Zhang et al p 1

The impact of analytical variability on clinical interpretation of multiplex pneumococcal serology.

Xiaochun Zhang, MD, PhD, Kelly Simmerman, Belinda Yen-Lieberman, PhD, and Thomas M Daly, MD.

Robert J Tomsich Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, OH

#Corresponding author.

Mailing address:
Clinical Pathology, LL3
9500 Euclid Avenue
Cleveland, OH 44195.

Phone: 216-444-4547.
FAX: 216-444-4414
Email: dalyt@ccf.org

Running title: Multiplex pneumococcal serology interpretation.
Response to pneumococcal vaccination can be used to assess a patient’s humoral immune response to polysaccharide antigens. Multiplex assays measuring serotype-specific levels of pneumococcal antibodies are often used for this purpose, and clinical algorithms have been published to assist in the definition of an adequate immune response. We evaluated whether inter-laboratory variability in multiplex pneumococcal serology assays would affect clinical classification of the immune response. Specimens from 57 patients were analyzed at three reference laboratories using different multiplex assays to measure pneumococcal serology. Analytical correlation and clinical agreement in the classification of a patient’s vaccination status was compared between the three methods. Although substantial variation was seen in the quantitative antibody levels measured by different laboratories, the qualitative classification of individual serologic results showed a high degree of agreement between labs and the ultimate classification of a patient as “protected” vs “non-protected” was identical for most patients. The majority of discordant classifications were driven by a systematic bias in results from one of the assays rather than by random error. These data suggest that the use of integrated assessments based on multiple serotypes can compensate for much of the analytical variability seen between laboratories. Knowledge of the analytical performance characteristics of a particular assay is most important when evaluating patients with results near clinical cutpoints.
Introduction

Pneumococcal serology can be used to measure vaccination response to polysaccharide antigens in patients who are being evaluated for defects in humoral immunity. Pneumococcal vaccines such as Pneumovax 23 and Prevnar 13 contain a mixture of antigens from multiple pneumococcal serotypes to provide coverage against a broad range of commonly-encountered strains (1). The earliest measurements of serotype-specific antibodies were performed using individual ELISAs for each serotype (2). However, the increasing availability of multiplex immunoassay techniques has led to the development of assays that can provide quantitative, serotype-specific measurements of multiple anti-pneumococcal antibodies in a single reaction (3, 4). Although this level of detail was initially used primarily for research into the nature of the immune response to pneumococcal vaccines (5, 6), multiplex pneumococcal serology assays are increasingly being measured in clinical practice to assess a patient’s response to pneumococcal vaccination.

Clinical interpretation of pneumococcal serology results can be complex, and several guidelines have been proposed to define an adequate vaccination response when using serotype-specific measurements (7-10). Although the details of these proposals vary, most approaches incorporate two primary factors: the level of antibody response achieved against a given pneumococcal serotype, and the number of serotypes that reach this level. A combination of these factors is then used to determine if the patient has achieved an adequate response to vaccination (10). A key parameter in such approaches is defining the clinically-relevant cutoff that indicates an adequate antibody response. One strategy is to define a “protective” level of antibody that has been associated with
reduced rates of a given clinical outcome based on epidemiologic data. In adult patients, a level of 1.3 μg/ml has been proposed as a relevant cutoff based on correlation to a reduced likelihood of primary infection and pneumococcal colonization in early studies (10, 11). A lower cutpoint of 0.35 μg/ml is often cited for pediatric patients, based on studies correlating this level with prevention of invasive pneumococcal disease in infants (12, 13).

The earliest multiplex assay for measuring pneumococcal serology was developed on a bead-based flow cytometric assay platform (14), and most currently-available clinical assays are based on this approach. At present, there is only one FDA-cleared in vitro diagnostic (IVD) kit for multiplex pneumococcal serology available in the United States. This assay measures antibodies against 14 pneumococcal serotypes (xMAP Pneumo14, Luminex Corporation, Austin TX). In addition to the IVD assay, several commercial reference laboratories have developed laboratory-developed tests (LDTs) which are also based on the Luminex platform and utilize reagents and assay conditions developed and validated by the individual laboratory under CLIA regulations. A recent inter-laboratory comparison evaluated the performance of three multiplex pneumococcal assays in a small series of reference standards provided by WHO (15). The three assays showed between 42-55% agreement with WHO-assigned values, with varying levels of correlation between serotypes. However, in practical terms it was not clear whether the level of variation seen was large enough to have an impact on clinical utilization of these results when evaluating vaccine response. To address this issue, we performed multiplex pneumococcal serology testing on a large series of clinical samples to determine whether
92  inter-laboratory variability could impact the clinical classification of patients when using
93  published clinical algorithms.
94
95
Methods

Sample acquisition. Specimens were obtained using residual material from clinical testing sent to the laboratory for multiplex pneumococcal serology. Initial sample selection was performed without knowledge of clinical history or vaccination status. Samples were acquired from a total of 57 patients (generating 741 individual serology results per laboratory) for use in the final validation set. The majority of these specimens were single, unpaired specimens from adult patients, reflecting the population that our laboratory serves. All samples were collected according to local IRB policies governing the use of residual material for assay validation.

Pneumococcal serology testing. Samples were tested using three different multiplex assays for measuring serotype-specific pneumococcal IgG antibodies. Samples were tested on-site at the Cleveland Clinic using the xMAP Pneumo14 pneumococcal immunity panel (Luminex Corporation, Austin TX). This IVD assay detects antibodies against the following 14 pneumococcal serotypes; 1, 3, 4, 6B, 7F, 8, 9N, 9V, 12F, 14, 18C, 19A, 19F and 23F. Standard values for the assay are calibrated to FDA-89 reference serum, and assay buffers incorporate cell wall polysaccharide and polysaccharide type 22 as blocking agents. The performance characteristics of this assay have previously been reported (14, 16, 17). In addition, aliquots of each sample were sent in a blinded fashion to two commercial reference laboratories which perform multiplex pneumococcal serologic testing. Both reference laboratories utilize LDT assays based on the Luminex platform, but the specific analytical details of these assays have not been published. Both LDT A and LDT B measure antibodies against the same
119 14 serotypes (1, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 12F, 14, 18C, 19F and 23F) which differ from
120 the IVD assay by one serotype, replacing serotype 19A with serotype 5.
121
122 Statistical analysis. Method comparisons and correlation statistics for quantitative results
123 were calculated using EP Evaluator software (Release 9). Comparison of population
124 medians between assays was performed by one-way ANOVA on ranks using SigmaPlot
125 11.2 software. For serotypes where significant differences were identified by ANOVA,
126 pairwise multiple comparison testing (Tukey test) was performed to identify the group
127 that differed from the others.
128
129
Results

Inter-laboratory variation in quantitative results. We initially compared the absolute levels of pneumococcal antibodies measured for each serotype in order to determine agreement in quantitative results between laboratories. A total of 2223 serotype results were generated from the 57 samples in the test set. The 13 serotypes which were common to all assays were evaluated independently, and correlation statistics were calculated for all possible pairwise combinations of assays. Substantial variation was seen between laboratories, with only 61% of the comparisons achieving r values > 0.8 (Table 1). However, the degree of variation differed widely between serotypes. While some serotypes showed good correlation across all laboratories, others showed very little agreement between assays (Figure 1). In addition to this sporadic serotype-specific variability, a systematic negative bias was seen for LDT B relative to the other two assays across multiple serotypes (Table 1). This bias was most evident when comparing the median values for each serotype in the sample set across the three assays, where the values generated by LDT B were lower for most serotypes (Figure 2).

Inter-laboratory variation in qualitative serotype classification. In order to determine whether the variations in pneumococcal antibody measurements would impact clinical decision-making, we examined how each assay classified serotypes as “protective” or “non-protective” versus a clinically-defined cutpoint. Because our sample set was comprised primarily of single timepoints from adult patients, we chose to use the classification scheme suggested by Paris et al which defines an adequate response as an antibody level greater than 1.3 μg/ml (10, 18). This type of qualitative classification
improved the overall agreement between laboratories, with rates of concordance near 80% for all pairwise comparisons between laboratories (Table 2). When comparing IVD and LDT A, discrepancies between the two assays were fairly evenly distributed in terms of which assay fell above the “protective” cutoff. In contrast, results were more skewed for comparisons involving LDT B, with 93% of the discrepancies involving samples where LDT B gave a result below the cutoff. This is consistent with the systematic negative bias in results from LDT B seen in the quantitative comparison.

Inter-laboratory variation in clinical classification of patients as “protected” vs “non-protected”. Because interpretation of vaccination status in a patient incorporates information from multiple serotypes, we evaluated whether variations in serotype classifications between assays would impact the final categorization of a patient’s immune response. Based on the previously referenced criteria, we classified a patient as “protected” if 70% of the serotypes (9/13) achieved a level > 1.3 μg/ml. Using these criteria, 82% of the patients (47/57) were identically classified by all three assays, with 96% agreement (55/57) between IVD and LDT A. The majority of discordant results were cases where patients were classified as non-protected by LDT B and protected by the other two assays (Figure 3).
Discussion

Multiplex assays for pneumococcal serology are among the most analytically complex assays used in clinical practice, and the use of differing methods introduces the potential for analytical variability between laboratories. In this study, we evaluated whether this variability is significant enough to impact the final clinical interpretation when assessing vaccination response. Substantial variation was seen in the absolute values of pneumococcal antibodies reported using the three different assays. Both serotype- and assay-specific variation were observed. The underlying causes for the differences are not clear, but may relate to differences in the specific reagent formulations and calibration materials used by the different laboratories to validate the assays. A more consistent approach to standardization of pneumococcal antibody assays could potentially help to reduce these differences. Detailed recommendations in this regard have previously been published by the WHO for standardization of ELISA assays (www.vaccine.uab.edu), and reference standards are available for this purpose.

Despite the variability seen in the quantitative results reported by the various assays, the overall classification of patients’ pneumococcal immune status was remarkably similar between assays. This appears to be primarily due to two factors. First, although pneumococcal serology assays provide quantitative results, many clinical algorithms use these results in a qualitative fashion by converting results into “protective” or “non-protective” based on a clinically-defined cutpoint. The data shown here suggests that interpreting results in this manner helps to mitigate much of the analytical variability seen between assays. In addition, the evaluation of multiple serotypes limits the impact that any one serotype has on the overall classification of the patient status. Because
multiple serotypes are considered in the final classification, isolated differences in individual serotypes are usually offset by the larger agreement across the remaining serotypes, leading to an overall concordance in the classification.

LDT B provided an interesting illustration of these benefits. Despite the systematic negative bias seen with this assay, results from LDT B still agreed with the consensus classification for >80% of the patient samples. In addition, most of the discordant samples were patients given an overall classification of “non-protective” by LDT B, but achieving protective levels for eight serotypes (narrowly missing the nine required for 70%). Clinicians should be aware of the specific characteristics of the assay they are using, as an awareness that an assay tends to give lower results might lead them to give further consideration to patients with “borderline” results.

A potential limitation of this study is that we did not examine the impact of analytical variability on algorithms which use changes between pre- and post-vaccination levels to define a successful immune response (19), which many authors have suggested is the most sensitive way to evaluate an immune response to pneumococcal variation (9). Unfortunately, such samples are uncommon in the patient population we serve, as less than 5% of the samples our laboratory receives are paired specimens (data not shown). One would hypothesize that the use of paired samples might further improve concordance between labs, since comparing the ratio of pre- and post- results would reduce the significance of the actual quantitative values. However, this would need to be further examined for individual assays.

The goal for any diagnostic test is to maximize analytical accuracy and precision to levels sufficient for clinical use, and multiplex assays for pneumococcal serology have...
a high level of inherent complexity which could potentially affect clinical interpretation of results. This study suggests that although individual and systematic bias does exist between assays, this variability has a relatively minor impact on the final classification of patients as “protected” or “non-protected” when using diagnostic algorithms based on the qualitative comparison of serotypes against a predefined threshold. Clinicians should be most aware of the potential for discordant results in patients with values near the clinical cutpoint, or in patients with an intermediate number of protective serotypes. Ultimately, improved assay standardization traceable to qualified reference materials could help minimize this issue, particularly for laboratories using LDTs.
References


Figures

Figure 1: Quantitative correlation for representative serotypes. Examples are shown for a serotype showing good (serotype 4, r > 0.9) and poor (serotype 19F, r < 0.6) correlation between assays.

Figure 2: Comparison of pneumococcal antibody levels. For each assay, the median value of the 57 samples is shown for each serotype. Serotypes marked with an asterisk were statistically different (p < 0.05) when compared to the median results from the other assays.

Figure 3: Classification of vaccination status for individual patients. Serotypes with protective results are green, while non-protective ones are red. Each row represents results from an individual patient. Patients below the horizontal bar have nine or more positive serotypes and were classified as protected by the consensus of the three assays. The ten patients with discrepant classifications are outlined in blue.
Table I: Inter-assay correlation statistics

<table>
<thead>
<tr>
<th>Serotype</th>
<th>IVD vs LDT A</th>
<th>IVD vs LDT B</th>
<th>LDT A vs LDT B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R</td>
<td>Slope</td>
<td>Intercept</td>
</tr>
<tr>
<td>1</td>
<td>0.73</td>
<td>2.12</td>
<td>-1.72</td>
</tr>
<tr>
<td>3</td>
<td>0.71</td>
<td>0.41</td>
<td>0.78</td>
</tr>
<tr>
<td>4</td>
<td>0.93</td>
<td>1.31</td>
<td>-0.01</td>
</tr>
<tr>
<td>6B</td>
<td>0.89</td>
<td>0.62</td>
<td>0.65</td>
</tr>
<tr>
<td>7F</td>
<td>0.83</td>
<td>2.19</td>
<td>-2.29</td>
</tr>
<tr>
<td>8</td>
<td>0.88</td>
<td>1.12</td>
<td>-0.59</td>
</tr>
<tr>
<td>9N</td>
<td>0.89</td>
<td>1.26</td>
<td>-0.74</td>
</tr>
<tr>
<td>9V</td>
<td>0.71</td>
<td>1.52</td>
<td>-0.88</td>
</tr>
<tr>
<td>12F</td>
<td>0.92</td>
<td>0.29</td>
<td>0.88</td>
</tr>
<tr>
<td>14</td>
<td>0.74</td>
<td>0.95</td>
<td>-0.71</td>
</tr>
<tr>
<td>18C</td>
<td>0.79</td>
<td>3.30</td>
<td>-3.25</td>
</tr>
<tr>
<td>19F</td>
<td>0.39</td>
<td>4.97</td>
<td>-1.69</td>
</tr>
<tr>
<td>23</td>
<td>0.90</td>
<td>1.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

All possible pair-wise comparisons of the three assays are shown for each serotype. Regression statistics were calculated by Deming analysis, while bias is presented as the average absolute bias for each comparison.
Table II: Inter-assay agreement in qualitative classification of serotypes.

<table>
<thead>
<tr>
<th>Serotypes</th>
<th>IVD vs LDT A</th>
<th></th>
<th>IVD vs LDT B</th>
<th></th>
<th>LDT A vs LDT B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agreem</td>
<td>% Agreem</td>
<td>Agreem</td>
<td>% Agreem</td>
<td>Agreem</td>
</tr>
<tr>
<td>1</td>
<td>49</td>
<td>3</td>
<td>5</td>
<td>86%</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>48</td>
<td>6</td>
<td>3</td>
<td>84%</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>3</td>
<td>9</td>
<td>79%</td>
<td>51</td>
</tr>
<tr>
<td>6B</td>
<td>49</td>
<td>6</td>
<td>2</td>
<td>86%</td>
<td>50</td>
</tr>
<tr>
<td>7F</td>
<td>48</td>
<td>7</td>
<td>2</td>
<td>84%</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>49</td>
<td>6</td>
<td>2</td>
<td>86%</td>
<td>46</td>
</tr>
<tr>
<td>9N</td>
<td>55</td>
<td>1</td>
<td>1</td>
<td>96%</td>
<td>53</td>
</tr>
<tr>
<td>9V</td>
<td>49</td>
<td>2</td>
<td>6</td>
<td>86%</td>
<td>49</td>
</tr>
<tr>
<td>12F</td>
<td>36</td>
<td>11</td>
<td>10</td>
<td>63%</td>
<td>42</td>
</tr>
<tr>
<td>14</td>
<td>50</td>
<td>6</td>
<td>1</td>
<td>88%</td>
<td>42</td>
</tr>
<tr>
<td>18C</td>
<td>49</td>
<td>1</td>
<td>7</td>
<td>86%</td>
<td>48</td>
</tr>
<tr>
<td>19F</td>
<td>44</td>
<td>7</td>
<td>6</td>
<td>77%</td>
<td>34</td>
</tr>
<tr>
<td>23</td>
<td>54</td>
<td>2</td>
<td>1</td>
<td>95%</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>625</td>
<td>61</td>
<td>55</td>
<td>84%</td>
<td>587</td>
</tr>
</tbody>
</table>

Serotype results were classified as protective or non-protective for each sample using a cutpoint of 1.3 μg/ml. Results which were classified identically by both assays are listed under “Agree”, while samples classified as protective by one assay and non-protective by the other are listed under the assay giving the protective result (IVD +, LDT A +, or LDT B +).