Note: Andrew Onderdonk, editor of The Journal of Clinical Microbiology has refused to ask Dumler to dislose the results of his study or explain his erroneous arithmetic. Dumler has not provided results after numerous requests. The rejected inquiry follows:

EVALUATION OF LYME DISEASE TESTS

I would like to thank Dr. Dumler for his reply to my letter expressing concerns regarding the Coulter et. al. study evaluating Lyme disease testing (4). Although he raises many subjective issues, I would prefer to focus on the study in question and the authors' presentation of results. I am uncertain why he states it was not the study's intention to determine sensitivity of Lyme disease tests when sensitivity figures are stated four times in the abstract alone.

This is not an issue of overdiagnosis or underdiagnosis; instead, this is a basic issue of proper presentation of results and their analysis.

Dr. Dumler claims that no history or physical findings can define a true positive for Lyme disease, yet the accepted case definition for an infection with *Borrelia burgdorferi* is the Centers for Disease Control definition that relies heavily on history and physical findings. This case definition is designed to determine near certain cases of Lyme disease and is already biased toward missing clinically significant cases (1). Its intent is to identify only nearly-certain surveillance cases. Instead of using this established definition, it appears Dumler is claiming that his study's only recourse is to define true positives by using the same tests he is attempting to evaluate. Obviously, this methodology is invalid and

produces a statistic that conveys little information. Sensitivity will always be 100%. Further clarification of the study is highly relevant at this time because of the inappropriate application of this study to support published clinical guidelines (6).

The underlying fallacy of using positive test results to define their own true positives is so apparent it hardly needs explanation. If one test had been evaluated using this methodology, sensitivity would obviously be 100%. Any larger number of evaluated tests will always produce a sensitivity of 100% in one combination, and highly inflated sensitivities in other combinations. It is only mildly informative to know how many tests are required to reach the 100% sensitivity figure, regardless of the researchers' pre-observation subjective confidence in the tests.

Dumler's study provides an indication only of how well the evaluated tests agree with each other within their own universe of tests with positive results, completely disassociated from any real world correlations. A test being evaluated should never be used in any manner to determine its own true positive. The results will always be biased and vary only in degree of bias. When only six tests are evaluated, with some contributing few unique positives as in Dumler's study, bias is fairly extreme.

The sensitivity and selectivity of the evaluated tests have not been determined with any convincing level of confidence (2, 5). A properly conducted and presented study regarding testing for Lyme disease could contribute to our base of knowledge. Dumler's

study, as presented, unfortunately, contributes little (except for making a weak case for perhaps not ordering plasma PCR tests) when it could have contributed much more. Very few studies validating Lyme tests have been conducted in the last ten years, particularly in the U.S.

If the authors had compared tests, such as serologies, to the accepted gold standards, positive culture or PCR (3), we would have an estimate of the specificities of these other tests. Of course, this does nothing to estimate sensitivity of any of the tests studied. At our current level of knowledge and technique, our estimates of sensitivity must depend on the accepted case definitions, instead of an artificial definition of infection based on the same tests we are attempting to evaluate.

Rather than produce highly biased, nearly meaningless statistics of test correlation I would encourage the study's authors to publish their study's results in terms of true positives, false positives, true negatives, and false negatives using the source data they summarized in their Table 1 where true positives and true negatives were lumped together to produce an "agreement" figure. If the message of their study is that Lyme disease tests have little clinical relevance, this is the fact that should be emphasized rather than the published statements of highly-biased sensitivity and selectivity figures.

As presented, the study does have one important message: if a physician required a positive test for diagnosis and only used the two-tier ELISA followed by Western Blot

testing procedure (currently these are the only tests routinely ordered in clinical settings), 11 cases of certain infection would be entirely missed in the 86 subjects. Those 11 (13%) were positive by our present day gold standards, culture and PCR results, yet had negative serologies. This statistic furthermore ignores all subjects not positive on any test, a number not stated in the study as presented, but quoted in the abstract as 49/86 (57%). If only 20% of these subjects (who were selected because they exhibited symptoms suggestive of Lyme disease) were truly infected, the physician relying on two-tier serology testing would miss 21/86 (24%) cases of a potentially disabling or fatal disease.

If the study's authors cannot bring themselves to use the established case definition of Lyme disease to define a true positive, they could perhaps use terms like positive-agreement, negative-agreement, and so on. The authors' audience wants to know the test results for the cohorts of probable, possibly, and unlikely infected subjects so readers can form their own opinions as to the validity and usefulness of the tests in question. The study's discussion scenario of a clinician requiring a positive test for diagnosis and searching for the most efficient set of tests to produce a possible positive result is irrelevant, especially when *B. burgdorferi* cultures are available only in research settings.

The authors are encouraged to carefully review their figures. Dumler's reply to my earlier letter states that the agreement figure of 32 subjects came from, '8 subjects initially seropositive with probable Lyme disease and also 43 initially seronegative assessed as NOT "probable" for Lyme disease.' The sum of 8 plus 43 does not equal 32.

Joel Spinhirne

Verim Research

Hood River, Oregon 97031

Phone: (815) 846-1642

Fax: (815) 642-0056

Email: joel@verimresearch.com

REFERENCES

1. 1997. Case definitions for infectious conditions under public health surveillance. Centers for Disease Control and Prevention. MMWR Recomm Rep **46:**1-55.
2. **Brown, S. L., S. L. Hansen, and J. J. Langone.** 1999. Role of serology in the diagnosis of Lyme disease. Jama **282:**62-6.
3. **Bunikis, J., and A. G. Barbour.** 2002. Laboratory testing for suspected Lyme disease. Med Clin North Am **86:**311-40.

4. **Coulter, P., C. Lema, D. Flayhart, A. S. Linhardt, J. N. Aucott, P. G. Auwaerter, and J. S. Dumler.** 2005. Two-year evaluation of Borrelia burgdorferi culture and supplemental tests for definitive diagnosis of Lyme disease. J Clin Microbiol **43:**5080-4.

5. **Mouritsen, C. L., C. T. Wittwer, C. M. Litwin, L. Yang, J. J. Weis, T. B. Martins, T. D. Jaskowski, and H. R. Hill.** 1996. Polymerase chain reaction detection of Lyme disease: correlation with clinical manifestations and serologic responses. Am J Clin Pathol **105:**647-54.

6. **Wormser, G. P., R. J. Dattwyler, E. D. Shapiro, J. J. Halperin, A. C. Steere, M. S. Klempner, P. J. Krause, J. S. Bakken, F. Strle, G. Stanek, L. Bockenstedt, D. Fish, J. S. Dumler, and R. B. Nadelman.** 2006. 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 **43:**1089-134.