CLINICAL AND DIAGNOSTIC LABORATORY IMMUNOLOGY, Nov. 1996, p. 689–700 Vol. 3, No. 6
1071-412X/96/$04.00+0

Collaborative Study for the Evaluation of Enzyme-Linked Immunosorbent Assays Used To Measure Human Antibodies to *Bordetella pertussis* Antigens

FREYJA LYNN,[1]* GEORGE F. REED,[2] AND BRUCE D. MEADE[1]

Laboratory of Pertussis, Division of Bacterial Products, Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, Maryland,[1] and Biometry Branch, Division of Microbiology and Infectious Disease, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland[2]

Received 2 August 1996/Accepted 30 August 1996

Acellular pertussis vaccines are being evaluated in multiple clinical studies, and human immunogenicity data will likely be pivotal in the appraisal of vaccine responses between populations and the responses to different vaccine combinations. Antibody response to pertussis antigens is also used in the diagnosis of pertussis. An international study was designed to assess the comparability of data generated in different laboratories by enzyme-linked immunosorbent assays (ELISAs). Thirty-three participating laboratories were asked to quantitate specific antibody to pertussis toxin (PT), filamentous hemagglutinin (FHA), pertactin (PRN), or fimbrial proteins (FIM) in 21 samples. Samples were to be assayed in triplicate in five independent assays by each ELISA routinely performed in the laboratory to assess intra-assay, interassay, and population variability. The mean sample values were used to compare quantitative results among the laboratories. Thirteen of the 32 laboratories which submitted evaluable data for an assay to measure antibodies to PT, 12 of 30 laboratories with assays for FHA, 10 of 17 laboratories with assays for PRN, and 6 of 13 laboratories with assays for FIM maintained a coefficient of variation below 20% for 75% of the samples tested. Assays that measure antibodies to FIM appear to be less precise than the other assays. Precision varied among laboratories that used similar methods. The relative values of intra- and interassay variabilities were not consistent for a given assay within a laboratory, indicating that the sources of these variability components may be unrelated. Precision and agreement appeared less reliable for samples with low antibody levels. Ranking and regression analyses suggest that some laboratories generated comparable quantitative results, although direct comparison or combination of results from different laboratories remains difficult to support. Calibration to the U.S. Reference Pertussis Antisera appears to have been successful at standardizing the results in some laboratories. Statistical analyses are affected by assay precision and are not necessarily reliable sole predictors of biologically relevant differences in quantitative results. If results from different laboratories must be compared, appropriate studies of precision and quantitative agreement should be conducted to support the specific comparisons.

Enzyme-linked immunosorbent assays (ELISAs) have been used extensively for evaluation of the serologic response to vaccination and infection by measuring the changes in antibody levels induced by exposure to relevant antigens. In some analyses, samples from an individual, either vaccinated or exposed to a pathogen, are examined to determine the response to the vaccination or to confirm a diagnosis. In other analyses, the responses of groups of individuals are compared to detect differences due to vaccine composition, vaccination schedule, or study population. The appropriate use of ELISAs in these studies requires characterization of assay accuracy and precision.

Assays to detect antibody to *Bordetella pertussis* antigens either to diagnose disease or to evaluate immune responses to vaccines are used by many laboratories worldwide. These assays undergo close scrutiny, because the reliance placed on them is high. Pertussis diagnosis by usual culture methods is insensitive (15) and underestimates the incidence of the disease (18). Serology has been shown to have improved diagnostic sensitivity in recent pertussis vaccine efficacy studies (10). The immunogenicity of the acellular pertussis components in vaccines has become an important means of evaluating differences between vaccines, populations, and proposed vaccine combinations (7) because no satisfactory laboratory or clinical correlate of immunity has been identified for acellular pertussis vaccines (1).

In recent years a large number of phase II and III studies of acellular and whole-cell pertussis vaccines have used ELISAs to quantitate serologic responses to the pertussis antigens. In order to evaluate assays currently being performed, an international collaborative study was organized. Demonstration of the ability of laboratories to generate the same results is usually addressed by standardization studies such as those conducted for assays to detect antibodies to *Haemophilus influenzae* or *Neisseria meningitidis* polysaccharides (3, 11). Systematic standardization of ELISAs for pertussis antigens has not been conducted for most laboratories, although some limited studies have been performed (13, 14). The present study did not attempt to standardize assays but instead focused on evaluating the comparability of existing assays. The assays are designed, in general, to measure immunoglobulin G (IgG) antibodies to pertussis toxin (PT), filamentous hemagglutinin (FHA), pertactin (PRN; 69-kDa outer membrane protein), or fimbrial proteins (FIM). The ELISAs performed in many of the par-

* Corresponding author. Mailing address: Center for Biologics Evaluation and Research, HFM-490, Food and Drug Administration, 1401 Rockville Pike, Rockville, MD 20852. Phone: (301) 443-8901. Fax: (301) 443-8306. Electronic mail address: lynn@a1.cber.fda.gov.

ticipating laboratories have been used to evaluate the immunogenicity of pertussis vaccines in humans. ELISAs that detect IgA antibodies have also been used in the diagnosis of pertussis; the performance of these assays was not evaluated here. The present study was designed to address both assay precision and quantitative agreement by requesting that each participating laboratory perform multiple estimates of antibody concentrations for a panel of samples.

(This study was presented in part at the International Symposium on Pertussis Vaccine Trials, Rome, Italy, 30 October to 1 November 1995, and at the 96th General Meeting of the American Society for Microbiology, New Orleans, La., 19 to 23 May 1996.)

MATERIALS AND METHODS

Study protocol. Each participating laboratory was requested to perform pertussis ELISAs according to the methods established in its laboratory, to include its controls and references, and to calculate results according to its standard procedures. Each of the 21 samples was to be run in triplicate on the same assay plate in a total of five independent assays, to generate 15 values per sample per antigen. Only values derived from valid assays as determined by the individual laboratories were to be reported to the coordinating center (Center for Biologics Evaluation and Research [CBER]). Laboratories were also asked to provide details about the methodology and reagents used in their assay(s) so that the effects of these factors could be assessed. The information requested included the following: methodology or standard operating procedure; source, characterization, and optimization of coating antigens; source and calibration of reference and control sera; characterization and optimization of enzyme-conjugated antibodies; calculation method; and determination of the minimum level of detection.

Samples. The majority of the samples were selected from processed sera, individual serum samples prepared by recalcifying plasma drawn from healthy blood donors (provided by C. H. Wirsing von König, Institut für Hygiene und Labormedizin, Krefeld, Germany). Four samples were prepared by pooling sera drawn from adults immunized with acellular pertussis vaccines (provided by David Klein, National Institute of Allergy and Infectious Diseases, and Wendy Keitel, Baylor College of Medicine). Samples were screened in the ELISAs performed at CBER (13) and were selected to cover a wide range of antibody levels, including some samples shown to be negative in the CBER assays. The samples were numbered from 1 to 21 and were shipped frozen to the participants. The laboratories were directed to test the samples in numerical order. Absolute values for the samples cannot be provided because of the lack of a "gold standard" for pertussis assays. However, samples 3 and 19 were identical samples prepared from a healthy adult donor. This donor was selected to have at least four times the minimum level of detection (MLD) in the CBER assays, which has been shown to be within the working ranges of these assays (13). The study also included two sets of paired samples with known twofold differences in antibody concentration: samples 2 and 21 were prepared by diluting samples 1 and 20, respectively, with an equal volume of a sample with almost no antibody detectable by the CBER assays. The measured value for the serum sample used as the diluent was below the MLD for PT, PRN, and FIM and 3 ELISA units/ml for FHA. MLDs for the CBER assays are 2, 6, 3, and 2 ELISA units/ml for PT, PRN, FIM, and FHA, respectively. At least one of the two paired samples included had a high enough antibody concentration that low levels of antibody in the diluent should have made a negligible contribution to the total antibody concentration of the diluted sample. The U.S. Reference Pertussis Antisera (human), lots 3 and 4 (HRP3 and HRP4, respectively), were also included as samples. Samples to be run at CBER were recoded before analysis to maintain blinding of CBER staff until completion of the assays.

Participating laboratories. Laboratories known to be involved in the assessment of the immunogenicity of acellular pertussis vaccines were invited to participate in the study, including vaccine manufacturers, national control laboratories, and research facilities. The panel of sera was shipped to 33 laboratories that agreed to participate in the study. We received evaluable data from 32 laboratories for PT, 30 laboratories for FHA, 16 laboratories for PRN, and 12 laboratories for FIM type 2, FIM type 3, or FIM types 2 and 3 combined. Laboratory 30 submitted data from separate assays for FIM type 2 and FIM type 3; thus, data from a total of 13 FIM assays were used in the precision analyses. For direct comparison of sample mean values between laboratory 30 and the other laboratories, the results for FIM type 2 and FIM type 3 from laboratory 30 were added together for a total FIM result. For analyses involving the use of the variance of the sample mean, the variance could not be determined for this combined mean, and data from laboratory 30 were not included in the FIM analyses. The number of laboratories contributing data for each antigen reflects the assays run in each of the laboratories. All laboratories that were included submitted data for PT, but not all laboratories submitted data for the other antigens. Table 1 lists the laboratories that submitted the data included in the analyses.

TABLE 1. Participating laboratories

Laboratory
Division of Bacteriology, NIBSC, Hertfordshire, United Kingdom
Massachusetts Publich Health Biologic Laboratories, Boston
Department of Pediatrics, Division of Infectious Diseases, University of California, Los Angeles, School of Medicine
Lederle Praxis Biologicals, Pearl River, N.Y.
Department of Pediatrics, Cedars-Sinai Medical Center, Los Angeles, Calif.
Centers for Disease Control and Prevention, Atlanta, Ga.
Department of Pediatrics, Vanderbilt University, Nashville, Tenn.
The Biocine Company/Chiron Corporation, Emeryville, Calif.
Research Serology Section, Connaught Laboratories, Inc., Swiftwater, Pa.
Division of Infectious Diseases, Saint Louis University School of Medicine, St. Louis, Mo.
St. Marianna University, Yokohama City SEIBU Hospital, Yokohama City, Japan
Pediatric Department, School of Medicine, Keio University, Tokyo, Japan
Clinical Trials Research Center—Infectious Diseases, Dalhousie University, Halifax, Nova Scotia, Canada
Swedish Institute for Infectious Disease Control, Stockholm, Sweden
Department of QC Clinical Serology, Connaught Laboratories Limited, North York, Ontario, Canada
Pasteur Merieux Serums & Vaccins, Val de Rueil, France
Centre National de Reference des Bordetelles, Institut Pasteur, Paris, France
Human Vaccine Production Department, The Chemo-Sero-Therapeutic Research Institute (KAKETSUKEN), Kumamoto, Japan
Centre for Applied Microbiology and Research/Microbiology Research Authority, Wiltshire, United Kingdom
Laboratory for Bacteriology and Antimicrobial Agents, Rijksinstituut voor Volksgezondheid en Milieuhygiene Bilthoven, The Netherlands
The Kitisato Institute, Research Center for Biologics, Saitama, Japan
CSL Limited, Victoria, Australia
Institut für Hygiene und Labormedizin, Krefeld, Germany
University of Göteborg, Göteborg, Sweden
Takeda Chemical Industries, Ltd., Osaka, Japan
Laboratory of Bacterial Toxins, Department of Bacteriology, National Institute of Health, Tokyo, Japan
Section of Microbiology, Department of Hygiene and Microbiology, University of Palermo, Palermo, Italy
Clinical Serology-Bacterial Vaccines, SmithKline Beecham Biologicals, Rixensart, Belgium
Department of Clinical Microbiology, Karolinska Hospital, Stockholm, Sweden
Swiss Serum and Vaccine Institute, Bern, Switzerland
BIKEN, Kanonji Institute, Kanonji City, Kagawa, Japan
CBER, Food and Drug Administration, Rockville, Md.

Twenty-eight laboratories submitted at least some information concerning their assay methodology, reagents, and calculation methods. Most of the laboratories that submitted methods used an indirect ELISA: passive adsorption of the antigens onto microtiter plates followed by the addition of the serum samples and then detection of antigen-specific antibody by enzyme-conjugated anti-human antibodies and substrate. Alternative methods, each used by at least one laboratory, included a fetuin capture system to assay antibodies to PT, the Falcon Assay Screening Test with a double sandwich, and a commercial kit with coated polystyrene balls. At least 23 laboratories used HRP3 and/or HRP4 as primary references or had tested them in their assay(s). At least 24 laboratories used an enzyme-conjugated anti-human antibody reported to be IgG specific, although evaluation of the isotype specificity of this reagent was not included in most submissions. At least 14 laboratories used a parallel line or reference line method to calculate sample values against values for a reference serum sample (12, 17). Other calculation methods used included calculation of a sample's value from single point estimates with either single or multiple dilutions quantitated against a standard curve. Other laboratories reported optical densities or titers, in some cases with a correction factor. A group of 10 laboratories (laboratories 1, 2, 3, 6, 12, 16, 17, 19, 26, and 31) performed assays using closely related methodologies. These laboratories essentially used the methodology described by Meade et al. (13), with little modification, and used HRP3 and/or HRP4 as primary references.

Statistical analysis. Values reported as less than the MLD, as defined by a given laboratory and assay method, were set to one-half that MLD for that laboratory. If a laboratory reported no numerical value for measurements that exceeded a fixed maximum for that laboratory and assay, then all values exceeding that maximum were set equal to the maximum.

For each sample in a given laboratory, variance of measurements has two sources in the present study: variation due to differences between assay runs (interassay) and variation due to differences in measurements within an assay plate (intra-assay). These components of the overall variance of measurements, denoted by σ^2_A for the between-assay (interassay) component and σ^2 for the within-assay (intraassay) component, were estimated from a one-way random effects analysis of variance model of the data (21). Negative estimates of σ^2_A, indicating that the interassay component is negligible relative to the intra-assay component, were set equal to zero. These components were used to estimate the variance of a single measurement, i.e., the population variance, $(\sigma^2_A + \sigma^2)$, and the variance of the mean of the 15 measurements $(\sigma^2_A/5 + \sigma^2/15)$. If one or more measurements were missing, the formula for the variance of the mean was modified accordingly.

Since the scale of measurements varied among samples and laboratories, variances were converted to coefficients of variation (CVs; standard deviation divided by the mean) in the percent scale to permit comparison across samples and laboratories.

No true antibody concentration was known for any of the samples, so that no direct assessment of accuracy defined as closeness to the known true value was possible. Agreement among laboratories was considered one surrogate for accuracy, under the assumption that in aggregate the laboratories will tend toward the true level. Agreement within a laboratory was assessed by determining how well the ratios of samples with known relative values estimated the dilution factors; these ratios are referred to as "recovery".

(i) Ranking. Since some laboratories used different measurement scales, ranking of the 21 samples, from those with the lowest to those with the highest antibody concentration, was one approach used in the analysis of agreement. For each sample, a mean rank was determined by averaging ranks assigned to it by each laboratory. These averages themselves were then ranked to produce a pooled rank for each sample. To assess an individual laboratory's closeness to the pooled ranking, the correlation coefficient (Spearman's rank correlation [5]) between that laboratory's ranking and the pooled ranking was calculated. A laboratory in perfect agreement with the pooled ranking would have a correlation of 1, a laboratory whose ranks bore no relation to the pooled ranking would have a correlation of 0, and a laboratory with the completely reversed ranking of the pooled ranking would have a correlation of -1.

(ii) Quantitative agreement with laboratory 12. To address the issue of how well the measurements of one laboratory were converted to the scale of another laboratory, a simple linear regression between laboratory 12 and each other laboratory was conducted with all data except for points that had been imputed a maximum value as described above, since these points tended to vitiate the linearity of the relationship. Laboratory 12 was chosen because it has routinely performed all four assays and, in the present study, maintained a CV of less than 20% for at least 75% of the samples in all assays. It has also analyzed samples from different immunogenicity studies and from infants immunized with vaccines from different manufacturers. Slopes, intercepts, and correlation coefficients were computed, and the hypothesis of whether the intercept was equal to zero was tested by the *t* test (6).

(iii) Recovery. The internal agreement for each laboratory's assay was determined by calculating the confidence interval of the ratios of the mean values for the paired samples with known relative antibody levels. For the two sets of samples paired with their 1:2 dilution, ratios of the sample means were calculated. If the diluent contained no specific antibody, then the calculated ratio should be 0.5. If the diluent contains low levels of antibody, the ratio would be close to, but slightly higher than, 0.5. The variance of the means, computed as described above, were used in the application of the delta method (2) to obtain the approximate standard error of the ratio of the means. The 95% confidence limits on the ratio were computed as the ratio ± 1.96 standard error. The ratio and confidence intervals of the means of samples known to be identical were similarly calculated.

RESULTS

Variability, expressed as CV, was examined as a function of the mean sample value for each laboratory for each antigen. In Figure 1, the population CV is plotted versus the mean value for each sample. Samples are plotted in order of increasing mean as determined for that antigen by that laboratory. Two examples of these precision profiles from participating laboratories are included. Figure 1A, the profile for the PRN assay run in laboratory 17, illustrates several features seen for some laboratories. Data for only 20 of the 21 samples are plotted because the value for sample 5 reported by laboratory 17 was above the assay limit. For samples that had concentrations

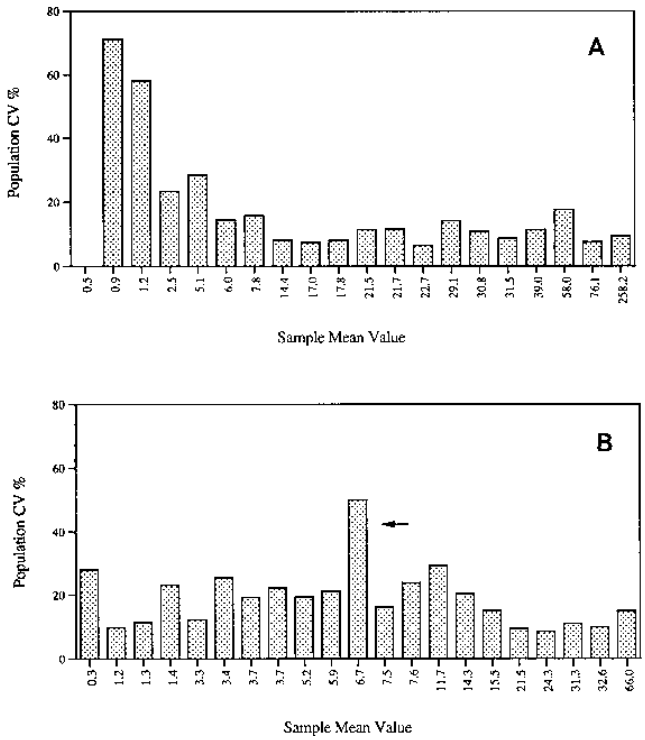

FIG. 1. Precision profile examples. CV as a percent is plotted against the mean value for each sample. Samples are displayed in order of increasing mean. (A) CV may increase with decreasing mean as the detection limit is approached. (B) Occasional samples with midrange values may have a high CV (indicated by the arrow) relative to the CV for other samples because of the value of a single replicate differing from those of the other 14 determinations.

outside the reporting limits of the assays, CVs were necessarily equal to zero. In some instances, the CVs increased for samples whose concentrations were approaching the upper and lower limits of the assay. Figure 1B, the profile for the FHA assay run in laboratory 8, illustrates another deviation from consistent CVs observed for some laboratories. In this case, a high CV was seen for a single midrange sample; this appeared to be due to only 1 of the 15 values for that sample lying substantially outside the range of the other 14 values. This single value appeared to be random and did not indicate systematic errors.

The intra-assay, interassay, and population precision, expressed as CVs, varied dramatically among laboratories. In most cases the intra-assay CV was lower than the interassay CV. In the case of laboratory 24, the interassay variability did not appear to be consistently greater than the intra-assay variability: in this analysis the inter-assay variability was less than the intra-assay variability for several samples for all antigen-specific assays performed (data not shown). This laboratory performed the assays slightly differently from the way that the other laboratories performed the assays, in that triplicates run in a given assay were not run on the same plate, but replicate plates with singlets on each plate were run. In effect, the intra-assay component reflected a between-plate rather than a within-plate variability. This variation in the study design could account for the increased intra-assay variability relative to the interassay variability.

The contribution of the intra-assay variability to the overall precision of an assay was explored. When the distributions of the ratio of intra-assay variability to population variability for

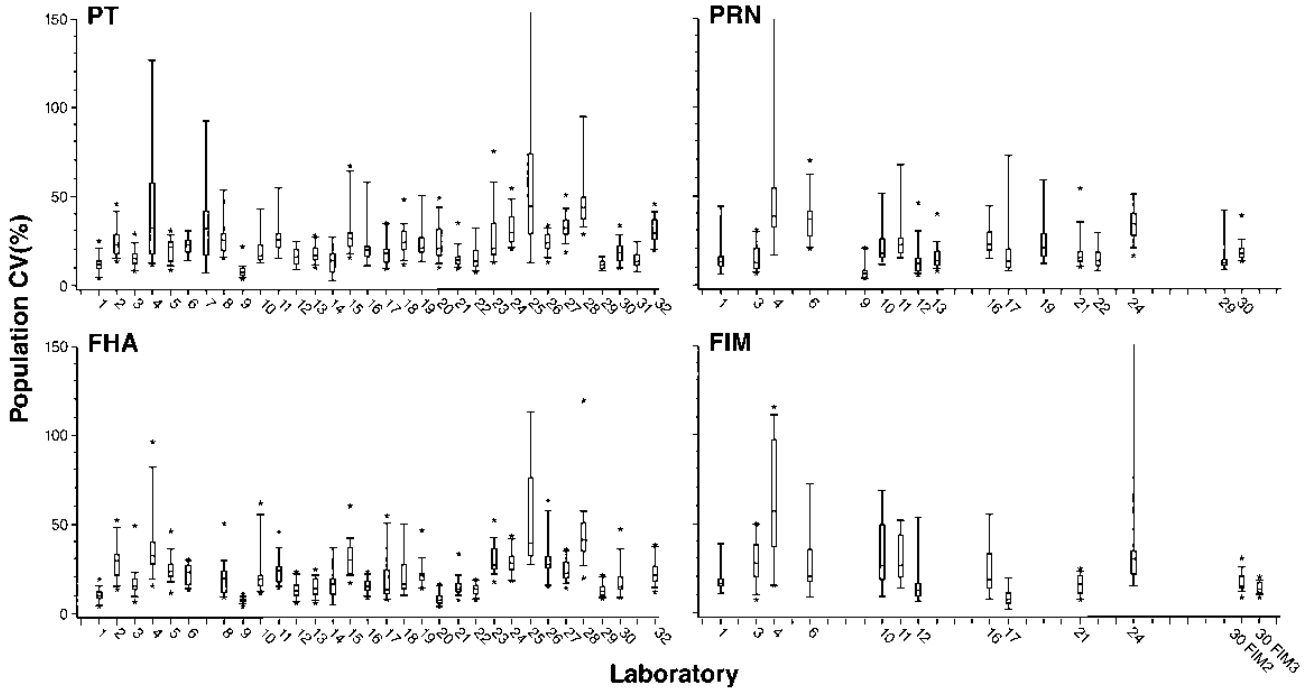

FIG. 2. Box plots of the population CVs by laboratory for each antigen. The top and bottom edges indicate the upper and lower quartiles, with the median marked in each box. The hinges indicate the upper 95th and lower 5th percentiles.

each sample was examined by laboratory, the relative contribution of the intra-assay variability component could vary substantially among samples within a laboratory (data not shown). The contribution of the intra-assay variability did not appear to be related to antibody concentration or population CV. These analyses indicate that the relationship between intra-assay and population variability cannot be easily assessed or estimated.

In order to assess the overall performance of the assays within each laboratory, the analyses compared the population CVs across laboratories. For concise presentation, the precision profiles were condensed into box plots (4). Figure 2 shows box plots of the population CVs by laboratory for each assay run in that laboratory. The central line within each box represents the median population CV for that laboratory. The upper and lower quartiles are indicated by the top and bottom edges of the boxes, respectively. The lines that extend from the box end with a hinge, a short horizontal line that indicates the 95th percentile at the top and the 5th percentile at the bottom. The asterisks represent samples above the 95th percentile or below the 5th percentile. All plots are arbitrarily cut off at 150% for consistent presentation. Not included in this analysis were any samples for which the intra-assay CV was equal to zero. Because of the increased CV observed in some laboratories for samples with low levels (see Fig. 1A), the upper quartile (75th percentile [CV_{75}]) of the population CV for each laboratory was used to compare precision among laboratories. Use of the upper quartile should eliminate misleadingly high CVs and reflect the precision in the working ranges of the assays. Several laboratories had CVs below 20% for at least 75% of the evaluable samples: 13 of 32 laboratories for PT, 12 of 30 laboratories for FHA, 10 of 17 laboratories for PRN, and 6 of 13 laboratories for FIM (laboratory 30 submitted separate data for FIM type 2 and FIM type 3, and thus, data for each assay are counted separately). With few exceptions, laboratories with a CV_{75} of less than 20% for any assay maintained the CV_{75} below 20% for all assays. In general, laboratories maintained a fairly consistent CV among all assays performed, although occasional laboratories generated substantially different population CV ranges in different assays. Laboratory 6 maintained CV_{75}s of 24 to 26% for PT and FHA assays but had CV_{75}s of 40.5 and 35.1% for its PRN and FIM assays, respectively. Laboratory 20 had a CV_{75} of 28.7% for PT assays but a CV_{75} of 8.3% for FHA assays. Precision did not appear to be directly dependent on the method used, in that laboratories using similar methods did not necessarily have comparable CVs. Some evidence suggests that variability may be antigen dependent: the FIM assay was the most variable assay for 7 of the 12 laboratories performing these assays, while the PRN assay was the most variable assay for only 2 of the 17 laboratories performing these assays.

The effect of the sample matrix was evaluated by comparing the variability for the processed sera to the variability for the pooled sera. With only four pooled serum samples and no matched samples to be evaluated, only gross evaluations of matrix effects are possible from this data set. Three laboratories were arbitrarily selected to look for any obvious effects. When samples were ranked in order of increasing CV for laboratories 1, 12, and 19, the CVs for the four pooled serum samples appeared to be distributed randomly throughout the samples, with no pattern of lower or higher CVs apparent for any given sample for any antigen (data not shown). No further analyses were performed because no indication of gross effects was seen in this limited look at the data.

Because laboratories varied in their quantitative procedures, the order in which laboratories ranked samples was used as an analysis independent of methods. Sample ranks as determined by each laboratory were compared with the pooled sample ranking of all laboratories and the Spearman correlation coefficient was calculated. Figure 3 plots the correlation coefficients, by increasing value, for each laboratory when data for

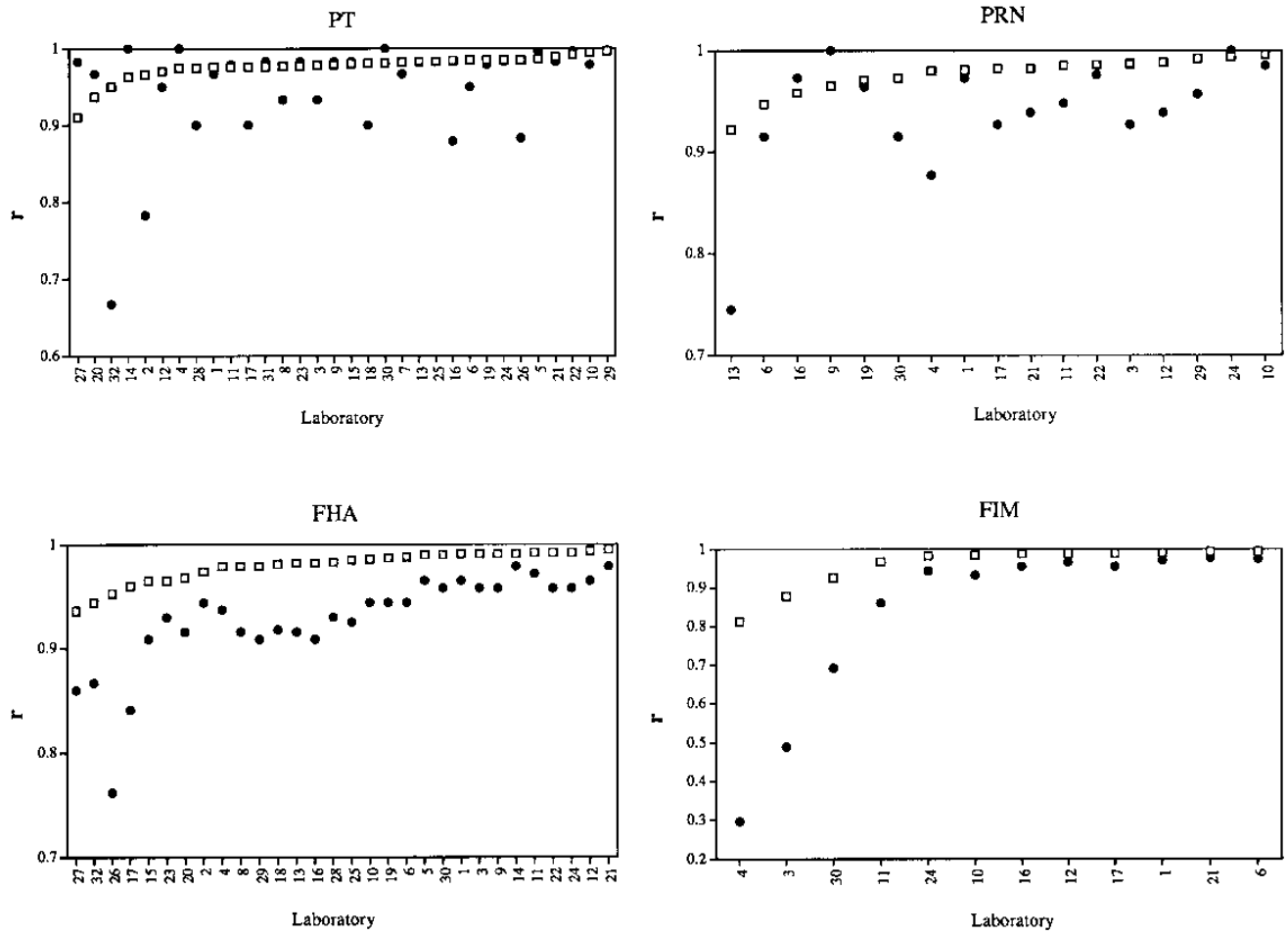

FIG. 3. Correlation of individual laboratory ranking to the pooled ranking. The correlation coefficients (r) are plotted for each laboratory by increasing r value when data for all samples are used in the analysis (open squares). Also plotted are the correlation coefficients when only data for the samples with the lowest antibody levels are used in the analysis (closed circles). Note that graphs use a different scale for the vertical axis.

all 21 samples are used in this analysis. Overlaid on Fig. 3 are the correlation coefficients for each laboratory when only the samples with the lowest antibody levels are used in the analysis (the 9 samples with the lowest values for PT, the 12 samples with the lowest values for FHA, the 10 samples with the lowest values for PRN, and the 13 samples with the lowest values for FIM). When data for all samples are used in the analysis, most laboratories had coefficients that were fairly consistent with one another and close to 1.0. For each antigen, however, some laboratories appeared to have a slightly lower coefficient than that seen for the majority of laboratories. The coefficients for these laboratories, shown as the first values plotted on the left of each graph, appeared as a distinct drop off the nearly horizontal line delineated by most laboratories. When the coefficients calculated by using only data for the samples with the lowest values are overlaid on the same plot, these coefficients appear to be more widely spread and lower than when data for all samples are used. The laboratories with the lowest coefficients vary depending on the analysis: for PT assays, laboratories 27, 20, and 32 have the lowest coefficients if data for all samples are used in the ranking, while laboratories 32 and 2 have the lowest coefficients if only the data for the nine samples with the lowest coefficients are used. For the FHA assay, four laboratories consistently had the lowest coefficients whether data for all samples or just the data for samples with the lowest values are used, although the order among the four laboratories varies. For the PRN assays, the order of increasing agreement is substantially different depending on the samples used for the analysis, although laboratory 13 has the lowest coefficient in both analyses. For the FIM assays, the same laboratories have the lowest coefficients in either analysis.

Regression of the mean value for each sample from one laboratory versus another laboratory was calculated to assess the utility of direct comparison of sample values between laboratories. Laboratory 12 was chosen as the basis for comparison. The slope, intercept, and correlation coefficient calculated by comparing results from each laboratory with the results from laboratory 12 are presented in Table 2. For laboratories whose assays are calibrated with respect to the same reference serum sample, the slope should be equal to 1 and the intercept should be equal to zero. For laboratories using different calculation methods, the intercept should still be at or near the origin, regardless of the scale of the measurement. The slope for these laboratories, however, will be different from 1. The intercepts which were statistically different from zero at $P = 0.05$ are indicated.

Laboratories that reportedly used similar assay methods, that used comparable calculation methods, and that used HRP3 and/or HRP4 as primary references were selected to evaluate how well the CBER references have unified value

TABLE 2. Regression coefficients for each laboratory compared to laboratory 12

Laboratory	PT			FHA			PRN			FIM		
	Intercept	Slope	Correlation coefficient	Intercept	Slope	Correlation coefficient	Intercept	Slope	Correlation coefficient	Intercept	Slope	Correlation coefficient
1	−5.2	1.2	0.992	**10.9**[a]	1.0	0.990	0.4	1.6	0.991	−0.6	1.1	0.990
2	**39.5**	1.6	0.990	−3.8	1.4	0.940						
3	−0.7	1.1	0.980	**8.0**	1.0	0.996	3.4	1.4	0.993	16.7	0.7	0.948
4	**38.6**	1.1	0.841	**51.9**	0.9	0.888	20.3	1.5	0.856	2.0	0.7	0.768
5	102.7	10.1	0.975	**89.1**	5.5	0.988						
6	0.1	1.3	0.985	**14.5**	1.1	0.989	3.3	1.5	0.991	3.0	1.1	0.999
7	−16.8	3.1	0.905									
8	0.8	0.1	0.954	0.7	0.1	0.994						
9	4.5	2.3	0.939	**14.6**	1.3	0.993	**6.6**	2.9	0.997			
10	−13.6	1.5	0.949	4.2	1.5	0.993	−7.9	3.0	0.993	−0.2	0.1	0.982
11	0.0	1.3	0.938	**21.1**	1.1	0.977	0.4	1.5	0.995	1.9	0.8	0.901
13	−2.1	1.4	0.973	4.6	1.3	0.951	−3.0	1.7	0.979			
14	**22.4**	1.1	0.829	**20.0**	0.9	0.938						
15	−0.5	0.9	0.947	−14.5	1.3	0.939						
16	−1.3	1.0	0.978	**24.2**	1.0	0.966	2.3	0.9	0.988	2.8	0.7	0.980
17	−6.0	1.1	0.946	3.3	1.0	0.984	−0.3	1.1	0.993	6.6	0.7	0.992
18	−0.7	1.3	0.974	5.4	1.3	0.986						
19	1.0	0.9	0.985	5.3	1.0	0.978	**5.9**	1.2	0.997			
20	**379.1**	3.1	0.766	**898.8**	2.8	0.796						
21	−7.1	1.2	0.937	1.5	1.1	0.985	−2.8	1.1	0.989	5.4	1.1	0.978
22	1.8	1.3	0.953	**10.2**	1.2	0.995	−1.7	1.3	0.995			
23	1.4	0.9	0.894	11.1	1.3	0.964						
24	−233.5	118.2	0.950	**1758.1**	138.0	0.990	−86.9	159.9	0.998	1,053.3	78.3	0.899
25	9.1	0.7	0.926	**22.8**	0.5	0.903						
26	8.1	1.1	0.957	12.8	1.1	0.984						
27	**25.2**	0.2	0.762	**23.0**	0.2	0.882						
28	−7.8	1.3	0.904	−2.1	2.2	0.987						
29	1.2	1.3	0.937	23.7	1.9	0.985	−1.7	1.9	0.997			
30	0.6	1.0	0.947	−1.0	1.0	0.997	**−10.3**	1.6	0.987	17.8	0.5	0.928
31	−14.9	1.6	0.891									
32	0.7	0.1	0.882	**3.3**	0.1	0.958						

[a] Intercepts in boldface type are significantly different from zero ($P < 0.05$).

estimates. HRP3 has been assigned a unitage of 200 ELISA units/ml for IgG anti-PT and anti-FHA, and HRP4 has been assigned a unitage of 90 ELISA units/ml for IgG anti-PRN. No unitage has been assigned for HRP4 in the PT and FHA assays or for HRP3 in the PRN assay. No unitage has been assigned to either reference for FIM because of the variety of FIM preparations, mostly mixtures of type 2 and type 3, used as antigens in these assays. Figure 4 presents the mean estimated values for HRP3 and HRP4 for PT, FHA, and PRN and the values for HRP3 for FIM plotted by laboratory. Most estimates are within 25% of the nominal values for each reference. Within a laboratory, the values generated for HRP3 in the PT and FHA assays appeared to be either consistently higher or consistently lower than the nominal value for HRP3 for both assays. Laboratory 16 appeared to underestimate the value of HRP4 in its PRN assay. However, that laboratory did not underestimate the PRN antibody concentration of HRP3 relative to the values estimated by the other laboratories. Laboratory 31 appeared to overestimate the value of HRP4 in its PT assay relative to the values generated by most other laboratories. Variability of HRP4 values among laboratories appears to be greater than that of HRP3 values for PT, FHA, and PRN assays.

Plotted in Fig. 5 are the means and 95% confidence intervals of the recoveries (ratios of estimated means for the diluted and undiluted samples) for each laboratory for each assay. If 1 or more of the 15 values reported for either sample of a pair was either greater than the upper limit or less than the lower limit of the assay, data for that pair for that laboratory were not used. Thus, the ratios should be based on estimates within the reported ranges of the assays. The confidence intervals for most laboratories include 0.5. However, the length of the confidence interval reflects the precision of the assay, and some laboratories with mean estimates close to 0.5 did not include 0.5 in the confidence interval because of a high degree of assay precision. Three laboratories appeared to have high recovery values in both assays (PT and FHA) performed in those laboratories. Laboratory 20 reported optical density values, corrected for background, for each sample. Laboratory 32 reported values converted by first multiplying the absorbance by the reciprocal dilution and then correcting the value to that for a reference serum. The calculation method used by laboratory 27 was not included in its data submission. Laboratory 13 appeared to have high recovery values for both pairs in the PRN assay but not for the PT or FHA assay. Laboratory 2 had a high recovery value in the PT assay for the pair with lower antibody levels but not for the pair with higher antibody levels. This phenomenon was also observed for the FIM assay run in laboratory 3. Most laboratories had wider confidence intervals for the pair with the lower antibody concentrations than for the pair with higher antibody concentrations, and the mean ratios were further from 0.5.

Figure 6 provides the 95% confidence intervals for the ratio of the mean values for the two samples which were identical. If the assays report the same value for each of these two samples, then the ratio should be 1.0. Several laboratories reported a result for the second sample that differed from the result for the first sample. For some laboratories, the 95% confidence

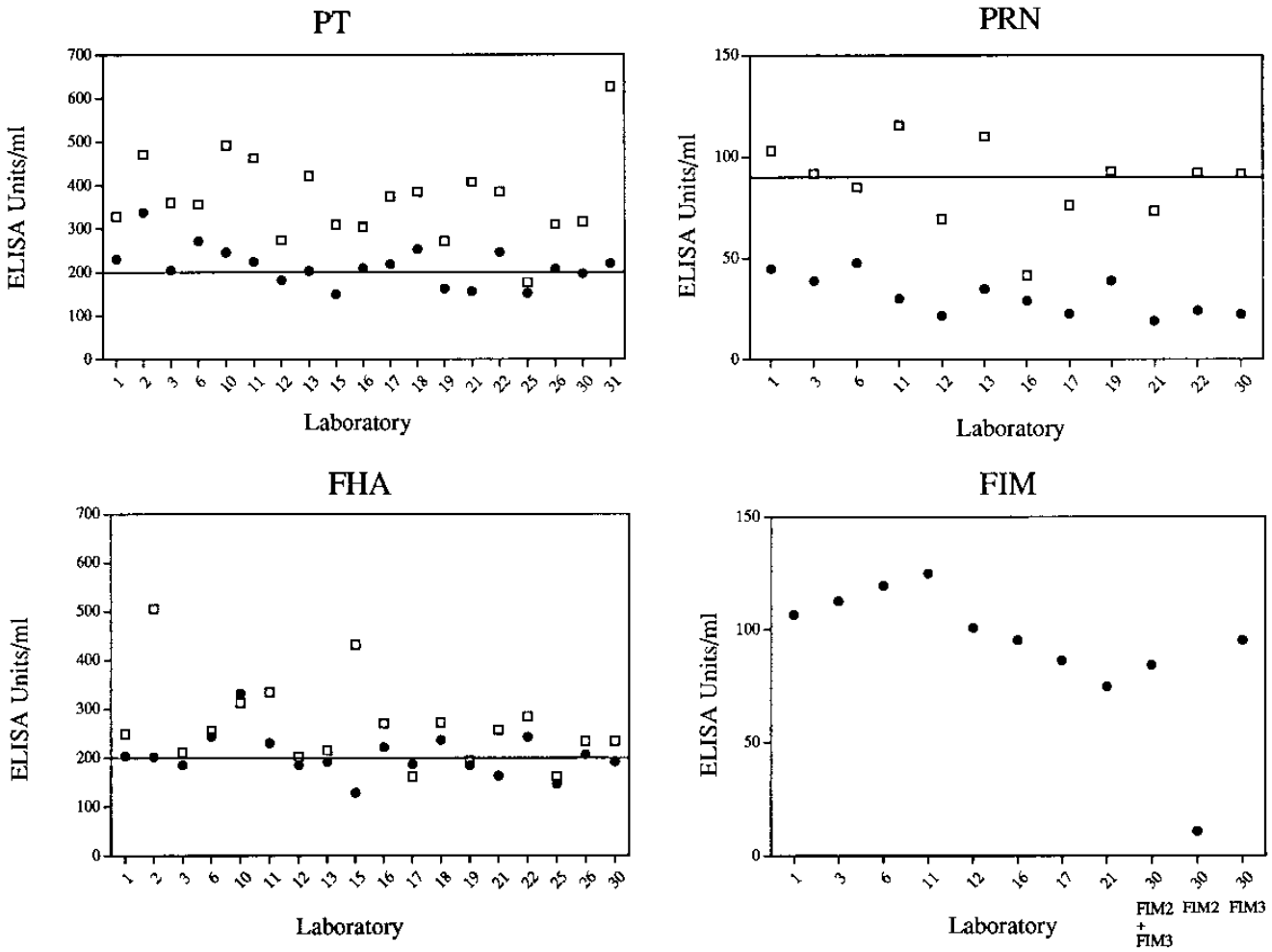

FIG. 4. Mean estimated values for HRP3 (closed circles) and HRP4 (open squares) by laboratory for each assay. The assigned unitages for HRP3 (200 ELISA units/ml for both PT and FHA) and HRP4 (90 ELISA units/ml for PRN) are indicated by the horizontal lines.

interval did not include the target of 1.0, nor were they within 20% of the target. Some laboratories, such as laboratories 5, 7, 25, 27, and 32, had consistently high or low ratios in all assays performed in that laboratory. Other laboratories, such as laboratories 2, 6, 11 and 24, had high or low ratios in only one or two of the assays that they performed.

DISCUSSION

Many studies of acellular pertussis vaccines, both efficacy and immunogenicity studies, report data derived from analysis of samples by ELISA. Comparisons of ELISA data among studies can be complicated by a wide range of factors, including differences among the laboratories generating the data. The assumption that similar methods applied in different laboratories will give similar results has only been tested in very specific cases between limited numbers of laboratories (13, 14). With the successful completion of efficacy trials (8, 9, 19, 22), the further development of acellular pertussis vaccines as well as their inclusion as components of new combination vaccines will likely depend on immunogenicity studies since no satisfactory correlate of immunity has been demonstrated (1). The comparability of results from the different laboratories generating data in support of these studies has been unclear. In the present study, the performance of assays in different laboratories has been evaluated for precision by a variety of methods. We have also explored analyses that can be used to assess quantitative agreement among laboratories around the world. These analyses have defined critical areas of concern when conducting interlaboratory validations or comparing results from multiple laboratories.

To fully evaluate assay precision, the study design addressed both intra-assay and interassay precision components. Although the intra-assay CVs were generally lower than the interassay CVs, as expected, the relative contribution of the intra-assay variability to the population variability was not necessarily consistent even within a laboratory. This supports the supposition that the sources of intra- and interassay variability can be independent, and thus, both intra- and interassay variabilities should be evaluated to determine the performance of an assay and the comparability of assays between laboratories. However, population variability, defined here as a summation of intra- and interassay variabilities, does appear to be a useful tool for comparing assay precision among laboratories. Using population variability, we have clearly identified several factors affecting precision in the participating laboratories.

Assay precision varied among laboratories. Some laboratories using essentially the same methods maintained different

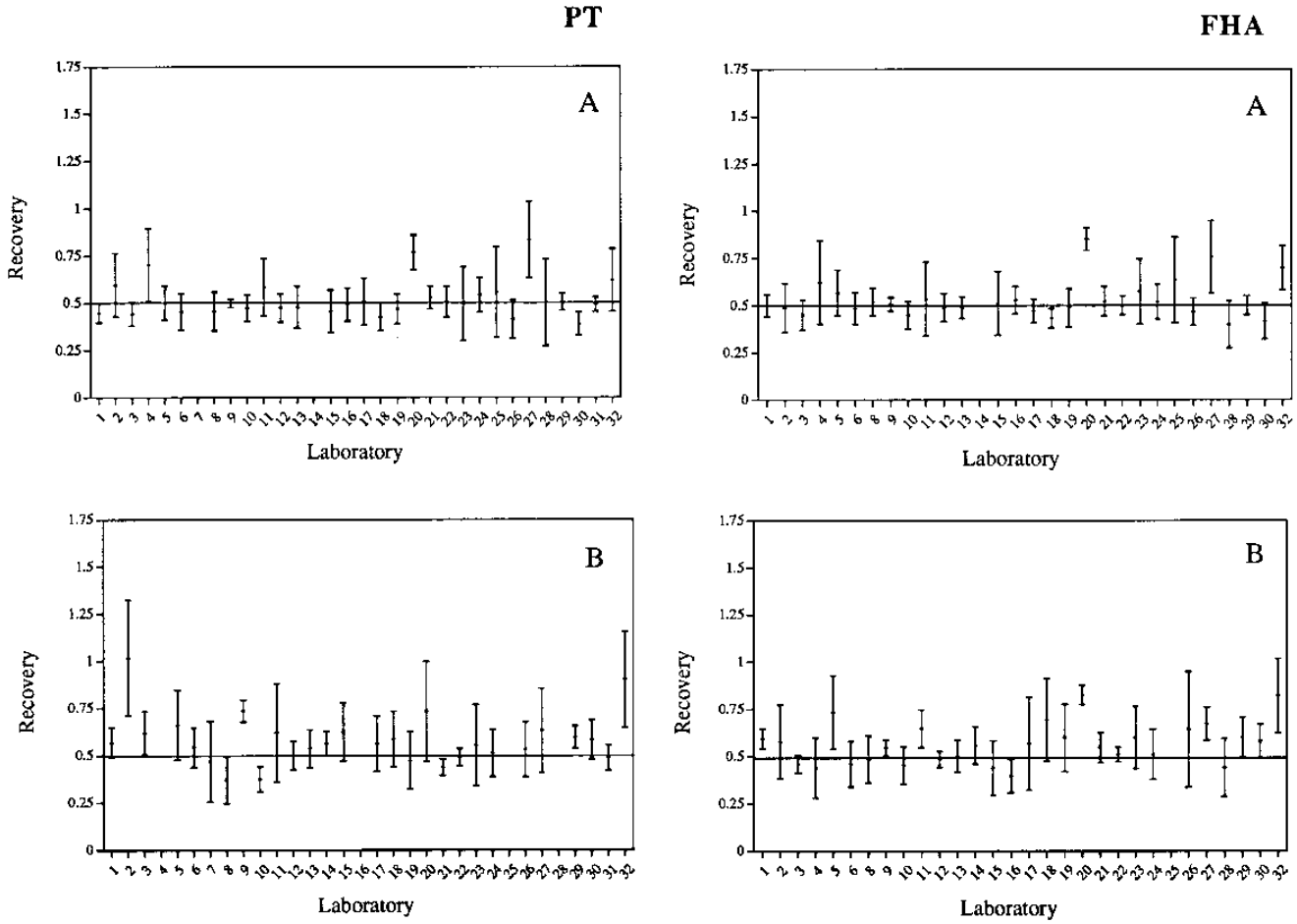

FIG. 5. Mean and confidence intervals by laboratory for the paired samples with relative values of 0.5 for each assay. (A) Values for the sample pair with higher antibody level to that antigen. (B) Values for the sample pair with lower antibody level to that antigen.

levels of precision. Increased precision could not be attributed to a particular method in the present study, but the different levels of precision seen in laboratories using similar methods implies that some of the variability is laboratory dependent. We did not have sufficient details about the methods, reagents, or equipment used to identify the specific causes of increased or decreased precision, but the differences indicate that assay precision is an essential part of interlaboratory validations, even among laboratories using the same methods.

Assay precision varied among the antigen assays. The FIM assays appeared to be the least precise assay in most of the laboratories that perform FIM assays. In most of these laboratories, the FIM antigen used was a mixture of fimbrial types, and FIM remains the least well characterized antigen. Laboratory 30 was the only laboratory reported to perform different assays for each of the two fimbrial types. It is notable that these two FIM assays were not more variable than the other assays run in laboratory 30 and that the CV_{75}s for these assays were among the lowest for the FIM assays. The use of a particular type of FIM preparation may affect the assay, and laboratory validation of this assay will require careful standardization of the antigen. The differences in precision seen among the other assays within a laboratory are not surprising because each antigen may require slightly different assay conditions, each antigen may bind to plastic differently, and each antigen may bind slightly different antibody populations in terms of cross-reacting antibodies or antibodies of different isotypes. Nonspecific binding of serum components and conjugates may also affect each assay differently.

Assay precision varied with antibody concentration. As expected, the CVs were greater for samples whose values were at the upper and lower extremes of the assays. Even though an assay can detect low levels of antibody, it may not be able to precisely quantitate low concentrations of antibody. The working range of an assay should define not only the minimum concentration that can be detected but also the precision of the measurement. This lower limit of quantitation should reflect the precision required for application of the assay results and be determined with actual samples when possible because dilutions of references or control samples may not necessarily reflect the behavior of the sample population (20).

The limited comparisons done for three laboratories between samples derived from processed plasma and samples derived from pooled serum provide no evidence that assay precision is affected by the processing of plasma to produce serum. The pooled sera did not appear to have CV_{75}s substantially different from those for the processed sera in the laboratories in which they were examined. These laboratories used fairly high starting dilutions of samples: greater than or equal to 1:60. Matrix effects at lower dilutions (higher serum concentrations) have not been addressed in these analyses. Laboratories using low sample dilutions should determine the ef-

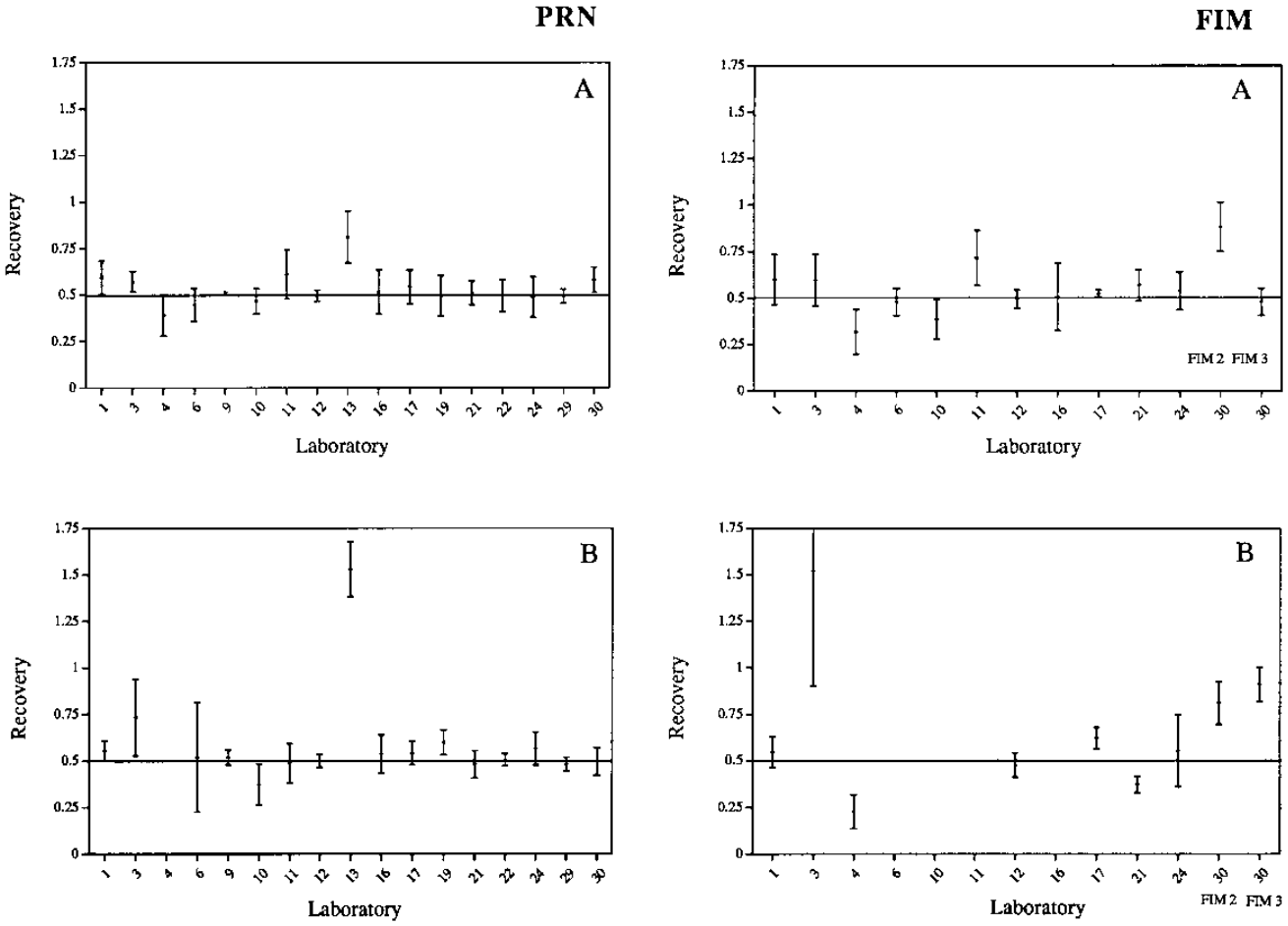

FIG. 5—*Continued.*

fects of the sample matrix on their assays. All samples in the present study were derived from adult donors and therefore cannot address the potential differences that may be seen when analyzing samples from infants due to either the sample matrix or antibody character.

Data similar to those presented here are necessary to define the ability of an assay to determine if samples or groups of samples have equivalent antibody concentrations. For example, when comparing paired samples within a plate in an assay with a known CV of 20%, the probability of detecting a twofold difference between identical samples, assuming that the measurements are normally distributed, would be 0.025. Sampling and population variability also need to be considered when determining the sample sizes required to detect relevant differences between groups. However, the inherent precision of the assays needs to be controlled to allow appropriate application. Separation and evaluation of the sources of variability are also necessary for the development of improved methods for serologic evaluations.

A variety of methods were used to assess quantitative agreement between laboratories. The analyses performed explored the comparability of data generated among laboratories which had not attempted formal standardization. In order to evaluate the plausibility and the limitations of data comparisons, the present study used a series of analyses that were either independent or dependent on the measurement scale.

The use of ranking analyses in the present study allowed a comparison among all laboratories, independent of the calibration and calculation method. In general, laboratories agreed fairly closely on the ranking of samples. The identification of laboratories that ranked samples slightly differently was partially dependent on the antibody concentration of the samples, and overall, agreement was not as good when only samples with low antibody concentrations were used in the analysis. Ranking analyses appear to be particularly relevant in the evaluation of certain diagnostic assays that require a test sample to be compared with known negative or positive samples. However, ranking of the relative responses of study groups in clinical trials is also important when determining differences in responses between populations or new vaccine combinations.

Regression analysis was another approach used to assess quantitative agreement between laboratories. The results from each laboratory were compared with those from laboratory 12, and the slope, intercept, and correlation coefficient of the best-fit linear regression line were determined. Regression analyses provided information somewhat different from that provided by the ranking analysis. For example, the correlation coefficients do not necessarily group the laboratories in the same order of agreement. Other parameters evaluating the agreement between laboratories, such as slope and intercept, were also evaluated. For laboratories with exact agreement on the same scale, a slope of 1 and an intercept of zero would be expected. Because of the variety of calculation methods and

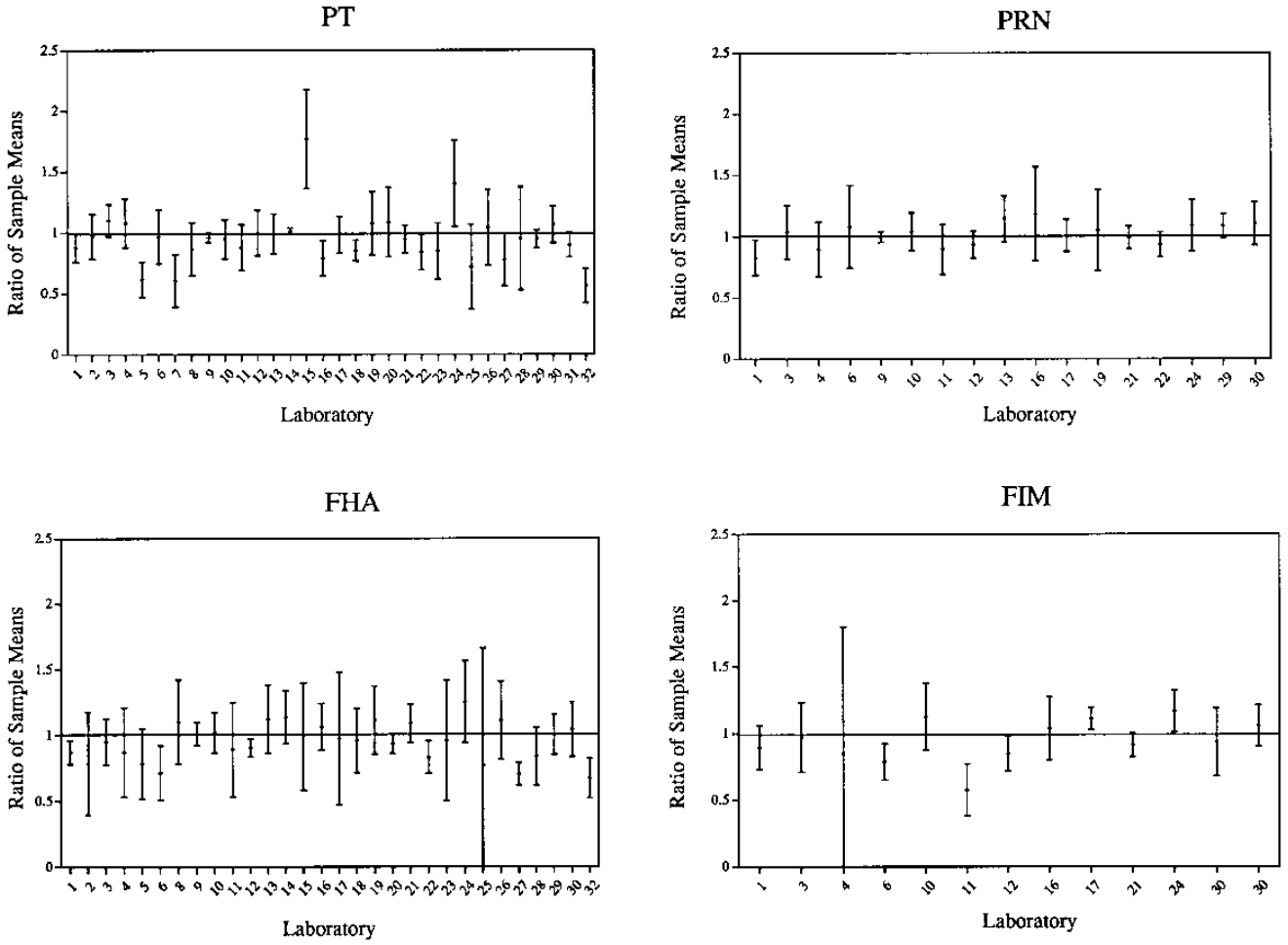

FIG. 6. Mean and confidence intervals for the identical samples for each assay.

measurement scales used by participating laboratories, no statistical tests on the slope were performed. However, independent of the reporting scale and calculation method, all laboratories measuring antibody to the same antigen should agree on a result close to zero for negative samples, and the intercepts of all regression lines should be at or near the origin. Accordingly, all intercepts were tested to see if they were significantly different from zero. Statistical testing, however, depends on the precision of the estimate as well as the magnitude of the difference from zero. There may be cases in which the statistical power was inadequate to detect biologically significant differences or, conversely, cases in which statistically significant differences were not of practical importance. Review of the regression plots (data not shown) and comparison of results from the regression analyses with those of the other analyses described in this report study indicate that the slope, intercept, and correlation coefficient, although informative, do not provide a complete assessment of quantitative agreement. Analytical tools beyond these three indicators are needed. The use of regression lines offers an attractive approach to normalizing data from different laboratories or transforming values from different laboratories to equivalent scales. However, the use of regression coefficients to transform or normalize results has not been adequately tested here and needs further study. Normalization of low antibody values could have profound effects on estimates of geometric mean concentrations or fold rises and thus requires a thorough evaluation. Transformation of values by using regression coefficients might force results into an appearance of equivalence in terms of the scale of measurement. The appearance of equivalence could be misleading, however, if the variability of measurements between the laboratories is high or the fit of the data to the linear model is not good.

Several laboratories used HRP3 and/or HRP4 as primary references and used similar assay methods. For those laboratories, the mean value determined for HRP3 or HRP4 was compared between laboratories and with the nominal value when applicable. HRP3 and HRP4 appear to have been appropriately used for the calibration of assays in most of the participating laboratories. The data from laboratory 30, which ran separate FIM type 2 and FIM type 3 assays, indicate that the reactivity of the reference is different for the two types of FIMs and that standardization of FIM assays may not be practical until purified and well-characterized FIM preparations are readily available. Quantitation of HRP4 is less consistent among laboratories, but the potential effects of the sample, the antibody concentration, or other factors on variability of the sample value estimates have not been thoroughly investigated. We conclude that the use of a single primary reference may be part of a successful strategy for unifying assays but does not guarantee comparable results.

Another approach to an accuracy assessment is the evalua-

tion of internal accuracy through determination of recovery of known concentrations of the analyte. The use of agreement in quantitation between known relative values for serially diluted samples has been proposed as a method for determining parallelism in bioassay curves (16). We have applied the concept of recovery by preparing paired samples, the second of which should have approximately half as much antibody as the first. These samples were produced by using antibody-negative serum as the diluent and thus were designed to mimic samples with twofold differences. If the diluent was not truly negative, then the ratio of the samples might be slightly higher than 0.5. The inclusion of a sample with a relatively high antibody concentration as one of the samples to be diluted should mitigate this effect for at least this pair of samples. The sample and diluent were both taken from processed serum samples to ensure that the matrix was consistent in both samples. The samples were arranged in the overall sample order to give the highest likelihood that the pairs would be analyzed on the same plate. The plots of confidence intervals of the recoveries allow one to evaluate the ability of laboratories to estimate a twofold difference. Statistical testing alone can be misleading in that laboratories with recoveries close to 0.5 but with highly precise assays may test as significantly different from 0.5. Recoveries substantially different from 0.5 were more commonly observed with pairs containing low antibody levels. Problems in estimating values for samples with low antibody levels include potential deviation from the curve-fitting model at the extremes of the titration curve or increased variability below the lower limit of quantitation. Occasional laboratories showed recoveries substantially different from 0.5 for both pairs of samples within an assay, which may indicate a problem in parallelism between sample and reference titration curves, in the sample matrix, or in the calculation method specific to that assay. When recoveries differed from 0.5 for all assays performed by a laboratory, a systematic problem may be indicated. Because methodologic information for these laboratories was limited, potential causes could not be thoroughly investigated. The use of recovery experiments such as this can be helpful in the assessment of parallelism between samples and references, the evaluation of calculation methods, and the definition of the working ranges.

The panel of samples included two identical samples that were ordered to maximize the change that they would be assayed on different plates within a single assay run. For each laboratory, these data indicate the agreement within a given assay run but not necessarily within the same plate. Again, statistical testing must be approached with caution for the reasons described above, so both the ratio of the identical samples and the confidence intervals were used to illustrate the performance of the laboratories. Most laboratories achieved results within a 20% difference and with fairly small 95% confidence intervals, reflecting good intralaboratory agreement. The occasional consistent differences observed suggest systematic drift within a single test; possible causes should be investigated.

The collection of samples used in the present study allowed for the evaluation of the assays over a wide range of antibody levels. The expected performance of these samples is now defined in many laboratories, and a subset of these samples can be rationally selected for future investigations.

Review of all data submitted indicates that some laboratories maintained high levels of precision for all assays; however, the level of precision did vary substantially among some laboratories and assays. Some laboratories appeared to agree fairly well on sample quantitation; however, the level of quantitative agreement can vary substantially among laboratories and may depend on the comparative analysis used as well as the assay precision. In all of the analyses performed here, interlaboratory comparisons revealed more discrepant results for samples with low antibody concentrations. The higher CVs observed in some laboratories for samples with low antibody levels may account for some of these results. However, another important factor may be the high relative impact of nonspecific assay background on the lower optical densities expected for these samples. Differences between minimum levels of detection and lower limits of quantitation need to be specifically examined during interlaboratory comparisons. Before correction factors to normalize the data generated by different laboratories are used, the effect that adjustment will have on low antibody values must be assessed. The present study did not address the evaluation of assay performance over extended periods of time.

The method of evaluation for interlaboratory validations should depend on the level of agreement required and should specifically address the types of comparisons desired among data generated in the different laboratories. Evaluations of assay precision should address all aspects of variability and should specify the precision obtained for each antigen and antibody level. Differences among study populations, study designs, and vaccine lots must be distinguished from the measurement error. Transfer of methods among laboratories does not guarantee comparable results; extensive initial validation is still required. For critical comparisons between samples, reliance on data from multiple laboratories will require careful evaluation of the comparability of an assay between the laboratories in question, regardless of the relatedness of the methods used in each laboratory. Ideally, immunogenicity studies should be designed so that comparisons are made between samples assayed in a single laboratory during a controlled period of time. As vaccine development continues, this may not always be possible, but any comparison between studies or laboratories should not be made until a thorough interlaboratory validation has been performed.

ACKNOWLEDGMENTS

We thank William Blackwelder, Steven Wassilak, Hans Hallander, and Elisabet Reizenstein for help in the design and review of the study and C. H. Wirsing von König for help with study design and review and for providing the majority of the samples used. David Klein and Wendy Keitel provided adult serum samples. Theresa Romani processed all samples and performed the sample analysis for CBER. We also thank all the collaborating laboratories for their participation and extraordinary cooperation.

REFERENCES

1. **Ad Hoc Group for the Study of Pertussis Vaccines.** 1988. Placebo-controlled trial of two acellular vaccines in Sweden—protective efficacy and adverse effects. Lancet **i:**955–960.
2. **Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland.** 1975. Discrete multivariate analysis. MIT Press, Cambridge, Mass.
3. **Carlone, G. M., C. E. Frasch, G. R. Siber, S. Quataert, L. L. Gheesling, S. H. Turner, B. D. Plikaytis, L. O. Helsel, W. E. DeWitt, W. F. Bibb, B. Swaminathan, G. Arakere, C. Thompson, D. Phipps, D. Madore, and C. V. Broome.** 1992. Multicenter comparison of levels of antibody to the *Neisseria meningitidis* group A capsular polysaccharide measured by using an enzyme-linked immunosorbent assay. J. Clin. Microbiol. **30:**154–159.
4. **Chambers, J. M., and W. S. Cleveland.** 1983. Graphical methods for data analysis. Wadsworth International Group, Belmont, Calif., and Duxbury Press, Boston.
5. **Conover, W. J.** 1971. Practical nonparametric statistics. John Wiley & Sons, Inc., New York.
6. **Draper, N., and H. Smith.** 1966. Applied regression analysis. John Wiley & Sons, Inc., New York.
7. **Edwards, K. M., B. D. Meade, M. D. Decker, G. F. Reed, M. B. Rennels, M. C. Steinhoff, E. L. Anderson, J. A. Englund, M. E. Pichichero, M. A. Deloria, and A. Deforest.** 1995. Comparison of 13 acellular pertussis vac-

cines: overview and serologic response. Pediatrics **96**(Suppl.)**:**548–557.

8. **Greco, D., S. Salmaso, P. Mastrantonio, M. Giuliano, A. E. Tozzi, A. Anemona, M. L. Ciofi degli Atti, A. Giammanco, P. Panei, W. C. Blackwelder, D. L. Klein, and S. G. Wassilak.** 1996. A controlled trial of two acellular vaccines and one whole-cell vaccine against pertussis. Progetto Pertosse Working Group. N. Engl. J. Med. **334:**341–348.
9. **Gustafsson, L., H. O. Hallander, P. Olin, E. Reizenstein, and J. Storsaeter.** 1996. A controlled trial of a two-component acellular, a five-component acellular, and a whole-cell pertussis vaccine. N. Engl. J. Med. **334:**349–355.
10. **Hallander, H. O., J. Storsaeter, and R. Möllby.** 1991. Evaluation of serology and nasopharyngeal cultures for diagnosis of pertussis in a vaccine efficacy trial. J. Infect. Dis. **163:**1046–1054.
11. **Madore, D. V., P. Anderson, B. D. Baxter, G. M. Carlone, K. E. Edwards, R. G. Hamilton, P. Holder, H. Käyhty, D. C. Phipps, C. C. A. Peeters, R. Schneerson, G. R. Siber, J. I. Ward, and C. E. Frasch.** 1996. Interlaboratory study evaluating quantitation of antibodies to *Haemophilus influenzae* type b polysaccharide by enzyme-linked immunosorbent assay. Clin. Diagn. Lab. Immunol. **3:**84–88.
12. **Manclark, C. R., B. D. Meade, and D. G. Burstyn.** 1986. Serologic response to *Bordetella pertussis*, p. 388–394. *In* N. R. Rose, H. Friedman, and J. L. Fahey (ed.), Manual of clinical laboratory immunology, 3rd ed. American Society for Microbiology, Washington, D.C.
13. **Meade, B. D., A. Deforest, K. M. Edwards, T. A. Romani, F. Lynn, C. H. O'Brien, C. B. Swartz, G. F. Reed, and M. A. Deloria.** 1995. Description and evaluation of serologic assays employed in a multicenter trial of acellular pertussis vaccines. Pediatrics **96**(Suppl.)**:**570–575.
14. **Meade, B. D., H. O. Hallander, T. A. Romani, L. Mardin, C. B. Swartz, M. Ljungman, B. Reizenstein, R. Möllby, W. C. Blackwelder, and C. R. Manclark.** 1990. A standard enzyme-linked immunosorbent assay procedure for quantitation of immunoglobulin G antibodies to pertussis toxin and filamentous hemagglutinin. *In* C. R. Manclark (ed.), Abstracts of the Sixth International Symposium on Pertussis DHHS Publication (FDA) 90-1162. U.S. Public Health Service, Bethesda, Md.
15. **Mink, C. M., J. D. Cherry, P. Christenson, K. Lewis, E. Pineda, D. Shlian, J. A. Dawson, and D. A. Blumberg.** 1992. A search for *Bordetella pertussis* infection in university students. Clin. Infect. Dis. **14:**464–471.
16. **Plikaytis, B. D., P. F. Holder, L. B. Pais, S. E. Maslanka, L. L. Gheesling, and G. M. Carlone.** 1994. Determination of parallelism and nonparallelism in bioassay dilution curves. J. Clin. Microbiol. **32:**2441–2447.
17. **Reizenstein, E., H. O. Hallander, W. C. Blackwelder, I. Kühn, M. Ljungman, and R. Möllby.** 1995. Comparison of five calculation modes for antibody ELISA procedures using pertussis serology as a model. J. Immunol. Methods **183:**279–290.
18. **Rosenthal, S., P. Strebel, P. Cassiday, G. Sanden, K. Brusuelas, and M. Wharton.** 1995. Pertussis infection among adults during the 1993 outbreak in Chicago. J. Infect. Dis. **171:**1650–1652.
19. **Schmitt, H. J., C. H. von Konig, A. Neiss, H. Bogaerts, H. L. Bock, S.-W. H., M. Gahr, R. Schult, J. U. Folkens, W. Rauh, and R. Clemens.** 1996. Efficacy of acellular pertussis vaccine in early childhood after household exposure. JAMA **275:**37–41.
20. **Shah, V. P., K. K. Midha, S. Dighe, I. J. McGilveray, J. P. Skelly, A. Yacob, T. Layloff, C. T. Viswanathan, C. E. Cook, R. D. McDowall, K. A. Pittman, and S. Spector.** 1992. Analytical methods validation: bioavailability, bioequivalence, and pharmacokinetic studies. J. Pharm. Sci. **81:**309–312.
21. **Snedecor, G. W., and W. G. Cochran.** 1980. Statistical methods, 3rd ed. The Iowa State University Press, Ames.
22. **Trollfors, B., J. Taranger, T. Lagergard, L. Lind, V. Sundh, G. Zackrisson, C. U. Lowe, W. Blackwelder, and J. B. Robbins.** 1995. A placebo-controlled trial of a pertussis-toxoid vaccine. N. Engl. J. Med. **333:**1045–1050.