A Reappraisal of the U.S. Clinical Trials of Post-Treatment Lyme Disease Syndrome

Brian A. Fallon*[,1], Eva Petkova[2], John G. Keilp[3] and Carolyn B. Britton[4]

[1]*Division of Clinical Therapeutics, Department of Psychiatry, Columbia University, USA*

[2]*Division of Biostatistics, Department of Child and Adolescent Psychiatry, New York University, USA*

[3]*Division of Neuroscience, Department of Psychiatry, Columbia University, USA*

[4]*Department of Neurology, Columbia University, USA*

Abstract: Four federally funded randomized placebo-controlled treatment trials of post-treatment Lyme syndrome in the United States have been conducted. Most international treatment guidelines summarize these trials as having shown no acute or sustained benefit to repeated antibiotic therapy. The goal of this paper is to determine whether this summary conclusion is supported by the evidence.

Methods: The methods and results of the 4 U.S. treatment trials are described and their critiques evaluated.

Results: 2 of the 4 U.S. treatment trials demonstrated efficacy of IV ceftriaxone on primary and/or secondary outcome measures.

Conclusions: Future treatment guidelines should clarify that efficacy of IV ceftriaxone for post-treatment Lyme fatigue was demonstrated in one RCT and supported by a second RCT, but that its use was not recommended primarily due to adverse events stemming from the IV route of treatment. While repeated IV antibiotic therapy can be effective, safer modes of delivery are needed.

In 2006, the Infectious Diseases Society in the United States updated its treatment guidelines for Lyme and other tick-borne diseases [1]. Since then other national and international organizations have published similar treatment guidelines, including the American Academy of Neurology [2], the British Infection Association [3], and the European Federation of Neurological Societies [4]. Based on the results from the U.S. clinical trials on post-treatment Lyme disease, these guidelines provide parameters for the treatment of patients with chronic persistent symptoms. These published guidelines state "antibiotic therapy has not proven to be useful" [1], "American trials have demonstrated that additional prolonged antimicrobial treatment is ineffective in Post Lyme Disease Syndrome" [4], and "studies of prolonged antimicrobial treatments of patients with Post Lyme Syndrome have not shown sustained benefit" [3].

It is the contention of the authors of this paper that these national and international treatment guidelines, while accurate and informative in most areas, overlook or inappropriately dismiss some of the key lessons learned from the NIH-funded U.S. Treatment trials on chronic post-treatment Lyme

syndrome. Expert committees have carefully reviewed and critiqued these studies, highlighting both their limitations and the adverse events associated with retreatment [1,5,6]. We will address these critiques in detail, focusing in particular on the study of post-treatment Lyme syndrome which enrolled patients based on fatigue [7] as it was this study that had the clearest positive findings. We conclude this paper by recommending issues to consider in the design of future clinical trials based on the lessons learned from these Lyme disease treatment studies.

Terms. Post-treatment Lyme disease syndrome (a.k.a. post-Lyme disease syndrome, post-treatment chronic Lyme disease) is a term used to describe the clinical experience of patients who have symptoms that persist for months or years despite having received recommended courses of antibiotic therapy for well-documented Lyme disease [8,9]. Typically these patients report problems with fatigue, musculoskeletal pain, and cognition [9]. The word "syndrome" reflects an acknowledgment that the cause of the persistent symptoms is unclear. The term "chronic Lyme disease" is less widely favored based on the argument that acute Lyme disease is caused by a known infection whereas the chronic post-treatment symptoms are of uncertain etiology [5]. For the purposes of this paper, the term post-treatment Lyme disease syndrome (PTLDS) will be used. Exact criteria for PTLDS vary. The most inclusive approach encompasses all patients

*Address correspondence to this author at Columbia University 1051 Riverside Drive, Unit 69, NYC, NY 10032, USA; Tel: 212-543-5487; Fax: 212-543-6515; E-mail: baf1@columbia.edu

with persistent significantly distressing or impairing symptoms that emerge within a defined period after having acquired and been treated for well-documented Lyme disease. Operationalized criteria for PTLDS have been proposed that are more restrictive, excluding the diagnosis of PTLDS if objective signs of disease are present [1]. These operational criteria do not clarify how extensive an evaluation is needed to conclude that an 'objective' deficit is not present. For example, for an individual with subjective cognitive complaints, does one require neurocognitive testing and, if so, what level of deficit is required to qualify for objective impairment? Or, if a patient complains of subjective paresthesias, does one then need to conduct skin biopsies on all patients to determine if a small fiber neuropathy is present? These questions remain unresolved. The U.S. clinical trials on chronic symptoms therefore vary in the criteria for enrollment of patients with chronic persistent symptoms. While some trials included all patients with symptoms of certain severity after well-documented Lyme disease regardless of whether objective signs of disease were present [7], other studies required objective operationalized criteria [10] while others excluded those with objective evidence of disease on physical examination (e.g., synovitis) [11].

US CLINICAL TRIALS 1 & 2 (KLEMPNER *et al.*, NEW ENGLAND JOURNAL OF MEDICINE 2001)

These two placebo-controlled randomized trials [11] used the same study design and outcome measures with the distinguishing feature that one trial enrolled seropositive patients while the other trial enrolled seronegative patients. The total sample size was 129 patients. The treatment period was 90 days: 30 days of IV ceftriaxone (2 gms/day) followed by 60 days of oral doxycycline (100 bid) or 30 days of IV placebo followed by 60 days of oral placebo. After treatment, the patients were followed for 90 days. The primary outcome measures at 6 months were categorical variables based on change from baseline on the physical and mental health-related quality of life composite indices of the Short form general health survey (SF-36)(trichotomized to "worsening", "no change" or "improvement") [12]. Secondary outcome measures assessed change in neurocognition and in psychiatric symptoms. For the seropositive patients entrance criteria did not require documentation of classic symptoms of Lyme disease, but did require documentation of having received recommended treatment for Lyme disease. For the seronegative patients, entrance criteria required documentation of an erythema migrans rash. Additionally, patients had to report functional impairment related to musculoskeletal or cognitive symptoms that had begun within 6 months of the infection and that had been present for at least 6 months (but less than 12 years). Patients were excluded if they had received more than 60 days of prior IV antibiotics, had known hypersensitivity to study medications, had active synovitis or positive PCR for Bb in the CSF or blood.

At six months, no placebo-drug differences were noted for change on the primary outcome measure of the SF-36 using a preset cutoff for improvement; this was true for the seropositive patients, the seronegative patients, and both groups combined. In addition, no placebo-drug differences were noted for change on the secondary outcome measures of cognition and depression. In these studies, the treatment was reported as ineffective and it was not clinically recommended.

Limitations of this study include: a) not requiring documentation of objective manifestations of Lyme disease for enrollment of all study patients; b) not requiring that patients meet a specific severity level for impairment on the primary measure of interest at the start of the study; c) inadequately sensitive data analytic methodology: (i) not including in the analytic method an adjustment for baseline levels of impairment (even though in the seronegative sample there was a significant baseline difference on the SF-36), (ii) no assessment for the potential moderating effect of baseline level of impairment on treatment outcome, and (iii) using a statistically inefficient test for comparison between the groups (e.g., categorizing a variable that is measured continuously); and d) a low dose of oral doxycycline that would not be adequate for CNS penetration. By not requiring documentation of the signs of Lyme disease, it cannot be certain that all patients in the study had had objective manifestations of Lyme disease previously. The exclusion of patients with post-treatment active synovitis would be consistent with the IDSA's definition of PTLDS [1] as this would be an objective finding on clinical exam, however patients with other objective findings (e.g., cognitive impairment on testing or elevated protein in the CSF) weren't excluded. Although patients who enrolled in this study were required to report that their symptoms led to functional impairment and the patients as a group did report substantial physical impairment, this study differed from the subsequent studies by not requiring a predetermined severity level for enrollment on the primary outcome measure of interest. While the mean impairment on the physical composite measures of the SF-36 was moderately severe for the group as a whole, the distribution may have included some patients who had such low levels of impairment that a treatment effect would have been hard to demonstrate even with a potent treatment. The sample of patients in this study, unlike the two subsequent studies [7,10], was therefore a heterogeneous collection of patients with respect to severity and baseline level of impairment. Not adjusting for baseline levels in the main analyses may have decreased the power of the test, since baseline levels of impairment typically are related to post treatment scores and change scores. Power was further decreased by categorizing the change scores and premature study discontinuation which lowered the final sample size. Because the analytic method did not include an assessment of the potential interaction effect of baseline severity and treatment group on outcome, true treatment effects distinguishing the drug and placebo groups (e.g., larger drug effects among more impaired subjects) may have been missed. Two patients given active treatment had a serious adverse event – pulmonary embolus in one and fever, anemia, and GI bleeding in the other. Particular strengths of these two studies were the relatively large sample size and the requirement that the seronegative patients have a history of physician documented erythema migrans.

U.S. CLINICAL TRIAL 3 – "STOP-LD" STUDY. (KRUPP *et al.*, NEUROLOGY, 2003)

This randomized double-masked placebo-controlled trial of post-Lyme syndrome [7] enrolled 55 patients with persistent severe fatigue at least 6 or more months after antibiotic

therapy. Patients had to provide physician documentation of previously having met objective criteria for Lyme disease (EM or CDC-defined late manifestation of Lyme disease confirmed by ELISA and Western blot serology) and of having been previously treated with at least 3 weeks of antibiotics six months or more before study entry. Patients were excluded who had concurrent disorders not related to Lyme disease that could cause fatigue. For study entry, patients had to have at least moderate fatigue as assessed by the score on the Fatigue Severity Scale-11 (FSS-11) that occurred in conjunction with onset of Lyme disease. Patients were not excluded if they had observable neurologic findings or other objective deficits on routine general or neurologic examination. Patients received 28 days of IV ceftriaxone or placebo following by 5 months of no treatment. The primary outcome time point was 6 months. Although there were two primary clinical measures (fatigue and cognitive response time) and one biological measure (an experimental measure of CSF infection - Osp A), at the time of enrollment impairment was required only on the fatigue measure. Meaningful improvement on the FSS-11 was assessed categorically with responder status determined by a decrease of 0.7 points or more. Meaningful improvement on the response time measure required a change of 25% or more. The results demonstrated that significantly more patients assigned to ceftriaxone showed improvement in disabling fatigue compared to the placebo group (RR 3.5, p<0.001). No drug-placebo difference was noted in cognitive response time or in Osp A. Four patients (3 on placebo) had adverse events that required hospitalization. The authors, while concluding that ceftriaxone did result in improvement in severe fatigue, concluded that the risks associated with treatment and the lack of benefit in other outcome measures mitigated against recommending repeated antibiotic therapy.

Among the 4 post-treatment Lyme syndrome studies, the STOP-LD study is the one that most clearly demonstrated efficacy for drug over placebo on a primary outcome measure. Perhaps because of this, the study results have been carefully scrutinized and critiqued [1,5]. In the comments below, we review each of these critiques to evaluate their merit in either supporting or dismissing the favorable efficacy result of the study.

Critique #1 :"Only 1 of 3 Primary Outcome Measures Showed a Treatment Effect"

The 3 outcome measures were: a) Fatigue Severity Scale; b) a cognitive test of mental processing speed; and c) clearance of Borrelial Osp A antigen.

a. Fatigue. Improvement was noted on the only primary outcome measure on which patients were enrolled. The Fatigue Severity Scale (FSS-11) was employed as the primary enrollment measure and as the primary outcome measure for fatigue, as this is a measure that had been psychometrically validated by the authors. On this fatigue measure, the percentage of responders to drug was high (64%) compared to placebo (18.5%) (p<0.001). In other words, based on this measure, the drug treatment was effective. To appropriately assess efficacy of a treatment, patients in a controlled trial must have meaningful impairment on the measure of interest at the start of the study. Given that all patients were required to have prominent fatigue at the start of the STOP-LD study, it is reasonable for fatigue to have been a primary outcome measure to assess efficacy. On the other outcome measures however (cognitive processing speed or Osp A antigen), patients were not enrolled based on presence of impairment or abnormality; (e.g., only 9 of the 55 patients were positive for Osp A antigen at baseline). This study therefore was not able to adequately test efficacy on these latter measures.

The therapeutic effect of additional IV ceftriaxone therapy on the fatigue measure was demonstrated even more robustly within the subgroup of study participants who at the time of enrollment were serologically IgG Western blot positive. In this more homogenous subgroup, the responder rate for the drug treated group was 6 times higher than in the placebo treated group (80% vs 13%, p<0.01). See Fig. (**1**).

A critique of the fatigue study result has been that while there was a responder rate difference on the primary fatigue severity scale measure, there was no significant improvement in fatigue as assessed by a secondary visual analog
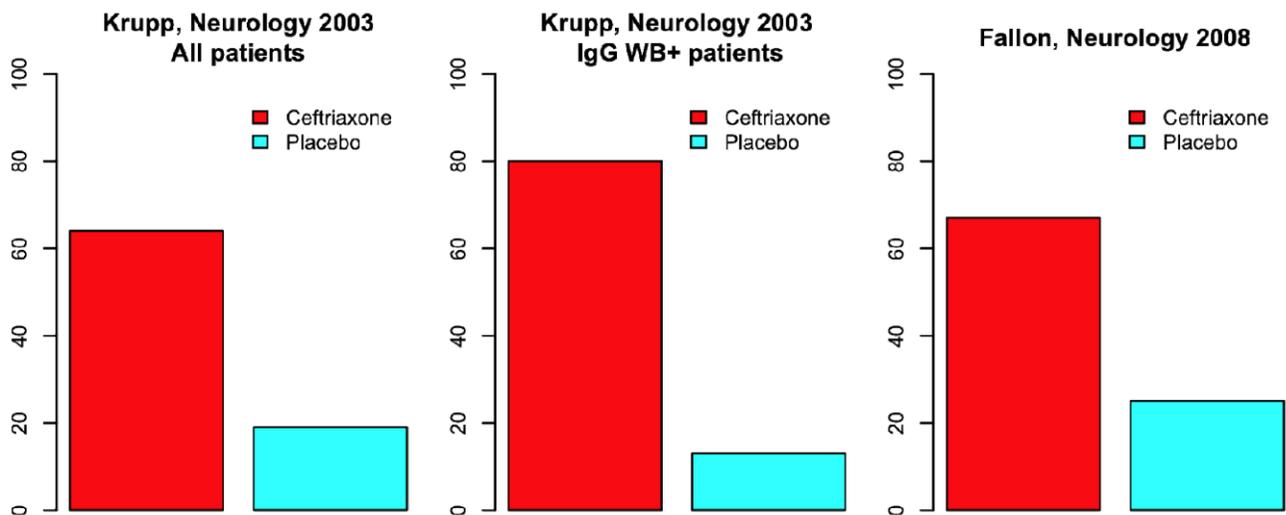

Fig. (1). Percentage of Responders on Fatigue Severity Scale in 2 placebo controlled trials of post-treatment Lyme disease.

scale (VAS) of fatigue intensity. The VAS, however, was a secondary measure whose validity was not clearly established, and may have simply been a less sensitive scale. Never the less, the authors of this study themselves note that the VAS results were consistent with the FSS-11 results: "The change in the fatigue VAS showed a similar pattern to the FSS-11, with a trend toward more improvement in the ceftriaxone vs the placebo group at 6 months (p=0.08). In addition, at 6 months, the mean (SD) VAS scores were lower in the ceftriaxone than in the placebo group (4.8 (2.5) vs 6.5 (2.1); *p*=0.01)".

b. Cognitive slowing. The authors of this study defined a 25% improvement in cognitive response time as a clinically meaningful response. Although the authors note that the enrolled Lyme patients had worse response time scores than historical normative controls, they do not clarify how many patients in this study actually were impaired on this measure. In the discussion, they emphasize that the observed cognitive deficits were "relatively mild, which may have contributed to the lack of a treatment effect on cognition". This important point highlights a key issue: if patients aren't sufficiently impaired, a treatment effect is unlikely to be seen even with a very potent treatment. Given the small sample size of this study and the lack of major impairment on this outcome measure at the time of enrollment, statistical power is limited to demonstrate a treatment effect on this measure. This study most likely was not adequately powered to assess the efficacy of drug vs. placebo on cognitive slowing.

c. Clearance of Osp A antigen. There are problems with this as an outcome measure. First, this is an experimental test of uncertain sensitivity and specificity. Second, presence of the Osp A antigen was detected in only 9 of 55 patients; a primary efficacy assessment of change in response to treatment cannot be based on a sample size of 9. The study was under-powered to detect any realistic change in clearance of OspA antigen. As noted by the authors, "The primary biologic outcome measure, CSF OspA antigen, was present in only 16% of patients at baseline and was not a useful marker of outcome". The study results therefore using this primary outcome measure should not be included in the final determination of whether IV ceftriaxone is effective as a treatment for PTLDS.

Summary. This study was adequately powered to test only one of the 3 primary outcome measures: the only clinical measure on which patients were enrolled (ie, fatigue) and on which they were definitively impaired. For the disabling fatigue commonly seen among patients with PTLDS, sustained efficacy using IV ceftriaxone was demonstrated.

Critique #2: "The Favorable Results on Fatigue may have Occurred due to Inadvertent Unmasking of the Treatment"

In this study, when patients were asked at the 6 month time points to guess the treatment to which they had been randomized, significantly more patients in the ceftriaxone group correctly guessed their treatment assignment compared to the placebo group. While this result certainly raises the possibility that masking was compromised, the following points are worth noting.

a. "Unmasking" effects are inherently ambiguous. This same result could lead to the possible explanation that people guessed correctly because the drug had a positive effect on their symptoms. If patients experience meaningful improvement, they will more likely guess that they are on active treatment. In other words, if an active treatment works, the patients on that treatment should guess their treatment assignment correctly more often than patients given a treatment that doesn't work. Since there was no significant difference between the two treatments with respect to adverse effects (another possible source of unblinding), this may lead one to hypothesize that the higher correct guess rate in the ceftriaxone group was due to the beneficial impact of the drug rather than unmasking.

b. The authors in the results note that: "at 6 months, among patients in both groups who believed they were on active therapy, those in the ceftriaxone group still showed a higher frequency of improvement in fatigue scores compared to those in the placebo group (83% vs 44%)." In other words, even among those who believed that they had been given active drug, the higher responder rate still persisted among those who had in fact received active drug rather than placebo.

c. Comparable results were obtained in a second study - the Lyme encephalopathy trial [10] (see Fig. **1**). When the Lyme encephalopathy results were reanalyzed using comparable methods as in the STOP-LD study, the responder rate on the fatigue measure was 66.7% for the ceftriaxone treated group vs 25% for the placebo-treated group (p=0.05) at week 24. This is nearly identical to the 64% vs 19% results for drug vs placebo comparison at week 24 (6 months) noted in the STOP-LD study. The Lyme encephalopathy study results are therefore consistent with the STOP-LD findings, suggesting that the STOP-LD results are valid and not due to inadvertent unmasking.

d. As previously noted, patients who were IgG Western blot positive in this study were 6 times more likely to show improvement in fatigue if given ceftriaxone than if they were given placebo (80% vs 13%, p<0.01) (Fig. **1**). This subgrouping, based on a biological marker, enhanced the responder rate for drug and diminished the responder rate for placebo. If the higher responder rate in response to ceftriaxone were primarily due to "correct guessing", then there shouldn't be a biologically homogenous subgroup that responded even better to the active treatment than to placebo. This biologically based result speaks to the greater likelihood that a true treatment effect was occurring rather than unmasking.

STOP-LD Critique #3. "The Magnitude of the Effect on Improvement in Fatigue was Small"

The responder status in this study for the fatigue scale was determined based on a change score on the FSS of 0.7 units. Based on prior work, the authors determined that a 0.7 unit change was "clinically meaningful", representing "an improvement approximately three times as large as that observed in a prior placebo-treated group". Therefore, based on the study design which specified in advance the criteria for a clinically meaningful improvement, it is reasonable to con-

clude that the responders in this study did in fact experience clinically meaningful change.

In Table 2 of the STOP-LD paper, the mean percentage change from baseline was 22.1% for the ceftriaxone group and 9.1% for the placebo group. Prior critiques [5] have noted that the difference between active drug and placebo was only 13% and then question whether a difference of only 13% is clinically meaningful. This issue can be further examined in two ways:

a. What was the effect size for the difference in improvement between drug and placebo at 6 months? In inferential statistics, an effect size helps to determine whether a statistically significant difference is a difference with clinical importance. Based on Cohen [13], an effect size of 0.20 is considered mild, 0.50 is medium and anything greater than 0.80 is large. The effect size on the Fatigue Severity Scale in the STOP-LD study was 1.0 using their baseline SD values for the FSS. (A meaningful analogy from the cognitive realm is that an effect size of 1 would represent a 15 point improvement in IQ.) A more conservative approach toward evaluating effect size would use the baseline SD values for a post-Lyme sample that had not been recruited specifically for fatigue. The Lyme encephalopathy sample [10] would meet this requirement. Using the baseline FSS SD values therefore from the encephalopathy sample, the effect size on the FSS in the STOP-LD study was 0.63. In other words, a more conservative estimate of effect size never the less demonstrates that the beneficial impact of ceftriaxone ranged from moderate to large. In comparison to drug trials for FDA approved anti-depressants in which effect sizes are often only mild to moderate [14], this would be considered a robust behavioral improvement and a valuable treatment.

b. Another way to examine this issue is to assess the magnitude of improvement in those patients who were responders vs. those patients who were non-responders. That data is not presented in the STOP-LD paper. However, to try to shed light on this question, we reexamined the Lyme encephalopathy Fatigue Severity Scale data to determine the magnitude of improvement at six months among responders if the enrollment were restricted to those who had FSS at the start of the study of 4 or higher, as had been done in the STOP-LD study. Using the same criteria for responders as had been used in the STOP-LD study, our data analysis revealed the following for the percentage improvement at week 24 compared to baseline on the fatigue severity scale: 0.34 (+/-0.20) improvement in fatigue for the responders vs -0.02 (+/-0.12) change (worsening) in fatigue for the non-responders. An improvement in fatigue of 34% is likely to be clinically meaningful for the two-thirds of the sample who experienced this improvement, although it is certainly not curative. It is also likely that the magnitude of improvement for responders in the STOP-LD study would be even greater than in our Lyme encephalopathy study given that patients in their sample had received less antibiotic therapy prior to entering the trial.

Based on the comments above, therefore, it is unreasonable to discount the improvement in fatigue in the STOP-LD study as having been "too small" in magnitude to be of clinical significance. Based on currently accepted methods for the analysis of clinical trial data, the effect size for improvement in fatigue in the STOP-LD study was moderate to large, and would be considered clinically meaningful.

STOP-LD Critique #4. "The High Drop-out Rate in the Placebo Group May have Led to Misleading Results."

According to the published paper, there were 5 drop-outs in the placebo-treated group who were not assessed at 6 months and 2 drop-outs in the drug-treated group who were not assessed at 6 months. This is a modest drop-out rate. If one conducts a "worst case scenario analysis" which would assume that the 5 drop-outs in the placebo-treated group would have been responders and the 2 drop-outs in the ceftriaxone-group would have been non-responders, then the X^2 analysis continues to demonstrate significantly more responders in the drug-treated compared to the placebo-treated group (18/28 vs 10/27, X^2=4.1, p= 0.04) Based on this analysis, it does not appear that the drop-outs led to misleading results.

U.S. CLINICAL TRIAL 4 – POST TREATMENT LYME ENCEPHALOPATHY. (FALLON *et al.*, NEUROLOGY, 2008)

This placebo-controlled randomized trial enrolled 37 patients with post-treatment Lyme encephalopathy. Patients were assigned to treatment with 10 weeks of IV ceftriaxone or placebo in a 2:1 randomization followed by no treatment for 14 weeks. The primary end-points were week 12 for efficacy and week 24 for longer-term durability. As this was a study of Lyme encephalopathy, the primary measure of interest assessed cognition, with a specific focus on memory. Patients had to have both subjective cognitive complaint and objectively confirmed deficits in memory; patients with other known causes of cognitive impairment were excluded. The criteria for the diagnosis of Lyme disease required: a) documentation of physician-diagnosed erythema migrans or a later manifestation of Lyme disease meeting CDC surveillance criteria with a positive or equivocal ELISA confirmed by positive Western blot serology; and b) a positive IgG Western blot at the time of study entry. All patients also had to have had at least 3 weeks of prior IV ceftriaxone or IV cefotaxime treatment for Lyme disease. Because repeated neurocognitive testing can result in improvement merely because of the practice effect, healthy controls were included to control for this effect. There were therefore 3 groups (patients on drug, patients on placebo, and healthy controls). Over the course of the 24 weeks of this study, there was a significant group difference (p= 0.04) in cognitive change across time among the 3 groups; this difference was attributed to the initial improvement in overall cognition at week 12 in the drug group followed by the loss of that improvement by week 24. On the drug vs healthy control comparison at week 12, the improvement in cognition for the drug-treated group was greater than the improvement (presumably from the practice effect) noted in the healthy control group (p< 0.01). On the specific drug vs placebo comparison at week 12, the preferential improvement for drug in overall cognition fell at the margin of significance (p= 0.053); because this fell at the margin of significance and because the sample size was small, this result must be viewed with caution, and requires replication for confirmation of a treatment effect, as

there is a slightly increased risk that this result occurred by chance. In this study, a statistically significant domain by treatment interaction effect was not seen; this indicates that the improvement in cognition was broadly distributed across the six cognitive domains and not specific to the primary domain of memory on which the study was powered. On the planned analysis of the secondary measures of pain, fatigue, and physical functioning, an interaction effect was noted with baseline severity such that the benefit of drug over placebo increased as baseline severity increased; the improvement observed at week 12 was sustained to week 24 for pain and physical functioning. Several patients had potentially serious side effects (e.g., 2 had thrombus formation at the PICC line, 1 had a staphylococcal infection, and 1 had biliary pain and stones that required cholecystectomy). On balance, given the marginal cognitive benefit at week 12, the lack of sustained cognitive benefit to week 24, and the risks, the concluding clinical recommendation from this study was that repeated IV antibiotic therapy followed by no treatment for 14 weeks was not recommended for sustained improvement in cognition.

It has been suggested that the initial improvement in cognition during the time of active drug treatment from baseline to week 12 may have been a "regression to the mean" effect. Fig. (**2**) demonstrates the change in cognition over time for the 3 groups in the Lyme encephalopathy study, indicating a comparable slope of improvement in cognitive performance over 24 weeks for the placebo and healthy control groups vs. a steeper slope of improvement for the drug group from baseline to week 12 followed by a decline to week 24. Regression to the mean is unlikely to account for the preferential improvement in the drug group between baseline and week 12 primarily because after drug was discontinued there was a regression *away* from the mean on the cognitive scores. While regression *toward* the mean is a well-established phenomenon, regression *away* from the mean is not. The improvement during the first 12 weeks (on antibiot-

ics for 10 of these weeks) followed by worsening in the second 12 weeks (off antibiotics) would be more consistent with an active drug treatment effect that was not sustained. Also worth noting is that the differences at baseline in the drug and placebo groups on cognition were not statistically significant.

The planned analysis of the secondary outcome measures of pain, fatigue, and physical impairment produced interaction effects at week 12 favoring drug over placebo as a function of baseline severity, with the drug effect increasing with higher baseline impairment. As a model-based illustration of the study results, Fig. (**3**) illustrates the improvement in pain, fatigue, and physical functioning in the Lyme encephalopathy study as a function of baseline severity. (The middle panel ("Pain") from Fig. (**3**) is reprinted here with permission from the journal Neurology [10]). Significant improvement was sustained among these more impaired patients to week 24 for the measures of pain and physical functioning. These results suggest that for a subgroup of patients with more severe impairment in pain or physical functioning, a course of repeated IV antibiotic therapy can result in sustained long-term benefit.

Limitations of the Lyme encephalopathy study include the small sample size which would limit the ability to detect less robust treatment effects (n=37 for Lyme patients and 18 for healthy controls), the relative treatment refractoriness of the sample (mean amount of prior therapy: IV 2.3 months, oral 7.7 months) which would reduce the likelihood of finding a treatment effect for repeated treatment, and the limited generalizability of the study findings given the extremely rigorous but narrow criteria used to enter patients.

Because this study sample comprised patients with objectively confirmed evidence of cognitive deficits, the distinction between Lyme encephalopathy and post-treatment Lyme disease syndrome was blurred. Although these patients certainly had the typical post-treatment symptoms of pain, fa-
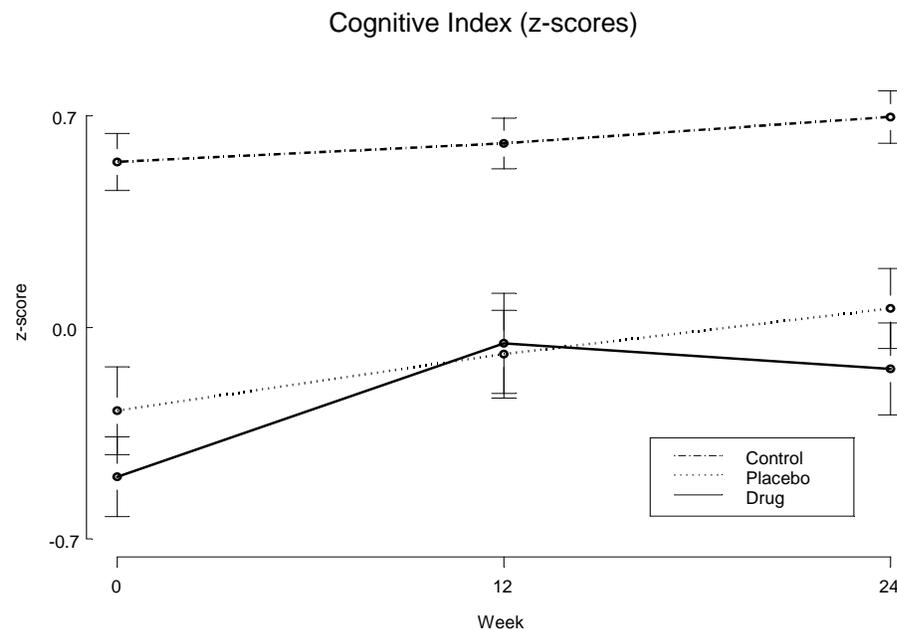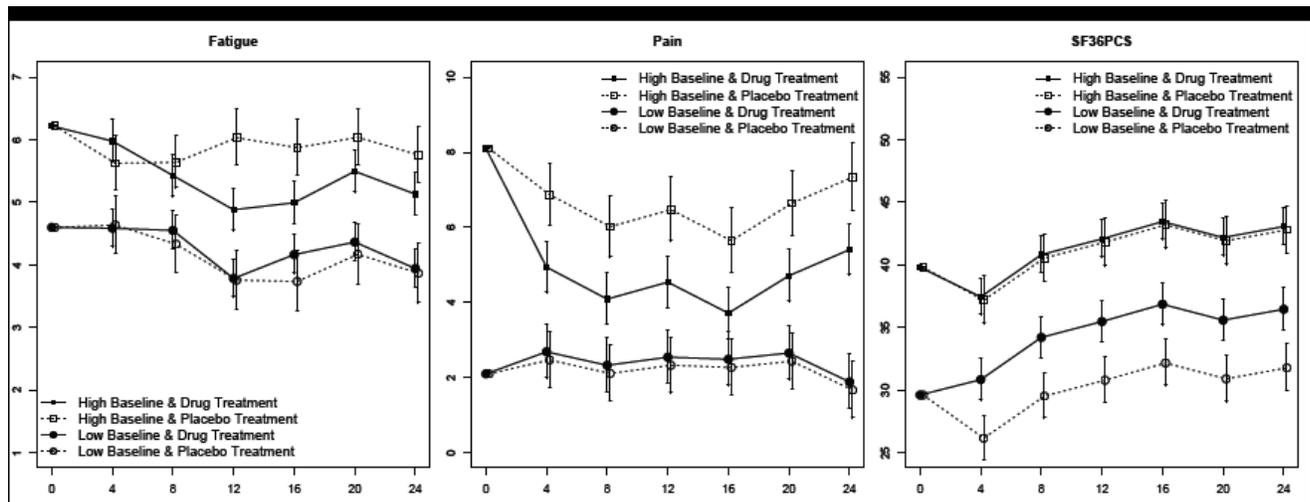

Fig. (2). Change in Cognition over time for 3 groups in Lyme Encephalopathy Study: patients were given 10 wks of IV ceftriaxone or placebo and healthy volunteers were evaluated to assess the practice effect.

Note: High scores on fatigue and pain and low scores on physical functioning indicate greater severity.

Fig. (3). Model-based Illustrations from Lyme Encephalopathy Study: change over time as a function of baseline severity on measures of Fatigue, Pain, and Physical Functioning.

tigue, and physical dysfunction, the presence of objective impairment in memory raises the question of whether these patients would be better defined as having "late Lyme disease" rather than "post-treatment Lyme disease syndrome".

Issues to Consider in Evaluating and Planning Clinical Trials

1. Efficacy vs. Clinically Recommended

Each of the U.S. treatment trials on PTLDS have concluded with the recommendation that the course of therapy tested in each specific trial was not recommended for amelioration of the symptoms indicated by the primary outcome measures. There is a difference however between whether or not a treatment is effective and whether or not a treatment is recommended. For example, in the STOP-LD trial of post-treatment Lyme disease, the treatment was shown to be effective on the primary outcome measure (Fatigue Severity scale) on which the patient's were enrolled; 64% of those who received drug were responders compared to 18.5% of those who received placebo (p<.001). However, the STOP-LD article concluded by not recommending repeated antibiotic treatment, "particularly in light of the frequency of serious adverse events". In other words, the treatment was effective but it carried risks. Presumably if an alternative antibiotic treatment of similar efficacy but improved safety profile could be found, then the authors would have concluded that repeated antibiotic treatment would be recommended. Treatment guidelines that dismiss the research findings as showing no efficacy do an injustice to the evidence and are not helpful to clinicians and patients. For example, a clinician with a patient suffering from disabling post-Lyme fatigue would want to know that the clinical trials provided divergent results to enable a thoughtful discussion of the risks and benefits of repeated antibiotic therapy, as a patient with severe fatigue may decide that the potential benefit of sustained improvement outweighs the risks. Treatment guidelines then should make clear the distinction between *efficacy* demonstrated in a placebo controlled trial and *clinical recommendation* based on a composite consideration which includes efficacy and adverse effects.

2. Treatment Refractoriness of the Sample

Consideration needs to be given to the amount of prior antibiotic therapy received by the recruited sample. For example, a study composed of patients who had 4+ months of prior antibiotic therapy for Lyme disease may come up with negative results. However, if that same study had recruited a less treatment refractory sample (e.g., composed of patients who had received no more than 2 months previously), a treatment effect may be detectable.

3. Sample Size and Power

Efficacy studies are powered based on the degree of expected improvement in the outcome measure of interest. In other words, a sufficient sample size is needed to test a hypothesis and the sample size is determined based on the magnitude of the expected effect. Clinical trials containing only a small sample size would therefore only be able to detect drug effects that are large. As the sample size increases, the study then becomes able to differentiate drug vs placebo differences of lesser magnitude. With this in mind, it is worth noting that the sample sizes employed in the chronic Lyme disease trials have been relatively small and thus would only be able to detect large drug vs placebo differences: Fallon *et al.,* (37 Lyme patients); Krupp *et al.,* (55 patients); Klempner *et al.,* (115 patients). That one (STOP-LD) of these studies had results indicating a significant clinical benefit for repeated antibiotic therapy on a key primary outcome measure is remarkable given the small sample size and reinforces the conclusion that the treatment was an agent of change (rate ratio>3.5).

4. Study Design and Analytic Method

In general, the more homogeneous a sample, the more sensitive will be the trial to test a hypothesis regarding treatment effects. Sample heterogeneity, on the other hand, may result in low signal to noise ratio that makes signal detection more difficult. This issue is of particular concern when sample sizes are small. This is a well known design issue in research studies. When patients are not recruited

based on a particular severity level for the outcome measure of primary interest, it is then possible that the study will incorrectly show no difference between drug and placebo – a Type II error. To address this potential weakness in the study design, if an outcome variable does not have a minimum severity cutoff for enrollment, data analytic methods should assess the moderating effect of baseline severity on outcome. This is usually done by examining the potential interaction between baseline severity and treatment group with regard to outcome. The two studies that demonstrated a treatment effect [7, 10] did address this study design issue, while the two studies that showed no effect [11] did not.

SUGGESTIONS FOR FUTURE TREATMENT GUIDELINES

Based on the evidence cited above, one cannot conclude that repeated antibiotic therapy is ineffective in improving certain symptoms associated with post-treatment Lyme disease syndrome. Nor can it be concluded that repeated antibiotic therapy is robustly effective. One can conclude however that approximately 60% of patients with persistent post-treatment Lyme fatigue may experience meaningful but partial clinical improvement in fatigue with antibiotic retreatment. Guidelines for Lyme disease that address patients with chronic symptoms therefore need to clarify that the controlled trials of additional antibiotic therapy for post-treatment Lyme symptoms have revealed conflicting results, with some studies demonstrating efficacy and others not showing benefit to repeated treatment.

Specifically, the re-analysis of the results of the chronic Lyme trials leads to the following recommendation for the text of treatment guidelines. "IV ceftriaxone therapy is moderately efficacious for patients with chronic ($\geq$6 months) subjective fatigue after recommended antibiotic treatment regimens, but the risk associated with IV antibiotic therapy requires careful discussion with the patient of the cost-benefit ratio. Sustained improvement from IV ceftriaxone therapy for other PTLDS symptoms such as physical dysfunction and pain is uncertain, with positive results suggested by one study but not by other studies."

FINAL COMMENTS

The conclusions of this analysis of the chronic Lyme trials emphasize the benefits of repeated antibiotic therapy for patients with specific chronic symptoms. This is done as a counterbalance to the majority of published guidelines which overlook and/or dismiss the evidence that demonstrates that additional antibiotic therapy can lead to sustained benefit. We hope that our review will lead to more carefully detailed and balanced summaries in future guidelines. However, we also wish to emphasize that while some patients do improve with repeated antibiotic therapy, other patients with persistent symptoms do not. Further, as the clinical trials also demonstrate, antibiotic therapy particularly when given intravenously can put the patient at serious risk.

Biomarkers are needed that can help clinicians to discriminate in advance which patients are more likely to benefit from repeated antibiotic therapy vs. those for whom such treatment is unlikely to be beneficial. Future studies must also begin to address non-antibiotic strategies to help improve persistent symptoms. Recent serologic and CSF studies of patients with post-treatment Lyme disease syndrome suggest that a persistently activated immune response may play a role in the pathophysiology of chronic symptoms [15, 16] . Clarification of whether these findings are of pathogenic relevance and whether this immune activation is due to persistent antigenic stimulation (as might occur from persistent Borrelia) or from a post-infectious autoimmune process would be quite beneficial to clinicians seeking to identify more effective and appropriately targeted treatments for these patients.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1] Wormser GP, Dattwyler RJ, Shapiro ED, *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(9): 1089-134.

[2] Halperin JJ, Shapiro ED, Logigian E, *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(1): 91-102.

[3] British Infection Association. The epidemiology, prevention, investigation and treatment of Lyme borreliosis in United Kingdom patients: a position statement by the British Infection Association. J Infect 2011; 62(5): 329-38.

[4] Mygland ALU, Fingerle V, Rupprecht T, Schmutzhard E, Steiner I. European Federation of Neurological Societies. EFNS guidelines on the diagnosis and management of European Lyme neuroborreliosis. Eur J Neurol 2010; 17(1): 8-16.

[5] Feder HM Jr, Johnson BJ, O'Connell S, *et al*. A critical appraisal of "chronic Lyme disease". N Engl J Med 2007; 357(14): 1422-30.

[6] Lantos P, Charini W, Medoff G, *et al*. Final report of the Lyme disease review panel of the Infectious Diseases Society of America. Clin Infect Dis 2010; 51(1): 1-5.

[7] Krupp LB, Hyman LG, Grimson R, *et al*. Study and treatment of post Lyme disease (STOP-LD): a randomized double masked clinical trial. Neurology 2003; 60(12): 1923-30.

[8] Gaudino EA, Coyle PK, Krupp LB. Post-Lyme syndrome and chronic fatigue syndrome. Neuropsychiatric similarities and differences. Arch Neurol 1997; 54(11): 1372-6.

[9] Cairns V, Godwin J. Post-Lyme borreliosis syndrome: a meta-analysis of reported symptoms. Int J Epidemiol 2005; 34(6): 1340-5.

[10] Fallon BA, Keilp JG, Corbera KM, *et al*. A randomized, placebo-controlled trial of repeated IV antibiotic therapy for Lyme encephalopathy. Neurology 2008; 70(13): 992-1003.

[11] Klempner MS, Hu LT, Evans J, *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(2): 85-92.

[12]   Ware JF, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston: The Health Institute 1994.

[13]   Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Academic Press 1988.

[14]   Turner E, Matthews A, Linardatos E, Tell R, Rosenthal R. Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy. N Engl J Med 2008; 358(3): 252-60.

[15]   Chandra AWG, Marques AR, Latov N, Alaedini A. Anti-Borrelia burgdorferi antibody profile in post-Lyme disease syndrome. Clin Vaccine Immunol 2011; 18(5): 767-71.

[16]   Schutzer SE, Angel TE, Liu T, *et al*. Distinct cerebrospinal fluid proteomes differentiate post-treatment lyme disease from chronic fatigue syndrome. PLoS One 2011; 6(2): e17287.