

# Computed Tomography for Subarachnoid Hemorrhage: What Should We Make of the “Evidence”?

## Computed Tomography for Subarachnoid Hemorrhage: What Should We Make of the “Evidence”?

Jerome R. Hoffman, MA, MD

Departments of Medicine and Emergency Medicine

UCLA School of Medicine

Los Angeles, CA

Reprints not available from the author.

**Address for correspondence:** Jerome R. Hoffman, MA, MD, UCLA Emergency Medicine Center, 924 Westwood Blvd, Suite 300, Los Angeles, CA 90095; 310-794-0573; E-mail [jrh@ucla.edu](mailto:jrh@ucla.edu).

Copyright © 2001 by the American College of Emergency Physicians.

0196-0644/2001/\$35.00 + 0

47/1/113924

doi:10.1067/mem.2001.113924

## Computed Tomography for Subarachnoid Hemorrhage: What Should We Make of the “Evidence”?

**Note from the Editor in Chief:** For the past 2 years, the Evidence-Based Emergency Medicine (EBEM) series has been published in *Annals of Emergency Medicine*. Our goal has been to introduce readers to a more formalized approach to information analysis. Such analysis is obviously important because, as physicians, we must carefully consider the tremendous amount of information accessible given today's technology. Valid and relevant information is the critical construct in our decisions regarding patients and their care.

The process of EBEM, including the essential elements, has been well described in *Annals*.<sup>1</sup> Furthermore, the application of this process is evident in each of the discussions on relevant topics subsequently published. Certainly, by applying this approach, we can improve our critical thinking. However, no approach, like no human, is without flaw. In the following Feedback article, Dr. Jerome Hoffman constructively critiques the process of EBEM, as well as its application. *Annals* has chosen to foster this discussion because it is our responsibility to present a balanced assessment of the benefits and liabilities of any information published in the journal, especially when so critical to patient care.

*Annals* will continue to publish the EBEM series; the approach is solid and the outcome is beneficial when carefully considered. Any process that helps us, as clinicians, approach 100% assurance is obviously desirable. However, only one thing in medicine is certain—there is no certainty. Caveats exist for any analytical approach, and EBEM is no exception. Information obtained through this approach is most useful when it is considered along with our knowledge base, experience, and gestalt regarding the patient. In the end, though, the ultimate test is if

you or your loved one is the patient. Statistical analysis based on the population as a whole becomes relatively meaningless. It becomes very personal; the N is 1. Only 100% certainty is acceptable. Approaching this level requires all of our knowledge, skills, and instincts, and EBEM is another tool for us to use properly.

1. Rowe BH, Klassen T, Wyer PC. One is the only number that you'll ever need! *Ann Emerg Med.* 2000;36:520-523.

[Hoffman JR. Computed tomography for subarachnoid hemorrhage: what should we make of the "evidence"? *Ann Emerg Med.* March 2001;37:345-349.]

Long before it acquired the unimpeachable name of "evidence-based medicine" (EBM), the discipline generally referred to as "clinical epidemiology" offered practitioners important tools to help understand the analysis and application of information in the decisionmaking process. Sophistication about prior probability, criterion standards, likelihood ratios, Bayes' theorem, or treatment thresholds could help a clinician with all sorts of clinical decisions, ranging from whether a new drug is better than an older one, to when to order a CBC count or a computed tomographic (CT) scan, to how to narrow a differential diagnosis. With the advent of the "information revolution," allowing for rapid access to all manner of published data, these important tools and concepts have developed and transformed somewhat into the EBM movement, which also appropriately emphasizes the pitfalls associated with making judgments in the absence of real evidence.

Although the value of proper application of knowledge derived from research seems straightforward and self-evident, the issues become less clearcut when we try to define both what constitutes "evidence," and the methodology for applying it to individual patients, as several articles in *Annals* have made clear. Rowe et al<sup>1</sup> nicely presented the need for an EBM construct in clinical medicine, and some of the reasons why old-fashioned reliance on experts, or anecdote, or personal experience, can be terribly misleading. They stress the notion of "best evidence," with the randomized controlled trial at the top of a "hierarchy of evidence." But as Schriger<sup>2</sup> noted in his counterpoint, the idea that "evidence" is binary—either 100% grade A true, or 0% useless—makes no sense, and any system that treats it as such is doomed to failure.

Without revisiting their discussion of EBM in detail, "evidence" from (even excellent) research can be quite misleading, and anecdote, experience, expertise, and judgment, for all their imperfections, must always play a fundamental role in clinical reasoning. Furthermore,

although the principles of EBM can help us be better clinicians, its limitations are most obvious when we try to use it as a path to the "truth," replete with precise mathematical calculations. Yet this is exactly what the series of exercises in *Annals* called "Evidence-Based Emergency Medicine," with literature searches to hunt down sensitivity and specificity, and instructions on "how to" calculate posterior probability, is all about. The November 2000 article by Edlow and Wyer<sup>3</sup> was a wonderful example, because it is about as good an attempt to make sense of this process as is possible, and yet the accuracy of its calculations can and should be challenged at multiple points along the way.

Edlow and Wyer<sup>3</sup> began by suggesting that their hypothetical patient has a prior probability of subarachnoid hemorrhage (SAH) of 15%. They base this claim on the selective citation of 2 (out of many) published case series—although nonsystematic use of the literature is itself a violation of one of the core principles of EBM. The authors are forced to do this, because there is in fact no good way to know the truth. The many and various reports (including the 2 cited) are easily challenged on methodologic bases, and the numbers they provide vary widely, even including a prior probability of 68% in 1 of the 4 articles ultimately chosen by Edlow and Wyer for analysis of CT accuracy.<sup>4</sup> If we change the prior probability of SAH in the hypothetical patient to 5%, or conversely to 25%, all subsequent calculations are dramatically altered.

Next, Edlow and Wyer<sup>3</sup> tell us that the patient wishes to set a threshold of less than 1% for ruling out SAH. The concept of diagnostic and treatment thresholds is critical, but extremely difficult for most of us (no less most laypeople) to understand. How many patients know what a 1% chance really means? Would a patient reach a different conclusion if we framed the issue differently ("You have a 99% likelihood that everything's fine")? Should we change our approach if a different patient said, "All I want is a better than even chance that I'm OK"? And what if a patient insisted on the impossible threshold of "certainty" (0% chance of SAH)?

Edlow and Wyer<sup>3</sup> then go on to gather the "evidence," but reject almost all the articles that they find. They give us good reasons why the rejected articles do not meet their a priori criteria, but does that mean there is nothing we can learn from the many other articles published on this topic? Or from the numerous small case series of missed SAH after a false-negative CT scan report (or one initially interpreted as negative, only to be reinterpreted as "actually showing a small amount of blood" once the

terrible rebleeding occurred)? Or from the many similar anecdotal cases in clinical practice, with at least one of which most readers are personally aware?

Edlow and Wyer<sup>3</sup> do a wonderful job of highlighting the problems with the extremely limited evidence on which they do rely (ultimately limited to a single article), but sometimes proceed to ignore their own cautions. They discuss the fact that their hypothetical patient may be different from those included in the cited studies, although a great deal of their screening of the literature was an attempt to make the match as close as possible. Nevertheless, after acknowledging that almost half of Morgenstern et al's<sup>5</sup> patients had meningismus (but ignoring that 40% also had altered mental status—another factor that makes them extremely different from the hypothetical patient to whom the “evidence” is being applied), they forge ahead as if this does not matter. The truth is, it matters greatly, because of spectrum bias.

Edlow and Wyer<sup>3</sup> do basically the same thing regarding the important observation that CT scan interpretation is operator-dependent. First, they note that the cited studies used faculty neuroradiologists, whose accuracy in diagnosing SAH is almost certainly far better than that of most typical community general radiologists. But then, they ignore the impact this undoubtedly has, in doing their calculations. (Edlow and Wyer do not acknowledge other related biases, such as that the authors all set out with the explicit intent to “prove” how well CT scanning performs, or that in the other study on which they somewhat rely,<sup>4</sup> a CT scan result was defined as true-positive if any 1 of 3 neuroradiologists identified the SAH. What this means is that if 2 of 3 neuroradiologists did not detect a hemorrhage, the authors still counted that as a case where CT scanning worked.)

Regarding Edlow and Wyer's<sup>3</sup> calculations, they ask the reader to assume that the CT scan is 100% specific, because it makes the calculations easier, and because that is how we are likely to interpret a positive CT scan finding in actual practice. Given the complex nature of what they have asked us to do, anything that makes the process easier is welcome. But because the entire rationale for this exercise is to calculate a precise mathematical (posterior) probability of disease, such a glib sacrifice of accuracy seems strange indeed. It may be reasonable in clinical practice to act as if all positive CT scan findings are truly positive, but that does not mean we can pretend that there are no false-positive CT scan results when we are doing these calculations. Although there is no available evidence regarding the actual specificity of CT scanning for

SAH, it is surely not 100% (especially in the hands of general radiologists); if we use a specificity of 90%, our calculations, once again, are very different.

Another major threat to the accuracy of the studies cited is verification bias, or the uncertainty that all the negative CT scan findings were actually true-negatives. The gold standard in Morgenstern et al<sup>5</sup> was the lumbar puncture (LP), and Edlow and Wyer<sup>3</sup> make a reasonable case for why this is appropriate, because there are essentially no false-negative results of LPs for SAH. But the definition of a “negative” LP result in Morgenstern et al is not a completely normal LP result, but rather one that failed to meet arbitrary and unproven cutoffs for total and relative RBC counts. There were 20 patients with a negative CT scan interpretation and an LP result that was not normal, but did not meet the threshold considered to represent SAH; if even 1 or 2 of these patients actually had a warning leak, and thus a false-negative CT scan finding, the calculations become much worse.

Finally, there is the matter of spectrum bias, the nature of which Edlow and Wyer<sup>3</sup> address very well, but the critical implications of which, regarding the attempt to apply evidence from research to individual patients, in a precise mathematical manner, they overlook. One of the fundamental principles of EBM is that sensitivity and specificity (unlike predictive values) are independent of the patient population to whom a test is applied. This is an important concept, theoretically—but unfortunately, it is not quite true. The sensitivity of a test may be stable in the entire universe of patients with the disease in question, but it changes greatly as the test is applied to subsets of patients along the spectrum of disease. Thus, a test that detects most patients with an advanced disease or condition usually performs far less well in a group with early disease. (Similarly, specificity changes if the “spectrum” of “nondiseased” subjects changes in the control group—more controls “without appendicitis” will have a high WBC count if the controls are emergency department patients with acute abdominal pain than if they are a sample of healthy patients donating blood.)

A serum human chorionic gonadotropin (hCG) test (one of the most accurate in all of medicine) is about 100% sensitive for identifying a 4-month normal gestation, but it is about 0% sensitive at the moment of fertilization. If a patient came to the ED complaining of 3 missed periods, morning sickness, and abdominal distention, we would assign her a fairly high prior probability of a viable intrauterine pregnancy. If her hCG test result is negative, however, we would have to reconsider—because the

hCG test is so extremely accurate for a (normal) second-trimester pregnancy that the posterior probability is now virtually zero. Conversely, if another woman came to the ED and said she had just had unprotected intercourse 10 minutes ago and was concerned about being pregnant, we would assign a low prior probability, but a negative hCG test result would not rule out early pregnancy—in fact, it would not change the prior probability at all—given the complete insensitivity of the (very same) test at that stage in the process.

We do not need “evidence” to know that this is true. Although there is no evidence about the impact of spectrum bias on the sensitivity of CT scanning in SAH, CT scanning must be far better at detecting lots of subarachnoid blood than it is at detecting a few milliliters. A lot of blood causes significant clinical effects, which leads us to assign a high prior probability of SAH. Consider, for example, a patient brought to the ED comatose and hemiparetic after having suddenly clutched his head, cried out in pain, vomited, and collapsed. A negative CT scan finding in this patient would be persuasive, because hemorrhage would not cause such a severe presentation unless there were a lot of blood—and CT scanning does not miss a lot of blood. Conversely, we appropriately assign a low prior probability of SAH to someone who is awake and alert, without focal findings, and with only a sudden-onset headache. Nevertheless, if that patient has an SAH, it is almost certainly a small one, and not surprisingly it is precisely in such patients that all the false-negative CT scan findings are concentrated—a negative CT scan result in this patient may lower the prior probability hardly at all. Thus, applying a “sensitivity” of CT scanning based on how it performed in patients further along the spectrum of disease severity (with meningismus or altered mental status) to patients with just a headache must lead to a drastic overestimate of the use and meaning of a negative test result. And if that is true, then all the calculations in this exercise lose their meaning.

Given the many demands on our time, we must ask how much benefit we derive from any given undertaking. Slawson et al<sup>6</sup> suggested that “usefulness” is equivalent to the relevance and validity of an activity, divided by the effort it takes to perform. Readers must surely have been struck by the massive effort that went into Edlow and Wyer’s<sup>3</sup> analysis, and the fact that it would be suicide for any of us to try to accomplish this, while working in an ED, for any single patient. We understand, of course, that they do not really mean for us to go through the process while in the ED, or for each of the endless clinical decisions we face. Their real goal, I suspect, is to help us

understand the process that is involved and reach a reasonable conclusion about the clinical question being addressed. Both of these are worthy goals, but in each case, my interpretation is a bit different than theirs.

As to the process, it is clear that “the best available evidence” is flimsy, and the little that exists is not very good. While waiting for better evidence, an EBM literature search such as this can be valuable in teaching us about the clinical topic, and identifying what remains unknown. The erudite discussion by Edlow and Wyer<sup>3</sup> also highlighted the tremendous limitations of “the evidence” and “the literature.” Together, these should convince us of the folly of trying to do a precise mathematical calculation, applying largely speculative numbers to a questionable classic Bayesian formulation, in either the hypothetical case in their article, or any such case in our own practice.

Absent definitive evidence of the “truth” regarding the evaluation of sudden-onset headache, what should we do about CT scanning and LP the next time we see such a patient? It is clear that CT scanning is not extremely sensitive for SAH overall, and is almost certainly far less sensitive for small SAH—even with the latest scanners, and even with neuroradiologists as readers, and even in the context of the controlled environment of a clinical study, and even with inadequate follow-up and nonuniversal gold standard tests distorting (favorably) the measured results, and even with publication bias and proponent authors trying to prove a point.

So, how should we counsel patients? First, we should not pretend to know that after a negative CT scan result “she still has a roughly 2% . . . chance of having a SAH. . . [and] the actual figure could be as high as 4%,” based on “evidence-based” calculations that are unreliable and almost certainly inaccurate. We should instead admit that we cannot provide her with any precise estimates. Then, based on the really best information we currently have, we should tell her the following: the only way we could provide any real assurance that she does not have SAH, a condition that could soon maim or kill her if we fail to find it, but which with identification and treatment has an excellent prognosis, is by doing an LP.

People sometimes argue that any information (evidence) is better than none at all, but they are wrong, if the information misleads us. Similarly, “a little knowledge can be a dangerous thing,” especially if we pretend it is a lot of knowledge.

## REFERENCES

1. Rowe BH, Klassen T, Wyer PC. One is the only number that you'll ever need. *Ann Emerg Med.* 2000;36:520-523.

- 
2. Schriger DL. One is the loneliest number: be skeptical of evidence summaries based on limited literature reviews. *Ann Emerg Med.* 2000;36:517-519.
  3. Edlow JA, Wyer PC. How good is a negative cranial computed tomographic scan result in excluding subarachnoid hemorrhage? *Ann Emerg Med.* 2000;36:507-516.
  4. van der Wee N, Rinkel GJE, Hasan D, et al. Detection of subarachnoid haemorrhage on early CT: is lumbar puncture still needed after a negative scan? *J Neurol Neurosurg Psychiatry.* 1995;58:357-359.
  5. Morgenstern LB, Luna-Gonzales H, Huber JC, et al. Worst headache and subarachnoid hemorrhage: prospective computed tomography and spinal fluid analysis. *Ann Emerg Med.* 1998;32:297-304.
  6. Slawson DC, Shaughnessy AF, Bennett JH. Becoming a medical information master: feeling good about not knowing everything. *J Fam Pract.* 1994;38:505-513.