Statistical Issues When Searching for Predictors of Post–Lyme Disease Symptoms

To the Editor—This letter is in regard to the recently published article by Strle et al [1]. The article covers the important topic of post-Lyme symptoms and it claims that “high T17-associated immune responses, often accompanied by autoantibodies, correlated with post-Lyme symptoms, providing a new paradigm for the study of post-infectious symptoms in a subset of patients with Lyme disease.” Their study compared levels of 15 cytokines and 11 chemokines in 45 subjects with post-Lyme symptoms versus corresponding levels in 41 subjects without post-Lyme symptoms. In addition, the 26 potential predictors were also measured for 22 healthy subjects and further comparisons were made between these subjects and the subjects with Lyme disease. There are 2 major issues with the statistical methods chosen by the authors for analyzing the study data.

The first major issue is the lack of multiplicity adjustments: “Of the 15 cytokines and 11 chemokines tested, the only significant differences between groups were in the levels of CXCL9, CXCL10, and IL-23 [interleukin 23]. Therefore, only the results of these mediators are presented here.” Even if we consider only the comparisons done for the levels at the initial visit between the 45 subjects with and the 41 subjects without post-Lyme symptoms, there would be a total of 26 different comparisons. Using a significance level of .05, random data are expected to yield 1 falsely significant result (5% of 26 tests). To adjust for a large false discovery rate, P values are adjusted in cases with many comparisons: see, for instance, Holm [2]. For Holm’s procedure, the 26 P values would be ordered from smallest to largest and the smallest P value would be multiplied by 26, the second largest by 25, and so on. The authors’ claim that “IL-23 levels were significantly higher in the post-Lyme group at study entry” is based on a P value of .02 (Figure 3A in [1]), which was the only significant P value <.05 out of the 26 performed comparisons. The right analysis would have adjusted that P value to .52, which would be far away from significance. The data graphs themselves show a large overlap in IL-23 levels between the groups, and one can reasonably assume that the difference could be due to simple random variation when testing 26 different mediators.

The second major issue is the fact that “samples with undetectable cytokine levels were not included in the analysis.” A visual count of the IL-23 levels in Figure 2B shows only 18 of the 45 of the erythema migrans (EM) subjects without symptoms graphed. The authors then claim that the median IL-23 of subjects with symptoms is statistically higher than the median IL-23 of subjects without symptoms. When the number of undetectable results is large, one cannot simply ignore those values which are close to 0; in fact, according to the count, the median IL-23 level of the EM subjects with symptoms should be undetectable as 27 of 45 subjects had undetectable levels. Censored data analysis methods such as the ones described in Amemiya [3] should be employed with data of this type.

Note

Potential conflicts of interest. Both authors: No reported conflicts.
Both authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

Mihaela Parvu1 and Valentin Parvu2
1Infectious Disease Department, Stony Brook University Hospital, Stony Brook, New York; and 2Corporate Statistics, Becton Dickinson, Franklin Lakes, New Jersey

References


Correspondence: Mihaela Parvu, MD, Division of Infectious Disease, Stony Brook University Hospital, HSC T 15, RM 080, Stony Brook, NY 11794-8153 (mihaela.parvu@stonybrookmedicine.edu).