

## NOTES

# Interlaboratory Reliability of Microimmunofluorescence Test for Measurement of *Chlamydia pneumoniae*-Specific Immunoglobulin A and G Antibody Titers

Alyson J. Littman,<sup>1,2\*</sup> Lisa A. Jackson,<sup>1,3</sup> Emily White,<sup>1,2</sup> Mark D. Thornquist,<sup>2</sup>  
Charlotte A. Gaydos,<sup>4</sup> and Thomas L. Vaughan<sup>1,2</sup>

Department of Epidemiology, University of Washington,<sup>1</sup> Division of Public Health Sciences, Fred Hutchinson Cancer Research Center,<sup>2</sup> and Center for Health Studies, Group Health Cooperative,<sup>3</sup> Seattle, Washington, and Division of Infectious Diseases, Johns Hopkins University, Baltimore, Maryland<sup>4</sup>

Received 17 September 2003/Returned for modification 29 October 2003/Accepted 23 January 2004

**To evaluate the reliability of *Chlamydia pneumoniae*-specific immunoglobulin G (IgG) and IgA antibody titers as measured by the microimmunofluorescence (MIF) test, we compared results from 392 individuals using a standard MIF protocol at two academic laboratories. The kappas for dichotomous titers ( $\geq 16$  versus  $< 16$ ) were 0.39 for IgA and 0.53 for IgG. Measurement error likely attenuates *C. pneumoniae*-disease associations; the magnitude of attenuation can be estimated from results of studies such as this one.**

The microimmunofluorescence (MIF) test measures *Chlamydia pneumoniae*-specific antibodies quantitatively and is the serologic assay of choice to detect antibodies to *C. pneumoniae* (5, 11). However, the MIF test is technically demanding and requires expertise in fluorescence microscopy. Findings can vary with reagents, antigens, incubation time and temperature, the microscope, and experience of the technician. Studies have linked infection with *C. pneumoniae* (as defined by titers of immunoglobulin G [IgG] or IgA antibodies) with an increased risk of several chronic diseases. However, results have been inconsistent (3, 4). It is important to understand if the weak associations observed in some studies are due to measurement error in detection of IgA or IgG antibody levels or to true weak or null effects. Because no “gold standard” for antibody measurement exists, we are unable to measure validity directly. Thus, a key measure of error in IgA and IgG antibody levels is the reliability of repeated testing of specimens on the same subjects at different laboratories.

We are investigating the relationship between *C. pneumoniae* IgA titers of  $\geq 16$  and subsequent lung cancer risk. Circumstances required us to switch laboratories after about 40% of the samples had been assayed. We chose to redo all of the assays at the second laboratory, providing us with the opportunity to assess the reliability of IgA and IgG antibody titers at two laboratories.

We selected 392 serum samples from participants in a randomized trial designed to study the effectiveness of supplemental beta-carotene and retinyl palmitate in preventing lung cancer incidence and mortality (10). Participants in the parent

study ranged in age from 50 to 74 years at entry and were current or former smokers (women and men), persons with significant occupational exposure to asbestos (men only), or both. Sixty percent of reliability study participants were male, and the mean age was 58.5 years; 72% were current smokers, and the median pack-years among current and former smokers was 43. Informed consent was obtained from patients, and human experimentation guidelines of the U.S. Department of Health and Human Services and/or those of the authors' institutions were followed in the conduct of the clinical research.

MIF tests to detect *C. pneumoniae*-specific IgA and IgG antibodies were conducted at two academic laboratories, A and B, that are leaders in the development and implementation of the MIF test. The MIF test is an indirect fluorescent antibody test that measures antibodies to epitopes in the cell wall of *C. pneumoniae* elementary bodies, and it is described in detail elsewhere (11). Antigens used in both laboratories were obtained from Washington Research Foundation. Titers are expressed as reciprocals of serum dilutions. Samples were tested in serial twofold dilutions, from 8 or 16 (varying by laboratory) to 1,024. For consistency, titers of 8 were categorized as  $< 16$ , because we assumed that samples with antibodies measured at 8 (but not 16) would be categorized as  $< 16$  at the other laboratory. Titers of 16 or greater were considered seropositive. A positive reaction required typical fluorescence associated with evenly distributed *Chlamydia* organisms. A single technician from each laboratory tested all samples at that laboratory and was blinded to the results from the other laboratory. Positive and negative controls were included in all test series. Ten percent of samples were retested at laboratory B; 90% were found to be within one dilution of the first test.

We calculated percent agreement that was exact or within a twofold dilution (e.g., titer values of 32 and 128 would be considered within a twofold dilution of 64), in order to be

\* Corresponding author. Mailing address: Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. North, M4 B402, Seattle, WA 98109-1024. Phone: (206) 667-7707. Fax: (206) 667-7850. E-mail: alittman@fhcrc.org.

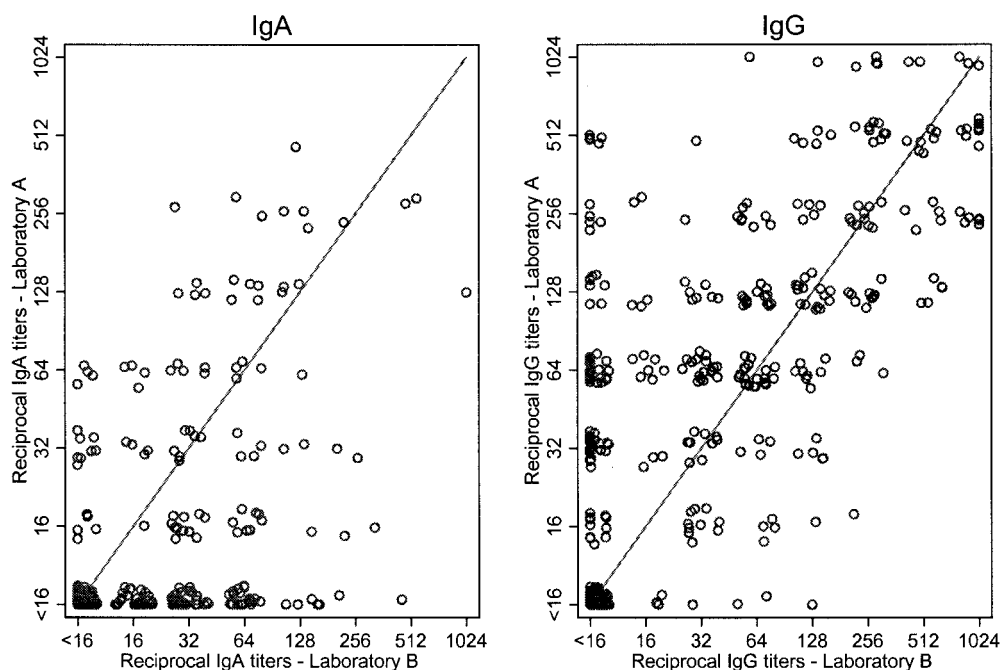


FIG. 1. Comparison of IgA and IgG antibody titers to *C. pneumoniae* at laboratories A and B ( $n = 392$ ). The lines indicate perfect agreement. The symbols have been shifted slightly to show detail.

compatible with the results from other studies (2, 7, 8). To evaluate the amount of agreement observed beyond that expected by chance, we also calculated the weighted kappa statistic, with 95% confidence intervals (1). Kappa equals one when there is exact agreement for the two measures for all subjects. Weighted kappa gives a higher reliability, i.e., “partial credit,” to titers that are close. Since we planned to analyze our data as a dichotomous variable ( $\geq 16$  versus  $< 16$ ), we calculated kappa using these categories as well.

While laboratory B was more likely to detect IgA antibody titers of  $\geq 16$  (45% versus 24% at laboratory A), laboratory A was more likely to detect IgG antibody titers of  $\geq 16$  (77% versus 59% at laboratory B) (Fig. 1). Kappas were higher for IgG than IgA and for the categorical measures than the dichotomous measures (Table 1). Specifically, kappas for the dichotomous measures were 0.39 for IgA antibodies and 0.53 for IgG antibodies.

One laboratory screened at a dilution of 1:8, while the other screened at 1:16. Because we chose to combine titers of 8 with those that were not detectable at a titer of 16, we investigated

whether this may have affected reliability. Nearly 5% ( $n = 18$ ) of IgA antibodies and 3.6% ( $n = 14$ ) of IgG antibodies had titers of 8. When retested at the other laboratory, 12 (66.7%) and 11 (78.6%) specimens tested for IgA and IgG antibodies, respectively, had titers of  $< 16$ . If we had instead chosen to exclude samples with values of 8, the weighted kappa would have been slightly higher for IgA (0.52) and slightly lower for IgG (0.66).

We are unable to explain the reason for the differences in the results from the two laboratories. Both technicians had participated in the same training workshop. The reagents and antigens were of the same type and purchased from the same supplier. However, a noted limitation of the MIF test is that it is difficult to perform, and reading of slides is subjective. Regardless, in the absence of a gold standard, it is not possible to adjudicate differing results. Thus, we assumed that both laboratories measured *C. pneumoniae*-specific antibodies with equivalent error. Because laboratories A and B are leaders in the conduct of the MIF assay, we anticipate the kappa between them is higher than it would be between less-experienced laboratories.

TABLE 1. Agreement and kappa statistics for IgA and IgG titers in serum specimens from 392 individuals evaluated by laboratories A and B

Antibody measurement <sup>a</sup>	IgA			IgG		
	% Exact agreement	% Agreement within a twofold dilution	Kappa (95% CI)	% Exact agreement	% Agreement within a twofold dilution	Kappa (95% CI)
Categorical	55	75	0.50 (0.41–0.59) <sup>b</sup>	38	66	0.68 (0.58–0.77) <sup>b</sup>
Dichotomous	71		0.39 (0.30–0.48) <sup>c</sup>	79		0.53 (0.44–0.62) <sup>c</sup>

<sup>a</sup> Categorical measurements were, for example,  $< 16$ , 16, 32, 64, etc., whereas the dichotomous measurements were  $\geq 16$  versus  $< 16$ .

<sup>b</sup> Weighted kappa using quadratic weights. CI, confidence interval.

<sup>c</sup> Unweighted kappa.

Only one other interlaboratory reliability study of *C. pneumoniae*-specific titers using MIF testing has been published (8). The investigators of that study focused on the reliability of detecting stable low IgG antibody titers and IgM antibodies in a single sample or fourfold changes in IgG antibody titers in paired samples. Each of 14 participating laboratories analyzed up to 10 serum sets, where each set included two or three samples from the same individual at different times (8). The overall percent agreement, defined as differing by less than a twofold dilution with the reference titers, was 80%. However, because expected agreement might also be high, percent agreement alone may present an overly optimistic picture of the reliability between laboratories. Suppose, for example, that one laboratory classified 75% of samples as positive and 25% as negative, but a second laboratory, which had problems with reagents, called all samples positive. The agreement would be 75%, which does not reflect the poor reliability of the measure. Observed agreement for IgA in our study was nearly as high (75%) yet kappa was 0.50, indicating only 50% agreement beyond that expected by chance alone (6).

Strengths of our study include its size, examination of the reliability of IgA antibody titers, and inclusion of laboratories that conduct MIF tests frequently. Limitations include the use of only two laboratories and the fact that we assessed reliability, when ideally we would have assessed validity (i.e., the ability of a test to measure the true exposure of interest). Regardless, a measure which lacks reliability also must lack validity, and information on reliability can be used to determine attenuation of measures of association in epidemiologic studies. For example, if the true relative risk (RR) for the association between lung cancer and *C. pneumoniae*-specific antibody titers of  $\geq 16$  was 2.5, reliability of the magnitude observed in this study (kappa = 0.39) would result in an attenuated RR of at most  $2.5^{\sqrt{0.39}}$ , or an RR equal to or less than 1.8 (9).

In conclusion, we found sufficient lack of reliability of IgA and IgG titer levels from MIF tests to obscure moderate epidemiologic associations. Inconsistent results or false null findings of *C. pneumoniae*-disease relations may be explained in part by errors in measuring antibody titers from MIF testing. Further research to develop reliable and accurate measures of infection is warranted. Development of accurate commercial kits or additional technician training and calibration of slide reading may help to standardize MIF tests. Furthermore, au-

tomated techniques, which will eliminate the subjective reading of test results, may also provide an important improvement.

This study was made possible through support from the National Cancer Institute, grant U01 CA63673-07S2.

We thank the laboratory technicians, especially Billie Jo Wood, Jeff Holden, and Elsie Lee, whose efforts made this study possible. Cim Edelstein, Matt Barnett, Stacy Cannon, Paula Filipic, Grace Powell, and Marty Stiller provided administrative and technical support.

We have no commercial or other association that might pose a conflict of interest (e.g., pharmaceutical stock ownership or consultancy).

#### REFERENCES

1. **Armstrong, B. K., E. White, and R. Saracci.** 1992. Principles of exposure measurement in epidemiology. Oxford University Press, Oxford, England.
2. **Bennedson, M., L. Berthelsen, and I. Lind.** 2002. Performance of three microimmunofluorescence assays for detection of *Chlamydia pneumoniae* immunoglobulin M, G, and A antibodies. *Clin. Diagn. Lab. Immunol.* **9**:833–839.
3. **Danesh, J., P. Whincup, S. Lewington, M. Walker, L. Lennon, A. Thomson, Y. K. Wong, X. Zhou, and M. Ward.** 2002. *Chlamydia pneumoniae* IgA titres and coronary heart disease; prospective study and meta-analysis. *Eur. Heart J.* **23**:371–375.
4. **Danesh, J., P. Whincup, M. Walker, L. Lennon, A. Thomson, P. Appleby, Y. Wong, M. Bernardes-Silva, and M. Ward.** 2000. *Chlamydia pneumoniae* IgG titres and coronary heart disease: prospective study and meta-analysis. *BMJ* **321**:208–213.
5. **Dowell, S. F., R. W. Peeling, J. Boman, G. M. Carlone, B. S. Fields, J. Guarner, M. R. Hammerschlag, L. A. Jackson, C. C. Kuo, M. Maass, T. O. Messmer, D. F. Talkington, M. L. Tondella, and S. R. Zaki.** 2001. Standardizing *Chlamydia pneumoniae* assays: recommendations from the Centers for Disease Control and Prevention (USA) and the Laboratory Centre for Disease Control (Canada). *Clin. Infect. Dis.* **33**:492–503.
6. **Landis, J. R., and G. G. Koch.** 1977. The measurement of observer agreement for categorical data. *Biometrics* **33**:159–174.
7. **Messmer, T. O., J. Martinez, F. Hassouna, E. R. Zell, W. Harris, S. Dowell, and G. M. Carlone.** 2001. Comparison of two commercial microimmunofluorescence kits and an enzyme immunoassay kit for detection of serum immunoglobulin G antibodies to *Chlamydia pneumoniae*. *Clin. Diagn. Lab. Immunol.* **8**:588–592.
8. **Peeling, R. W., S. P. Wang, J. T. Grayston, F. Blasi, J. Boman, A. Clad, H. Freidank, C. A. Gaydos, J. Gnarp, T. Hagiwara, R. B. Jones, J. Orfila, K. Persson, M. Puolakkainen, P. Saikku, and J. Schachter.** 2000. *Chlamydia pneumoniae* serology: interlaboratory variation in microimmunofluorescence assay results. *J. Infect. Dis.* **181**(Suppl. 3):S426–S429.
9. **Tavare, C. J., E. L. Sobel, and F. H. Gilles.** 1995. Misclassification of a prognostic dichotomous variable: sample size and parameter estimate adjustment. *Stat. Med.* **14**:1307–1314.
10. **Thornquist, M. D., G. S. Omenn, G. E. Goodman, J. E. Grizzle, L. Rosenstock, S. Barnhart, G. L. Anderson, S. Hammar, J. Balmes, M. Cherniack, et al.** 1993. Statistical design and monitoring of the Carotene and Retinol Efficacy Trial (CARET). *Control. Clin. Trials* **14**:308–324.
11. **Wang, S.** 2000. The microimmunofluorescence test for *Chlamydia pneumoniae* infection: technique and interpretation. *J. Infect. Dis.* **181**(Suppl. 3):S421–S425.