

Verim Research Medical Study Analysis

Study Analyzed:

Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease, *New England Journal of Medicine*, 2001. 345(2):85-92.

Klempner, M. S., Hu, L. T., Evans, J., Schmid, C. H., Johnson, G. M., Trevino, R. P., Norton, D., Levy, L., Wall, D., McCall, J., Kosinski, M., Weinstein, A.

This analysis is prepared to support litigation or administrative proceedings where this study is cited as evidence or support of a scientific position. The intent is to present the best arguments supporting and refuting the positions stated in the study, with the goal that experts and advocates on both sides of the issue be aware of the study's strengths and weaknesses. Analysts make every attempt to disregard status quo positions and evaluate study data and conclusions on their own merits.

Table of Contents

Summary	3
Scientific Strength Rating	4
Impression	5
Background	6
Major Strength Arguments.....	8
Other Strength Arguments.....	8
Details of Strength Arguments	8
Major Weakness Arguments	9
Other Weakness Arguments	9
Details of Weakness Arguments	10
Bottom line	11
Supplemental Information.....	12
Double-blind Trials	12
Herxheimer reaction.....	12
Statistical Power	12
The SF-36 Questionnaire	13
Published Abstract of Study.....	14
References	15

Summary

Question posed: Are additional antibiotics beneficial to patients previously treated for Lyme disease who continue to exhibit symptoms?

Methods: Patients with persisting symptoms, previously treated for Lyme disease, were selected and treated either with antibiotics or placebo. Patients' mental and physical health status were assessed with the SF-36, a patient-completed questionnaire and the Fibromyalgia Impact Questionnaire.

Results: No significant differences were reported between the treated and placebo groups.

Implication: This study did not provide evidence that additional antibiotics are effective in treating patients with persisting symptoms of Lyme disease after initial short-course antibiotic treatment.

Scientific Strength Rating

A=Extremely Strong, B=Strong, C=Support, D=Little Support, F=No support

Design D

The SF-36, the primary outcome measure, is not designed to measure treatment improvements for a specific condition. Participants selected for study were biased toward failing the treatment studied. There were no objective measures reported post-treatment.

Sample Size C

Best estimate is that there is a 24% chance the study reached the wrong conclusion solely because of the small number of participants.

Execution C

There are indications of poor randomization into antibiotic and placebo groups. PCR testing did not produce a single positive indicating a methodology that was highly insensitive (see below).

Analysis/Conclusions D

The authors make a highly questionable generalization: that since this specific treatment does not seem effective, no long-term antibiotic treatment would be effective for treating continuing symptoms of Lyme disease. They support this generalization with the highly debatable statement that longer treatments have not been shown effective in other diseases.

Impression

Researchers had a strong bias toward reporting a negative outcome. They completed the minimal requirements necessary to present their study as a valid, randomized, double-blind clinical trial but provided minimal post-treatment data, using insensitive measures.

Background

Two opinions have evolved for the treatment of Lyme disease. The first, a **narrow view**, promoted by professional organizations such as the American Medical Association, the American College of Physicians, and the Infectious Disease Society of America¹ says that Lyme disease is hard to catch and easy to cure. The second, a **broader view**, promoted by individual treating clinicians, says short course antibiotic treatments are effective only early in the course of the disease and later-stage Lyme disease requires longer courses of antibiotics, often administered intravenously. The broader view also states contracting Lyme disease may not fit the conventional profile of transmission by tick bite, followed by an obvious rash.

The debate has been rancorous, especially since the late 1980's². Three factors seem to fuel this: 1) IV antibiotic treatment can cost \$50,000 to \$100,000, 2) there are no laboratory tests that prove or disprove infection, and 3) the narrow-view camp has a long history of taking strong positions without supporting hard evidence and it would be difficult for them to retract earlier positions.

Lyme disease, like syphilis, is caused by a spirochetal (spiral-shaped) bacteria. In 1977, epidemiologists reported on a disease outbreak near Lyme, Connecticut where 51 patients were identified with an unusual arthritis, 25% of which reported an expanding round rash later referred to as erythema migrans (**EM**)³. Although residents had reported unusual clustering of neurological problems, extreme fatigue, and recurring skin rashes in the locale for at least twenty years, investigators concentrated on the two most visible symptoms, the EM rash and arthritis, and treated the outbreak as a new disease entity. The EM rash eventually became the predominant defining characteristic of the disease⁴, although there has never been an accurate assessment of how often infection occurs without the EM rash.

Borrelia burgdorferi (**Bb**) was identified as the causative bacteria in 1981⁵ and the deer tick was found to carry and transmit the bacteria to humans. A similar disease had been described

in Europe at least fifty years earlier⁶. Narrow-view physicians assumed that the common antibiotic treatments for treating syphilis, largely unproven to eliminate infection, would be effective for Lyme disease. There is much evidence that Lyme disease symptoms return and worsen after shorter antibiotic treatments⁷. Narrow-view theory is that this is an autoimmune condition that persists after *Bb* is eliminated, although this is highly speculative and there is little evidence to support this.

The absence of a good test for Lyme disease makes it difficult for either side to prove their point. The study analyzed here, commonly referred to as the "Klempner study", reinforces the failure of Lyme antibody tests to discriminate. The "Two controlled trials" in the study title refers to separate trials with participants who tested positive and negative in an antibody test. There were no significant differences observed between the seropositive and seronegative groups before or after treatment.

The bacteria *Bb* is difficult to culture⁸ and is often present in the blood at extremely low counts. It typically sequesters in poorly oxygenated fibrous tissues such as nerves, ligaments, and the skin. Common **tests for antibodies** to *Bb* (usually ELISA's or Western blots) present many problems⁹. Detailed studies of antibody response in humans and animal models show a confusing host immune response difficult to relate to what is commonly accepted as response to other infections¹⁰. Further, the tests detect only free antibodies, problematic because an ill patient's antibodies may be entirely, or almost entirely, bound to bacteria, making the antibodies undetectable by the tests¹¹.

Estimates of the accuracy of the antibody tests are highly speculative. Existing studies evaluating test accuracy typically compare patients who have had the EM rash to patients without a history of EM. Because the prevalence of infected patients who never exhibit EM has never been determined, the calculations of accuracy of the antibody tests in detecting infected individuals are highly suspect. Basing diagnosis on symptoms is also difficult because symptoms are variable and can appear and

reappear unpredictably. No set of symptoms is definitive for all infected individuals.

PCR testing for *Bb* detects nucleic acid sequences in DNA strands. The small number of *Bb* in the bloodstream and difficulties in technique make PCR testing fairly insensitive. Best estimates are that PCR testing of serum or whole blood will detect less than 30% of infected individuals¹². There is no general consensus as to which nucleic acid sequences best indicate the presence of *Bb*. Probably because of the predominance of the narrow-view concept of Lyme disease, PCR testing is not routinely accepted as a definitive test for Lyme disease as it is for other infectious diseases.

Highly **variable symptoms** with much overlap with other diseases make diagnosis subjective. Experienced diagnosticians recognize that uncommon symptom combinations, such as arthritis and neurological problems, are highly suggestive of *Bb* infection. Partial or complete remission of multiple symptoms after antibiotic treatment also strongly indicates *Bb* as the causative agent. These subtleties of diagnosis are not typically emphasized in medical literature. Instead, great emphasis is placed on the EM rash and a history of tick bite.

Another complication in evaluating symptoms and treatment results of Lyme disease is that an effective treatment can produce a temporary worsening of symptoms. This variable reaction, known as a Jarisch-Herxheimer reaction, or simply **Herxheimer reaction** is recognized in several infectious diseases, particularly those caused by spirochetal bacteria, such as Lyme disease and syphilis (see Supplemental Information).

The broader view has been censored both at conferences and in medical journals. Speakers with views not adhering to the narrow view are not permitted to present at conferences and their articles are routinely rejected by journals.

The Klempner study received more attention than it deserved. The *New England Journal of Medicine* made a pre-publication version available on the Internet implying the study produced information that was immediately important to clinicians. The study's limited measure of one treatment protocol is often

quoted as evidence "long-term antibiotics are ineffective for treating chronic Lyme disease". Actually, what the study shows is that the specific treatment of 30 days of IV ceftriaxone followed by 60 days of oral doxycycline at specific dosages did not improve patient's health to the extent that there was a fairly dramatic improvement in the patient's self-rating on the SF-36 questionnaire.

This clinical trial is quite unusual in that it reports no objective measures post-treatment, only subjective patient self-assessments, highly summarized before statistical analysis, making detection of a significant difference unlikely.

Major Strength Arguments

This is a double-blind clinical trial conducted at a reputable institution and published in a prestigious medical journal.

In support of the conclusion, there were no indications of self-reported improvements in the treated group that exceeds the group who received a placebo.

There is no reason to suspect gross errors in study execution.

the authors stated in a response that they conducted over 1800 cultures and PCR tests, both for participant selection and for assessment during the study, and all results were negative¹³. This calls into question the researchers' techniques in performing these tests, which were not specifically described in the case of the PCR and poorly described for the culture. Performing 1800 tests on participants from locales highly endemic for Lyme disease would be expected to produce some positives, however few, from new or reinfections, if nothing else. Additionally, there was no statement that positive controls were used to validate test procedures.

Other Strength Arguments

If the studied treatment produced significant, dramatic improvement in physical or mental health, this would have been detected by the SF-36 questionnaire. This questionnaire is well validated and has been used in over 4,000 peer-reviewed published studies.

This study was conducted at a respected institution, funded by the NIH. It was published in one of the world's most respected medical journals, *The New England Journal of Medicine*, and was fully peer-reviewed.

This study was a randomized, double-blinded study, the most objective type of clinical testing, with carefully selected participants.

Details of Strength Arguments

For this study, other than the support of NIH funding and the basic design as a double-blind study, there are no apparent detailed strength arguments.

The authors attempted to attribute significance to the fact that none of the participants had positive cultures or PCR (genetic) tests for *Bb*. The original paper stated that subjects testing positive in these tests were excluded from the study, making their absence less than surprising. When later challenged on this point,

Major Weakness Arguments

The study reported no post-treatment objective measures, only patient self-assessments. Overall, there is a significant lack of post-treatment information presented. Although many tests and assessments were made prior to treatment, the only results reported post-treatment were the results of participants' questionnaire responses of how they perceived their health. Scores were highly summarized before statistical analysis. The primary questionnaire used, the SF-36, is recognized as being insensitive to measuring an improvement in a specific health condition^{14, 15}.

Instead of reporting the eight subscales of the SF-36, which are scored on a 0 to 100 scale, the researchers use only the SF-36's two summary scores, the physical component summary (PCS) and mental component summary (MCS) scores. The researchers' then lumped scores into "Improved", "Worse", and "Unchanged" categories, further reducing sensitivity to detecting changes. Another self-rating instrument, the Fibromyalgia Impact Questionnaire, was also administered, but scores were not reported, only that changes were not significant, again after summarization.

There was a strong bias toward selecting only participants that were highly likely to have already failed a treatment similar to the one studied. The participants were required to have already had at least one course of antibiotics and to have continuing symptoms.

Neither of the two antibiotics used in the treatment studied, IV ceftriaxone or oral doxycycline, have intracellular penetrability. There is a great deal of evidence that *Bb* is a invades cells (including one report by Klempner in 1993¹⁶) where it is protected from antibiotics lacking cellular penetrability. In any case, 2 grams daily of IV ceftriaxone for 30 days followed by 100mg of doxycycline twice daily for 60 days is not definitive or aggressive treatment for patients who have had earlier treatment failures.

The small number of participants does not yield great statistical power.

Other Weakness Arguments

This study only presents data from one post-treatment measure, the SF-36, and then only after the results are highly summarized. Although there were a number of baseline pre-treatment evaluations and tests performed, none of these were reported post-treatment.

At baseline, there were assessments of six symptom classes, plus fibromyalgia tender points, and objective measures of white-cell counts and protein levels in cerebrospinal fluid. These assessments are not reported post-treatment, although the authors state additional clinical and laboratory evaluations were performed on days 3, 5, 13, 21, 30, 45, 75, 90, and 180.

Antibiotic selection for treatment of Lyme disease should be tailored to individual patients and often requires mixes of different antibiotics, after evaluation of where a patient is in their disease course¹⁷. This study applied the same treatment to all participants: 30 days of IV ceftriaxone followed by 60 days of doxycycline.

Details of Weakness Arguments

The participants selected had already received a mean of 3.0 antibiotic treatments in the antibiotic group and 2.7 treatments in the placebo group. If a participant had received parenteral (essentially IV) treatment for more than 60 days previously, they were excluded. Other studies report wide variation in patient response to antibiotics¹⁸. The Klempner study selected participants who had a high likelihood of previous failure with a treatment similar to the one being evaluated.

The SF-36 is a generic health utility questionnaire generally recognized as insufficient for measuring progress in resolving a particular condition. (Most observers, when encountering the questionnaire the first time will be surprised at its perceived superficiality and would hardly regard it as a detailed medical assessment of a complex condition.) Its primary use is to compare the impact of two or more medical conditions or to compare patients with a condition to the general population of healthy individuals. (In this respect, the study did indicate the participants were seriously impaired: their pre-treatment physical component summary scores were comparable to patients with congestive heart failure and participants had greater impairment than patients with a recent myocardial infarction [heart attack]). The study did not report participant's response in the eight specific subscale areas the SF-36 measures such as physical functioning or bodily pain. SF-36 scoring uses detailed formulae to calculate summary measures of physical and mental health that can obscure specific area responses.¹⁹

There are indications that at baseline the researchers administered the complete 116-question Medical Outcomes Study (MOS) Core Survey questionnaire upon which the SF-36 is based, because a baseline "Cognition" score is reported, a result not produced by the SF-36. No post treatment results of the 116-question MOS were reported.

This study did not report mean (average) changes in scores after treatment, highly unusual in a study of this type²⁰. While reporting

scores in terms of each individual participant's change can add an additional aspect to analysis, this should not replace reporting the group's mean change. Because mean change was not reported in this study, there cannot be an estimate of whether the difference in mean change between treated and placebo groups was significant. This is a basic statistical test for determining the significance of difference between two samples.

The researchers made arbitrary decisions about scoring. Any participant who withdrew from the study was recorded as having "Worse" health, regardless of the withdrawal reason. If a participant had a significant decline in either mental or physical health, their overall health was recorded as "Worse". (In other words, a "Worse" score on either mental or physical health trumped an "Improved" score on the other measure.) This scoring scheme could obscure a treatment improvement.

The small sample of participants in this study indicates a deficiency in **statistical power** (see "Supplemental Information"). Using a theoretical (and highly debatable) assumption that the SF-36, when scored according to the authors' criteria, is a perfect measure of improved health, an estimate of statistical power indicates how likely the sample of participants would produce the right conclusion. The authors do not make their own estimate of statistical power of the group actually studied. Instead, they state what they calculate the statistical power would have been if they had enrolled their full goal of 194 seropositive patients and 66 seronegative patients. As executed, the study only enrolled 70 seropositive and 45 seronegative patients so their goal of 90% power for the seropositive patients and 80% for the seronegative patients is substantially unrealized.

Independently estimating the study's actual statistical power requires some assumptions. If, in the population of participants meeting criteria for study, 25% improve on placebo and 50% improve on treatment, this study would have been expected to have a statistical power of 76% (Fisher Exact Test). In other words, we would expect testing this small of a group to produce the correct conclusion 76% of the time, using the gross assumption the SF-36, as utilized, was a perfect test. Stated another way, if the study

were conducted perfectly, with the antibiotic group getting all 90 days of treatment and the SF-36 perfectly identifying every improved patient, there is still a 24% chance the study would have come to the wrong conclusion, solely because the study did not have enough participants. (Participants were, in this study, only required to have 75% compliance with the medication protocol, another factor biased toward obscuring treatment success.)

Although somewhat theoretical, an argument can be made that this study's conclusion violates one of the most elementary tenets of experimental research. Basically, when two samples are evaluated, this is a test of the null hypothesis: that there is no difference between the two populations from which the samples were chosen. In this study, the researchers took a post-treatment sample of 64 from a theoretically unlimited population of antibiotic-treated patients meeting study inclusion criteria. They compared this to a sample of 65 from a theoretically unlimited population of placebo-treated patients meeting study inclusion criteria.

These two samples did not show a statistically significant difference based on highly summarized SF-36 scores. Basic statistical analysis should produce the conclusion that, "the two samples did not produce evidence the populations differed". Elementary hypothesis testing says this should never be construed as evidence that, "the two populations are the same," which is how this study has been presented. It is much more difficult to produce evidence that two samples come from populations that do not differ, or differ very little. For researchers to present this study as evidence that the course of antibiotics studied produces no changes in patients continuing to have symptoms is a scientifically fallacious statement.

In summary, there are several possible, or even probable, reasons the Klempner study did not show significant differences between placebo and treatment groups:

- The primary measurement tool, the SF-36, might not have enough sensitivity to detect treatment difference even if results were completely reported²¹
- Reducing the SF-36's eight scales (each scored 0-100) to two summary scores (ranges 20-58 for PCS and 17-62 for MCS in about 85% of the general population²²) further reduced sensitivity to sample differences
- Reducing the PCS and MCS summary scores to three values ("Improved", "Unchanged", or "Worse") further reduced sensitivity
- Combining the SF-36's two summary scores to a total score still further reduced sensitivity
- The small number of participants reduced likelihood of detecting a difference
- Selecting participants who had a history of already failing similar treatments reduced sensitivity
- Testing a treatment considered marginal, at best, for patients exhibiting these symptoms also reduced sensitivity

Bottom line

This study reported no post-treatment results other than the results of a generic health questionnaire, the SF-36, which is self-reported by the patient. The results were repeatedly summarized before analysis, losing sensitivity to detect change at each step. The treatment tested would rarely be used by a clinician treating a patient who had already failed one or more previous courses of antibiotics. The participants selected for study did not completely represent the population potentially needing additional antibiotics for resolving continuing symptoms of Lyme disease. In conclusion, this study presents little evidence supporting or refuting the effectiveness of additional antibiotic treatment for Lyme disease patients.

Supplemental Information

Double-blind Trials

This trial was conducted with the procedures generally used in clinical trials to prove the efficiencies and safety of new drugs. “Double-blind” indicates medication containers were labeled with a code known only to personnel not involved with drug administration or patient interaction. The code, when broken, indicates which containers had medication and which had placebo.

Herxheimer reaction

Lyme disease antimicrobial treatment has long been identified as producing a temporary worsening of symptoms known as a Jarisch-Herxheimer, or simply, **Herxheimer** reaction. This reaction has been extensively reported in other spirochetal diseases such as syphilis and relapsing fever and obviously complicates evaluation of treatment results.

Although the cause is somewhat speculative, this temporary worsening of symptoms or manifestation of new symptoms, is thought to be due to toxins released from bacterial cell walls during cell death caused by the antibiotics. The Herxheimer reaction can cause confusion for the less-experienced clinician or, for the more experienced and perceptive clinician could be used as a diagnostic tool. The time lapse between treatment initiation and indications of a Herxheimer reaction with Lyme disease is highly variable, anywhere from a few hours to a few weeks. The duration and intensity of a Herxheimer reaction is also highly variable.

Measuring health status during treatment cycles for patients with Lyme disease could present a confusing picture to researchers as effects of the Herxheimer reaction could wax and wane.

Statistical Power

Statistical power estimates how likely a study would produce the “right” conclusion based on the likelihood the sample of participants selected accurately represents the population as a whole. Several factors affect statistical power. For example, larger samples increase statistical power, an element which the researcher can control. Greater differences in the populations from which the samples are selected also increase statistical power, a factor the researcher cannot control.

The SF-36 Questionnaire

The SF-36 is probably the world's most popular quality of life questionnaire. Patients check a box in response to 36 questions. They have from three to five response choices, generally of the nature of "Excellent", "Very Good", "Good", "Fair", or "Poor". The responses are scored and the results are eight subscale scores in the range of 0-100:

Physical Functioning (PF)
Role-Physical (RF)
Bodily Pain (BP)
General Health (GH)
Vitality (VT)
Social Functioning (SF)
Role-Emotional (RE)
Mental Health (MH)

These eight subscale scores are then summarized into two summary scores, the Physical Component Summary (PCS) and Mental Component Summary (MCS) which are scored using fairly elaborate techniques to produce norm-based scores (bell-shaped distribution) with a mean of 50 and standard deviation of 10. In the general population about 85% of the population scores between 20 and 58 for PCS and 17 to 62 for the MCS.

The eight subscale scores are straightforward but the PCS and MCS are subjects of continuing debate as to how well they measure physical and mental health and what score changes mean clinically. John Ware and Mark Kosinski, primary refiners of the SF-36, best summarize how the SF-36 scores should be used, "Because of the potential for information loss with summary health measures, we encouraged those who use them to interpret their results in parallel with the profile of SF-36 subscales..."²⁰.

[John Ware and Mark Kosinski are executives at QualityMetrix Corporation, a private company that licenses use of the SF-36. Mark Kosinski is a coauthor of the Klempner study, making the questionable use of the SF-36 results especially surprising.]

Published Abstract of Study

Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease.

Klempner MS, Hu LT, Evans J, Schmid CH, Johnson GM, Trevino RP, Norton D, Levy L, Wall D, McCall J, Kosinski M, Weinstein A.

New England Medical Center and Tufts University School of Medicine, Boston, MA, USA.
klempner@bu.edu

BACKGROUND: It is controversial whether prolonged antibiotic treatment is effective for patients in whom symptoms persist after the recommended antibiotic treatment for acute Lyme disease. **METHODS:** We conducted two randomized trials: one in 78 patients who were seropositive for IgG antibodies to *Borrelia burgdorferi* at the time of enrollment and the other in 51 patients who were seronegative. The patients received either intravenous ceftriaxone, 2 g daily for 30 days, followed by oral doxycycline, 200 mg daily for 60 days, or matching intravenous and oral placebos. Each patient had well-documented, previously treated Lyme disease but had persistent musculoskeletal pain, neurocognitive symptoms, or dysesthesia, often associated with fatigue. The primary outcome measures were improvement on the physical- and mental-health-component summary scales of the Medical Outcomes Study 36-item Short-Form General Health Survey (SF-36)--a scale measuring the health-related quality of life--on day 180 of the study. **RESULTS:** After a planned interim analysis, the data and safety monitoring board recommended that the studies be discontinued because data from the first 107 patients indicated that it was highly unlikely that a significant difference in treatment efficacy between the groups would be observed with the planned full enrollment of 260 patients. Base-line assessments documented severe impairment in the patients' health-related quality of life. In intention-to-treat analyses, there were no significant differences in the outcomes with prolonged antibiotic treatment as compared with placebo. Among the seropositive patients who were treated with antibiotics, there was improvement in the score on the physical-component summary scale of the SF-36, the mental-component summary scale, or both in 37 percent, no change in 29 percent, and worsening in 34 percent; among seropositive patients receiving placebo, there was improvement in 40 percent, no change in 26 percent, and worsening in 34 percent ($P=0.96$ for the comparison between treatment groups). The results were similar for the seronegative patients. **CONCLUSIONS:** There is considerable impairment of health-related quality of life among patients with persistent symptoms despite previous antibiotic treatment for acute Lyme disease. However, in these two trials, treatment with intravenous and oral antibiotics for 90 days did not improve symptoms more than placebo.

References

1. Wormser, G.P., et al., *Practice guidelines for the treatment of Lyme disease. The Infectious Diseases Society of America*. Clin Infect Dis, 2000. **31 Suppl 1**: p. 1-14.
2. Barinaga, M., *Furor at Lyme disease conference*. Science, 1992. **256**(5062): p. 1384-5.
3. Steere, A.C., J.A. Hardin, and S.E. Malawista, *Erythema chronicum migrans and Lyme arthritis: cryoimmunoglobulins and clinical activity of skin and joints*. Science, 1977. **196**(4294): p. 1121-2.
4. *Case definitions for infectious conditions under public health surveillance*. MMWR Morb Mortal Wkly Rep, 1997. **46**(RR-10): p. 20-22.
5. Burgdorfer, W., et al., *Lyme disease-a tick-borne spirochetosis?* Science, 1982. **216**(4552): p. 1317-9.
6. Lipschutz, B., *Über eine seltene Erythemform (erythema cronicum migrans)*. Arch Dermatol Syph, 1913. **118**: p. 349-356.
7. Oksi, J., et al., *Borrelia burgdorferi detected by culture and PCR in clinical relapse of disseminated Lyme borreliosis*. Ann Med, 1999. **31**(3): p. 225-32.
8. Wormser, G.P., et al., *Improving the yield of blood cultures for patients with early Lyme disease*. J Clin Microbiol, 1998. **36**(1): p. 296-8.
9. Tylewska-Wierzbanowska, S. and T. Chmielewski, *Limitation of serological testing for Lyme borreliosis: evaluation of ELISA and western blot in comparison with PCR and culture methods*. Wien Klin Wochenschr, 2002. **114**(13-14): p. 601-5.
10. Salazar, J.C., et al., *Coevolution of markers of innate and adaptive immunity in skin and peripheral blood of patients with erythema migrans*. J Immunol, 2003. **171**(5): p. 2660-70.
11. Brunner, M. and L.H. Sigal, *Immune complexes from serum of patients with lyme disease contain Borrelia burgdorferi antigen and antigen-specific antibodies: potential use for improved testing*. J Infect Dis, 2000. **182**(2): p. 534-9.
12. Dumler, J.S., *Molecular diagnosis of Lyme disease: review and meta-analysis*. Mol Diagn, 2001. **6**(1): p. 1-11.
13. Bransfield, R., S. Brand, and V. Sherr, *Treatment of patients with persistent symptoms and a history of Lyme disease*. N Engl J Med, 2001. **345**(19): p. 1424-5.
14. Hobart, J.C., et al., *Quality of life measurement after stroke: uses and abuses of the SF-36*. Stroke, 2002. **33**(5): p. 1348-56.
15. Leong, K.P., et al., *Why generic and disease-specific quality-of-life instruments should be used together for the evaluation of patients with persistent allergic rhinitis*. Clin Exp Allergy, 2005. **35**(3): p. 288-98.
16. Klempner, M.S., R. Noring, and R.A. Rogers, *Invasion of human skin fibroblasts by the Lyme disease spirochete, Borrelia burgdorferi*. J Infect Dis, 1993. **167**(5): p. 1074-81.
17. Donta, S.T., *Macrolide therapy of chronic Lyme Disease*. Med Sci Monit, 2003. **9**(11): p. P1136-42.
18. Donta, S.T., *Late and chronic Lyme disease*. Med Clin North Am, 2002. **86**(2): p. 341-9, vii.
19. Ware, J.E., Kosinski M., Keller S.D., *Physical and Mental Health Summary Scales: a user's manual*. 1994, Boston: Health Assessment Lab.
20. Ware, J.E., Jr., et al., *Differences in 4-year health outcomes for elderly and poor, chronically ill patients treated in HMO and fee-for-service systems. Results from the Medical Outcomes Study*. Jama, 1996. **276**(13): p. 1039-47.
21. Hobart, J., et al., *The SF-36 in multiple sclerosis: why basic assumptions must be tested*. J Neurol Neurosurg Psychiatry, 2001. **71**(3): p. 363-70.
22. Ware, J.E. and M. Kosinski, *Interpreting SF-36 summary health measures: a response*. Qual Life Res, 2001. **10**(5): p. 405-13; discussion 415-20.