25] The incidence of aneurysmal subarachnoid hemorrhage, however, is reported to be 3 to 25 per 100,000. Therefore, of the 10 to 15 million Americans living with an aneurysm, only 30,000 will have an subarachnoid hemorrhage annually. Studies of the natural history of unruptured aneurysms have defined 7 mm as the cutoff beyond which the risk of subarachnoid hemorrhage correlates with increased size of the aneurysm. The annual rupture rate for aneurysms less than 5 mm is about 0.5%.

Using the numbers we derived in answer 1B, we can calculate the number of patients with aneurysms whom the combination of CT and CT angiography would identify. We first multiply the annual number of patients treated by our emergency physician (3,500) by the percentage with acute headaches (2%) by the percentage who receive a CT (30%) by the percentage of the population who have an aneurysm (5%): $3{,}500\times0.02\times0.05=3.5$ patients found to have an aneurysm on CT/CT angiography. We have already calculated that this same emergency physician treats 1.75 patients with subarachnoid hemorrhage every year. Therefore, the combination of CT and CT angiography will increase by 100% the number of patients undergoing treatment for subarachnoid hemorrhage. This group of patients may incur the potential adverse effects of treatment (eg, surgical complications and financial and emotional burden), whereas only a few will benefit from the finding of an unruptured aneurysm on CT angiography (those with aneurysms >7 mm).

ANSWER 4

Q4. The authors performed a sensitivity analysis to assess the rule's stability using a 1,000-iteration bootstrap analysis.

Q4.a What is a sensitivity analysis? Why are sensitivity analyses often performed in observational studies?

All investigations that enroll a convenience sample are at risk for sampling bias. A sensitivity analysis is a statistical technique used to determine whether sampling bias may be of sufficient magnitude to alter the meaning of the results.[43] The March 2008 Journal Club answers provide a detailed example of how to perform a sensitivity analysis and a spreadsheet template for a reader to practice.[43] A sensitivity analysis examines how results might change if assumptions made in the primary analysis are relaxed or altered. Whereas classical statistical analysis examines how results might change due to chance alone, sensitivity analyses typically examine how non-random phenonmena (biases) might alter results. For example one might ask "In a study using convenience sampling in which enrollees' initial problem was classified as mild, moderate, or severe, how would results change if there were 15% more severe patients and 15% less mild ones. Performing a sensitivity analysis is critical to determine how potential biases might influence the results. A sensitivity analysis that provides results similar to the base values affirms the study's internal validity and minimizes the concern for sampling bias or other forms of bias confounding the results.[43]

Q4.b What is a bootstrap analysis? Why are bootstrap analyses especially important when evaluating clinical decision rule performance? What assumptions are invoked by bootstrap analyses?

Bootstrapping is a statistical resampling technique that can be used to internally validate a prognostic model (ie, clinical decision rule). Steyerberg et al recommend bootstrapping for estimating internal validity of a predictive logistic regression model.[44] The method is described as "[b]ootstrapping replicates the process of sample generation from an underlying population by drawing samples with replacement from the original data set, of the same size as the original data set."[44,45] What is unique about bootstrapping is the idea of select and replace. In each bootstrap repetition, the computer selects the same number of patients as were in the original sample, but the computer might select an individual patient 0, 1, or *n* times. The prediction

model is then derived in each bootstrap sample and applied to the bootstrap and original samples. The model's performance in the bootstrap sample and the original sample reflects the apparent validation and the internal validation, respectively. The difference in these 2 values indicates the estimated optimism of the model's discrimination.[46,47] This sampling procedure is typically repeated at least 500 to 1,000 times to obtain stable results.[48,49] This optimism value is then subtracted from the apparent performance of the original model derived from the original sample to obtain an optimism-corrected, stable estimate of model performance.[46,47] The bootstrap method is more efficient than the split sample analysis because it permits use of the entire sample for development of the decision rule rather than splitting the sample into 2 smaller derivation and validation groups.[44] Furthermore, the bootstrap validation method has been shown to appropriately reflect all sources of model uncertainty, including variable selection.[47]

The bootstrap technique does have some limitations in that only automated modeling strategies (eg, fitting a full model without selection or automated stepwise variable selection) can be chosen for the original model development.[47] Variable reduction techniques, testing both univariate and multivariate *P* values, and assessing proportional hazards for a Cox regression model may be difficult to replicate in the bootstrap testing.[47] The bootstrap method assumes that bootstrap sample's distribution of data is a reasonable representation of the distribution of data within the study population.[49] If this assumption is not met (eg, a poorly designed study with severely biased data), then the bootstrap resampling may result in additional sampling bias, leading to invalid results and conclusions.[49]

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

Author affiliations: From the University of California, Los Angeles, CA (Gupta, Schriger); and the Vanderbilt University Medical Center, Nashville, TN (Barrett).

REFERENCES

1. Mark DG, Hung YY, Offerman SR, et al. Nontraumatic subarachnoid hemorrhage in the setting of negative computed tomography results: external validation of a clinical and imaging prediction rule. *Ann Emerg Med*. 2013;62:1-10.
2. McCormack RF, Hutson A. Can computed tomography angiography of the brain replace lumbar puncture in the evaluation of acute-onset headache after a negative noncontrast cranial computed tomography scan? *Acad Emerg Med*. 2010;17:444-451.
3. Baumann BM, Chen EH, Mills AM, et al. Patient perceptions of computed tomographic imaging and their understanding of radiation risk and exposure. *Ann Emerg Med*. 2011;58:1-7.
4. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med*. 1989;320:1489-1491.
5. Katz DA, Williams GC, Brown RL, et al. Emergency physicians' fear of malpractice in evaluating patients with possible acute cardiac ischemia. *Ann Emerg Med*. 2005;46:525-533.
6. Mangione-Smith R, McGlynn EA, Elliott MN, et al. Parent expectations for antibiotics, physician-parent communication, and satisfaction. *Arch Pediatr Adolesc Med*. 2001;155:800-806.
7. Ong S, Nakase J, Moran GJ, et al. Antibiotic use for emergency department patients with upper respiratory infections: prescribing practices, patient expectations, and patient satisfaction. *Ann Emerg Med*. 2007;50:213-220.
8. Pearson SD, Goldman L, Orav EJ, et al. Triage decisions for emergency department patients with chest pain: do physicians' risk attitudes make the difference? *J Gen Intern Med*. 1995;10:557-564.
9. Spurling GK, Del Mar CB, Dooley L, et al. Delayed antibiotics for respiratory infections. *Cochrane Database Syst Rev*. 2007(3);CD004417.
10. Stearns CR, Gonzales R, Camargo CA Jr, et al. Antibiotic prescriptions are associated with increased patient satisfaction with emergency department visits for acute respiratory tract infections. *Acad Emerg Med*. 2009;16:934-941.
11. Stivers T, Mangione-Smith R, Elliott MN, et al. Why do physicians think parents expect antibiotics? what parents report vs what physicians believe. *J Fam Pract*. 2003;52:140-148.
12. Sun BC, Adams J, Orav EJ, et al. Determinants of patient satisfaction and willingness to return with emergency care. *Ann Emerg Med*. 2000;35:426-434.
13. Welschen I, Kuyvenhoven M, Hoes A, et al. Antibiotics for acute respiratory tract symptoms: patients' expectations, GPs' management and patient satisfaction. *Fam Pract*. 2004;21:234-237.
14. Wears RL. The error of counting "errors." *Ann Emerg Med*. 2008;52:502-503.
15. Gupta M, Schriger DL, Hiatt JR, et al. Selective use of computed tomography compared with routine whole body imaging in patients with blunt trauma. *Ann Emerg Med*. 2011;58:407-416.
16. Ginde AA, Delaney KE, Pallin DJ, et al. Multicenter survey of emergency physician management and referral for hyperglycemia. *J Emerg Med*. 2010;38:264-270.
17. Puri S, Hu R, Quazi RR, et al. Physicians' and midlevel providers' awareness of lifetime radiation-attributable cancer risk associated with commonly performed CT studies: relationship to practice behavior. *AJR Am J Roentgenol*. 2012;199:1328-1336.
18. Correia MJ, Hellies A, Andreassi MG, et al. Lack of radiological awareness among physicians working in a tertiary-care cardiological centre. *Int J Cardiol*. 2005;103:307-311.
19. Lee CI, Haims AH, Monico EP, et al. Diagnostic CT scans: assessment of patient, physician, and radiologist awareness of radiation dose and possible risks. *Radiology*. 2004;231:393-398.
20. Hermer LD, Brody H. Defensive medicine, cost containment, and reform. *J Gen Intern Med*. 2010;25:470-473.
21. Sloan FA. *Medical Malpractice*. Cambridge, MA: MIT Press; 2008.
22. Kahneman D. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus & Giroux; 2011.
23. Taleb N. *The Black Swan: The Impact of the Highly Improbable*. 2nd ed. New York, NY: Random House Trade Paperbacks; 2010.
24. Wears RL. Risk, radiation, and rationality. *Ann Emerg Med*. 2011;58:9-11.
25. Siddiq F. Nonaneurysmal subarachnoid hemorhhage. In: Biller J, ed. *UptoDate*. Waltham, MA, 2013. Available at: http://www.uptodate.com/contents/nonaneurysmal-subarachnoid-hemorrhage?source=search_result&search=noaneurysmal+subarachnoid+hemorhhage&selectedTitle=1~150(2013). Accessed May 9, 2013
26. Brown TB, Cofield SS, Iyer A, et al. Assessment of risk tolerance for adverse events in emergency department chest pain patients: a pilot study. *J Emerg Med*. 2010;39:247-252.
27. Brenner DJ, Hall EJ. Computed tomography—an increasing source of radiation exposure. *N Engl J Med*. 2007;357:2277-2284.

28. Pearce MS, Salotti JA, Little MP, et al. Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *Lancet*. 2012;380:499-505.
29. Evans RW. Complications of lumbar puncture. *Neurol Clin*. 1998;16:83-105.
30. Kainer MA, Reagan DR, Nguyen DB, et al. Fungal infections associated with contaminated methylprednisolone in Tennessee. *N Engl J Med*. 2012;367:2194-2203.
31. Kauffman CA, Pappas PG, Patterson TF. Fungal infections associated with contaminated methylprednisolone injections. *N Engl J Med*. 2013;368:2495-2500.
32. Pettit AC, Kropski JA, Castilho JL, et al. The index case for the fungal meningitis outbreak in the United States. *N Engl J Med*. 2012;367:2119-2125.
33. Gilbert JW, Johnson KM, Larkin GL, et al. Atraumatic headache in US emergency departments: recent trends in CT/MRI utilisation and factors associated with severe intracranial pathology. *Emerg Med J*. 2012;29:576-581.
34. Goldstein JN, Camargo CA Jr, Pelletier AJ, et al. Headache in United States emergency departments: demographics, work-up and frequency of pathological diagnoses. *Cephalalgia*. 2006;26:684-690.
35. Schriger DL, Newman DH. Medical decisionmaking: let's not forget the physician. *Ann Emerg Med*. 2012;59:219-220.
36. Cameron C, Naylor CD. No impact from active dissemination of the Ottawa Ankle Rules: further evidence of the need for local implementation of practice guidelines. *CMAJ*. 1999;160:1165-1168.
37. Stiell IG, Bennett C. Implementation of clinical decision rules in the emergency department. *Acad Emerg Med*. 2007;14:955-959.
38. Hoffman JR, Cooper RJ. Overdiagnosis of disease: a modern epidemic. *Arch Intern Med*. 2012;172:1123-1124.
39. Newman DH, Schriger DL. Rethinking testing for pulmonary embolism: less is more. *Ann Emerg Med*. 2011;57:622-627.
40. Huckins DS, Price LL, Gilley K. Utilization and yield of chest computed tomographic angiography associated with low positive D-dimer levels. *J Emerg Med*. 2012;43:211-220.
41. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008:175-181.
42. Barrett TW, Brywczynski JJ, Schriger DL. Annals of emergency medicine journal club. Is the golden hour tarnished? registries and multivariable regression: answers to the March 2010 Journal Club questions. *Ann Emerg Med*. 2010;56:188-200.
43. 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.
44. Steyerberg EW, Harrell FE Jr, Borsboom GJ, et al. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. 2001;54:774-781.
45. Efrom B, Tibishirani R. *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability*. New York, NY: Chapman & Hall; 1993.
46. Harrell FE Jr, ed. *Regression Model Strategies, With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York, NY: Springer; 2001.
47. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer; 2009.
48. Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol*. 2003;56:441-447.
49. Haukoos JS, Lewis RJ. Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. *Acad Emerg Med*. 2005;12:360-365.