Inter-laboratory variability in multiplexed pneumococcal antibody testing

David C. LaFon, MD[a], **Moon H. Nahm, MD**[a,b]

[a]Division of Pulmonary, Allergy, and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

[b]Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA

Summary

This pragmatic, updated assessment of variability among pneumococcal antibody assays suggests that variability may now be greater than previously reported, and potentially influential in clinical decision making.

Keywords

Streptococcus pneumoniae; vaccine; IgG; antibody; multiplex; assay; variability; diagnosis; immune; deficiency

To the Editor:

Assays that measure levels of IgG antibodies against specific serotypes of *Streptococcus pneumoniae* were initially developed and subsequently refined primarily for evaluation of vaccine efficacy (1). Since polysaccharide vaccines represent a well-defined immunogenic stimulus, measurement of serotype-specific antibody production in response to vaccination with the pneumococcal polysaccharide vaccine (PPV23) has also been used to evaluate humoral immune function in cases of suspected primary immune deficiency disease (PIDD) (2). For diagnosis of PIDD, the American Academy of Allergy, Asthma and Immunology (AAAAI) defines a normal response to PPV23 as the conversion of 70% of serotypes tested to 1.3 μg/ml with at least a 2-fold rise in antibody level (2). However, the 1.3 μg/ml threshold was developed based on limited evidence using an analytical method (radioimmunoassay) which is no longer in use (3), and different thresholds may be used elsewhere (4, 5). A serotype-specific IgG level of 0.35 μg/ml is used as a threshold for

comparison in vaccine studies, based on studies of invasive pneumococcal disease in infants who had received the pneumococcal conjugate vaccine (PCV) (6) but has been used for diagnostic evaluation as well (4). IgG antibody levels that correspond with protection from pneumococcal infections in adults are not well-known, and vary by serotype, age, and type of infection.

Although the standardized World Health Organization (WHO) enzyme-linked immunosorbent assay (ELISA) remains the gold standard for measurement of pneumococcal antibodies (1), performing ELISA for evaluation of suspected PIDD is impractical, since a separate assay must be performed for each serotype. Multiplex bead array assays are widely utilized for clinical applications since they allow for analysis of multiple serotypes in a single assay run. These assays are not typically performed locally by clinical laboratories, and samples are usually sent to one of relatively few large commercial laboratories in the US that provide clinical testing services. Prior studies have compared the results generated by different assays that included up to 14 serotypes. These studies found that assays performed by different laboratories can produce different results (7, 8) although one study suggested that this variability had little effect on overall diagnostic outcomes (7). In the time since these studies were performed, laboratories have expanded assays to measure antibodies to all 23 serotypes included in PPV23.

To gain a current and pragmatic assessment of inter-laboratory variability as it applies to clinical diagnostic evaluation, with approval from the Institutional Review Board of the University of Alabama at Birmingham we obtained serum samples from 10 healthy adults aged 19–48 years old. No participants reported ever having received PPV23, while one participant may have received PCV as a young child. Split serum samples were stored frozen, and sent to the three large commercial reference laboratories (referred to as A, B, and C) that perform testing for the vast majority of clinical samples in the U.S. Storage and shipping of samples followed each laboratory's published guidelines for handling of clinical specimens. Each laboratory analyzed the samples using its 23-serotype *Streptococcus pneumoniae* IgG multiplexed immunoassay. Although the three laboratories each developed and validated their assays in accordance with Clinical Laboratory Improvement Amendments (CLIA) requirements, none of the assays have been cleared or approved by the US Food and Drug Administration (FDA). Pre-absorption steps to remove non-functioning antibodies, as well as reference sera used to calibrate the assays could differ among the three laboratories.

When comparing antibody levels from each laboratory using Spearman's correlation coefficient (r), correlation was found to be variable by comparison and by serotype (Figure 1, Table I). Strength of correlation ranged from $r = -0.48$ to $r = 0.93$. The laboratory A vs C comparison had the greatest number of serotypes (10/23) with strongly positive correlation (r 0.70), with 2/23 and 5/23 serotypes meeting this criterion for the A vs B and B vs C comparisons, respectively. The A vs B comparison had the greatest number of serotypes (16/23) with r 0.50.

Comparison of pneumococcal antibody levels using the Wilcoxon Signed Rank Test demonstrated a systematic tendency for laboratory C to produce the lowest measurements,

with laboratory B generating the highest. Between laboratories A and B, laboratory B had significantly higher IgG levels than laboratory A for 13/14 of serotypes with a significant difference in IgG levels (P < 0.05). Between laboratories A and C, laboratory A had significantly higher IgG levels than laboratory C for 11/13 total serotypes with a significant difference in IgG levels. Between laboratories B and C, laboratory B had significantly higher IgG levels than laboratory C for 19/19 of the serotypes with a significant difference in IgG levels.

A threshold of 1.3 μg/ml was used to classify individual measurements as "protective" or "nonprotective" (Figure 2a). 83% of the discrepancies in classification between A and B were nonprotective by A and protective by B. 89% of the discrepancies between A and C were protective by A and nonprotective by C. 96% of the discrepancies between B and C were protective by B and nonprotective by C. Overall, there was agreement in "protective" versus "nonprotective" classification for 56% of all individual measurements. Overall, laboratory A classified 3/10 samples as "protected" (IgG   1.3 μg/ml for at least 70% of serotypes), compared to 6/10 for laboratory B and 0/10 for laboratory C (Figure 2a). Use of 0.35 μg/ml as a cutoff (Figure 2b) increased overall agreement in individual antibody measurements to 79%, with 89% agreement between A and B, 74% agreement between A and C, and 73% agreement between B and C. However, there was still evidence of strong directional bias among discordant classifications of individual values which was similar to that which was observed using a 1.3 μg/ml cutoff.

The results of our study indicate that there is substantial variability in the results of current 23-serotype multiplex pneumococcal antibody assays performed by different laboratories. This variability may be greater than previously reported, and has the potential to affect clinical decision making. As was noted in prior studies (7, 8), we found correlation between laboratories to be variable based on serotype and specific inter-laboratory comparison. However, our study found overall weaker inter-laboratory correlations than previously reported (7, 8). Comparison of IgG levels demonstrated systematic bias for each of the three laboratories to produce higher or lower results when compared with the others. This bias affected the classification of results using clinical diagnostic criteria. In 2013 Zhang *et al.* reported directional bias for comparisons involving one of the three assays included in that study (7); however we observed directional bias across all three comparisons. We also found a lower rate of concordance in "protective" antibody levels for individual measurements than previously observed (56% vs 81%) (7), and lower rate of agreement in overall classification of samples across all three assays (40% vs 82%) (7). Variability in results was observed for the serotypes unique to PPV23 which were recently added, as well as those included in 14-serotype multiplex assays. This suggests that the expansion from 14 to 23 serotypes did not mitigate the effects of variability in diagnostic vaccination.

A significant caveat to our analysis is the use of a 1.3 μg/ml as a "protective" threshold. Despite recent investigation of serotype-specific reference values (5), the ideal approach for applying thresholds across different assays remains unclear. Substantial inter-laboratory variability and lack of standard methods for assigning reference values currently preclude widespread use of such values. One of the three laboratories in our study provided serotype-specific reference values, while the other two indicated that 1.3 μg/ml could be used for all

serotypes. While use of 0.35 μg/ml improved inter-laboratory agreement, there is a paucity of evidence to support the use of any single threshold to define a "protective" or "normal" response to diagnostic vaccination with PPV23. We opted to use the AAAAI threshold to enable comparisons between all three laboratories and all 23 serotypes. Our sample was derived from a small number of healthy, predominantly pneumococcal vaccine-naïve individuals. While we did not examine post-vaccine sera, a study of paired pre and post-vaccine samples by Daly *et al.* found better agreement in diagnostic classification among the pre-vaccine samples when using a similar threshold-based system (8). A recent study by Hajjar *et al.* also found substantial inter-laboratory variability between two different 14-serotype multiplex assays, using sera from 48 subjects, some of whom had recurrent infections or impaired immune response and had received pneumococcal vaccines (9).

Our findings present a clear dilemma to clinicians who utilize multiplex bead array assays for the diagnosis of suspected immune deficiency syndromes. Inter-laboratory variability contributes to the existing uncertainty in defining normal versus abnormal responses to diagnostic vaccination. Establishing more uniform assay techniques, standard protocols for establishing assay-specific reference values, and standardization of results using reference sera could be investigated as potential means of mitigating the effects of variability. Clinicians must account for potential variability when applying generalized criteria to the results of diagnostic vaccination with PPV23. AAAAI guidelines recommend integration of vaccine response data with additional clinical findings and compatible history of recurrent infections in the diagnosis of PIDD (2). The results of this study underscore the importance of these recommendations in the evaluation of suspected immune dysfunction.

Acknowledgments

Abbreviations

PPV23	Pneumococcal polysaccharide vaccine
PIDD	primary immune deficiency disease
AAAAI	American Academy of Allergy, Asthma and Immunology
PCV	pneumococcal conjugate vaccine
WHO	World Health Organization
ELISA	enzyme-linked immunosorbent assay
CLIA	Clinical Laboratory Improvement Amendments
FDA	US Food and Drug Administration

References

1. Wernette CM, Frasch CE, Madore D, Carlone G, Goldblatt D, Plikaytis B, et al. Enzymelinked immunosorbent assay for quantitation of human antibodies to pneumococcal polysaccharides. Clin Diagn Lab Immunol. 2003;10(4):514–9. [PubMed: 12853378]

2. Orange JS, Ballow M, Stiehm ER, Ballas ZK, Chinen J, De La Morena M, et al. Use and interpretation of diagnostic vaccination in primary immunodeficiency: a working group report of the Basic and Clinical Immunology Interest Section of the American Academy of Allergy, Asthma & Immunology. J Allergy Clin Immunol. 2012;130(3 Suppl):S1–24. [PubMed: 22935624]

3. Sorensen RU, Leiva LE, Javier FC 3rd, Sacerdote DM, Bradford N, Butler B, et al. Influence of age on the response to Streptococcus pneumoniae vaccine in patients with recurrent infections and normal immunoglobulin concentrations. J Allergy Clin Immunol. 1998;102(2):215–21. [PubMed: 9723664]

4. Beck SC. Making sense of serotype-specific pneumococcal antibody measurements. Ann Clin Biochem. 2013;50(Pt 6):517–9.

5. Schaballie H, Bosch B, Schrijvers R, Proesmans M, De Boeck K, Boon MN, et al. Fifth Percentile Cutoff Values for Antipneumococcal Polysaccharide and Anti-Salmonella typhi Vi IgG Describe a Normal Polysaccharide Response. Frontiers in immunology. 2017;8:546. [PubMed: 28553290]

6. WHO. Recommendations for the production and control of pneumococcal conjugate vaccines. WHO technical report series. 2005;927; http://www.who.int/biologicals/publications/trs/areas/vaccines/pneumo/ANNEX%202%20PneumococcalP64-98.pdf:64-98.

7. Zhang X, Simmerman K, Yen-Lieberman B, Daly TM. Impact of analytical variability on clinical interpretation of multiplex pneumococcal serology assays. Clin Vaccine Immunol. 2013;20(7):957–61. [PubMed: 23677324]

8. Daly TM, Pickering JW, Zhang X, Prince HE, Hill HR. Multilaboratory assessment of threshold versus fold-change algorithms for minimizing analytical variability in multiplexed pneumococcal IgG measurements. Clin Vaccine Immunol. 2014;21(7):982–8. [PubMed: 24807051]

9. Hajjar J, Al-Kaabi A, Kutac C, Dunn J, Shearer WT, Orange JS. Questioning the accuracy of currently available pneumococcal antibody testing. J Allergy Clin Immunol. 2018.
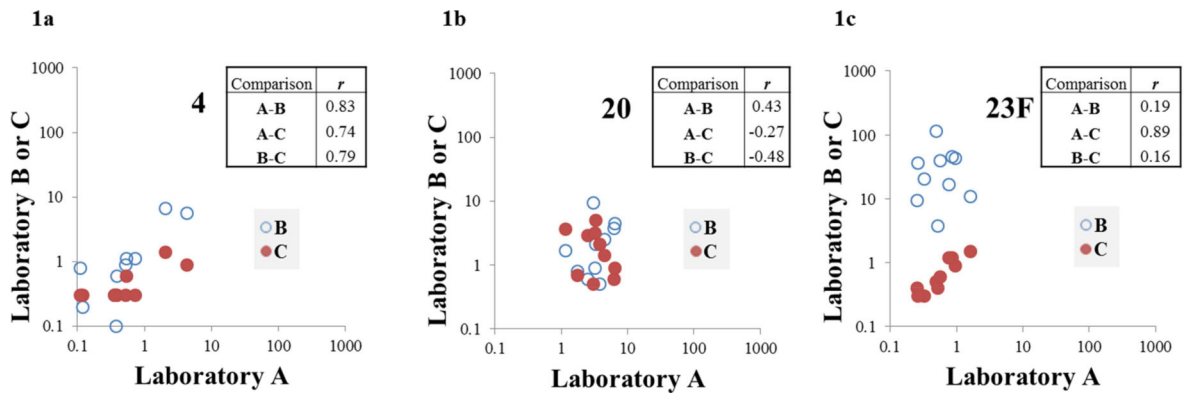
Figure 1. Inter-laboratory correlation results for serotypes 4 (1a), 20 (1b), and 23F (1c).
X axis values represent pneumococcal IgG antibody levels (µg/ml) for laboratory A, with Y axis values representing antibody levels from laboratories B (open circle) or C (solid circle). Boxes show correlation coefficients (*r*) for each inter-laboratory comparison. There was strong correlation (*r* 0.70) across all three comparisons for only one serotype (serotype 4, Figure 1a). There was weak correlation (*r* 0.50) for all three comparisons for four serotypes (including serotype 20, Figure 1b). There was strong correlation between A and C, with weaker correlation for other comparisons for eight serotypes (including serotype 23F, Figure 1c).
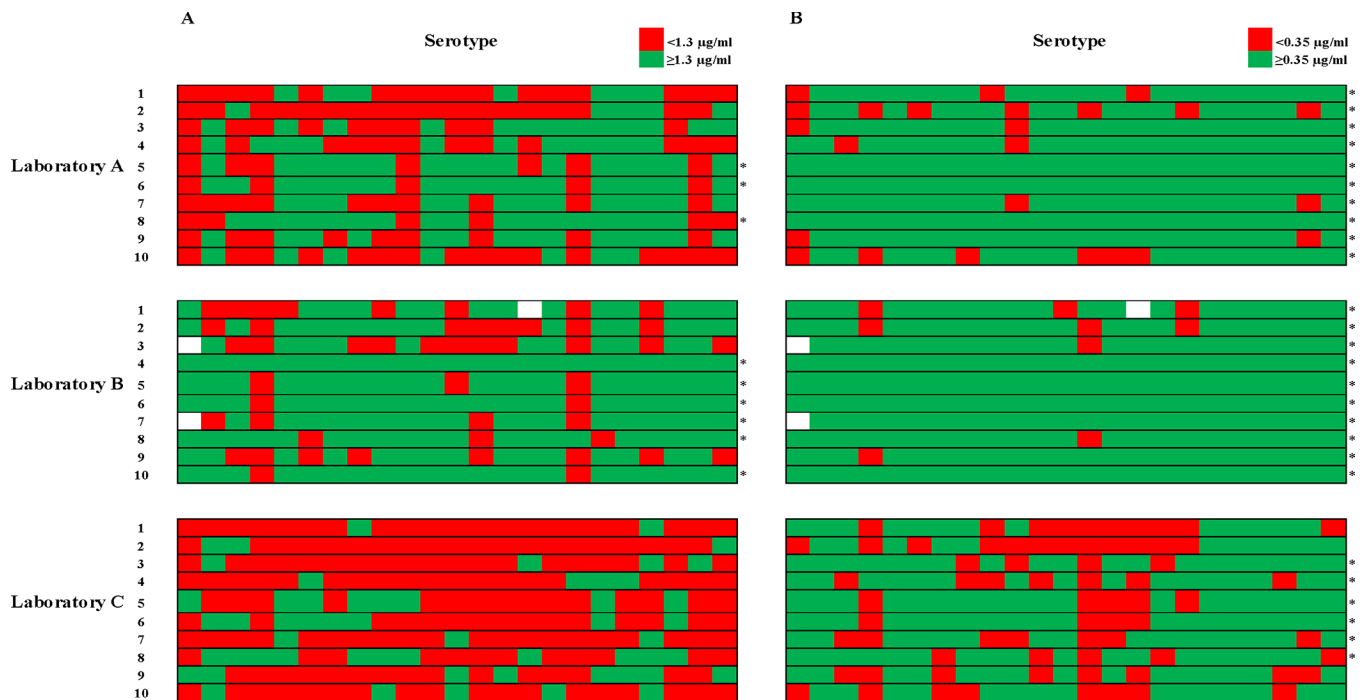
Figure 2. Protective versus nonprotective classification of pneumococcal IgG levels by different laboratories.

Individual measurements were categorized as nonprotective or protective on the basis of the specified thresholds, and are represented by red and green shaded boxes, respectively. Numbered rows represent serum samples, with columns representing serotypes. An asterisk to the right of a sample's row indicates overall "protected" status for that sample (at least 70% of serotypes at or above the protective threshold). Empty boxes indicate three missing measurements from laboratory B (serotype 1 from two samples, and 15B from another), which was unable to analyze these serotypes due to a non-linear dilution response for those samples.

2a. Classification of pneumococcal IgG levels based on AAAAI criteria. A threshold of 1.3 μg/ml was used to categorize individual measurements as nonprotective (<1.3 μg/ml, red shaded boxes) or protective (1.3 μg/ml, green shaded boxes).

2b. Classification of pneumococcal IgG levels based on the WHO recommendations for pneumococcal vaccine evaluation. A threshold of 0.35 μg/ml was used to categorize individual measurements as nonprotective (<0.35 μg/ml, red shaded boxes) or protective (0.35 μg/ml, green shaded boxes).

Table I.

Inter-laboratory correlation for pneumococcal antibody measurements.

	r (Spearman's Correlation)		
	A vs B	A vs C	B vs C
1	0.33	0.61	0.60
3	0.39	0.89	0.42
4	0.83	0.74	0.79
5	0.56	0.78	0.60
6B	0.53	0.90	0.63
7F	0.31	0.40	0.06
8	0.62	0.93	0.67
9N	0.48	0.92	0.31
9V	0.42	0.90	0.62
12F[*]	−0.03	-	-
14	0.36	0.48	0.28
18C	0.88	0.68	0.67
19A	0.19	0.24	0.64
19F	0.47	0.61	0.89
2	0.41	0.43	0.86
10A	0.28	0.36	0.38
11A	0.35	0.55	0.71
15B	0.28	0.80	0.21
17F	0.67	0.74	0.81
20	0.43	−0.27	−0.48
22F	0.55	0.68	0.10
23F	0.19	0.89	0.16
33F	−0.21	0.51	−0.18

The nine bottom rows contain serotypes unique to the 23-serotype assay which were not included in the 14-serotype assay.

[*] Statistical analysis could not be performed for comparisons involving Laboratory C for serotype 12F, as all 10 results for that laboratory and serotype fell below the lower limit of detection, which did not permit rank order determination.