

DIFFERENTIAL DIAGNOSIS

W. Scott Richardson, Mark Wilson, Jeroen Lijmer,
Gordon Guyatt, and Deborah Cook

The following EBM Working Group members also made substantive contributions to this section: Peter Wyer, C. David Naylor, Jonathan Craig, Luz Maria Letelier, and Virginia Moyer

IN THIS SECTION

Finding the Evidence

Are the Results Valid?

Did the Investigators Enroll the Right Patients? Was the Patient Sample Representative of Those With the Clinical Problem?

Was the Definitive Diagnostic Standard Appropriate? Was the Diagnostic Process Credible?

For Initially Undiagnosed Patients, Was Follow-up Sufficiently Long and Complete?

What Are the Results?

What Were the Diagnoses and Their Probabilities?

How Precise Are the Estimates of Disease Probability?

How Can I Apply the Results to Patient Care?

Are the Study Patients Similar to Those in My Own Practice?

Is It Unlikely That the Disease Possibilities or Probabilities Have Changed Since This Evidence Was Gathered?

Clinical Resolution



CLINICAL SCENARIO

A 33-Year-Old Man With Palpitations: What Is the Cause?

You are a primary care physician seeing a patient from your practice, a 33-year-old man who presents with heart palpitations. He describes the new onset as episodes of fast, regular chest pounding that come on gradually, last from 1 to 2 minutes, and occur several times per day. He reports no relationship of symptoms to activities and no change in exercise tolerance. You have previously noted that this patient tends to suffer from anxiety, and he now tells you that he fears heart disease. He has no other symptoms, no personal or family history of heart disease, and he takes no medications. You find his heart rate is 90 bpm and regular, and physical examinations of his eyes, thyroid gland, and lungs are normal. His heart sounds also are normal, without click, murmur, or gallop. His 12-lead ECG is normal, without arrhythmia or signs of preexcitation.

You suspect that anxiety explains this patient's palpitations, that they are mediated by hyperventilation, and that they may be part of a panic attack. Also, although there are no findings to suggest cardiac arrhythmia or hyperthyroidism, you wonder if these disorders are common enough in this sort of patient to warrant serious consideration. You reject pheochromocytoma as too unlikely to consider further. Thus, you can list causes of palpitations, but you want more information about the frequency of these causes to choose a diagnostic workup. You ask the question, "In patients presenting with heart palpitations, what is the frequency of underlying disorders?"

FINDING THE EVIDENCE

Your office computer networks with the medical library, where MEDLINE is on CD-ROM. In the MEDLINE file for current years, you enter three text words: “palpitations” (89 citations), “differential diagnosis” (7039 citations), and “cause or causes” (71,848 citations). You combine these sets, yielding 17 citations. Reviewing the titles and abstracts onscreen, you see a paper by Weber and Kapoor that explicitly addresses the differential diagnosis in patients presenting with palpitations.¹ With a keystroke and a mouse click, you review this article’s full text.

ARE THE RESULTS VALID?

Table 1C-1 summarizes the guides for an article about the diagnostic possibilities.

TABLE 1C-1

Users’ Guide for an Article About Differential Diagnosis

Are the results valid?

- Did the investigators enroll the right patients? Was the patient sample representative of those with the clinical problem?
- Was the definitive diagnostic standard appropriate? Was the diagnostic process credible?
- For initially undiagnosed patients, was follow-up sufficiently long and complete?

What are the results?

- What were the diagnoses and their probabilities?
- How precise are the estimates of disease probability?

How can I apply the results to patient care?

- Are the study patients similar to the one being considered in my own practice?
- Is it unlikely that the disease possibilities or probabilities have changed since this evidence was gathered?

Did the Investigators Enroll the Right Patients? Was the Patient Sample Representative of Those With the Clinical Problem?

This question asks about two related issues: defining the clinical problem and ensuring a representative population.

First, how do the investigators define the clinical problem under study? The definition of the clinical problem determines the population from which the study patients should be drawn. Thus, investigators studying hematuria might include patients with microscopic and gross hematuria, with or without symptoms. On the other hand, investigators studying asymptomatic, microscopic hematuria would exclude those with symptoms or with gross hematuria.

Differing definitions of the clinical problem will yield different frequencies of underlying diseases. Including patients with gross hematuria or urinary symptoms will raise the frequency of acute infection as the underlying cause relative to those without symptoms. Assessing the validity of an article about differential diagnosis begins with a search for a clear definition of the clinical problem.

Having defined the target population by clinical problem statement, investigators next assemble a patient sample. Ideally, the sample mirrors the target population in all important ways, so that the frequency of underlying diseases in the sample approximates that of the target population. We call a patient sample that mirrors the underlying target population *representative*. The more representative the sample, the more accurate the resulting disease probabilities.

Investigators seldom use the strongest method of ensuring representativeness, which is to obtain a random sample of the entire population of patients with the clinical problem. The next strongest methods are either (1) to include all patients with the clinical problem from a defined geographic area or (2) to include a consecutive series of all patients with the clinical problem who receive care at the investigators' institution(s). To the extent that a nonconsecutive case series opens the study to the differential inclusion of patients with different underlying disorders, it compromises study validity.

You can judge the representativeness of the sample by examining the setting from which patients come. Patients with ostensibly the same clinical problem can present to different clinical settings, resulting in different services seeing different types of patients. Typically, patients in secondary or tertiary care settings have higher proportions of more serious or more uncommon diseases than patients seen in primary care settings. For instance, in a study of patients presenting with chest pain, a higher proportion of referral practice patients had coronary artery disease than the primary care practice patients, even in patients with similar clinical histories.²

To further evaluate representativeness, you can note investigators' methods of identifying patients, how carefully they avoided missing patients, and whom they included and excluded. The wider the spectrum of patients in the sample, the more representative the sample should be of the whole population and, therefore, the more valid the results will be. For example, in a study of *Clostridium difficile* colitis in 609 patients with diarrhea, the patient sample consisted of adult inpatients whose diarrheal stools were tested for cytotoxin, thereby excluding any patients whose clinicians chose not to test.³ Including only those tested is likely to raise the probability of *C. difficile* in relation to the entire population of patients with diarrhea.

Weber and Kapoor¹ defined palpitations broadly as any one of several patient complaints (eg, fast heartbeats, skipped heartbeats, etc) and included patients with new and recurring palpitations. They obtained patients from three clinical settings (emergency department, inpatient floors, and a medical clinic) in one university medical center in a middle-sized North American city. Of the 229 adult patients presenting consecutively for care of palpitations at their center during the study period, 39 refused participation; the investigators included the remaining 190 patients, including 62 from the emergency department. No important subgroups

appear to have been excluded, so these 190 patients probably represent the full spectrum of patients presenting with palpitations.

Was the Definitive Diagnostic Standard Appropriate? Was the Diagnostic Process Credible?

Articles about differential diagnosis will provide valid evidence only if the investigators arrive at a correct final diagnosis. To do so, they must develop and apply explicit criteria when assigning each patient a final diagnosis. Their criteria should include not only the findings needed to confirm each diagnosis, but also those findings useful for rejecting each diagnosis. For example, published diagnostic criteria for infective endocarditis include both criteria for verifying the infection and criteria for rejecting it.^{4,5} Investigators can then classify study patients into diagnostic groups that are mutually exclusive, with the exception of patients whose symptoms stem from more than one etiologic factor. This allows clinicians to understand which diagnoses remain possible for any undiagnosed patients.

Diagnostic criteria should include a search that is sufficiently comprehensive to ensure detection of all important causes of the clinical problem. The more comprehensive the investigation, the smaller the chance that investigators will reach invalid conclusions about disease frequency. For example, a retrospective study of stroke in 127 patients with mental status changes failed to include a comprehensive search for all causes of delirium, and 118 cases remained unexplained.⁶ Since the investigators did not describe a complete and systematic search for causes of delirium, the disease probabilities appear less credible.

The goal of developing and applying explicit, credible criteria is to ensure a reproducible diagnosis, and the ultimate test of reproducibility is a formal agreement evaluation. Your confidence in investigators will increase if, as in a study of causes of dizziness,⁷ investigators formally demonstrate the extent to which they achieved agreement in diagnosis (see Part 2C, “Diagnosis, Measuring Agreement Beyond Chance”).

While reviewing the diagnostic criteria, keep in mind that “lesion finding” is not necessarily the same thing as “illness explaining.” In other words, using explicit and credible criteria, investigators may find that patients have two or more disorders that might explain the clinical problem, causing some doubt as to which disorder is the culprit. Better studies of disease probability will include some assurance that the disorders found actually did account for the patients’ illnesses. For example, in a sequence of studies of syncope, investigators required that the symptoms occur simultaneously with an arrhythmia before that arrhythmia was judged to be the cause.⁸ In a study of chronic cough, investigators gave cause-specific therapy and used positive responses to this to strengthen the case for these disorders actually causing the chronic cough.⁹

Explicit diagnostic criteria are of little use unless they are applied consistently. This does not mean that every patient must undergo every test. Instead, for many clinical problems, the clinician takes a detailed yet focused history and performs a problem-oriented physical examination of the involved organ systems, along



with a few initial tests. Then, depending on the diagnostic clues from this information, further inquiry proceeds down one of multiple branching pathways. Ideally, investigators would evaluate all patients with the same initial workup and then follow the clues, using prespecified testing sequences. Once a definitive test result confirms a final diagnosis, then further confirmatory testing is unnecessary and unethical.

You may find it easy to decide whether patients' illnesses have been well investigated if they were evaluated prospectively using a predetermined diagnostic approach. When clinicians do not standardize their investigation, this becomes harder to judge. For example, in a study of precipitating factors in 101 patients with decompensated heart failure, although all patients underwent a history and physical examination, the lack of standardization of subsequent testing makes it difficult to judge the accuracy of the disease probabilities.¹⁰

In the Weber and Kapoor study,¹ the investigators developed a priori explicit and credible criteria for confirming each possible disorder causing palpitations and listed their criteria in an appendix, along with supporting citations. They evaluated study patients prospectively and assigned final diagnoses using two principal means: a structured interview completed by one of the investigators and the combined diagnostic evaluation (ie, history, examination, and testing) chosen by the individual physician seeing the patient at the index visit. In addition, all patients completed self-administered questionnaires designed to assist in detecting various psychiatric disorders. Electrocardiograms were obtained in a majority of patients (166 of 190), and a large number underwent other testing for cardiac disease as well. Whenever relevant, the investigators required that the palpitations occurred at the same time as the arrhythmias before they would attribute the symptoms to that arrhythmia. However, they did not report on agreement for the ultimate decisions about the diagnoses attributed to each patient.

Thus, the diagnostic workup was reasonably comprehensive—although not exhaustive—for common disease categories. Since the subsequent testing ordered by the individual physicians was not fully standardized, some inconsistency may have been introduced, although it does not appear likely to have distorted the probabilities of common disease categories, such as psychiatric or cardiac causes.

For Initially Undiagnosed Patients, Was Follow-up Sufficiently Long and Complete?

Even when investigators consistently apply explicit and comprehensive diagnostic criteria, some patients' clinical problems may remain unexplained. The higher the number of undiagnosed patients, the greater the chance of error in the estimates of disease probability. For example, in a retrospective study of various causes of dizziness in 1194 patients in an otolaryngology clinic, about 27% remained undiagnosed.¹¹ With more than a quarter of patients' illnesses unexplained, the disease probabilities for the overall sample might be inaccurate.

If the study evaluation leaves patients undiagnosed, investigators can follow these patients over time, searching for additional clues leading to eventual diagnoses and



observing the prognosis. The longer and more complete this follow-up is, the greater will be our confidence in the benign nature of the condition in patients who remain undiagnosed yet unharmed at the end of the study. How long is long enough? No single answer would correctly fit all clinical problems, but we would suggest 1 to 6 months for symptoms that are acute and self-limited and 1 to 5 years for chronically recurring or progressive symptoms.

USING THE GUIDE

Returning to our earlier discussion, Weber and Kapoor¹ identified a diagnosable etiology of palpitations in all but 31 (16.3%) of 190 patients included in their study. The investigators followed nearly all of the study patients (96%) for at least a year, during which time one additional diagnosis (symptomatic correlation with ventricular premature beats) was made in those initially undiagnosed. None of the 31 undiagnosed patients had a stroke or died.

WHAT ARE THE RESULTS?

What Were the Diagnoses and Their Probabilities?

In many studies of disease probability, the authors display the main results in a table listing the diagnoses made, along with the numbers and percentages of patients found with those diagnoses. For some symptoms, patients may have more than one underlying disease coexisting with and, presumably, contributing to the clinical problem. In these situations, authors often identify the major diagnosis for such patients and separately tabulate contributing causes. Alternatively, authors sometimes identify a separate, multiple-etiology group.

Weber and Kapoor¹ present a table that tells us that 58 patients (31%) were diagnosed with psychiatric causes and 82 (43%) had cardiac disorders, while thyrotoxicosis was found in five (2.6%), and none had pheochromocytoma. This distribution differed across clinical settings. For instance, cardiac disorders were more than twice as likely to occur in patients presenting to the emergency department, compared to patients presenting to the outpatient clinic.

How Precise Are the Estimates of Disease Probability?

Even when valid, these disease probabilities are only estimates of the true frequencies. You can examine the precision of these estimates using the confidence intervals (CIs) presented by the authors. If the authors do not provide them for you, you can calculate them yourself using the following formula:

$$95\% \text{ CI} = P + 1.96 \sqrt{[P(1 - P)]/n},$$

where P is the proportion of patients with the etiology of interest and n is the number of patients in the sample. This formula becomes inaccurate when the number of cases is 5 or fewer, and approximations are available for this situation.

For instance, consider the category of psychiatric causes of palpitations in the Weber and Kapoor¹ study. Using the above formula, we would start with $P = 0.31$, $(1 - P) = 0.69$, and $n = 190$. Working through the arithmetic, we find the CI to be 0.31 ± 0.066 . Thus, although the most likely true proportion is 31%, it may range between 24.4% and 37.6%.

Whether you will deem the confidence intervals sufficiently precise depends on where the estimated proportion and confidence intervals fall in relation to your test or treatment thresholds. If both the estimated proportion and the entire 95% confidence interval are on the same side of your threshold, then the result is precise enough to permit firm conclusions about disease probability for use in planning tests or treatments. Conversely, if the confidence limit around the estimate crosses your threshold, the result may not be precise enough for definitive conclusions about disease probability. You might still use a valid but imprecise probability result, while keeping in mind the uncertainty and what it might mean for testing or treatment.

USING THE GUIDE

Weber and Kapoor do not provide the 95% CIs for the probabilities they found. However, as we just illustrated, if you were concerned about how close the probabilities were to your thresholds, you could calculate the 95% CIs yourself.

HOW CAN I APPLY THE RESULTS TO PATIENT CARE?

Are the Study Patients Similar to Those in My Own Practice?

As mentioned previously, we suggest you ask yourself whether the setting or patients are so different from those in your practice that you should disregard the results.¹² For instance, consider whether the patients in your practice come from areas where one or more of the underlying disorders are endemic, which could make the occurrence of these disorders much more likely in your situation than was found in the study.

USING THE GUIDE

Weber and Kapoor¹ recruited the 190 patients with palpitation from those presenting to the outpatient clinics, the inpatient medical and surgical services, and the emergency department (62 of the 190) in one university medical center in a middle-sized North American city. Thus, these patients are likely to be similar to the patients seen in your hospital emergency department, and you can use the study results to help inform the pretest probabilities for the patient in the scenario.

Is It Unlikely That the Disease Possibilities or Probabilities Have Changed Since This Evidence Was Gathered?

As time passes, evidence about disease frequency can become obsolete. Old diseases can be controlled or, as in the case of smallpox, eliminated.¹³ New diseases or, at least, new epidemics of disease can arise. Such events can so alter the spectrum of possible diseases or their likelihood that previously valid and applicable studies may lose their relevance. For example, consider how dramatically the arrival of human immunodeficiency virus (HIV) transformed the list of diagnostic possibilities for such clinical problems as generalized lymphadenopathy, chronic diarrhea, and unexplained weight loss.

Similar changes can occur as the result of progress in medical science or public health. For instance, in studies of fever of unknown origin, new diagnostic technologies have substantially altered the proportions of patients who are found to have malignancy or whose fevers remain unexplained.¹⁴⁻¹⁶ Treatment advances that improve survival, such as chemotherapy for childhood leukemia, can bring about shifts in disease likelihood because the treatment might cause complications, such as secondary malignancy years after cure of the disease. Public health measures that control such diseases as cholera can alter the likelihood of occurrence of the remaining etiologies of the clinical problems that the prevented disease would have caused—in this example, acute diarrhea.

USING THE GUIDE

The palpitations study was published in 1996 and the text states that the study period was 8 months during 1991. In this instance, you know of no new developments likely to cause a change in the spectrum or probabilities of disease in patients with palpitations.



CLINICAL RESOLUTION

Let us return to the patient in your practice. Considering the possible causes of his palpitations, your leading hypothesis is that acute anxiety is the cause of your patient's palpitations. You do not believe that the diagnosis of anxiety is so certain that you can rule out other disorders (ie, the pretest probability is below your threshold for treatment without testing). After reviewing the Weber and Kapoor¹ palpitations study, you decide to include in your list of "active alternatives" some cardiac arrhythmias (as common, serious, and treatable) and hyperthyroidism (as less common but serious and treatable) and you arrange testing to exclude these disorders (ie, these alternatives are above your threshold for treatment without testing). Finally, given that none of the 190 study patients had pheochromocytoma, and since your patient has none of the other clinical features of this disorder, you place it into your "other hypotheses" category (ie, below your test threshold) and decide to delay testing for this condition.

References

1. Weber BE, Kapoor WN. Evaluation and outcomes of patients with palpitations. *Am J Med.* 1996;100:138–148.
2. Sox HC, Hickam DH, Marton KI, et al. Using the patient's history to estimate the probability of coronary artery disease: a comparison of primary care and referral practices. *Am J Med.* 1990;89:7–14.
3. Katz DA, Bates DW, Rittenberg E, et al. Predicting *Clostridium difficile* stool cytotoxin results in hospitalized patients with diarrhea. *J Gen Intern Med.* 1997;12:57–62.
4. von Reyn CF, Levy BS, Arbeit RD, Friedland G, Crumpacker CS. Infective endocarditis: an analysis based on strict case definitions. *Ann Intern Med.* 1981;94:505–517.
5. Durack DT, Lukes AS, Bright DK, and the Duke Endocarditis Service. New criteria for diagnosis of infective endocarditis: utilization of specific echocardiographic findings. *Am J Med.* 1994;96:200–209.
6. Benbadis SR, Sila CA, Cristea RL. Mental status changes and stroke. *J Gen Intern Med.* 1994;9:485–487.
7. Kroenke K, Lucas CA, Rosenberg ML, et al. Causes of persistent dizziness: a prospective study of 100 patients in ambulatory care. *Ann Intern Med.* 1992;117:898–904.
8. Kapoor WN. Evaluation and outcome of patients with syncope. *Medicine.* 1990;69:160–175.

9. Pratter MR, Bartter T, Akers S, et al. An algorithmic approach to chronic cough. *Ann Intern Med.* 1993;119:977–983.
10. Ghali JK, Kadakia S, Cooper R, Ferlinz J. Precipitating factors leading to decompensation of heart failure: traits among urban blacks. *Arch Intern Med.* 1988;148:2013–2016.
11. Katsarkas A. Dizziness in aging—a retrospective study of 1194 cases. *Otolaryngol Head Neck Surg.* 1994;110:296–301.
12. Glasziou P, Guyatt GH, Dans AL, Dans LF, Straus SE, Sackett DL. Applying the results of trials and systematic reviews to individual patients [editorial]. *ACP Journal Club.* 1998;129:A15–A16.
13. Barquet N, Domingo P. Smallpox: the triumph over the most terrible of the ministers of death. *Ann Intern Med.* 1997;127:635–642.
14. Petersdorf RG, Beeson PB. Fever of unexplained origin: report on 100 cases. *Medicine.* 1961;40:1–30.
15. Larson EB, Featherstone HJ, Petersdorf RG. Fever of undetermined origin: diagnosis and follow up of 105 cases, 1970–1980. *Medicine.* 1982;61:269–292.
16. Knockaert DC, Vanneste LJ, Vanneste SB, Bobbaers HJ. Fever of unknown origin in the 1980s: an update of the diagnostic spectrum. *Arch Intern Med.* 1992;152:51–55.