The Need for Clinical Judgment in the Diagnosis and Treatment of Lyme Disease

Elizabeth L. Maloney, M.D.

ABSTRACT

Clinical practice guidelines are increasing in number. Unfortunately, when scientific evidence is uncertain, limited, or evolving, as is often the case, conflict often arises between guideline committees and practicing physicians, who bear the direct responsibility for the care of individual patients. The 2006 Infectious Diseases Society of America guidelines for Lyme disease, which have limited scientific support, could, if implemented, limit the clinical discretion of treating physicians and the treatment options available to patients.

Introduction

Clinical practice guidelines are now ubiquitous throughout the United States. The National Guidelines Clearing House, under the category “diseases,” currently lists 2,126 separate guidelines on its website. Clinical guidelines are intended to assist physicians in patient care by clearly communicating the results of the guideline committees’ evaluation of available therapeutic options. However, the processes by which individual guidelines are constructed may be less clear, leading to disagreements between the issuing committee and the physicians who treat patients—physicians who may well be as experienced and knowledgeable as the guideline committee.

The 2006 Infectious Diseases Society of America (IDSA) guidelines for Lyme disease were released in the fall of that year and were soon the focus of an antitrust suit brought by Connecticut’s attorney general. A settlement between the two sides was announced on May 1, 2008; it called for the seating of a new panel and a review of divergent medical points of view. This article reviews the 2006 IDSA Lyme guidelines regarding the impact various recommendations may have on the use of clinical judgment in the diagnosis and treatment of patients with Lyme disease.

Clinical Judgment in the Diagnosis of Lyme Disease

The IDSA in its 2006 Lyme disease guidelines states:

Clinical findings are sufficient for the diagnosis of erythema migrans, but clinical findings alone are not sufficient for diagnosis of extracutaneous manifestations of Lyme disease or for diagnosis of [human granulocytic anaplasmosis] HGA or babesiosis. Diagnostic testing performed in laboratories with excellent quality-control procedures is required for confirmation of extra cutaneous Lyme disease, HGA, and babesiosis.

Initially, the statement appears innocuous; laboratory confirmation of any diagnosis is always reassuring. But here the guidelines panel goes a step further. By requiring lab confirmation, it sets up a diagnostic hierarchy in which testing supersedes clinical judgment, negative results on indirect laboratory assessments of infection overrule carefully constructed clinical assessments, and tests are deemed infallible.

Yet, this diagnostic scheme is fallible. Consider the situation in which 100 patients with undiagnosed Lyme disease seek medical attention for evaluation of fever, headache, fatigue, and body aches occurring at the end of June. Recall that CDC data indicate that erythema migrans (EM) rashes are reported in 68% of patients meeting the surveillance case definition, and that the guidelines recommend two-tier serologic testing of patients lacking the diagnostic rash. In the two-tier scheme, patients are first tested with an enzyme-linked immunoabsorbant assay (ELISA) or indirect fluorescent antibody (IFA) test, and those with positive or equivocal results are then tested with Western blotting; patients who are negative on ELISA are not tested further. Steere et al. found it to be 38%. As Table 1 demonstrates, the laboratory confirmation requirement is problematic; as many as 22% of early Lyme disease patients would go untreated.

Clearly, this is unacceptable; patients would be left untreated at the stage when therapy is most efficacious. Owing to the potential for false negative results in these circumstances, Steere et al. suggested that physicians consider treating patients with “summertime flu” symptoms. The need for such a suggestion emphasizes the principal reason for this challenge—laboratory confirmation requirements undermine the value and primacy of clinical data and may impede care, as would be the case in this very common clinical scenario.

The same problem with laboratory confirmation holds true for late neurologic Lyme disease. Starting again with 100 patients who have undiagnosed Lyme disease and objective, non-EM findings, 43%-56% would be misdiagnosed because of deficits in laboratory capabilities, as shown in Table 2. In late Lyme, sensitivity of the testing procedure was found to be 44% by Ledue et al., and 57% by Dressler et al.

The low sensitivity of two-tier testing in late neurologic Lyme disease can be traced back to the original paper by Dressler et al., from which the Centers for Disease Control and Prevention (CDC)
took its IgG Western blot criteria. After identifying the 10 bands on Western blotting that yielded the highest specificity in a retrospective study, Dressler et al. then tested the criteria in a prospective study. In that study, the paper reports that 21 of 29 patients with neuroborreliosis had positive IgG Western blot results, yielding a sensitivity of 72%. The ELISA used by Dressler et al. had a sensitivity of 79%. Performing the tests sequentially, as is done in two-tier testing, results in an overall sensitivity of 57% (79% x 72%). With the two-tier sensitivity for late Lyme disease roughly 50%, a negative result does not inform physicians, but may easily lead them astray.

Other studies on the two-tier strategy yield different and higher values for sensitivity. Some studies speak of the “relative sensitivity” of a test rather than the true sensitivity. The disagreement between studies investigating the sensitivity of various testing methodologies for Lyme disease indicates a problem with test reliability, which has been the subject of other papers. If the serologic tests for Lyme disease were equally reliable, sensitivity would be nearly identical across studies of similar, and appropriate, design. (A full discussion on the limitations of serologic testing is beyond the scope of this paper.)

Other methods available to support or confirm a clinical diagnosis of Lyme disease in the absence of an EM have low sensitivity (polymerase chain reaction [PCR] of cerebrospinal fluid and blood), may be invasive, or are not clinically available.

With serologic testing being insensitive, clinical data—the history and physical examination—become even more important. Relying on clinical data to make a diagnosis is not unique to Lyme disease. One study on the relative values of history, physical examination, and diagnostic studies found that internists used history alone to establish the correct diagnosis in 76% of test cases. Another found that in distributing a 100% total relative value between these three types of data, clinical faculty valued history at 63.3%, physical examination at 19.2%, and laboratory/imaging data at 17.5%. Such evidence establishes that the diagnostic hierarchy proposed by the guidelines is inconsistent with the way medicine is practiced.

A Lyme disease history begins with the potential for exposure. This history, while a key element, is not always enlightening. Patients may be unaware of whether they live/work/recreate in a Lyme-endemic area; they may forget about vacations in endemic areas. Questions regarding tick bites may lead to inappropriately ruling out Lyme disease; in one study on erythema migrans, only 14% of the patients recalled being bitten by a tick.

Clinically, and in keeping with its multisystemic nature, Lyme disease has been described as being “symptom rich, exam poor.” Symptoms may be specific or nonspecific, mundane or unusual, acute or chronic; some are prognostic. Some physicians have been criticized for “seeing Lyme everywhere” in that they recognize scores of symptoms beyond EM rashes, Bell’s palsy, and arthritis as being associated with Lyme disease. Yet, early researchers also noted these symptoms. In a treatment trial on early Lyme disease, Massarotti et al. found that subjects reported the following symptoms: 56% had headache; 42%, stiff neck, with 19% having pain with neck flexion; 14%, dysestheasias; 11%, photophobia; and 4%, facial palsy. Consider these symptoms from Logigian et al., shown in Figure 1.

The wide array of Lyme disease symptoms is consistent with Borrelia burgdorferi’s ability to infect multiple organ systems; nervous system involvement creates the potential for varied and atypical symptoms. Common symptoms include: EM rash, fever, fatigue, headache, neck pain, joint or muscle pain, paresthesias, memory impairment, weakness of facial muscles, mood disorders, neuropathic pain. A compendium of manifestations by system is given in Table 3.

![Figure 1. Frequency of Various Signs and Symptoms in Late Neurologic Lyme Disease](image-url)
It is the multisystemic nature of the illness that provides physicians with useful diagnostic information. In fact, with the exception of an isolated EM rash or swollen joint, patients with symptoms restricted to a single system are unlikely to have Lyme disease. Recognizing the potential for disease is different from “seeing it everywhere.” Failure to recognize Lyme disease may lead to serious harm, as antibiotics are delayed and the infection is unchecked. 

The nonspecific nature of many Lyme disease symptoms leads some to suggest that such symptoms hold no diagnostic value. Lyme disease is like many other illnesses that present with nonspecific and often subtle symptoms—symptoms that may go unrecognized by physicians. Examples include hypothyroidism, ovarian cancer, and acute subendocardial myocardial infarction. What gives the individual symptoms of Lyme disease value is their occurrence in clusters; a single symptom means little but four or five may, for all practical purposes, make the case. Just as abdominal bloating, urinary urgency, and pelvic pain raise “red flags” for gynecologists, the combination of fatigue, paresthesias, arthralgias, and memory complaints presenting in a single patient commands the attention of physicians aware of these potential Lyme disease symptoms.

Steere et al. noted that patients with early Lyme disease who lacked an EM rash presented with an average of four or more symptoms. Fever, chills, malaise, and myalgia, all nonspecific, were present in 46%-71% of the patients with definite Lyme disease alone. In this group, it was the clustering of nonspecific symptoms in the appropriate setting that led to the correct diagnosis of Lyme disease. Logian et al. also noted the nonspecific nature of identifying symptoms: “The most common form of chronic central nervous system involvement in our patients was subacute encephalopathy affecting memory, mood, and sleep, sometimes with subtle disturbances in language. Diagnosis of this condition may be difficult because the typical symptoms are nonspecific” [emphasis added].

To provide a clinical level of diagnostic sensitivity higher than two-tier testing, physicians need to recognize the symptom clusters and maintain a high index of suspicion for Lyme disease.

Symptoms not only form the basis of disease identification, they may also inform on prognosis. Dysesthesias, paresthesias, multiple EM lesions, increased irritability, persistent fatigue, headache, stiff neck, and increased severity of the initial illness were associated by various investigators in the early Lyme disease treatment trials with an increased risk of treatment failure. Symptoms were also used in the trials as indicators that a strategy was working or needed to be altered.

Findings on physical exam are usually subtle and limited; they may be variably present. The more common findings include: solitary or multiple EM lesions, manifestations of cranial neuritis (such as extraocular palsy, ptosis, decreased facial sensation, facial nerve palsy, decreased hearing), swollen and tender joints, diminished sensation, and motor weakness. Cognitive deficits are usually not readily apparent on mental status testing, but patients may be disorganized or slow to respond to questions. A lack of physical findings does not necessarily indicate that the symptoms in those cases cannot be corroborated with objective evidence. Halperin et al. studied 14 patients with complaints of distal paresthesias; 10 had completely normal sensory, motor and reflex findings on examination, three had only mild sensory loss, and one had moderate sensory and motor loss coupled with decreased reflexes. All underwent EMG testing; 13 of the 14 had “significant neurophysiologic findings.” Logian et al. also found that detailed neuropsychometric testing could reveal cognitive deficits that were not apparent on routine mental status testing. Cost and time constraints do not allow for such complete testing in a community setting, but the studies suggests that with sufficiently detailed testing, objective evidence may be discovered and the subjective data supported. The absence of findings does not equal absence of disease.

Even the EM rash has a variable presentation that may cause less informed physicians to miss it. An EM lesion may have one or more of the following characteristics: homogeneously erythematous color, prominent central clearing, target-like appearance, central vesicles or pustules, partially purpuric, and not scaly, unless topical corticosteroid creams have been applied or the rash is old and fading. An EM rash must be distinguished from: tick bite hypersensitivity reactions, insect or spider bites, contact dermatitis, bacterial cellulitis, and tinea. An interesting study in JAMA compared responses from physicians in endemic and nonendemic areas with regard to what percentage of EM rashes in their practices had central clearing. Physicians from endemic areas thought it only 19%, while those from nonendemic estimated 80%. The authors did not give a reason for the disparity; possibilities include B. burgdorferi strain variation or physician experience. The variable presentation of the EM rash, coupled with the fact that it does not manifest in 32% of patients, makes it unwise to rely on EM as the only manifestation of Lyme disease that has clinical diagnostic utility.

Physicians use pattern recognition as a common diagnostic heuristic. These cognitive “shortcuts,” when used properly, allow physicians to move quickly to the correct diagnosis. Pattern recognition transforms exposure, individual symptoms, and the course of illness into a unified diagnosis; it is why some physicians specifically see “Lyme disease” when colleagues see only a generalized “positive review of systems.” For physicians unfamiliar with the pattern of Lyme disease, serologic testing, combined with clinical data, offers the potential for reaching the correct diagnosis. However, serology alone cannot confirm or deny presence of infection. In Lyme disease, there is no testing shortcut.

Furthermore, diagnostic criteria are situational. Clinical criteria are constructed to diagnose and treat ill patients. Research criteria are constructed to test a hypothesis in a uniform group of subjects; researchers have no duty to those excluded from the trial. Surveillance case definitions are created for the purpose of standardization, not patient care; they exist so that health officials can reasonably compare the number and distribution of “cases” over space and time. Whereas physicians appropriately err on the side of over-diagnosis, thereby assuring they don’t miss a case,
surveillance case definitions appropriately err on the side of specificity, thereby assuring that they do not inadvertently capture illnesses due to other conditions. Recognition of the differing goals allows knowledgeable physicians the discretion to diagnose Lyme disease in patients lacking the five of 10 bands required for admittance into the surveillance group. Failure to acknowledge the distinction results in many patients with Lyme disease remaining undiagnosed and untreated.

Mandatory laboratory confirmation of clinical diagnoses, as advanced in the 2006 IDSA guidelines, reverses the roles of clinical and laboratory data in the diagnostic process and hierarchy. Substituting laboratory tests for physician judgment is not clinically sound, particularly when laboratory tests lack sensitivity. This recommendation is a change from the 2000 IDSA guidelines on Lyme disease, but the 2006 panel did not discuss the reasons for this change nor cite any references from the literature to support it. Guideline developers have identified the need for reconciliation between new and former versions of the same disease guidelines; the IDSA, itself, endorsed the reconciliation process, yet it did not occur in this instance.

Correctly diagnosing extracutaneous Lyme disease can be difficult. The importance of clinically derived data has been demonstrated repeatedly, as have the weaknesses of serologic testing. At this time, Lyme disease should remain a clinical diagnosis, with testing playing a supportive role.

Clinical Judgment in Management of Patients with Lyme Disease

Clinical judgment is required to appropriately manage patient care. Patient management is an evolutionary process, not a static state; ongoing assessment allows for refinement of the original diagnosis or the search for new one. Lyme disease is no exception to this rule; yet the 2006 IDSA guidelines reduce clinical management to a one-size-fits-all approach quickly chosen from a table. Clinical judgment is especially important when the clinical picture is unclear and laboratory data unhelpful. After careful investigation of other potential diagnoses, physicians may need to perform an empiric treatment trial as a diagnostic modality. The use of such trials extends well beyond Lyme disease. For example, patients with nonspecific epigastric pain may be offered “GI cocktails” as a means to both diagnose and treat the condition.

Clinical decision-making in Lyme disease requires ongoing information; the longitudinal treatment trials on Lyme disease demonstrated the value of this data. Historical and physical examination data were gathered at defined points; on some occasions this information was used to alter the treatment protocol (investigators withdrew or re-treated some subjects). Follow-up visits in many of the studies on Lyme disease demonstrated a positive correlation between reported symptomatic changes and subsequent physical findings or test results. Long-term follow-up extending beyond the active treatment phase provides researchers, as well as physicians in clinical practice, the ability to discern the difference between placebo and treatment effects.

Clinical judgment in Lyme disease requires physicians to weigh risk-benefit concerns with individual patients. Treatment risks for the patient include potential adverse effects from antibiotic therapy (including risks associated with medication administration), costs associated with therapy, and lifestyle changes to accommodate treatment. Patient benefits include improved health with attendant improvement in quality of life and lower medical costs following recovery. Antibiotic therapy, including long-term oral antibiotics, is generally safe and well tolerated. A meta-analysis on the risks associated with intravenous (IV) access of various types found that peripheral intravenous catheters cause 0.5 bloodstream infections per 1,000 intravascular device (IVD) days while surgically implanted long-term central venous devices—cuffed and tunneled catheters—cause 1.6 infections per 1,000 IVD-days. Data from Lyme disease treatment trials can inform on the risk of IV antibiotic therapy in this patient population. Table 4 reports the complication rates in the treatment groups of Lyme disease studies which used IV ceftriaxone for a minimum of 30 days. Significant adverse events included medication-related events (severe allergic reactions, gall bladder toxicity, Clostridium difficile enterocolitis, renal failure) and catheter-related events (skin infiltration, infection, and thrombosis).

Adverse events in the Fallon study are considerably higher than in the others; reasons are unknown, and the small sample size makes it difficult to draw conclusions. There were three cases of ceftriaxone allergy in the 23 patients; this 13% allergic rate is higher than expected. Thrombi developed in two patients, but the paper does not provide details of the site of the peripherally inserted central catheter (PICC) or its specific type. Additional studies are needed to delineate the risk of IV antibiotic therapy extending beyond 30 days in better detail, and to determine whether there would be opportunities to minimize those factors contributing to the total risk.

There are also risks to the patient associated with failure to treat a continuing infection. These include declining health, decreased productivity, a potential for increased costs as more health-related services are required, and costs related to palliative medications (including their potential adverse effects). The IDSA guidelines raise concerns about the impact longer treatment regimens may have on society. While these concerns should not sway treating physicians who are entrusted with the care of individual patients, the concerns merit some comments. The guidelines authors focus attention on treatment risks to society, citing additional costs and the potential for increased bacterial resistance in the community. However, the authors ignored potential benefits to society from such treatment regimens. These benefits include improved health in the community, increased production from previously ill patients, and potential for success in this patient population to inform treatment decisions in other groups.
Additionally, there are societal risks from not treating; these include ever increasing expenses for a chronically ill subpopulation and lost productivity from ill workers. In the individual patient, the decision to treat or to prolong treatment may depend on the length of time between onset of illness and diagnosis; severity of the patient’s presenting symptoms; presence of neurological symptoms; whether the course of the illness is progressive; whether the illness significantly affects the patient’s quality of life or functional abilities; presence of untreated co-infections; the patient’s immune system status; whether diagnostic tests, symptoms or treatment response suggest ongoing infection; the patient’s response to treatment; which medications the patient can tolerate; the specifics of prior treatment regarding antibiotic type, dose, and duration; whether the patient relapses when treatment is withdrawn; the risks/benefits of the treatment approach under consideration; and availability of any alternative treatment approaches and their attendant risks balanced against the risks associated with failing to treat. These highly individualized decisions are best made by the treating physician and the patient.

The controversy over antibiotic treatment duration for patients with Lyme disease exists because there is no test of cure, and individual patient responses to specific therapeutic approaches have been highly variable. Lyme disease, in many patients, is marked by periods when the illness is relatively quiescent. Lacking a test of cure, physicians who do not rely on arbitrary cut-off points are faced with a difficult decision when attempting to determine an appropriate stopping point. Mixed results from the treatment trials add to the uncertainty.

The variable response to treatment has been well documented, and the causes remain unclear, as scientific evidence in this area is still evolving. Early hypotheses of autoimmune processes have not been substantiated; persistent infection, however, has been demonstrated in case reports and animal studies. Patients with Lyme disease are a heterogeneous group. Genetic variation may play a role in pathogenesis and treatment response. Just as HLA status may be related to treatment response, which medications the patient can tolerate; the specifics of prior treatment regarding antibiotic type, dose, and duration; whether the patient relapses when treatment is withdrawn; the risks/benefits of the treatment approach under consideration; and availability of any alternative treatment approaches and their attendant risks balanced against the risks associated with failing to treat. These highly individualized decisions are best made by the treating physician and the patient.

Variation in infecting strains of B. burgdorferi certainly is a factor. More than 100 strains of B. burgdorferi have been identified. Certain strains are more virulent and pathogenic than others; instances of antibiotic susceptibility varying between strains is well documented. Coinfections and comorbidities also contribute to the heterogeneity of treatment response seen in Lyme disease. *Ixodes scapularis* is able to carry multiple known bacterial, viral, and parasitic pathogens, and evidence for additional tick-borne pathogens continues to emerge. Different combinations of pathogens require different treatment regimens; failure to identify and treat the specific pathogens causing an illness may partially explain variations in treatment responses.

As explained by Kravitz et al., “[h]eterogeneity of treatment effects reflects patient diversity to risk of disease, responsiveness to treatment, vulnerability to adverse effects, and utility for different outcomes.” Kravitz et al. discuss the application of generalized, or averaged, results from treatment trials to the care of an individual patient, and pitfalls inherent in applying them too strictly, noting that “misapplying averages can cause harm, by either giving patients treatments which do not help or denying patients treatments that would help them.” The individual patient is not a numeric average but, rather, falls somewhere on the continuum of the bell curve and, hence, requires individualized care.

Clinical guidelines should not suppl Ant-9ant the judgment of treating physicians. Quality care requires the physician to consider management decisions in light of the details unique to each patient. When guideline recommendations are substituted for carefully derived, individualized decisions, there is a potential for harm. The American Academy of Pediatrics policy statement on guideline development recognizes this principle. The document outlines how evidentiary strength and risk-benefit analyses are integrated to yield a specific recommendation level. For example, strongly positive recommendations require benefits to clearly exceed risks, and supporting evidence must be of excellent quality.

In this scheme, strong recommendations are not made based on low-quality evidence or expert opinion. Options identify treatment alternatives. Options recognize patient preferences and respect the clinician’s decision-making process. The U.S. Preventive Services Task Force also recognizes scenarios in which the certainty of the evidence is low. In those situations, no recommendation is made, regardless of the perceived net magnitude of benefit or harm. Additionally, the Task Force advocates shared decision-making between individual patients and their physicians, instead of population-based recommendations, when issues under consideration are highly sensitive to patient utilities.

Guideline committees are not in a position to perform risk-benefit analyses for specific patients. Patient-specific risk-benefit analyses are the essence of clinical judgment. Such judgments are the domain of individual treating physicians; guideline committees may inform judgments through their evaluation of therapeutic options, but they may not substitute their judgments for those of the treating physicians. A recent *JAMA* editorial by Shaneyfelt and Centor said as much: “Guidelines are not patient-specific enough to be useful and rarely allow for individualization of care. Most guidelines have a one-size-fits-all mentality and do not build flexibility or contextualization into the recommendations.” While the 2006 IDSA guidelines contain the typical legal disclaimer that “they are not intended to supplant physician judgment with respect to particular patients or special clinical situations,” formulaic disclaimers cannot overcome the failure of the guidelines to provide treatment options and to recognize the role of clinical judgment in individualized care. These shortcomings cannot be addressed in boilerplate disclaimers; they can only be addressed in the substance of the guidelines.

Available laboratory tests for Lyme disease have poor sensitivity. Treatment trials cited in the guidelines for early Lyme disease were dissimilar, making it hard to compare outcomes; those for late neurologic Lyme disease involved only 96 patients whose treatment responses can be analyzed. Both the early and late treatment trials yielded poor outcome rates for complete recovery. The prophylaxis recommendation is based on a single
study performed under conditions unlikely to be reproduced in community practices, and the list of “not recommended” therapeutic modalities is apparently based on panel opinion. 4-23  Given the limits of guidelines in general, and the specific shortcomings of the 2006 IDSA guidelines on Lyme disease, patients and their physicians should be free to act without interference; many may justifiably decide to decide for themselves which strategy to embrace.

Elizabeth L. Maloney, M.D., is a family physician from a Lyme disease endemic area in Minnesota. Contact: P.O. Box 84, Wyoming, MN 55092, tel. (651) 462-0192; email bettymal2003@yahoo.com.

Acknowledgments: The author would like to acknowledge Lorraine B. Johnson, J.D., M.B.A., and Dr. Bea Szantyr for their input and thoughtful editorial comments.

Potential Conflicts of interest: The author receive speaker’s fees for creating and presenting continuing education courses on Lyme disease. She submitted six challenges to the IDSA Lyme disease guideline review panel and is a member of the International Lyme and Associated Disease Society (ILADS).

REFERENCES


