Contemporary Clinical Trials 33 (2012) 1132–1142

Contents lists available at SciVerse ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Antibiotic retreatment of Lyme disease in patients with persistent symptoms: A biostatistical review of randomized, placebo-controlled, clinical trials

Allison K. DeLong [a,*], Barbara Blossom [b], Elizabeth L. Maloney [c], Steven E. Phillips [d]

[a] Center for Statistical Sciences, Department of Biostatistics, Brown University, Providence, RI, USA
[b] Department of Statistics, Colorado State University, Fort Collins, CO, USA
[c] Partnership for Healing and Health, Ltd., Wyoming, MN, USA
[d] Greenwich Hospital, Greenwich, CT, USA

ARTICLE INFO

Article history:
Received 4 May 2012
Revised 16 July 2012
Accepted 14 August 2012
Available online 19 August 2012

Keywords:
Lyme disease
Neuroborreliosis
Randomized controlled trial
Statistical power
Minimum clinically important difference (MCID)
Treatment guidelines

ABSTRACT

Introduction: Lyme disease (Lyme borreliosis) is caused by the tick-borne spirochete *Borrelia burgdorferi*. Long-term persistent illness following antibiotic treatment is not uncommon, particularly when treatment is delayed. Current treatment guidelines for persistent disease primarily rely on findings from four randomized, controlled trials (RCTs), strongly advising against retreatment.

Methods: We performed a biostatistical review of all published RCTs evaluating antibiotic retreatment, focusing on trial design, analysis and conclusions.

Results: Four RCTs met the inclusion criteria; all examined the efficacy of intravenous ceftriaxone versus placebo at approximately 3 or 6 months. Design assumptions for the primary outcomes in the two Klempner trials and two outcomes in the Krupp trial were unrealistic and the trials were likely underpowered to detect clinically meaningful treatment effects. The Klempner trials were analyzed using inefficient statistical methods. The Krupp RCT was well-designed and analyzed for fatigue, finding statistically significant and clinically meaningful improvement. Fallon corroborated this finding. Fallon also found improvement in cognitive functioning, a primary outcome, at 12 weeks which was not sustained at 24 weeks; improvements in physical functioning and pain were demonstrated at week 24 as an interaction effect between treatment and baseline symptom severity with the drug effect increasing with higher baseline impairment.

Discussion: This biostatistical review reveals that retreatment can be beneficial. Primary outcomes originally reported as statistically insignificant were likely underpowered. The positive treatment effects of ceftriaxone are encouraging and consistent with continued infection, a hypothesis deserving additional study. Additional studies of persistent infection and antibiotic treatment are warranted.

1. Introduction

Reporting bias in clinical trials, particularly with respect to publishing bias toward significant findings [1,2] and interpretive "spin" to overemphasize a possible benefit while de-emphasizing non-significant findings [3] is receiving increased attention within the statistical and medical communities. A variation on interpretive bias deserves concern as well, namely the interpretation of statistically insignificant findings from small, underpowered, or poorly executed clinical trials as evidence of treatment inefficacy. Such trials may lead to the premature and erroneous conclusion that the treatment is ineffective, constituting a type II error. Concerns about such

* Corresponding author at: Center for Statistical Sciences, Department of Community Health, Brown University, Providence, RI 02912, USA. Tel.: +1 401 863 9697; fax: +1 401 863 9182.
E-mail address: adelong@stat.brown.edu (A.K. DeLong).

1551-7144/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cct.2012.08.009

errors may arise when disagreement and uncertainty exists in the medical community, as is the case with Lyme disease (Lyme borreliosis).

Lyme disease, caused by the tick-borne spirochete *Borrelia burgdorferi* sensu lato, is classified as an emerging infectious disease by the U.S. National Institute of Allergy and Infectious Diseases (NIAID) due to the relatively recent discovery of its causal agent (1982) [4] and its rapidly increasing incidence over the last two decades in the U.S. [5] and much of Europe [6]. The infection is multi-systemic, resulting in diverse physical and neuro-psychiatric symptoms and manifestations and causing mild to severe disease [7–13]. Although many patients respond to antibiotic treatment regimens of 2 to 4 week duration [9], it is well recognized that long-term persistent illness can occur following a 30-day course of treatment, particularly when treatment is delayed [7,9,14,15]. Multiple randomized trials found significant morbidity in their study populations, similar to that of multiple sclerosis or congestive heart failure. Although the trials employed different entrance criteria, none required this degree of physical disability as a condition of enrollment [16,17].

The management of patients with ongoing debilitating symptoms following antibiotic treatment for Lyme disease has generated debate within the medical community. The primary questions concern whether or not infection persists after standard antibiotic treatment and whether additional antibiotic treatment is of benefit [18,19]. Until a sensitive laboratory test for active infection is clinically available, clinical trials evaluating retreatment in persistently symptomatic Lyme disease patients provide the cornerstone of treatment guideline recommendations. Most guidelines for the diagnosis and management of Lyme disease [20–23] direct clinicians to limit the duration of antibiotic treatment, even in cases where ongoing symptoms compatible with a *B. burgdorferi* infection are present. These publications base their recommendations on a similar interpretation of the four randomized, blinded, placebo-controlled antibiotic retreatment trials funded by the U.S. National Institutes of Health (NIH) for patients with ongoing symptoms following standard Lyme disease treatment [16,17,24].

For this reason, a rigorous, independent evaluation of the findings from these trials is needed. The present study is a biostatistical review of the four NIH-funded clinical trials. By focusing on the trial design and analyses of primary and secondary outcomes in each trial, the review demonstrates weaknesses which limit the ability to draw strong conclusions regarding retreatment. This review will likely be of broad interest to medical practitioners, researchers, medical ethicists, and treatment guideline developers in Europe and North America.

2. Methods

The four NIH-funded Lyme disease retreatment trials were initially selected for evaluation in January 2009 through a review of current Lyme disease treatment guidelines, which identify these trials as the only published RCTs relevant to the question of retreatment [21,22]. To ensure that other relevant RCTs to date were not missed, a Cochrane Library search of the published literature was conducted on September 10, 2010, setting the limits of study type to “clinical trial” and requiring the use of “Lyme” or “Borrelia” in the title, abstract or in the manuscript's keywords. Additional studies were sought by searching ClinicalTrials.gov, a registry of both federally and privately funded clinical trials. The title and abstract of each selected publication were read by two authors (AKD and BB) and coded as a clinical trial and if it was a clinical trial evaluating retreatment of Lyme disease patients with persistent symptoms despite receipt of a standard course of antibiotics. The full text of all articles evaluating retreatment was read by all authors and eligibility was determined by consensus. All primary and secondary outcomes were tabulated for each clinical trial, including, where possible, the treatment effect and 95% confidence interval (CI) overall and by trial arm.

A review was conducted of each trial's design, execution, statistical analysis and conclusions. For trial design, attention was paid to the enrolled patient population, the definitions and measurements of primary and secondary outcomes, and the definition of clinically meaningful changes in those outcomes which determine power of the sample sizes to detect clinically meaningful treatment effects. For trial execution, patient dropout, masking of study medication, and interim analyses were considered. We evaluated the appropriateness of the statistical method chosen to estimate the treatment effect and the handling of patient dropouts. Since our objective is to place the findings from these trials within the current framework of Lyme disease as of 2012, the present review is also informed by research conducted after the retreatment trials were designed, executed, and/or published. Three important statistical concepts are used throughout the review: statistical power, interim analysis and stopping rules, and non-inferiority trials.

2.1. Statistical power

When designing a clinical trial, the sample size can only be calculated after researchers determine an appropriate and plausible design treatment effect δ, which is a hypothetical value of the effect of the treatment under investigation. In addition to selecting δ, trial design also requires an acceptable probability of declaring treatment effectiveness if δ is true (i.e. power, typically 80–90%). For a fixed power, a smaller δ would necessitate a study design with a larger sample size, and vice versa. Ideally δ should correspond to the minimum clinically important difference (MCID) for the disease and outcome measure studied. If the *true underlying treatment effect* is greater than the MCID, yet less than the design treatment effect δ, then the study is underpowered with an insufficient sample size, and thus inadequately designed to meet its stated goals, and the power may be far less than the nominal value set in the trial design. Such studies are likely to conclude an insignificant result although a true, clinically relevant treatment effect exists. Although MCID values are context-specific and difficult to ascertain, reasonable estimates are identified based on published knowledge of the disease studied or, when disease-specific data are not available, of studies of other similar diseases [25].

2.2. Interim analyses and stopping rules

Interim analyses are commonly used to gauge the success of a clinical trial, by analyzing outcome data at pre-defined points during the study instead of waiting until all patients

have completed follow up. An interim analysis can trigger one of three possible actions: (1) conclude that the treatment is effective and stop the trial early, (2) continue the trial until the next interim 'look', and (3) stop the trial early for "futility". If action (1) is triggered, trial findings can be published and disseminated quickly and effective treatments can be provided to patients sooner. Action (3) implies that at the study terminus, the authors will most likely fail to reject the null hypothesis that the outcomes in the two arms are the same. This action is often triggered when the designed sample size is too small to detect the true treatment effect, which may occur as a result of underestimation of patient variability in the study design, use of an unrealistically large design δ (greater than the MCID), or because the treatment is, indeed, ineffective. Many have argued that conducting underpowered trials is unethical; therefore stopping such trials is desirable. Stopping a trial for statistical insignificance or futility does not necessarily indicate treatment is ineffective and it would be incorrect to conclude that this was the case.

2.3. Non-superiority trials

To examine whether a treatment is ineffective, statistical tests using non-superiority hypotheses are required. In such trials, the null hypothesis is that the treatments differ, with rejection of the null hypothesis indicating that the treatment effects in the two arms are similar, i.e. the difference lies within a certain small but acceptable window. None of the Lyme disease retreatment trials was designed as a non-superiority trial. However, if 95% confidence intervals (CIs) on the treatment effects exclude and are below the MCID, then the trial has essentially shown the treatment to be ineffective.

3. Results

The literature search found 105 clinical trials using the word "Lyme" or "Borrelia" in the title, abstract or keyword (Fig. 1). Of these, 100 papers were eliminated from consideration for the following reasons: did not assess antibiotic efficacy (49); evaluated antibiotic prophylaxis after a tick bite (4); evaluated first-line antibiotic treatment of early or late Lyme disease (39), including a study evaluating longer-term treatment which enrolled patients with and without a history of prior treatment [26]; evaluated treatment of coinfection of Lyme disease and babesiosis (1); and involved treatment of relapsing fever (7). The full text of the remaining 5 publications was read. One clinical trial was excluded because it did not present an intention-to-treat analysis of primary outcomes due to an excessive dropout rate in the placebo arm [27]. Klempner et al. [16] presented two primary and one secondary outcome from two trials which enrolled patients from two different populations. Kaplan et al. [28] presented an analysis of several additional secondary outcomes from the Klempner trials. Henceforth, these trials are collectively referred to as the Klempner trials. The publications by Krupp et al. [24] and Fallon et al. [17] present primary and secondary outcomes from two additional clinical trials. As a result, the primary outcomes from four clinical trials were presented in three publications.

Participants in all four trials had a confirmed history of Lyme disease for which they received at least one standard course of antibiotic therapy, and had persistent symptoms thought to be consistent with Lyme disease beginning at or within 6 months of disease onset, with symptoms persisting at least 4 months following the cessation of therapy. The studies enrolled different subpopulations of patients with persistent symptoms, but all examined intravenous (IV) ceftriaxone for a minimum of 4 weeks and evaluated various primary and secondary treatment effects at approximately 3 and/or 6 months as described (Table 1).

3.1. Klempner et al. trials [16]

3.1.1. Trial summary

Klempner et al. conducted two multicenter trials; the designs differed only in that one enrolled IgG-seropositive and the other IgG-seronegative patients. Patients received either IV placebo followed by 2 months of oral placebo or 1 month of IV ceftriaxone followed by 2 months of oral doxycycline. Clinical inclusion criteria were broad, including any of: widespread musculoskeletal pain, cognitive impairment, radicular pain, and paresthesias that interfered with functioning per patient self-report. The primary outcomes were changes in SF-36 summary scores, which are commonly used subjective measures of health-related quality of life (HRQoL). The SF-36 physical component summary (PCS) and mental component summary (MCS) scores represent numeric composites of eight subcategories, scaled such that the means and standard deviations (SD) for the general U.S. population are 50 and 10 respectively with lower scores representing poorer health.

Klempner classified patients as "improved," "worsened," or "the same" based on changes in their summary scores from baseline to the 180-day evaluation. Positive and negative cutoffs for classification were set at 6.5 units for the SF-36 PCS and 7.9 for the SF-36 MCS; these values represent twice the standard error of measurement (SEM). A chi-square test of proportions was used to evaluate the treatment effect, which was taken to be whether the proportion of patients in each class differed by treatment arm. For sample size estimation, the researchers set the design treatment effects to be 25% and 35% for the difference in percent improved in the seropositive and seronegative trials, respectively. The calculated sample sizes were 194 participants in the seropositive trial and 66 in the seronegative trial. Interim analyses using O'Brien-Fleming boundaries were performed after 107 of 260 (41%) planned participants in both trials combined completed follow-up, and the trials were stopped for futility. No statistically significant treatment benefit was reported for either trial. The authors concluded that the trial regimen did not result in a significant treatment effect and also stated that other antibiotic regimens were unlikely to result in a different finding.

3.1.2. Trial critique

3.1.2.1. Design. In order to evaluate the Klempner trials' design in light of all available evidence, a literature search for studies that determined MCIDs for SF-36 summary scores (PCS and MCS) was conducted. We were unable to find any studies evaluating MCIDs for the SF-36 in patients with Lyme disease. Studies evaluating MCIDs in patients with other chronic illnesses causing a level of disability similar to that of

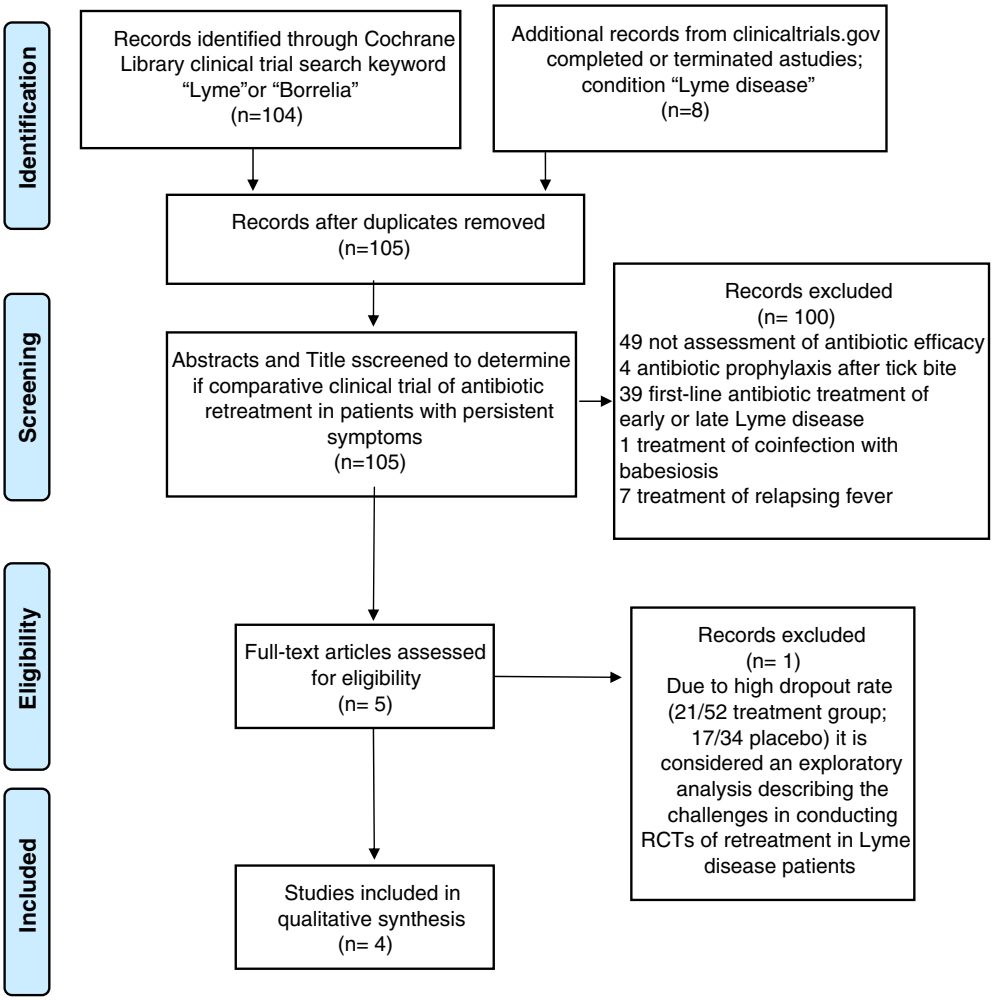

Fig. 1. Flow diagram of the literature search for randomized, controlled trials evaluating antibiotic retreatment in Lyme disease patients with persistent symptoms following a standard course of treatment.

the Klempner subjects were published after the Klempner trials were conducted. These studies identified clinically meaningful changes on the SF-36 summary scores to be in the range of 2 to 5 points (Table 2) [29–33]. Changes of this magnitude align with the SF-36 developers' recommendations [34,35] and with studies identifying 1 SEM or 0.5 standard deviation in baseline scores as appropriate statistical benchmarks of clinical relevance for health-related quality of life (HRQoL) measures including the SF-36 [36–38].

The Klempner trial design assumed that a δ of absolute 25% or 35% difference between arms in the percent improved would correspond to a valid threshold for clinically relevant treatment effects. Since observed changes in SF-36 outcomes were not reported in the manuscript, we mapped Klempner's δs to their corresponding δ^* on the continuous SF-36 scale as follows. Let s_i be the observed 6-month treatment effect in the PCS for participant i in the placebo arm and let t_j be the observed 6-month treatment effect for participant j in the antibiotic arm and assume $s \sim N(\mu,\sigma)$ and $t \sim N(\mu+\delta^*, \sigma)$. We can use the observed percentages (quantiles) of patients classified as having "improved", "stayed the same", and "worsened" in the placebo arm to estimate μ and σ. The expected difference in the "% improved" on the PCS for pertinent values of δ^* can be estimated as $\Pr(t>6.5)-\Pr(s>6.5)$, using the estimates $\hat{\sigma}$ and $\hat{\mu}$. A similar calculation can be conducted for the MCS, with δ^* estimated as $\Pr(t>7.9)-\Pr(s>7.9)$.

A $\delta = 25\%$ corresponds to mean differences in SF-36 scores between the two arms of 6.7 and 9.1 points on the PCS and MCS, respectively, and a $\delta = 35\%$ corresponds to 9.3 and 12.8 points (Table 3). Thus the trials, as designed, called for treatment effects considerably larger than the 2 to 5 point MCIDs identified in other chronic illnesses, suggesting the sample sizes were inadequate and the trials were very likely underpowered to detect the true underlying MCIDs. The importance of this finding becomes clear when one considers the scale of the SF-36 instrument. For example in the antibiotic arm of the seronegative trial, adding the estimated treatment effect of 12.8 points on SF-36 MCS to the baseline mean MCS score of 46.7 points would require the average participant to achieve a score essentially one standard deviation (SD) above

Table 1
Available measures of treatment effects for each trial and outcome.

Trial	Measurement	Outcome	Primary or secondary outcome	Meas. time months	Effect or "Success" rate by arm		Treatment effect
Klempner et al. Seronegative [16]: Antibiotic (n = 25), Placebo (n = 26)					Placebo	Antibiotic	Effect
	SF-36 physical component summary (PCS)[a]	Success = change in PCS from baseline to 180 days > 6.5	Primary	6	5/23 (22%)	9/22 (41%)	19 (−7 to 46)
	SF-36 mental component summary (MCS)[a]	Success = change in MCS from baseline to 180 days > 7.9	Primary	6	6/23 (26%)	8/22 (36%)	10 (−17 to 37)
	Fibromyalgia impact questionnaire[b]	Success > 25% improvement from baseline	Secondary	6	–	–	NS
	Medical outcome study symptom checklist[d]	Pain, cognitive functioning, performance of daily activities	Secondary	3 and 6	–	–	NS
	Neuropsychological tests[d]	Common battery	Secondary	3 and 6	–	–	NS
	Mood[d]	BDI and MMPI-2	Secondary	3 and 6	–	–	NS
Klempner et al. Seropositive [16]: Antibiotic (n = 39), Placebo (n = 39)					Placebo	Antibiotic	Effect
	SF-36 physical component summary (PCS)[a]	Success = change in PCS from baseline to 180 days > 6.5	Primary	6	10/35 (29%)	11/35 (31%)	3 (−19 to 24)
	SF-36 mental component summary (MCS)[a]	Success = change in MCS from baseline to 180 days > 7.9	Primary	6	16/35 (46%)	11/35 (31%)	−14 (−37 to 8)
	Fibromyalgia impact questionnaire[b]	Success > 25% improvement from baseline	Secondary	6	–	–	NS
	Medical outcome study symptom checklist[d]	Pain, cognitive functioning, performance of daily activities	Secondary	3 and 6	–	–	NS
	Neuropsychological tests[d]	Common battery	Secondary	3 and 6	–	–	NS
	Mood[d]	BDI and MMPI-2	Secondary	3 and 6	–	–	NS
Krupp et al. [24]: Antibiotic (n = 28), Placebo (n = 27)					Placebo	Antibiotic	Effect
	Fatigue severity scale (FSS-11)	Success = Improvement of > 0.7 points from baseline	Primary	6	5/22 (23%)	18/26 (69%)	p < 0.01
	Alphabet arithmetic test	Success = improvement > 25% from baseline	Primary	6	2/22 (9%)	2/26 (8%)	p = 0.99
	Osp A antigen to Borrelia Burgdorferi	Success = clearance of Osp A antigen from baseline	Primary	6	4/4 (100%)	3/4 (75%)	p = 1.0
Fallon et al. [17]: Antibiotic (n = 23), Placebo (n = 14)					Placebo[c]	Antibiotic[c]	Effect[c]
	Multivariate outcome measured across 6 cognitive domains[a]	Standardized to represent z-scores	Primary (efficacy)	3	0.16 (−0.6, 0.38)	0.43 (0.27, 0.61)	0.28 (−0.01, 0.56) p = 0.05
			Primary (durability)	6	0.31 (0.09, 0.53)	0.35 (0.18, 0.53)	0.04 (−0.24, 0.33) p = 0.76
	Fatigue severity scale (FSS-11)[b]	Continuous measure, interaction with baseline score	Secondary	3	−0.2 (−1, 0.6)	−1.3 (−1.9, −0.7)	−1.1 (−2.1, −0.1) p < 0.05
				6	−0.4 (−1.4, 0.6)	−1.1 (−1.7, −0.5)	−0.7 (−1.8, 0.4)
	Fatigue (FSS-11)	Krupp et al. analysis	Secondary		25%	67%	p = 0.05

Pain (McGill) VAS[b]	Continuous measure, interaction with baseline score	Secondary	3	−1.6 (−3.2, 0)	−3.6 (−5, −2.2)	−2 (−4.1, 0.1) p<0.05
			6	−0.8 (−2.6, 1)	−2.7 (−4.1, −1.3)	−1.9 (−4.1, 0.3) p<0.05
Total pain[b]	Continuous measure	Secondary	3	−5.3 (−8.6, −2)	−6.7 (−9.6, −3.8)	−1.4 (−5.8, 3)
			6	−6.4 (−9.7, −3.1)	−7.7 (−10.6, −4.8)	−1.3 (−5.7, 3.1)
SF-36 physical component summary (PCS)[a]	Continuous measure, interaction with baseline score. Significant without interaction	Secondary	3	1.2 (−2.3, 4.7)	5.9 (2.6, 9.2)	4.7 (−0.2, 9.6) p<0.05
			6	2.2 (−1.5, 5.9)	6.9 (3.6, 10.2)	4.7 (−0.3, 9.7) p<0.05
SF-36 mental component summary (MCS)[a]	Continuous measure	Secondary	3	8.8 (3.7, 13.9)	7.2 (3.3, 11.1)	−1.6 (−8, 4.8)
			6	8.1 (2.8, 13.4)	6.5 (2.6, 10.4)	−1.6 (−8.2, 5)
# joints with pain on exam[b]	Continuous measure	Secondary	3	−1.2 (−3.7, 1.3)	−2.9 (−4.7, −1.1)	−1.7 (−4.8, 1.4)
			6	−3.8 (−5.6, −2)	−2.7 (−4.3, −1.1)	1.1 (−1.3, 3.5)
Depression (Beck)[b]	Continuous measure	Secondary	3	−3.9 (−7, −0.8)	−2.5 (−5, 0)	1.4 (−2.6, 5.4)
			6	−3.9 (−7, −0.8)	−2.5 (−5, 0)	1.4 (−2.6, 5.4)
Anxiety (Zung)[b]	Continuous measure	Secondary	3	−5.3 (−8.8, −1.8)	−3.9 (−6.8, −1)	1.4 (−3.2, 6)
			6	−6.3 (−9.8, −2.8)	−5 (−7.9, −2.1)	1.3 (−3.3, 5.9)
Global Psycho-Pathology (GSI SCL-90)[b]	Continuous measure, interaction with baseline score	Secondary	3	−3.6 (−9.7, 2.5)	−7.6 (−11.7, −3.5)	−4 (−11.3, 3.3)
			6	−5.1 (−12, 1.8)	−7.7 (−12, −3.4)	−2.6 (−10.7, 5.5)

NS effect not given, reported as not statistically significant.
"–"Within-arm effects were not reported for each trial.
[a] Higher score implies better health.
[b] Higher score implies worse health.
[c] All secondary outcomes presented as the mean of participants with worse baseline scores (75th percentile), estimated using values from the manuscript. Some estimated confidence intervals for statistically significant effects cross zero because data were available only to 1 decimal place.
[d] Presented in Kaplan et al. [28].

the mean score for the U.S. general population. As such, the chosen design treatment effects were unrealistic. Additionally, treatment effects of 2 to 5 points correspond to expected differences in the percent "improved" of 7 to 18% on the PCS (Table 3). These differences are within the 95% confidence intervals for both trials, indicating the trials did not show the treatment to be ineffective.

Our estimated standard deviations, $\hat{\sigma}$, defined in an earlier paragraph, were 10.1 for the PCS and 14.2 for the MCS. Although one may speculate that this large variability is due to a "placebo effect", this terminology should be used cautiously; other possible explanations for the large standard deviation could be regression to the mean or higher variability in chronically ill populations. Ware et al. [34], who used the same cutoffs as Klempner et al. in their evaluation of the SF-36 in chronically ill patients, called attention to the fact that outcome variations in their categorical analysis were substantially larger than expected.

Using these values for $\hat{\sigma}$, the sample sizes required per arm to have 80% power to detect average treatment effects from 2 to 5 points using a *t*-test and a two-sided alpha of 0.05 are 66 and 128 patients for an assumed MCID of 5 points on the PCS and MCS, respectively and 179 and 354 patients for an assumed MCID of 3 points. Detecting a treatment effect of 2 points on the PCS would require about 400 patients per arm and detecting differences of 2 points on the MCS would require about 800 participants per arm. An analysis incorporating repeated, longitudinal measurements per participant would require a smaller number of participants.

3.1.2.2. Analysis. The trials' use of a chi-square test on categorized, continuous data collected at four study time points is not an efficient use of data, yet its use can provide unbiased results in certain circumstances. If the study design is simple, if missing outcomes are non-informative (i.e. missing completely at random), and if randomization is successful in balancing patient arms by pertinent characteristics, the chi-square test can be used. The Klempner studies did not meet these criteria. The trials were multicenter, the observed baseline outcomes differed by treatment arm, and authors provided insufficient information about patient dropout to determine whether or not missing outcomes were uninformative; therefore, in this setting a chi-square test is not recommended and its use may have produced biased results. Lastly, combining the data from the two trials is not valid without accounting for the stratification of patients in the analysis, which was not done using a chi-square test.

The analysis of the secondary outcomes is compromised in the same manner as the primary outcomes. In addition, the results for the secondary outcomes were not presented by trial. Instead, the seronegative and seropositive patients were combined, disregarding the fact that these were designed as two distinct trials (ClinicalTrials.gov Identifier No. NCT00001101 and NCT00000938).

3.1.2.3. Interpretation. We found that the Klempner trials were designed using excessive treatment effect sizes (much greater than minimum clinically meaningful) making it likely that the trials were underpowered to detect MCIDs. Although the trials had adequate power to detect the large changes in SF-36 scores used for outcome categorization (equal to 2*SEM for the general population), it is important to note that while 2*SEM is an appropriate benchmark to ensure statistical significance for an individual, it is not necessarily the appropriate cutoff to identify clinically meaningful and statistically significant differences at a group level. Thus it is not surprising that the MCIDs for diseases causing similar levels of disability, discussed above, and to which δ should correspond, are less than 2*SEM.

The authors noted in the discussion that their antibiotic regimen did not lead to improved outcomes, and, given the "*in vitro and in vivo activity of both of these antibiotics against B. burgdorferi*" and experience with other chronic infections, they concluded that it was unlikely that other antibiotic regimes would be useful. These trials do not support such a broad statement. Based on our findings, we conclude that the Klempner trials are uninformative with regards to the potential benefits of antibiotic retreatment utilizing 1 month of ceftriaxone followed by 2 months of doxycycline (or any other regimen) in patients with persistent symptoms of Lyme disease.

3.2. Krupp et al. STOP-LD trial [24]

3.2.1. Trial summary

Krupp et al. enrolled 55 patients with a history of Lyme disease and ongoing symptoms of severe fatigue validated by a Fatigue Severity Scale (FSS-11) score $\geq$4.0. Patients were randomized to receive 4 weeks of IV ceftriaxone versus placebo, and three primary outcomes were evaluated: fatigue measured by the FSS-11, mental speed using an alphabet arithmetic (A-A) test, and clearance of outer surface protein A (OspA) from the CSF. At 6 months follow-up, the authors found a significant treatment effect on fatigue. Clinical improvement, defined as a decrease $\geq$0.7 FSS-11 points, was seen in 18.5% on placebo versus 64% on ceftriaxone ($p<0.01$). Treatment effects on the other two primary outcomes were not statistically significant. The authors noted six significant adverse events. Four were serious; three of these involved sepsis in placebo subjects while the fourth was anaphylaxis in a ceftriaxone subject. The other two events were minor allergic reactions in ceftriaxone subjects.

Krupp et al. concluded that their findings did not support antibiotic retreatment. The authors noted the positive effect on fatigue but thought it may be due to unmasking of the study medication. They also concluded that the beneficial effect on fatigue was outweighed by the lack of effect on the other primary endpoints and the high number of adverse events.

3.2.2. Trial critique

3.2.2.1. Design. The trial was well-designed for the primary endpoint of fatigue, with clearly defined inclusion criteria and 80% power. However, it was inadequately designed with regard to mental processing speed. The authors defined a clinically meaningful change in the mental speed outcome as a 25% improvement on the A-A test, and designed their study with low (74%) power to detect a $\delta = 25\%$ difference in the percent improved between the arms. An earlier study by the same author [39] found that patients with a history of Lyme disease and continued fatigue or cognitive symptoms had an overall deficit of less than 25% on 7 of the 8 measures comprising the A-A test when compared with matched

Table 2
SF-36 summary score changes found to be clinically and statistically significant for chronic diseases of similar severity to Lyme disease.

Reference	Disease	Increase in PCS	Increase in MCS	Verification of clinical significance
Kosinski et al. [29]	Rheumatoid arthritis	4.4, 4.3, 3.0, 2.6, 3.2	4.7, 3.1, 2.2, 3.1, 2.3	1 level of improvement across five clinical RA measures[a]
Angst et al. [30]	Osteoarthritis	2	‡	Improvement in global health self-assessments
Coteur et al. [31]	Crohn's disease	4.1	3.9	IBDQ improvement[b]
Regensteiner et al. [32]	Peripheral artery disease	2	§	Increased maximum treadmill walking distance
Okamoto et al. [33]	Asthma	≤5.0	§	Increased FEV_1[c]

‡ Not determined; § Not significant.
[a] Values presented in order: Patient global assessment, physician global assessment, pain assessment, joint swelling, and joint tenderness.
[b] Inflammatory Bowel Disease Questionnaire (authors considered this the "best" MCID estimate among several clinical measures in this study because it correlated most closely with SF-36 scores).
[c] Forced expiratory volume in 1 s.

healthy controls (Table 4). Cognitive impairment was not an entrance criterion in the STOP-LD study and the authors noted that participants had only mild deficits in baseline processing speed. Therefore, the expected 25% increase in speed may have required the average STOP-LD subject to perform better than a matched healthy control. Coupled with the low power, this expectation renders the insignificant treatment effect on mental processing speed uninformative.

The third primary endpoint, clearance of OspA antigen from the CSF, was an experimental laboratory marker of treatment outcome. Previous studies documented the presence of OspA in the CSF of some Lyme disease patients [40]; the investigators were attempting to determine if its absence, post-treatment, could be used a surrogate marker of treatment success. The prevalence of OspA in the CSF of Lyme disease patients is unknown. Only 16% of the Krupp subjects were positive for the OspA antigen at baseline (Table 1), making its clearance an unsuitable surrogate of treatment outcome and the lack of a positive effect here is uninformative.

3.2.2.2. Analysis. With regard to fatigue, the authors performed a careful sensitivity analysis of loss to follow-up, demonstrating that the finding on fatigue was robust to patient dropout. After adjustment for baseline measures of psychiatric disorder, depressive symptoms, pain and age, the treatment benefit on fatigue remained significant.

Krupp et al. suggested that the finding of improved fatigue may have been biased due to unmasking of the study medication. This suggestion was based on their observation that the proportion of participants correctly guessing treatment assignment at 1 and 6 months was significantly higher in the antibiotic arm ($p<0.05$). This observation alone, however, is not indicative of unmasking. Consider an example in which patients were randomly assigned to a treatment or placebo group, and then guessed with equal probability of 0.8 in both arms that they were receiving treatment. In such a case, 80% of patients on treatment and only 20% on placebo would be expected to correctly guess their treatment assignment, yet masking was not corrupted. Instead of comparing the proportion in each arm that correctly guessed assignments, Krupp et al. should have compared the proportions that guessed they were on active therapy. In the STOP-LD trial, this proportion did not differ by arm at 1 month (57% placebo, 71% antibiotic, $p=0.37$, Fisher exact test) or at 6 months (68% and 69%, $p=1.0$). Therefore, there is no evidence demonstrating that masking was compromised.

3.2.2.3. Interpretation. The benefits of retreatment were significant and clearly demonstrated for fatigue, the sole outcome for which the study was properly designed and analyzed; the authors' suggestion that this positive finding was due to unmasking is unfounded.

Aspects of the trial's design with regard to the clearance of OspA from the CSF and improvements in mental processing speed made it unlikely that a positive treatment effect on these endpoints would be found. Thus, the lack of demonstrable benefits on these endpoints is uninformative and the

Table 3
Estimated differences in the proportion of patients expected to be classified as improved using Klempner et al.'s categorization for various mean treatment effects consistent with published MCIDs. Klempner et al.'s results are provided and confirm clinically meaningful mean differences of 2 to 5 points fall within Klempner et al.'s 95% confidence intervals [16].

PCS (physical component)		MCS (mental component)	
Expected treatment effect (δ*)	Difference in % improved (treatment vs. placebo)	Expected treatment effect (δ*)	Difference in % improved (treatment vs. placebo)
2	7%	2	5%
3	10%	3	8%
4	14%	4	10%
5	18%	5	13%
6.7	25%	9.1	25%
9.3	35%	12.8	35%
PCS-observed results (95% CI)		MCS-observed results (95% CI)	
[a]Seropositive trial	3% (−19 to 24%)	[a]Seropositive trial	−14% (−37 to 8%)
[a]Seronegative trial	19% (−7 to 46%)	[a]Seronegative trial	10% (−17 to 37%)

[a] Actual results reported by Klempner et al.

Table 4
Mean response times of Lyme patients and controls on the Alphabet Arithmetic test (Pollina et al., Table 3) [39] and the differences in the two groups presented as the percentage faster that healthy participants completed the task compared to the Lyme patients.

Question type	Lyme patients (msec)	Healthy participants[2] (msec)	% faster for healthy participants vs. Lyme patients
Letter match (true)	1012	896	11.5%
AA + 2 (true)	3022	2256	25.3%
AA + 3 (true)	3631	2813	22.5%
AA + 4 (true)	4180	3256	22.1%
Letter match (false)	1088	990	9.0%
AA + 2 (false)	3572	2696	24.5%
AA + 3 (false)	4074	3178	22.0%
AA + 4 (false)	4324	3588	17.0%

[1] In the STOP-LD study design, Krupp et al. [24] assumed a 25% improvement as the MCID.
[2] Age- and education-matched controls.

authors were wrong to recommend against retreatment on this basis.

Given the clear benefit on severe fatigue, the uninformative findings on OspA clearance and mental processing speed, and despite the potential for significant antibiotic-associated adverse events, we conclude the trial by Krupp et al. demonstrates that retreatment with ceftriaxone may be helpful for patients with ongoing severe fatigue after a standard course of Lyme disease treatment.

3.3. Fallon et al. trial [17]

3.3.1. Trial summary

The Fallon et al. trial enrolled subjects who had memory impairment on subjective and objective assessment tools (Wechsler Memory Scale-III) despite having previously received a minimum of 3 weeks of IV ceftriaxone; IgG seropositivity on entrance was an inclusion criterion. Thirty-seven subjects were randomly assigned 2-to-1 to receive 10 weeks of ceftriaxone or placebo. The primary outcome was cognitive change over time as measured across six domains to assess multiple aspects of cognition, with memory being the domain hypothesized as showing greatest change. The 3-month outcome measured treatment efficacy and the 6-month outcome measured treatment durability. Secondary measures included the SF-36 PCS and MCS scores, fatigue (FSS-11), pain (VAS), depression (Beck), anxiety (Zung), and global psycho-pathology (GSI SCL-90). For the primary outcome, a healthy control group was also enrolled. Fallon found improvement in cognitive functioning at 12 weeks, with a significance level of 0.053 that falls just above the margin of significance demonstrating treatment efficacy; however, it was unsustained at 24 weeks. Among the secondary outcomes, none of the psychiatric or mental outcomes were significant. However, there was a significant interaction effect between treatment and baseline scores, confirming that those with worse baseline scores had sustained improvement in the physical component score (SF-36 PCS) and decreases in VAS pain score to 24 weeks. In addition, a post hoc analysis of the subgroup meeting Krupp's STOP-LD enrollment criteria and using the same definition for a positive treatment response on the FSS-11 as Krupp found that retreatment was beneficial (66.7% in the ceftriaxone arm vs 25% in the placebo arm). There were 7 significant treatment-related adverse events (18.9%); 6 occurred in subjects on active treatment.

Due to the lack of durable cognitive improvement and the risk of adverse events, Fallon et al. concluded that 10 weeks of ceftriaxone was not an effective strategy; the authors encouraged searching for more effective and safer retreatment strategies.

3.3.2. Trial critique

3.3.2.1. Design. The trial was designed with a planned enrollment of 45 participants but recruited only 37 subjects; 23 randomized to active treatment and 14 to placebo. Under-enrollment could have resulted in the cognitive functioning outcome becoming underpowered.

3.3.2.2. Analysis. While the study was under-enrolled, 32/37 (86%) of enrolled patients completed the protocol at 12 and 24 weeks. Detection of a significant treatment effect on pain and physical functioning among those with worse baseline scores can likely be attributed to an efficient statistical analysis incorporating monthly measures of these secondary outcomes, and incorporating effect modification due to baseline disease severity.

3.3.2.3. Interpretation. This trial, with its small sample size and extensive secondary outcome analysis, is more reminiscent of a pilot study than a definitive clinical trial. The conclusions were fittingly cautious. Noting a positive treatment effect on fatigue, similar to that seen in the Krupp trial, and a high rate of adverse events, the authors highlighted the need for additional studies and safer antibiotic regimens.

4. Discussion

This biostatistical review of the four NIH-sponsored Lyme disease retreatment trials highlights the need for close scrutiny of all clinical trials, including those which emphasize findings of insignificant treatment effects. Our careful examination of the trials suggests that, for some patients with Lyme disease, retreatment can, in fact, be beneficial. Krupp's study was properly designed and analyzed with regard to fatigue, detecting significant, sustained and clinically meaningful improvement in this primary endpoint, and the Fallon trial demonstrated treatment efficacy on cognition at the margin of statistical significance at 3 months. And, although these were secondary outcomes, the Fallon trial corroborated Krupp's finding on fatigue and, further, found that patients with worse baseline pain and physical functioning had significant and sustained improvement in these measures.

Unfortunately, misinterpretation of insignificant findings from underpowered or poorly designed trials can have profound ramifications on treatment guideline recommendations, patient care and the direction of future research. In Lyme disease, the lack of demonstrable improvement in persistent symptoms in the Klempner trials and the absence of an antibiotic effect on mental processing speed in the Krupp trial do not provide evidence against the efficacy of antibiotic retreatment. Our analysis reveals that these outcome measures were not well

designed and lacked statistical power. Therefore, contrary to the conclusions of some Lyme disease guidelines panels [20,21], the inability of these trials to demonstrate a statistically significant finding provides neither proof of the absence of a clinically meaningful treatment effect nor evidence that patients with persistent symptoms suffer from a post-infectious syndrome.

The findings from the Krupp and Fallon trials imply a causal link between antibiotic treatment and physical improvement, which would be consistent with the hypothesis that persistent symptoms may be the result of a persistent *B. burgdorferi* infection. A recent uncontrolled, observational study of Lyme disease patients treated with ceftriaxone appears to support such a link, with patients experiencing long-term benefits in fatigue, pain and cognition [41]. While one may speculate about the potential neuro-protective effects of ceftriaxone in Lyme disease [17], the presence of such effects would not explain findings of sustained benefits on fatigue or pain, or disprove the existence of persistent infection.

Conclusions favoring post-infectious processes as the explanations for persistent symptoms may be premature. Several authors, using a variety of accepted laboratory tests, conclusively demonstrated persistent *B. burgdorferi* infection in humans and animals following antibiotic therapy appropriate for their stage of illness [42–51]. The most recent of these was a primate study, in which investigators recovered intact *B. burgdorferi* spirochetes by xenodiagnosis from rhesus macaques treated with the Infectious Diseases Society of America (IDSA)-recommended regimen for disseminated infection [48]. Additionally, spirochetal DNA and RNA were detected post-treatment in this group and in a second group of rhesus macaques that were given the treatment used in the trial by Klempner et al.: "the animals were treated at the late disseminated phase of infection and the treatment regimen was chosen to correspond to the regimen used to treat human PTLDS patients in a clinical evaluation of treatment for this population" [48].

It is incorrect to draw strong conclusions regarding antibiotic retreatment in patients with persistent symptoms of Lyme disease based on the four NIH-sponsored randomized controlled trials discussed in this review. Inadequacies in trial designs and the small sample sizes leave many questions unanswered, and underscore the need for additional clinical research on this question. Those who wrongly conclude that the trials found no benefit from retreatment commit an even greater error, as such a statement is demonstrably false.

Future RCTs should investigate oral antibiotic regimens which are likely to be safer [52–54] and less costly than ceftriaxone. Such trials should also avoid enrolling patients according to a proposed definition of "post-Lyme syndrome" [21], not only because the terminology prematurely assumes a post-infectious process, but also because this broader grouping may mask significant treatment effects in specific patient subsets, such as the fatigue subset identified in the Krupp and Fallon trials. Instead, future trials should consider a stratified design ensuring good balance by arm with respect to disease symptom clusters (such as joint or CNS involvement) and with sufficient power to detect realistic treatment effects within strata. Until evidence from such trials becomes available, it would be wise for clinicians to disregard generalized and unsupported recommendations against retreatment and instead rely on their clinical judgment to manage patients with persistent symptoms of Lyme disease.

References

[1] Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. Cochrane Database Syst Rev 2009:MR000006.
[2] Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. PLoS Med 2008;5:e217 [discussion e].
[3] Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA 2010;303:2058-64.
[4] Burgdorfer W. Arthropod-borne spirochetoses: a historical perspective. Eur J Clin Microbiol Infect Dis 2001;20:1-5.
[5] Centers for Disease Control and Prevention (CDC). Reported CASES of Lyme disease by year, United States, 1994–2008. Available: http://www.cdc.gov/ncidod/dvbid/LYME/ld_UpClimbLymeDis.htm Accessed 2009 Sept 20.
[6] Smith R, Takkinen J, Editorial team. Lyme borreliosis: Europe-wide coordinated surveillance and action needed? Euro Surveill 2006;11 (Available: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=2977. Accessed 2010 Feb 10).
[7] Shadick NA, Phillips CB, Logigian EL, Steere AC, Kaplan RF, Berardi VP, et al. The long-term clinical outcomes of Lyme disease. A population-based retrospective cohort study. Ann Intern Med 1994;121:560-7.
[8] Fallon BA, Nields JA, Burrascano JJ, Liegner K, DelBene D, Liebowitz MR. The neuropsychiatric manifestations of Lyme borreliosis. Psychiatr Q 1992;63:95–117.
[9] Asch ES, Bujak DI, Weiss M, Peterson MG, Weinstein A. Lyme disease: an infectious and postinfectious syndrome. J Rheumatol 1994;21:454-61.
[10] Clarissou J, Song A, Bernede C, Guillemot D, Dinh A, Ader F, et al. Efficacy of a long-term antibiotic treatment in patients with a chronic Tick Associated Poly-organic Syndrome (TAPOS). Med Mal Infect 2009;39:108-15.
[11] Steere AC, Dhar A, Hernandez J, Fischer PA, Sikand VK, Schoen RT, et al. Systemic symptoms without erythema migrans as the presenting picture of early Lyme disease. Am J Med 2003;114:58-62.
[12] Logigian EL, Kaplan RF, Steere AC. Chronic neurologic manifestations of Lyme disease. N Engl J Med 1990;323:1438-44.
[13] Pachner AR. *Borrelia burgdorferi* in the nervous system: the new "great imitator". Ann N Y Acad Sci 1988;539:56-64.
[14] Cairns V, Godwin J. Post-Lyme borreliosis syndrome: a meta-analysis of reported symptoms. Int J Epidemiol 2005;34:1340-5.
[15] Ljostad U, Mygland A. Remaining complaints 1 year after treatment for acute Lyme neuroborreliosis; frequency, pattern and risk factors. Eur J Neurol 2010;17:118-23.
[16] Klempner MS, Hu LT, Evans J, Schmid CH, Johnson GM, Trevino RP, et al. Two controlled trials of antibiotic treatment in patients with persistent symptoms and a history of Lyme disease. N Engl J Med 2001;345:85-92.
[17] Fallon BA, Keilp JG, Corbera KM, Petkova E, Britton CB, Dwyer E, et al. A randomized, placebo-controlled trial of repeated IV antibiotic therapy for Lyme encephalopathy. Neurology 2008;70:992–1003.
[18] Auwaerter PG. Point: antibiotic therapy is not the answer for patients with persisting symptoms attributable to Lyme disease. Clin Infect Dis 2007;45:143-8.
[19] Stricker RB. Counterpoint: long-term antibiotic therapy improves persistent symptoms associated with Lyme disease. Clin Infect Dis 2007;45:149-57.
[20] Stanek G, Fingerle V, Hunfeld KP, Jaulhac B, Kaiser R, Krause A, et al. Lyme borreliosis: clinical case definitions for diagnosis and management in Europe. Clin Microbiol Infect 2011;17:69-79.
[21] Wormser GP, Dattwyler RJ, Shapiro ED, Halperin JJ, Steere AC, Klempner MS, et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis 2006;43:1089-134.
[22] Halperin JJ, Shapiro ED, Logigian E, Belman AL, Dotevall L, Wormser GP, et al. Practice parameter: treatment of nervous system Lyme disease (an evidence-based review): report of the Quality Standards Subcommittee of the American Academy of Neurology. Neurology 2007;69:91–102.
[23] Mygland A, Ljostad U, Fingerle V, Rupprecht T, Schmutzhard E, Steiner I. EFNS guidelines on the diagnosis and management of European Lyme neuroborreliosis. Eur J Neurol 2010;17(8–16):e1-4.
[24] Krupp LB, Hyman LG, Grimson R, Coyle PK, Melville P, Ahnn S, et al. Study and treatment of post Lyme disease (STOP-LD): a randomized double masked clinical trial. Neurology 2003;60:1923-30.
[25] Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? Pharmacoeconomics 2000;18:419-23.

[26] Oksi J, Nikoskelainen J, Hiekkanen H, Lauhio A, Peltomaa M, Pitkaranta A, et al. Duration of antibiotic treatment in disseminated Lyme borreliosis: a double-blind, randomized, placebo-controlled, multicenter clinical study. Eur J Clin Microbiol Infect Dis 2007;26:571-81.
[27] Cameron D. Severity of Lyme disease with persistent symptoms. Insights from a double-blind placebo-controlled clinical trial. Minerva Med 2008;99:489-96.
[28] Kaplan RF, Trevino RP, Johnson GM, Levy L, Dornbush R, Hu LT, et al. Cognitive function in post-treatment Lyme disease: do additional antibiotics help? Neurology 2003;60:1916-22.
[29] Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware Jr JE. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum 2000;43:1478-87.
[30] Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001;45:384-91.
[31] Coteur G, Feagan B, Keininger DL, Kosinski M. Evaluation of the meaningfulness of health-related quality of life improvements as assessed by the SF-36 and the EQ-5D VAS in patients with active Crohn's disease. Aliment Pharmacol Ther 2009;29:1032-41.
[32] Regensteiner JG, Ware Jr JE, McCarthy WJ, Zhang P, Forbes WP, Heckman J, et al. Effect of cilostazol on treadmill walking, community-based walking ability, and health-related quality of life in patients with intermittent claudication due to peripheral arterial disease: meta-analysis of six randomized controlled trials. J Am Geriatr Soc 2002;50:1939-46.
[33] Okamoto LJ, Noonan M, DeBoisblanc BP, Kellerman DJ. Fluticasone propionate improves quality of life in patients with asthma requiring oral corticosteroids. Ann Allergy Asthma Immunol 1996;76:455-61.
[34] Ware Jr JE, Bayliss MS, Rogers WH, Kosinski M, Tarlov AR. 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;276:1039-47.
[35] Ware JE, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston: Health Assessment Lab; 1994.
[36] Wolinsky FD, Wan GJ, Tierney WM. Changes in the SF-36 in 12 months in a clinical sample of disadvantaged older adults. Med Care 1998;36:1589-98.
[37] Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. Med Care 1999;37:469-78.
[38] Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care 2003;41:582-92.
[39] Pollina DA, Sliwinski M, Squires NK, Krupp LB. Cognitive processing speed in Lyme disease. Neuropsychiatry Neuropsychol Behav Neurol 1999;12:72-8.
[40] Coyle PK, Deng Z, Schutzer SE, Belman AL, Benach J, Krupp LB, et al. Detection of *Borrelia burgdorferi* antigens in cerebrospinal fluid. Neurology 1993;43:1093-8.
[41] Stricker RB, Delong AK, Green CL, Savely VR, Chamallas SN, Johnson L. Benefit of intravenous antibiotic therapy in patients referred for treatment of neurologic Lyme disease. Int J Gen Med 2011;4:639-46.
[42] Oksi J, Marjamaki M, Nikoskelainen J, Viljanen MK. *Borrelia burgdorferi* detected by culture and PCR in clinical relapse of disseminated Lyme borreliosis. Ann Med 1999;31:225-32.
[43] Schmidli J, Hunziker T, Moesli P, Schaad UB. Cultivation of *Borrelia burgdorferi* from joint fluid three months after treatment of facial palsy due to Lyme borreliosis. J Infect Dis 1988;158:905-6.
[44] Preac-Mursic V, Weber K, Pfister HW, Wilske B, Gross B, Baumann A, et al. Survival of *Borrelia burgdorferi* in antibiotically treated patients with Lyme borreliosis. Infection 1989;17:355-9.
[45] Haupl T, Hahn G, Rittig M, Krause A, Schoerner C, Schonherr U, et al. Persistence of *Borrelia burgdorferi* in ligamentous tissue from a patient with chronic Lyme borreliosis. Arthritis Rheum 1993;36:1621-6.
[46] Strle F, Preac-Mursic V, Cimperman J, Ruzic E, Maraspin V, Jereb M. Azithromycin versus doxycycline for treatment of erythema migrans: clinical and microbiological findings. Infection 1993;21:83-8.
[47] Hunfeld KP, Ruzic-Sabljic E, Norris DE, Kraiczy P, Strle F. In vitro susceptibility testing of *Borrelia burgdorferi* sensu lato isolates cultured from patients with erythema migrans before and after antimicrobial chemotherapy. Antimicrob Agents Chemother 2005;49:1294-301.
[48] Embers ME, Barthold SW, Borda JT, Bowers L, Doyle L, Hodzic E, et al. Persistence of *Borrelia burgdorferi* in rhesus macaques following antibiotic treatment of disseminated infection. PLoS One 2012;7:e29914.
[49] Barthold SW, Hodzic E, Imai DM, Feng S, Yang X, Luft BJ. Ineffectiveness of tigecycline against persistent *Borrelia burgdorferi*. Antimicrob Agents Chemother 2010;54:643-51.
[50] Hodzic E, Feng S, Holden K, Freet KJ, Barthold SW. Persistence of *Borrelia burgdorferi* following antibiotic treatment in mice. Antimicrob Agents Chemother 2008;52:1728-36.
[51] Straubinger RK, Summers BA, Chang YF, Appel MJ. Persistence of *Borrelia burgdorferi* in experimentally infected dogs after antibiotic treatment. J Clin Microbiol 1997;35:111-6.
[52] Goulden V, Glass D, Cunliffe WJ. Safety of long-term high-dose minocycline in the treatment of acne. Br J Dermatol 1996;134:693-5.
[53] Tilley BC, Alarcon GS, Heyse SP, Trentham DE, Neuner R, Kaplan DA, et al. Minocycline in rheumatoid arthritis. A 48-week, double-blind, placebo-controlled trial. MIRA Trial Group. Ann Intern Med 1995;122:81-9.
[54] Cooper C. Safety of long term therapy with penicillin and penicillin derivatives. Bioterrorism and drug preparedness. U.S. Food and Drug Administration (FDA); 2001. Available: http://www.fda.gov/Drugs/EmergencyPreparedness/BioterrorismandDrugPreparedness/ucm072755.htm. Accessed 2010 Feb 27.