APPLYING EVIDENCE TO THE DIAGNOSTIC PROCESS FOR GIANT CELL ARTERITIS: HOW TO USE AN ARTICLE ON THE CLINICAL MANIFESTATIONS OF DISEASE

Case and Discussion: *Francis J. McBee Orzulak, MD, and Rajesh S. Mangrulkar, MD*

When patients seek care from physicians, they have 2 main concerns: What is wrong with me? What should be done about it? To address the first question we engage in the complex process of diagnosis, the outcome of which is essential for effective treatment (ie, addressing the second question). Four distinct forms of medical knowledge are recognized as key to the diagnostic process: knowledge of the clinical manifestations of disease, knowledge of diagnostic tests and how to interpret them, knowledge of which diseases cause clinical problems and their prevalence, and experience from taking care of actual patients [1]. While all are important to diagnosing conditions, knowledge of clinical manifestations is the only type that can be applied to all 3 steps of the diagnostic process: selecting a differential diagnosis and assigning pretest probability, selecting and interpreting diagnostic tests and assigning posttest probability, and verifying the final diagnosis [1–3]. Indeed, we spend much of our time in training learning how diseases present and making conclusions based on findings elicited directly from patients during the history and physical examination.

Making a clinical diagnosis requires knowledge about the clinical manifestations that characterize a disease and differentiate it from other diagnoses. Without complete and accurate knowledge of the clinical manifestations of disease, we risk missing clues to the correct diagnosis that may be found in the initial clinical evaluation of a patient. We also may be erroneously drawn to a faulty differential diagnosis, which could lead us to make an inaccurate diagnosis or to pursue inappropriate interventions. This article provides guidance on how to search for, interpret, and apply information contained in articles about the clinical manifestations of disease, specifically to augment the diagnostic process.

Francis J. McBee Orzulak, MD, Departments of Internal Medicine and Pediatrics and Communicable Diseases, University of Michigan Health System; and Rajesh S. Mangrulkar, MD, Department of Internal Medicine, University of Michigan Health System, Ann Arbor, MI.

Assessing the Clinical Problem and Formulating a Question

You are an intern in internal medicine seeing patients at continuity clinic. Your next patient is Mrs. Brown, a 73-year-old woman with a history of hypertension and hyperlipidemia. You last saw her 3 months ago and recall that her conditions are well controlled on hydrochlorothiazide and atorvastatin.

Mrs. Brown's chief complaint is worsening fatigue over the last month. While interviewing her, you learn that she has also been bothered by stiffness in her shoulders and hips, which is worse in the morning. On review of systems, she denies any headache or vision changes but does report jaw claudication for about 2 weeks. Physical examination reveals a blood pressure of 124/78 mm Hg and decreased range of motion in the neck, shoulders, and hips. Fundoscopic examination is normal. There are no visual field deficits. Both temporal arteries are nontender. The remainder of the examination is unremarkable.

You discuss the case with Dr. Manteau, the clinic attending, and tell her that you are concerned about polymyalgia rheumatica (PMR). She agrees that this is a consideration but also raises the possibility of giant cell arteritis (GCA) because of the patient's jaw claudication. You recall that headache is a common clinical feature of GCA and wonder if the diagnosis can be excluded in the absence of headache. Dr. Manteau reminds you that GCA is a disease with significant morbidity and should not be dismissed lightly. If GCA is seriously considered, she adds, you should initiate therapy prior to confirmatory diagnostic testing. You agree but note that definitive testing

for GCA, a temporal artery biopsy, is an invasive test that should not be undertaken without good reason. On further discussion, you and Dr. Manteau decide that you should investigate further the clinical manifestations of GCA while Mrs. Brown is having an erythrocyte sedimentation rate (ESR) checked in the lab next door.

To begin, it is crucial to recognize that this situation requires information about how the diseases under consideration manifest. Note that the differential diagnosis in this case seems short, including just 2 possibilities. The attending is entertaining GCA because of the presence of jaw claudication. It is likely that she has seen patients with this presenting complaint and that a few of them have been diagnosed with GCA. So, the question in her mind is: How often is jaw claudication present in patients with GCA? The resident is more concerned about PMR and wonders whether the absence of a finding, specifically headache, is enough to eliminate GCA from consideration. This likely stems from learning that temporal headache is a common finding in GCA. The exact prevalence of this finding, however, is uncertain and merits further investigation.

While both of these diagnoses are reasonable to entertain, several other crucial considerations must be weighed in the diagnostic decision making. First, the cost of missing GCA must be considered. As Dr. Manteau mentions, failure to treat this condition when it exists could be catastrophic, causing blindness and even death. Second, we need to ascertain the risk of the invasive temporal biopsy to definitively document GCA. Not mentioned is a third concern—falsely treating GCA with high-dose corticosteroids should PMR be the correct diagnosis. Because of the many issues at play—a common occurrence in the complex world of medical decision making—the pursuit of the question about the clinical manifestations of GCA may not lead to an obvious path in managing this patient.

Nevertheless, having identified a knowledge gap (ie, frequencies of jaw claudication and headache in GCA), we are compelled to seek guidance from the literature for this case. When searching the literature, we should articulate our knowledge gap in a focused question. Carefully defining the various PICO elements (P = population, I = intervention, C = comparison, O = outcome) will assist in the search process. In this case, the population could be defined as patients with symptoms of PMR. It is important to note that an intervention can be a treatment, risk factor, prognostic factor, or diagnostic test and that clinical manifestations are equivalent to diagnostic tests. Thus, in this case, the intervention could be defined as the absence of headache or the presence of jaw claudication. In this case, there is no comparison. Finally, the outcome of interest is the presence of GCA. Using these PICO elements, our question regarding Mrs. Brown would be:

In **patients with symptoms of PMR**, does the **absence of headache** or **presence of jaw claudication** significantly alter the **diagnostic probability of GCA**?

Where to Find Information About Manifestations of Disease

Many types of information resources are available that describe clinical manifestations of disease, including textbooks, case reports, review articles (including online resources such as *UpToDate*), and original studies on the frequencies of clinical manifestations. Although all of these types of resources can provide information regarding frequencies of clinical manifestations, which can then help guide the diagnostic process, most suffer from bias due to a lack of a systematic approach. Textbooks and review articles are widely considered to reflect the bias of the author and vary in the degree to which they use a systematic process in reviewing and analyzing the literature. Case reports often represent unusual manifestations of common or uncommon diseases, important in broadening our perspective and knowledge about a clinical condition but unhelpful in providing guidance in our daily practice. Unlike these resources, studies of the frequencies of clinical manifestations have the potential to provide guidance as to whether a particular sign or symptom is more or less frequently present than originally thought. In addition, a valid study of this type can give information about the sensitivity of a certain manifestation in the diagnosis of the disease in question.

Acquiring the Best Evidence

To search the literature on clinical manifestations, you access PubMed through the clinic's library interface and initiate a search. You begin by searching for the principal diagnosis in question ("giant cell arteritis"), followed by the word "diagnosis." The interface returns the following: "Did you mean 'giant cell arteritis/diagnosis'?" You click on this term because it allows you to retrieve articles that reflect the diagnostic findings of the disease, including physical examination, differential diagnosis, and prognosis. This strategy yields 2373 articles at first, a daunting and unhelpful list.

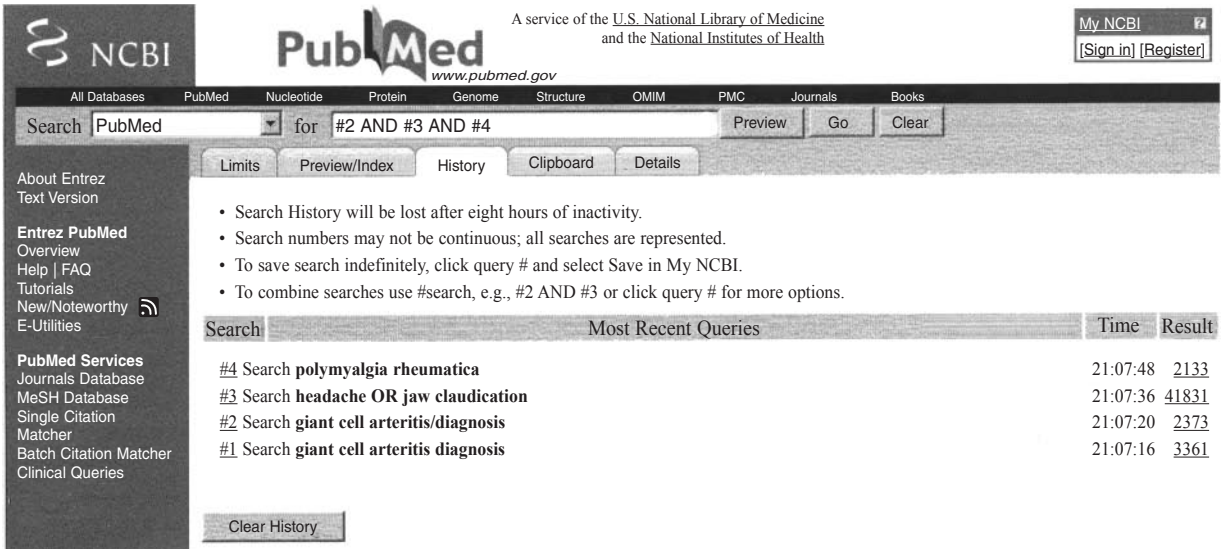

Figure 1. Using the "History" tab on PubMed allows you to see your previous searches. You can combine these searches using Boolean operators (OR, AND, and NOT).

You attempt to narrow the yield by searching for the clinical manifestations in question using "headache OR jaw claudication," but this yields 41,831 articles. Adding the term "polymyalgia rheumatica" to indicate the patient population of interest yields 2133 articles. Using the "History" tab, you combine these 3 searches, entering "#2 AND #3 AND #4" in the query box (**Figure 1**), which yields 85 articles. Finally, using the "Limits" tab, you further narrow your search to "English" and "All Adult" articles, which yields 51 articles.

You recall that your attending warned you that, given your quest for information about clinical manifestations, the yield should be sensitive enough so that you are not omitting potentially important studies, so you decide to stop here and scan the results. After eliminating isolated case studies, review articles, and basic science articles, you find 2 articles that seem to report on the frequencies of the clinical manifestations of GCA and provide a systematic focus: one by Gonzalez-Gay et al [4] and a second by Kobayashi et al [5]. You show Dr. Manteau the articles and she agrees that they look promising. She refers you to the *Users' Guides* [1,6s] to assist you in the interpretation of these articles.

Studies of the Frequencies of Clinical Manifestations

Studies of this type can guide us in several ways as we ponder the differential diagnosis for a cluster of clinical manifestations. However, these studies also have limitations that are important to understand.

Lessons. One important lesson that we frequently learn is that the "classic" clinical manifestations could be less frequent than we would expect based on what we learn from textbooks [1]. For example, we are taught that children with intussusception will present with the triad of intermittent abdominal pain, bloody stool, and vomiting [7]. However, when Sparnon et al [8] systematically studied and reported on the clinical manifestations of this disease, they found that this triad of symptoms and signs was present in only 20% of patients.

The inverse can hold true as well—a clinical manifestation can prove to be more common than previously believed [1]. Continuing with intussusception as an example, lethargy is usually taught to be an infrequent and late manifestation [9], but Justice et al [10] found it to be a presenting symptom in 87% of cases in a recent study. The discordant information must be questioned—especially in relation to the validity criteria that will be described later—but it raises the consideration that true lethargy may be more common than one might expect.

Finally, a sign or symptom may be found to be so common in a study of this type that the absence of the finding essentially may rule out the disease [1]. In a well-done study, when the presence of a clinical manifestation approaches 100%, we can be comfortable removing the disease from the differential diagnosis when the manifestation is absent. Returning to our intussusception example, we can see that abdominal pain is present in 100% of the cases in the study by Justice et al [10]. Therefore,

		Diagnosis		Totals
		Present	Absent	
Test (clinical manifestation)	Present	A = 90	B = ???	A + B = ???
	Absent	C = 10	D = ???	C + D = ???
Totals		A + C = 100	B + D = ???	A + B + C + D = ???

Figure 2. Using a 2 × 2 table to calculate sensitivities. Studies of the frequencies of manifestations of disease only include patients with the disease. The sensitivity of a manifestation can be calculated using the equation: sensitivity = A ÷ (A + C) = 90%. However, specificity (specificity = D ÷ [B + D]) cannot be calculated because patients who do not have the disease are not included in these types of studies.

we could conclude that if our patient does not have abdominal pain, we can feel comfortable removing intussusception from the differential diagnosis or not even entertaining it as a potential diagnosis in the first place.

Limitations. The first major limitation of studies of clinical manifestations relates to the concepts of test characteristics (ie, sensitivity and specificity). Although we typically associate sensitivity and specificity calculations with interpreting diagnostic tests, these concepts also relate to articles about the signs and symptoms of disease. Eliciting a complaint and the description of that complaint, as well as the detection of an abnormality (or normality) on physical examination, can be thought of as a series of diagnostic tests that are similar to those we order. All require interaction with the patient and interpretation, all have a range of "normal" and "abnormal," and all are combined to assist us in the diagnostic process. The degree of standardization may vary, but the underlying concepts apply equally.

As depicted in **Figure 2**, the sensitivity of a given clinical manifestation for a disease is simply the frequency of the clinical manifestation reported in the study. Studies of this type include only patients who are known to have the disease in question, and patients with other disorders are excluded. This may cause some bias in calculating sensitivity because, as will be discussed, it is important that the study adequately cover the spectrum of patients with this disease, but the sensitivity should be reasonably accurate. However, we cannot use data from these studies to derive specificities or likelihood ratios since these calculations require knowledge about patients without the disease. These patients, by definition, are excluded from studies of the frequencies of clinical manifestations.

The second major limitation of these studies involves the closely related concepts of validity and reliability. Validity reflects the degree to which the test result accurately measures the condition being examined. Reliability reflects how well a test reproduces its results on repeated trials. To illustrate these concepts, we can use the analogy of an archer's ability to hit the bull's-eye on a target (**Figure 3**). If each arrow fired by the archer reflects a result from the same test, a valid test would reflect the arrow's proximity to the bull's-eye. Reliability would reflect the consistency of the archer's attempts, indicated by the scatter of the results that the arrows leave on the target. Figure 3A shows the pattern of a valuable test, one with high validity and high reliability. Figures 3B and 3C show different examples of deficient tests. In Figure 3B, there is poor validity and good reliability. In Figure 3C, there is poor reliability and poor validity. Note that there is no scenario portrayed when reliability is poor but validity is high. This reflects the reality that reliability is a necessary precondition to a valid test. How would we know if a test is valid if the scatter of the test results is so wide as to impede consistent analysis of these results?

Although studies of the clinical manifestations of disease can provide good information on the frequencies of these findings, they often offer no information on how reliably clinicians gather this information [1]. We can illustrate this by looking at aspects of the physical examination as an example. Many studies use expert examiners as a method to report these observations, but the studies do not necessarily know the frequency with which the examiners miss specific examination findings. If the examiner did not reliably detect the findings, the study data would be fundamentally flawed. Additionally, as readers, we do not know the rate at which we, ourselves, miss specific examination findings. If we are not as skilled as the practitioners used in the study, we may not be able to reliably detect these findings and therefore may not be able to effectively use the information gleaned from the study in our clinical practice. To summarize, these types of studies can contribute to evidence of validity of a finding in the diagnosis of the disease, but they contribute little or no information about the reliability of these findings.

Appraising Evidence for Validity and Importance

With the *Users' Guides* criteria in hand [1,6], you embark on the analysis of the 2 studies that you retrieved. As you review the appraisal criteria, you note

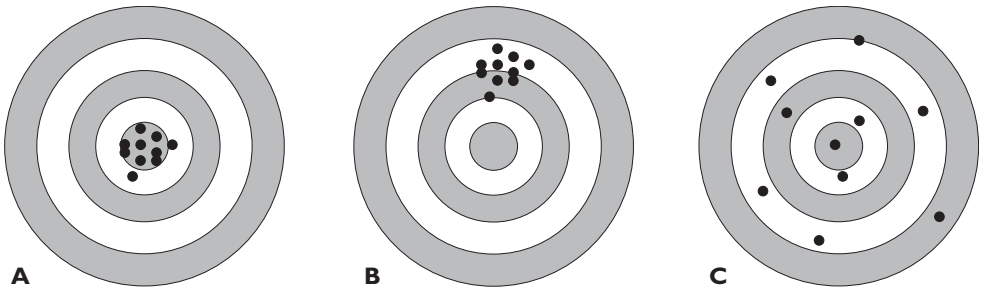

Figure 3. To understand the concepts of validity and reliability, it is useful to think of an archer's ability to hit a bull's-eye, with each arrow fired reflecting a result from the same test. Scenario A depicts a test with good validity and reliability—all of the arrows are closely clustered and center on the bull's-eye. Scenario B shows a test with poor validity and good reliability—the arrows are still closely clustered but are far away from the center of the target. Finally, scenario C shows a test with poor validity and poor reliability—the arrows are dispersed across the target, with most nowhere close to the center.

that there are multiple questions that need to be answered (**Table 1**). You recognize that many of the questions are similar to those asked about studies on diagnostic tests, such as whether a representative sample of patients was used in the study, whether an appropriate gold standard test was used to verify the diagnosis in the study, and whether there were any methodologic problems that led to bias (verification versus incorporation bias). You decide to use this as a paradigm for understanding the process to be described. Recognizing the first step, you realize the need to assess the validity of the articles before moving on to reviewing the results.

Table 1. Users' Guides Criteria for Appraising and Applying an Article About the Clinical Manifestations of Disease

Are the results valid?
Does the patient sample represent the full spectrum of those with the disease?
Was the definitive diagnostic standard appropriate?
Was the diagnosis verified using credible criteria that were independent of the clinical manifestations under study?
Were the clinical manifestations sought thoroughly, carefully, and consistently?
Were the clinical manifestations classified by how and when they occurred?
What are the results?
How frequently did the manifestations occur?
How precise are the estimates?
When and how did manifestations occur in the course of disease?
How can I apply the results to patient care?
Are the patients in the study similar to mine?
Is it unlikely that the disease manifestations have changed?

Adapted with permission from Guyatt G, Rennie D. Users' guides to the medical literature: a manual for evidence-based clinical practice. Chicago: AMA Press; 2002:449.

Are the Results Valid?

- **Does the patient sample represent the full spectrum of those with the disease?**

The issue of sampling is relevant to the analysis of many types of studies and is especially applicable here. In a perfect world, the patients sampled for a study would identically match the population of patients with the disease. If this were true, the characteristics of the patients in the study, in this case the clinical manifestations, would mimic the characteristics of all patients with the disease. A skewed population—for example, one that represents only patients with mild forms of the disease—might lead the study to miss certain findings or to misrepresent the frequency of findings in the disease.

The adequacy of sampling in a study can be assessed by looking at 3 characteristics of the study population: setting, demographics, and descriptions of disease [1,6].

1. What is the clinical setting from which the patients were taken? For instance, is the setting a referral or community center? If the setting is a referral center, the study may end up with a population that is biased toward atypical or complicated cases. Likewise, if the setting is a community center, the study may be biased toward more common manifestations.
2. What methods were used to identify or exclude patients in the study population? To assess this, we should ask whether all important demographic groups (eg, gender, age, ethnicity) are represented in the study, especially those that have influence

on the presence of the disease. We can get a sense of this issue by comparing the demographic data given by the authors with the epidemiologic data available for the disease and noting if there are any significant differences. For example, a study on the clinical manifestations of prostate cancer would be significantly deficient if the study underrepresented African Americans.

3. What descriptions of disease were included in the study? For instance, were patients with mild, moderate, and severe disease included? Additionally, if multiple patterns of the disease being studied are known, were all subtypes included?

By evaluating these 3 characteristics, we can decide whether the patients included in the study were sufficiently representative of those with the disease in order to draw valid conclusions on the clinical manifestations of the disease in question.

Turning our attention to the study by Gonzalez-Gay et al [4], we find that the patients were seen at a single hospital in Spain. The hospital is described as a referral center that also serves a mixed rural and urban population. Because it serves a wide variety of patients, the hospital likely sees both typical and atypical cases of GCA. The study population was primarily white Spanish, and there was a slight predominance of women and elderly patients. The study participants are described as having had disease of varying severity; 18 (7.5%) patients with subclinical disease and 131 (54.6%) patients with severe ischemic disease were included in the study. Of note, 96 (40%) patients had symptoms of PMR. Thus, we can conclude that, although the patient population was not ethnically varied, the study seems to have captured the entire spectrum of disease from typical to atypical and from mild to severe. Glancing at the study by Kobayashi et al [5], we find that these investigators used a similarly broad sample of patients with a lack of ethnic variability; the study included all patients diagnosed with GCA in Japan during 1997.

- **Was the definitive diagnostic standard appropriate?**

Another way to ask this question is: How sure are we that the patients actually have the disease in question? This is particularly important in studies of clinical manifestations of disease because including patients in the study with a tentative diagnosis or a different disease would inject significant bias into the study [1,6]. To minimize the bias that a tentative diagnosis would incur, it is strongly recommended that investigators use diagnostic criteria that are widely accepted in the field (eg, reference, gold, or criterion-based standard tests). A careful evaluation of the diagnostic standard used in the study, with an assessment of its reliability and validity, helps address this crucial step.

Looking at the study by Gonzalez-Gay et al [4], we see that the diagnosis of GCA was confirmed by unilateral temporal artery biopsy. This is reassuring, given that temporal artery biopsy is widely accepted as the best test for diagnosing GCA. On the other hand, unilateral biopsy could miss a significant number of cases of GCA, and thus bilateral biopsy is recommended in some cases [11,12]. It is reasonable to conclude that, while diagnosis by unilateral temporal artery biopsy may result in some patients who actually have GCA being excluded from the study, it does not result in patients being included who do not have the disease. It may, however, result in inaccurate frequencies of the clinical manifestations described because of possible differing rates of these signs and symptoms in patients who were excluded from the study due to a negative temporal artery biopsy.

- **Was the diagnosis verified using credible criteria that were independent of the clinical manifestations under study?**

Many diseases do not have a widely accepted gold standard test. In these cases, the investigators may rely on diagnostic criteria based, at least in part, on the clinical manifestations of the disease. This raises the problem of *incorporation bias* when the diagnosis in the study is made using clinical manifestations that are being studied [1,6]. As an example, the study by Kobayashi et al [5] contains incorporation bias because the investigators used the American College of Rheumatology (ACR) criteria for GCA as the diagnostic criteria for the study. The ACR criteria require the presence of 3 of the following 5 features: age 50 years or older at time of onset, localized headache of new onset, tenderness or decreased pulse of the temporal artery, ESR greater than 50 mm/hr, and biopsy revealing necrotizing arteritis [13]. Because of this bias, interpreting information on the frequency of headache in GCA becomes risky in this study.

- **Were the clinical manifestations sought thoroughly, carefully, and consistently?**

This question has 3 components [1,6]. First, how thoroughly were all patients evaluated and assessed at study enrollment as well as during the study? The more comprehensive the investigators were in their evaluation, the less likely it is that they missed a sign or

symptom and skewed the estimate of the prevalence of this finding. Second, were the clinical manifestations sought carefully to eliminate bias? If the authors include enough information, we should try to assess for bias in the data collection (eg, leading questions or unblinded examiners). Finally, how consistently were the manifestations assessed? Inconsistent evaluations can yield markedly different frequencies of clinical manifestations, even in the same patient.

The daunting nature of these criteria essentially forces us toward a fundamental observation—prospective studies of clinical manifestations, which employ an explicit standardized approach, are far more reliable and valid than retrospective studies of this type, in which control over the assessment is often relinquished. However, it is important to realize that most studies of clinical manifestations are retrospective, and we may have no choice but to use them.

Returning again to the study by Gonzalez-Gay et al [4], we see that these investigators used a retrospective chart review to determine which clinical manifestations were present, but they do not describe the diagnostic evaluation in any detail. The charts explaining the results are very detailed and reflect a careful approach to the investigators' assessment, but we cannot draw any conclusions about the consistency of the workup. Thus, this study has many of the shortcomings we would expect of a retrospective design, but this does not negate the results. The study by Kobayashi et al [5] also was retrospective. The investigators administered a survey to all physicians who treated the patients in the patient sample at various centers across Japan. The authors do not describe the diagnostic evaluation, leaving us to wonder about the inconsistency of having many physicians at many institutions perform the evaluations and, therefore, to question the validity of the frequencies reported, especially for uncommon manifestations.

- **Were the clinical manifestations classified by how and when they occurred?**

Clinical manifestations can occur at varying times during a disease process. Once present, these features can persist for the remainder of the disease, progress into another manifestation, or resolve entirely. For this reason, it is important for the investigators to describe when and how the manifestation was present during the course of the disease [1,6]. If the investigators fail to do this, they can miss early or transient findings and thereby underrepresent the prevalence of these findings in their study, making the data less useful in clinical practice. Sometimes, investigators are able to describe the qualitative aspects of these findings, which is very helpful in the clinical environment but not mandatory for the study's utility. For example, when considering chest pain, if the pain is described as a "pressure," we are led more toward myocardial infarction, but if it is described as a "tearing" sensation, we think more about aortic dissection.

In the study by Gonzalez-Gay et al [4], all clinical manifestations are categorized as presenting symptoms if they occurred any time from the start of GCA symptoms until 4 weeks after treatment was initiated. Knowing this, we can feel more confident that these investigators captured all the early and transient manifestations of GCA. The article by Kobayashi et al [5] provides even better information on the timing of clinical manifestations, describing the frequency that occurred both at the onset of the disease and at any time during the illness course (average follow-up was 39.9–54.6 months). Neither study includes any data on the qualitative aspects of the manifestations.

Upon completing your validity appraisal, you conclude that both studies have limitations and strengths. Both did an adequate job including representative samples of patients, although neither included an ethnically diverse population. However, you have read that GCA seems to affect all populations rather equally with the exception of those of Scandinavian descent, who are affected at a higher rate [13]. As such, you feel comfortable with the populations sampled in these studies, although you recognize that your patient population has little representation from Asian groups.

Both studies were retrospective, but the study by Kobayashi et al [5] likely suffers from more inconsistency because the patients were cared for in a variety of institutions across Japan, as compared with the other study [4], which included patients seen at 1 institution. However, you noted 2 important differences. The diagnostic standard used by Kobayashi et al [5] suffers from incorporation bias due to its use of the ACR criteria for GCA. Gonzalez-Gay et al [4] used an independent diagnostic standard, giving this study a clear validity advantage. On the other hand, the study by Kobayashi et al [5] gives more information on the timing of the clinical manifestations, making it somewhat more useful in the clinical setting but not more valid.

Considering all factors, you decide to extract and apply the results from the study by Gonzalez-Gay et al [4]. As you move on to examine the study's results, you return to the *Users' Guides* article [1,6] for guidance.

What Are the Results?

- **How frequently did the manifestations occur?**

The results of studies of frequencies of clinical manifestations are usually reported in tabular form and are generally easy to interpret. For example, the data from the study by Gonzalez-Gay et al [4] clearly show that of the 240 patients with GCA, 203 (84.6%) presented with headache (**Table 2**). Furthermore, of the 37 patients who presented without headache, only 27% had jaw claudication and almost 60% had symptoms of PMR. Similarly, data for the 96 patients with GCA and symptoms of PMR show that 77.1% of patients in this group had headache and 41.7% had jaw claudication (**Table 3**). Although the study is valid, we cannot use these frequencies as precise measurements, but rather must use them as estimates.

- **How precise are the estimates?**

To understand the precision of the estimates of frequency, the *Users' Guides* advocates that we look at the 95% confidence interval, which provides an estimate of the uncertainty in the data [1,6]. Although more precise definitions have been proposed, a simple and more understandable (albeit less precise) definition of the 95% confidence interval is the range of values within which 95% of the data lies [15].

Typically, the 95% confidence interval is provided for the reader, although this is not the case for the study by Gonzalez-Gay et al [4]. It is possible to calculate the range based on the sample size and the proportion of patients with the finding of interest, using the following equation:

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

where P is the proportion of the patients with the finding and *n* is the number of patients in the sample [1,6]. However, given that the frequency of headache in patients with GCA and symptoms of PMR is high (77.1%) but not approaching total penetrance, knowing the range will likely not help us use the results any differently.

- **When and how did manifestations occur in the course of disease?**

The *Users' Guides* advises that we look for specific descriptions of the clinical manifestations that would allow us to classify them as *presenting* (prompting the patient to seek care), *concurrent* (detected while seeking care), or *eventual* (occurring later in the course of the illness) [1,6]. The study by Gonzalez-Gay et al [4] classified all clinical manifestations as presenting, which limits how we can classify the symptoms based on timing, especially since time to diagnosis and treatment varied greatly among the patients. A more complete description of the timing of the clinical manifestations of GCA might reveal a pattern of clinical manifestations that changes as the disease progresses, but we recognized this flaw when selecting this study over the one by Kobayashi et al [5].

> In thinking about your patient, you recall that Mrs. Brown came to see you with the complaint of worsening fatigue for 1 month—her presenting clinical manifestation. She also had 2 concurrent clinical manifestations that were elicited during the history and physical examination—stiffness in her shoulders and hips that was worse in the morning and jaw claudication. Although the clinical manifestations described by Gonzalez-Gay et al [4] are a hodgepodge of presenting, concurrent, and early eventual, you still feel comfortable in identifying Mrs. Brown as early in the stage of her disease process.

Applying Evidence to Patient Care

The final step in the evidence-based medicine process is frequently the most difficult, because it requires the most judgment on the part of the reader-clinician. However, as emphasized in the *Users' Guides*, this step is also the most crucial, returning the reader to the clinical scenario and patient at hand [1,6].

- **Are the patients in the study similar to mine?**

The more the patients in the study population match the patients in our clinical practice, the more confidently we can apply the results. In the appraisal step, we considered many of these key factors while trying to address whether the patient sample represented the full spectrum of those with the disease. Most important is to ask whether differences in certain factors (eg, cultural group, geographic location, clinical risk factors) might lead to significant differences in the clinical manifestations of the disease. This question should have been asked early in the analytic process, prior to reviewing the data, because if the answer is "yes," perhaps we should not have read the study at all.

The population in the study by Gonzalez-Gay et al [4] is predominantly white, which is characteristic of the region of Spain in which the study was performed. This causes some problems with the generalizability of the data; it is safe to assume that the majority of clinicians do not practice in locations where the population is predominantly of white

Table 2. Characteristics of Patients with Biopsy-Proven GCA Who Presented with/without the Classic Pattern of Headache at Time of Disease Diagnosis

Variable	Pattern: With Headache (*n* = 203; 84.6%) No. (%)	Pattern: Without Headache (*n* = 37; 15.4%) No. (%)	*P*
Age at diagnosis (mean ± SD), yr	74.7 ± 6.7	75.2 ± 6.9	0.656
Women	108 (53.2)	22 (59.5)	0.482
Delay to diagnosis (mean ± SD), wk*	9.2 ± 9.9	16.6 ± 15	< 0.001
Scalp tenderness	79 (38.9)	2 (5.4)	< 0.001
Constitutional syndrome†	123 (60.6)	23 (62.2)	0.857
Abnormal temporal arteries‡	162 (79.8)	13 (35.1)	< 0.001
Jaw claudication	88 (43.3)	10 (27)	0.063
Dysphagia	12 (5.9)	0 (0)	0.222
Polymyalgia rheumatica	74 (36.5)	22 (59.5)	0.009
Fever (temperature ≥ 38°C)	18 (8.9)	5 (13.5)	0.377
Visual manifestations	49 (24.1)	7 (18.9)	0.490
Permanent visual loss	26 (12.8)	5 (13.5)	0.906
Cerebrovascular accidents	3 (1.5)	3 (8.1)	0.049
Limb claudication of recent onset	5 (2.5)	1 (2.7)	1.00
ESR (mean ± SD) mm/1st hr	92.6 ± 22.9	97.4 ± 20.2	0.235
Hemoglobin (mean ± SD) g/dL	11.8 ± 1.6	11.3 ± 1.6	0.092
Platelet count/mm^3 (mean ± SD)	406,000 ± 136,000	405,000 ± 118,000	0.996
Raised ALP§	54 (26.6)	6 (16.2)	0.175

ALP = alkaline phosphatase; ESR = erythrocyte sedimentation rate; GCA = giant cell arteritis; SD = standard deviation. (Adapted with permission from Gonzalez-Gay MA, Barros S, Lopez-Diaz MJ, et al. Giant cell arteritis: disease patterns of clinical presentation in a series of 240 patients. Medicine [Baltimore] 2005; 84:271.)

*From the onset of symptoms until the disease was diagnosed.

†Constitutional syndrome: asthenia, anorexia, and weight loss of at least 4 kg.

‡Abnormal temporal arteries on physical examination.

§Raised ALP: ALP values at diagnosis > 2 times above the upper normal range.

Table 3. Characteristics of Patients with Biopsy-Proven GCA Who Presented with/without PMR at Time of Disease Diagnosis

Variable	Pattern: With PMR (*n* = 96; 40%) No. (%)	Pattern: Without PMR (*n* = 144; 60%) No. (%)	*P*
Age at diagnosis (mean ± SD), yr	73.4 ± 6.3	75.6 ± 6.9	0.013
Women	58 (60.4)	72 (50)	0.113
Delay to diagnosis (mean ± SD), wk*	13.4 ± 12.2	8.3 ± 10	< 0.001
Headache	74 (77.1)	129 (89.6)	0.009
Scalp tenderness	32 (33.3)	49 (34)	0.911
Constitutional syndrome*	58 (60.4)	88 (61.1)	0.914
Abnormal temporal arteries*	65 (67.7)	110 (76.4)	0.138
Jaw claudication	40 (41.7)	58 (40.3)	0.83
Dysphagia	3 (3.1)	9 (6.3)	0.371
Fever (temperature ≥ 38°C)	7 (7.3)	16 (11.1)	0.325
Visual manifestations	15 (15.6)	41 (28.5)	0.021
Permanent visual loss	10 (10.4)	21 (14.6)	0.346
Cerebrovascular accidents	4 (4.2)	2 (1.4)	0.221
Limb claudication of recent onset	3 (3.1)	3 (2.1)	0.686
ESR (mean ± SD) mm/1st hr	91.3 ± 20.4	94.7 ± 23.8	0.261
Hemoglobin (mean ± SD) g/dL	11.7 ± 1.5	11.8 ± 1.7	0.632
Platelet count/mm^3 (mean ± SD)	410,000 ± 140,000	402,000 ± 129,000	0.658
Raised ALP*	19 (19.8)	41 (28.5)	0.121

ALP = alkaline phosphatase; ESR = erythrocyte sedimentation rate; GCA = giant cell arteritis; PMR = polymyalgia rheumatica; SD = standard deviation. (Adapted with permission from Gonzalez-Gay MA, Barros S, Lopez-Diaz MJ, et al. Giant cell arteritis: disease patterns of clinical presentation in a series of 240 patients. Medicine [Baltimore] 2005;84:272.)

*From the onset of symptoms until the disease was diagnosed.

Spanish descent. However, because there seem to be no large differences in the rate of GCA in most demographic groups, any bias in this study most likely comes from potential differences in the types of complaints people of different cultures would present with to a physician. This difference in presenting complaints probably does not change the overall rate of clinical manifestations.

- **Is it unlikely that the disease manifestations have changed since the evidence was reported?**

This question is often difficult to answer. Over time, diseases are redefined, new diseases are discovered, and new presentations can be described for old diseases, potentially changing previously reported frequencies of the clinical manifestations of a disease [1,6]. These shifts in disease taxonomy can cause such fundamental changes in the way we think about a disease that old descriptions of the manifestations of disease are no longer clinically relevant. A quick literature search on the disease, with a scan of the titles of the most recent articles, may point out fundamental new classification schema or research developments. It also may lead us to approach older studies about disease manifestations more cautiously. In the case at hand, the study by Gonzalez-Gay et al [4] was published in 2005 and describes patients seen from 1981 to 2004. Therefore, this is a reassuringly recent study, although we would want to confirm that there are no new diseases or changes in GCA that would alter its presentation since the article was published.

Based on your assessment, you believe that you can apply the results of the study by Gonzalez-Gay et al [4] to your patient. The reported rate of headache in GCA patients with symptoms of PMR does not seem to allow you to remove GCA from your differential diagnosis. You reflect that you would have been comfortable altering your differential diagnosis if the rate of headache had been near 100%.

Mrs. Brown returns from testing after an hour and you see that her ESR is markedly elevated at 73 mm/hr. After discussing the result with Dr. Manteau, you speak with Mrs. Brown and recommend an immediate temporal artery biopsy and initiation of empiric prednisone at 40 mg/day. The biopsy is scheduled for the following day. You discharge Mrs. Brown from clinic on the prednisone with this follow-up.

In 2 days, you receive the results of Mrs. Brown's biopsy and it is consistent with temporal arteritis. You telephone Mrs. Brown immediately to inform her of the diagnosis. She tells you that she is feeling better. Relieved, you schedule her for another appointment with you for the following week.

Conclusion

Articles on the frequencies of the clinical manifestations of disease are extremely valuable in augmenting all steps of the diagnostic process. Although these studies are frequently retrospective and difficult to read, researchers can often overcome these limitations, especially when the data are collected and presented in a systematic fashion. By reading and interpreting these studies, clinicians can substantially improve their diagnostic acumen over time.

Corresponding Author: Francis J. McBee Orzulak, MD, 1500 E. Medical Center Drive, Mott F-6882/SPC 5243, Ann Arbor, MI 48109 (email: fmcbeeor@med.umich.edu).

References

1. Richardson WS, Wilson MC, Williams JW Jr, et al. Users' guides to the medical literature XXIV. How to use an article on the clinical manifestations of disease. Evidence-Based Medicine Working Group. JAMA 2000;284:869–75.
2. Kassirer J. Diagnostic reasoning. Ann Intern Med 1989;110:893–900.
3. Richardson WS, Wilson MC, Guyatt GH, et al. Users' guides to the medical literature XV. How to use an article about disease probability for differential diagnosis. Evidence-Based Medicine Working Group. JAMA 1999;13:1214–9.
4. Gonzalez-Gay MA, Barros S, Lopez-Diaz MJ, et al. Giant cell arteritis: disease patterns of clinical presentation in a series of 240 patients. Medicine (Baltimore) 2005;84:269–76.
5. Kobayashi S, Yano T, Matsumoto Y, et al. Clinical and epidemiologic analysis of giant cell (temporal) arteritis from a nationwide survey in 1998 in Japan: the first government-supported nationwide survey. Arthritis Rheum 2003;49:594–8.
6. Guyatt G, Rennie D, editors. Users' guides to the medical literature: a manual for evidence-based clinical practice. The Evidence-Based Medicine Working Group. Chicago: AMA Press; 2002.
7. Marx JA, Hockenberger RS, Walls RM, et al, editors. Rosen's emergency medicine: concepts and clinical practice. 6th ed. Philadelphia: Mosby/Elsevier; 2006.
8. Sparnon AL, Little KE, Morris LL. Intussusception in childhood: a review of 139 cases. Aust N Z J Surg 1984;54:353–6.
9. Berhman RE, Kliegman RM, Jenson HB, editors. Nelson textbook of pediatrics. 17th ed. Philadelphia: Saunders; 2004.
10. Justice FA, Auldist AW, Bines JE. Intussusception: trends in clinical presentation and management. J Gastroenterol Hepatol 2006;21:842–6.
11. Salvarani C, Gabriel SE, O'Fallon WM, Hunder GG. The incidence of giant cell arteritis in Olmsted County, Minnesota: apparent fluctuations in a cyclic pattern. Ann Intern Med 1995;123:192–4.
12. Boyev LR, Miller NR, Green WR. Efficacy of unilateral versus bilateral temporal artery biopsies for the diagnosis of giant cell arteritis. Am J Ophthalmol 1999;128:211–5.
13. Pless M, Rizzo JF 3rd, Lamkin JC, Lessell S. Concordance of bilateral temporal artery biopsy in giant cell arteritis. J Neuroophthalmol 2000;20:216–8.
14. Hunder GG, Bloch DA, Michel BA, et al. The American College of Rheumatology 1990 criteria for the classification of giant cell arteritis. Arthritis Rheum 1990;33:1122–8.
15. Sackett DL, Straus SE, Richardson WS, et al. Evidence-based medicine: how to practice and teach EBM. New York: Churchill Livingstone; 2000.

How to cite this article:
McBee Orzulak FJ, Mangrulkar RS. Applying evidence to the diagnostic process for giant cell arteritis: how to use an article on the clinical manifestations of disease. Semin Med Pract 2008;11:1–10. Available at www.turner-white.com.