21 Differential Diagnosis

Andrew Jull and Gordon Guyatt

The following Editorial Board members also made substantive contributions to this chapter: Rien de Vos, Sharon Lock, and Ania Willman.

We gratefully acknowledge the work of Scott Richardson, Mark Wilson, Jeroen Lijmer, and Deborah Cook on the original chapter that appears in the Users' Guides to the Medical Literature, *edited by Guyatt and Rennie.*

In This Chapter

Clinical Scenario

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

You are training as an emergency department nurse practitioner. Your instructor presents the following clinical scenario to you. A 33-year-old man arrives in the emergency department 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. There is no relationship between symptoms and activity and no change in exercise tolerance. The patient works as a teacher and tends to have anxiety related to role demands. He has no other symptoms, no personal or family history of heart disease, and takes no medications. Physical examination reveals a regular heart rate of 90 beats per minute and normal eyes, thyroid gland, and lungs. His heart sounds also are normal, without click, murmur, or gallop, and his 12-lead electrocardiogram is normal, without arrhythmia or signs of pre-excitation. You are asked to list the likely causes of this man's palpitations.

You suspect that this patient's palpitations may be explained by anxiety, mediated by hyperventilation, and that they may be part of a panic attack. Cardiac arrhythmia and hyperthyroidism are also possibilities, although you wonder whether these disorders are common enough in this type of patient to warrant serious consideration. You reject pheochromocytoma (tumor of adrenal gland that causes excess production of adrenaline) as too unlikely to consider further. Thus, although you can identify several possible causes of palpitations, you want more information about the frequency of these causes as a basis for choosing a diagnostic workup. You ask the following question: In patients presenting with heart palpitations, what are the potential causes, and how do they guide the diagnostic workup?

FINDING THE EVIDENCE

Your computer networks with the health sciences library, where PubMed is online. In the left hand column, under "PubMed Services," you click onto "Clinical Queries," which has built-in research methodology filters. You then click on the category "diagnosis," with an emphasis on "sensitivity." In the "Enter Subject Search" field, you type in the following keywords: palpitations, causes, outcomes. On pressing "Go," you are presented with four citations, one of which explicitly addresses differential diagnosis in patients presenting with palpitations.[1] With a keystroke and a mouse click, you review the full text of the article by Weber and Kapoor.[1]

A *differential diagnosis* considers the active alternatives that can plausibly explain a patient's presentation. As clinicians learn and incorporate new information, they may modify the differential diagnosis. Two types of systematic investigations can inform the process of generating a differential diagnosis. One type of study addresses the presenting manifestations of a disease or condition (see Chapter 20, Clinical Manifestations of Disease). The second, and more important, type of study directly addresses the underlying causes of a presenting symptom, sign, or constellation of symptoms and signs. This chapter will focus on the second type of study. Table 21-1 summarizes the criteria for assessing a study about diagnostic possibilities.

Table 21-1 Users' Guides for an Article About Differential Diagnosis

Are the Results Valid?

- Were the right patients enrolled? 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 were the estimates of disease probability?

How Can I Apply the Results to Patient Care?

- Are the study patients similar to those in my clinical setting?
- Is it unlikely that the disease possibilities or probabilities have changed since this evidence was gathered?

ARE THE RESULTS VALID?

Were the Right Patients Enrolled? Was the Patient Sample Representative of Those With the Clinical Problem?

These questions address two related issues: defining the clinical problem and ensuring a representative population. First, how do the investigators define the clinical problem? The definition of the clinical problem determines the population from which the study patients should be drawn. Thus, investigators studying hematuria could include patients with microscopic and gross hematuria, with or without symptoms. Conversely, investigators studying asymptomatic, microscopic hematuria would exclude patients with symptoms or with gross hematuria. Differing definitions of a clinical problem will yield different frequencies of underlying diseases. Including patients with gross hematuria or urinary symptoms increases the frequency of acute infection as the underlying cause relative to patients without symptoms. Therefore, assessing the validity of a study about differential diagnosis begins with a search for a clear definition of the clinical problem.

Having identified the target population by first defining the clinical problem, investigators next assemble a patient sample. Ideally, the sample mirrors the target population in all important ways so the frequency of underlying diseases in the sample approximates that of the target population. A *representative* patient sample mirrors the underlying target population. The more representative a sample is, the more accurate the resulting disease probabilities will be.

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 to include all patients with the clinical problem from a defined geographic area or to include a consecutive series of all patients

with the clinical problem who receive care at the investigator's institution. Using a nonconsecutive case series opens the study to differential inclusion of patients with different underlying disorders or disease states and thus compromises the study's validity.

You can judge the representativeness of a sample by examining the setting from which patients are identified. Patients with ostensibly the same clinical problem can present to different clinical settings; as a result, different services see different types of patients. Typically, patients in secondary or tertiary care settings have higher proportions of more serious or uncommon diseases than do patients seen in primary care settings. For instance, in a study of patients presenting with chest pain, a higher proportion of patients from referral practices had coronary artery disease than did patients from primary care practices, even among those with similar clinical histories.[2]

To evaluate representativeness further, you can note how patients were identified, what measures were used to avoid missing patients, and which patients were included and excluded. The wider is the spectrum of patients in a sample, the more representative the sample should be of the whole population, and the more valid the results will be. For example, in a study of *Clostridium difficile* colitis in 609 patients with diarrhea, the sample comprised adult inpatients whose diarrheal stools were tested for cytotoxin, an approach that excluded patients whose clinicians chose not to perform this test.[3] Inclusion of only patients who had cytotoxin testing of stools is likely to increase the probability of *C. difficile* infection in relation to the entire population of patients with diarrhea.

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

An article 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 for assigning a final diagnosis to each patient. The criteria should include not only findings needed to confirm each diagnosis, but also findings useful for rejecting each diagnosis. For example, published diagnostic criteria for group A streptococcal pharyngitis include criteria for verifying the infection and criteria for rejecting it.[4,5] Investigators can then classify patients into diagnostic groups that are mutually exclusive, with the exception of patients whose symptoms stem from more than one etiologic factor. This approach allows clinicians to understand which diagnoses remain possible for any patients whose conditions are undiagnosed.

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 is, 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] Because 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 evaluation of agreement.

Your confidence in a study's findings will increase if investigators formally demonstrate the extent to which they achieved agreement in diagnosis. In a study by Kroenke and colleagues,[7] all study patients underwent a comprehensive assessment of their dizziness, including a history, physical examination, neuro-ophthalmologist examination, health status measurement, psychiatric assessment, and laboratory tests. All data for each patient were abstracted onto a standard form, and these data abstract forms were independently reviewed by three investigators: a general internist, a neurologist, and a neuro-ophthalmologist. Each recorded their opinion about the primary cause of a patient's dizziness. Disagreements were discussed, and a final cause was determined by consensus. The overall kappa of 0.39 approached only moderate agreement. Agreement was best for vertigo (0.52), followed by agreement for psychiatric disorders (0.42), presyncope (0.41), disequilibrium (0.31), and unknown cause (-0.06) (see Chapter 30, Measuring Agreement Beyond Chance). Raters tended to have diagnostic preferences, with the neurologist diagnosing vertigo more often than the other two raters, the neuro-ophthalmologist diagnosing presyncope more often, and the general internist diagnosing psychiatric disorders more frequently. The investigators concluded that multiple raters and a consensus process should be used in studies of diagnostic possibilities to counterbalance diagnostic biases.

When reviewing diagnostic criteria, keep in mind that "lesion finding" is not necessarily the same as "illness explaining." In other words, by using explicit and credible criteria, investigators may find that patients have two or more disorders that could explain the clinical problem, and this can cause some doubt about which disorder is the culprit. Better studies of disease probability 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 therapy to strengthen the case that these disorders actually caused 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, a clinician can take a detailed, focused history and perform 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 will proceed down one of several 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, further confirmatory testing is unnecessary and costly. It is also unethical because it may delay treatment and cause unnecessary discomfort.

It may be easier to judge whether patients' illnesses have been well investigated when investigators prospectively evaluate patients using a predetermined diagnostic approach than when they use an unstandardized approach. For example, in a study of precipitating factors in 101 patients with symptomatic heart failure, although all patients had a history and physical examination, subsequent testing was not standardized, making it difficult to judge the accuracy of the disease probabilities.[10]

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 is 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% were not assigned a diagnosis.[11] With more than one quarter of patients' illnesses unexplained, the disease probabilities for the overall sample could be inaccurate.

If the study evaluation leaves patients' conditions undiagnosed, investigators can follow up these patients and search for additional clues leading to eventual diagnoses and observe 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 whose illnesses remain undiagnosed yet who are unharmed at the end of the study. How long is long enough? No single answer applies to all clinical problems, but we suggest 1 to 6 months for acute and self-limited symptoms and 1 to 5 years for chronically recurring or progressive symptoms. For example, in a study of nonacute abdominal complaints in family practice, 933 patients were followed-up for at least 1 year (mean, 18 months) before a final diagnosis was assigned.[12]

Using the Guide

Weber and Kapoor[1] defined palpitations broadly as any one of several patient complaints (e.g., fast heartbeat, skipped heartbeats) and included patients with new and recurring palpitations. The investigators identified patients from three clinical settings (an emergency department, inpatient floors, and a medical clinic) in a university medical center in a mid-sized North American city. Of 229 adult patients presenting consecutively for care of palpitations, 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 the sample likely represents the full spectrum of patients presenting with palpitations.

The investigators developed a priori, explicit, and credible criteria for confirming each possible disorder that caused palpitations and listed their criteria in an appendix, along with supporting citations. They evaluated patients prospectively and assigned final diagnoses based on structured interviews completed by one of the investigators and the combined diagnostic evaluation (i.e., history, examination, and testing) chosen by the individual physician who saw 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 most patients (166 of 190), and many patients had other testing for cardiac disease as well. When 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, the investigators did not report on agreement for the ultimate decisions about the diagnoses attributed to each patient.

Using the Guide—cont'd

Thus, the diagnostic workup was reasonably comprehensive—although not exhaustive—for common disease categories. Because subsequent tests ordered by individual physicians were 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.

Weber and Kapoor[1] identified a diagnosable cause of palpitations in all but 31 (16.3%) of 190 patients. The investigators followed up 96% of patients for at least 1 year, during which time an additional diagnosis (symptomatic correlation with ventricular premature beats) was made in patients with initially undiagnosed conditions. None of the 31 patients with undiagnosed conditions 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 and the numbers and percentages of patients 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.

How Precise Were the Estimates of Disease Probability?

Even when valid, 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 report them, you can calculate them yourself by using the following formula:

$$95\%\ \text{CI} = P \pm 1.96 \times \sqrt{(P\{1-P\}/n)}$$

where P is the proportion of patients with the cause of interest and n is the number of patients in the sample. This formula becomes inaccurate when the number of cases is five or fewer, and approximations are available for such situations.[13,14]

For instance, consider the study by Weber and Kapoor in which 58 patients (31%) were diagnosed with psychiatric causes of palpitations.[1] Using the 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 ± 0.07. Thus, although the most likely true proportion is 31%, it may range from 24% (31% – 7%) to 38% (31% + 7%).

Whether you deem the CIs sufficiently precise depends on where the estimated proportion and CIs fall in relation to your test or treatment thresholds (see Chapter 6, Diagnosis). If both the estimated proportion and the entire 95% CI are on the same side of your threshold, the result will be precise enough to permit firm conclusions about disease probability for use in planning tests or treatments. Conversely, if the confidence interval around the estimate crosses your threshold, the result may not be precise enough for definitive conclusions about disease probability. You may still use a valid but imprecise probability result, but keep in mind the uncertainty and its implications for testing or treatment.

Using the Guide

In the study by Weber and Kapoor,[1] 58 patients (31%) were diagnosed with psychiatric causes, 82 (43%) had cardiac disorders, five (2.6%) had thyrotoxicosis, 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 than in patients presenting to the outpatient clinic.

The investigators did not provide the 95% CIs for the probabilities they found. However, as illustrated, if you are concerned about how close the probabilities are to your thresholds, you can calculate the 95% CIs.

HOW CAN I APPLY THE RESULTS TO PATIENT CARE?

Are the Study Patients Similar to Those in My Clinical Setting?

As mentioned previously, we suggest that you ask yourself whether the setting and patients are so different from your own that you should disregard the results.[15] For instance, consider whether the patients in your clinical setting come from areas where one or more of the underlying disorders are endemic, a factor that could result in higher frequencies of disorders in your setting than were found in the study.

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

As time passes, evidence about disease frequencies can become obsolete. Old diseases can be controlled or even eradicated. New diseases and 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, the emergence of human immunodeficiency virus dramatically transformed the list of diagnostic possibilities for such clinical problems as generalized lymphadenopathy, chronic diarrhea, and unexplained weight loss. More recently, severe acute respiratory syndrome has been added to the list of potential diagnoses for a traveler arriving with a fever of more than 38° C and a dry cough.

Similar changes can occur as the result of progress in medical science or public health. For instance, in studies of fever of unknown origin, newer diagnostic technologies have

substantially altered the proportions of patients who are found to have malignancy or whose fevers remain unexplained.[16-18] Treatment advances that improve survival, such as chemotherapy for childhood leukemia, can bring about shifts in disease likelihood because the treatment may cause complications, such as secondary malignant disease, years after the initial disease is cured. Public health measures that control such diseases as cholera can alter the likelihood of the remaining causes of the clinical problems that the disease would have caused—in this example, acute diarrhea.

Using the Guide

Weber and Kapoor[1] recruited 190 patients with palpitations from those presenting to outpatient clinics, inpatient medical and surgical services, and an emergency department (62 patients) in a university medical center in a mid-sized North American city. Thus, the study 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 presenting patient.

Considering the results reported in the study, you know of no new developments likely to cause a change in the spectrum or probabilities of disease in patients with palpitations.

Clinical Resolution

Considering the possible causes of the patient's palpitations, your leading hypothesis is that acute anxiety is the cause. However, you do not believe that the diagnosis of anxiety is so certain that you can rule out other disorders (i.e., the pretest probability is below your threshold for treatment without testing). See Chapter 6, Diagnosis, for a discussion of pretest probabilities. After reviewing the study on palpitations by Weber and Kapoor,[1] you decide to include in your list of active alternatives cardiac arrhythmias (common, serious, and treatable) and hyperthyroidism (less common but serious and treatable), and you suggest testing to exclude these disorders (i.e., these alternatives are above your threshold for treatment without testing). Finally, given that none of the 190 study patients had pheochromocytoma, and because the presenting patient has none of the other clinical features of this disorder, you place it into your "other hypotheses" category (i.e., below your test threshold), which would 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. McIsaac WJ, White D, Tannenbaum D, Low DE. A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *CMAJ*. 1998;158:75-83.

5. McIsaac WJ, Goel V, To T, Low DE. The validity of a sore throat score in family practice. *CMAJ*. 2000;163:811-815.
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 (Baltimore)*. 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. Muris JW, Starmans R, Fijten GH, Crebolder HFJM, Schouten HJA, Knotterus JA. Non-acute abdominal complaints in general practice: diagnostic value of signs and symptoms. *Br J Gen Pract*. 1995;45:313-316.
13. Hanley JA, Lippman-Hand A. If nothing goes wrong, is everything all right? Interpreting zero numerators. *JAMA*. 1983;249:1743–1745.
14. Newman TB. If almost nothing goes wrong, is almost everything all right? Interpreting small numerators. *JAMA*. 1995;274:1013.
15. 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 J Club*. 1998;129:A15-A16.
16. Petersdorf RG, Beeson PB. Fever of unexplained origin: report on 100 cases. *Medicine (Baltimore)*. 1961;40:1–30.
17. Larson EB, Featherstone HJ, Petersdorf RG. Fever of undetermined origin: diagnosis and follow up of 105 cases, 1970-1980. *Medicine (Baltimore)*. 1982;61:269-292.
18. 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.