Interlaboratory Comparison of Results of an Anthrax Lethal Toxin Neutralization Assay for Assessment of Functional Antibodies in Multiple Species[∇]

Kristian S. Omland,[1] April Brys,[2] David Lansky,[1] Kristin Clement,[3]
Freyja Lynn,[4]* and the Participating Laboratories

*Precision Bioassay, Burlington, Vermont[1]; Battelle Eastern Science and Technology Center, Aberdeen, Maryland[2];
Battelle Biomedical Research Center, West Jefferson, Ohio[3]; and Division of Microbiology and
Infectious Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland[4]*

The anthrax lethal toxin neutralization assay (TNA) will likely be used to correlate the protection offered by new anthrax vaccines in animal models to the immunogenicity that will be provided in humans. TNA data are being generated in several different laboratories to measure the immune responses in rabbits, nonhuman primates, and humans. In order to compare data among species and laboratories, a collaborative study was conducted in which 108 samples from the three species were analyzed in seven independent laboratories. Six of the seven laboratories had participated in an interlaboratory technology transfer of the TNA. Analysis of the titration curves generated by samples from each species indicated that the behaviors of the samples from all species were similar; the upper and lower asymptotes and the slopes of the curves were less than 30% divergent from those for human reference material. Dilutional linearity was consistent among samples from each species, with spike to effective dilution at 50% inhibition (ED_{50}) slopes of less than 1.2 for all species. Agreement among the laboratories with consensus values was within 10% of the ED_{50}s for all samples and within 7.5% of the quotients of the test sample ED_{50} and the reference standard ED_{50} (NF_{50}s) for all samples. The relative standard deviations obtained when data from all laboratories and for all species were combined were 45% for the ED_{50}s and 35% for the NF_{50}s. These precision data suggest that the NF_{50} readout may normalize the values generated by different laboratories. This study demonstrates that the TNA is a panspecies assay that can be performed in several different laboratories with a high degree of quantitative agreement and precision.

The bioterrorism-related anthrax cases in 2001 brought to the forefront the need for medical countermeasures against anthrax for both military and civilian populations. While a licensed vaccine exists for anthrax (BioThrax anthrax vaccine adsorbed [AVA]), modifications to the AVA vaccination regimen, as well as new anthrax vaccine candidates, are deemed to be an essential part of the countermeasure program of the United States (12). For ethical and logistical reasons, anthrax vaccines cannot be tested in conventional efficacy studies. When testing of a vaccine for efficacy in human clinical trials is neither feasible nor ethical, the U.S. Food and Drug Administration can utilize a new rule, 21 CFR 601, subpart H, usually referred to as the "animal rule," whereby the efficacy of such vaccines can be demonstrated through animal studies. Application of the animal rule to anthrax vaccines would be facilitated if an immunological correlate of protection were identified in an animal model. The appropriate immune parameter could then be measured in human immunogenicity trials to demonstrate that the vaccine elicits the same response in humans as that observed in animals protected against a lethal aerosol challenge of virulent *Bacillus anthracis* spores.

A predominant virulence factor in *B. anthracis* is a tripartite exotoxin consisting of a binding moiety, protective antigen

(PA), which combines with either lethal factor or edema factor to form lethal toxin (LT) and edema toxin, respectively (8). AVA and most experimental anthrax vaccines primarily consist of PA (2, 4, 19). PA has been shown to be immunogenic in both animals and humans and has been shown to confer protection against *B. anthracis* challenge in animal studies (2, 4, 7, 10, 14, 16, 18). The protection elicited by PA-based vaccines is believed to be mediated by PA-specific antibodies that neutralize the action of anthrax toxin (23, 27).

Two immunoassays are currently in routine use for the quantitation of antibodies against PA: an enzyme-linked immunosorbent assay (ELISA) that measures antibody to PA (21, 24) and an LT toxin neutralization assay (TNA) (3, 6, 17). The ELISA was developed in parallel for use with multiple species; however, the species specificity of the ELISA limits direct comparison of the responses between different animal species as well as direct comparison of the responses between animals and humans. The TNA was also developed for use with various species. In contrast to the ELISA, the TNA theoretically overcomes species specificity issues. The TNA measures the ability of antibodies to neutralize the cytotoxicity of LT rather than quantifying total antibody through a conjugated species-specific secondary antibody. TNA may provide a more relevant immunological measure, as it quantitates functional antibodies only rather than total PA-binding antibodies. TNA antibody levels have been shown to correlate with protection in rabbits and nonhuman primates (NHPs) (10, 14, 16). Because vaccination with PA-based vaccines induces neutralizing antibodies

---

* Corresponding author. Mailing address: OBRA, DMID, NIAID, 6610 Rockledge Dr., Room 5064, Bethesda, MD 20892. Phone: (301) 451-6811. Fax: (301) 480-1263. E-mail: lynnf@niaid.nih.gov.

in humans (2, 4, 18), TNA antibody levels are a likely choice for an immunological correlate between animals and humans for PA-based vaccines.

The data for human immunogenicity studies as well as pivotal animal efficacy studies will be generated in multiple laboratories. Human serum samples as well as serum samples from at least two different animal species will likely be assessed to generate the data needed to support the use of the animal rule for the licensing of anthrax vaccines. Therefore, the assay used to generate the data should optimally be reproducible in multiple laboratories and should be species independent. This interlaboratory study was designed to provide data from a single study that would address both the interlaboratory reproducibility of the TNA and the comparability of the behavior of sera from different species in the assay. The objective was to demonstrate that data from various laboratories and species generated during animal efficacy studies and human immunologic studies can be combined to support the efficacy of PA-based vaccines in humans. By eliminating concerns regarding which laboratory performed the assays and what species was under study, we hope to simplify the comparisons that will need to be made when data are compiled to demonstrate efficacy on the basis of the animal rule.

MATERIALS AND METHODS

Seven laboratories participated in the study. These laboratories used similar, but not necessarily identical, assay procedures. All laboratories are involved in the development and/or evaluation of anthrax vaccines as well as the identification of correlates of protection in animals and humans.

The laboratories performed the TNA essentially as described by Quinn et al. (20) and Li et al. (9). Briefly, serum samples were titrated by using twofold serial dilutions in a 96-well plate, followed by the addition of a constant amount of LT to each dilution. The concentration of LT added was that needed to kill approximately 95% of the cells in the absence of any neutralization. After preincubation of the test serum with the LT, the mixtures were transferred to another 96-well plate that had been seeded with J774A.1 cells in late log phase. LT that was not neutralized by anti-LT antibodies in the serum would intoxicate and kill the cells. Following intoxication, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was added to the plates, followed by the addition of a solubilization buffer to lyse the cells and solubilize the MTT. The cell plates were then incubated, and the optical density (OD) values were read with a microplate reader to determine cell viability (11). All incubations were carried out at 37°C in approximately 5% CO_2. Neutralization of anthrax LT was manifest as a suppression of cytotoxicity and, hence, the preservation of cell viability. A four-parameter logistic regression (4PL) model was used to analyze the OD versus the reciprocal of the serum dilution. The inflection point was reported as the effective dilution at 50% inhibition (ED_{50}). The quotient of the test sample ED_{50} and the reference standard ED_{50} (NF_{50}) was also calculated and is reported for each sample. A pooled human antiserum against AVA, known as AVR801 (25), was used as the reference serum sample in all laboratories.

For this study, each laboratory assayed a common panel of 108 samples using the TNA method as routinely performed in that laboratory. Although laboratories A, B, C, E, F, and G had performed an interlaboratory technology transfer prior to the study, laboratory-specific variations in the TNA method were noted. Most of the laboratories also used the SAS analysis code designated ED50.51 (Centers for Disease Control and Prevention [CDC] reference no. I-019-02) developed at the CDC (9). Only laboratory D used different data reduction software, which was a package that accompanied its plate reader. Laboratories A, B, E, F, and G used a high-throughput format of the TNA method, in which the titer in each serum sample was determined in singlet on the plate (which allowed the analysis of nine samples per plate). Laboratories C and D determined the titer in each serum sample in triplicate on each plate (which allowed the analysis of three samples per plate).

The serum samples were provided to each participating laboratory as a kit. The 108 samples included 36 samples from each species (rabbits, NHPs, and humans). Samples were taken from human subjects or animals immunized with either AVA or an experimental recombinant PA vaccine. The rabbit samples

included seven samples from pooled sera and two samples from individual animals. The NHP samples included three samples from serum pools and eight samples from individual animals. The human samples included 2 pooled samples and 11 samples from individuals. The remaining samples for each species were prepared by spiking either the pools or the individual samples into normal serum so that dilutional linearity could be assessed. Each species panel included samples that titers that were negative or in the low range (ED_{50}s = 20 to 100), midrange (ED_{50}s = 100 to 500), and high range (ED_{50}s > 500). Human samples were deidentified and were exempt from institutional review board approval.

Some serum specimens were diluted (spiked) into species-specific naïve serum to produce test samples with known relative ED_{50} ranges for the evaluation of dilutional linearity. In these cases, each dilution was an independent preparation. In addition to the kits containing the 108 samples, each site was also provided with AVR801, which had an assigned mean ED_{50} of 656 and a 2-standard-deviation range of 322 to 990.

The sample kits were prepared at a central laboratory in which one preparation of each sample was divided into eight aliquots, with one aliquot prepared for each of the testing laboratories. At each site, the samples were tested in numerical order (1 to 108) in at least two independent assays for a total of at least 216 sample values. The first and second rounds of testing were separated by a minimum of 3 weeks to ensure complete independence between the reportable values.

Statistical analysis. All but one laboratory (laboratory D) used CDC ED50.51 analysis software, which is an SAS program, to reduce raw ODs to reportable values (ED_{50}s, NF_{50}s, and the parameters of the 4PL model). Laboratory D used KC4, a commercial software package bundled with its plate reader, to implement 4PL model fitting and to estimate the ED_{50} for each sample. The CDC ED50.51 program used a 4PL model to fit a dose-response curve to the data. Samples that produced no signal of neutralization even at the minimal starting dilution were classified as negative. For the evaluation of sera with low reactivity, in which the upper asymptote was not observable even at the minimal starting dilution, the program used an algorithm to constrain the upper asymptote to either (i) the upper asymptote of the reference standard or (ii) the upper asymptote of a serum control sample assayed on the same plate. The choice of constraint was based on model fit. The resulting fitted curve was the best approximation of the full curve (9).

The results generated at each site were sent to a single statistical group for compilation and analysis. Prior to data analyses, the Box-Cox method was used to identify the appropriate transformation of each variable (ED_{50}, NF_{50}, asymptotes, 4PL model slope). We found that it was most appropriate to model both asymptotes on the original untransformed OD scale and to use $\log_2$ transformations for 4PL model slope, ED_{50}, and NF_{50}.

The data were modeled by using the linear mixed-effects function in R (15, 22). For all analyses, the factors that were likely to affect the performance of the assay were modeled as nested random effects. These factors were ordered by using a hierarchy of laboratory, then date-analyst, and then plate. The factors that were of the greatest interest in our analysis were modeled as the fixed effects in three families of models. To assess the similarities of the titration curves across species, the 4PL model curve slope and each (lower and upper) asymptote were modeled (in three separate models) by using a mean for the reference material and differences from the reference for each species. Species were also compared in terms of dilutional linearity on the basis of the ED_{50} and the NF_{50} data for specimens that were represented at more than one spike. Each specimen had its own intercept; and the data were modeled in three ways: (i) with the slope constrained to equal 1, (ii) with the slope common among all serum classes, and (iii) with the slope specific to each species. We used Akaike's information criterion for model selection (1). Finally, the agreement and the precision of the neutralizing activity estimates across laboratories were compared in models in which each sample had its own mean. Consensus estimates, which were maximum-likelihood estimates (those values would be called predicted values in the technical mixed models literature), of both ED_{50} and NF_{50} for each sample were derived after the omission of extreme outliers. Extreme outliers were defined as observations more than three residual standard errors from the sample mean and were omitted one at a time. Consensus estimates were used to define six strata representing low (stratum I) to high (stratum VI) neutralization activity (Table 1).

The similarities of the titration curves among species were analyzed by equivalence testing on the basis of two one-sided tests (5). The idea was to compare the magnitude of the difference attributable to each species to the baseline estimate for the reference material. The more extreme of the two one-sided 95% confidence limits of the 4PL model curve slope for each species was divided by the 4PL model curve slope estimate for the reference material and expressed as a percentage. To make the equivalence test comparable for the two asymptotes, the more extreme of the two one-sided 95% confidence limits for each asymptote

TABLE 1. Description of neutralization activity strata

Stratum	No. of samples	Consensus ED_{50} range[a]
Negative	6	$y = 0$
I	14	$y < 32$
II	17	$32 \le y < 64$
III	20	$64 \le y < 128$
IV	16	$128 \le y < 256$
V	16	$256 \le y < 512$
VI	19	$512 \le y$

[a] Maximum-likelihood estimates based on a mixed model of data from all seven laboratories but with the omission of extreme outliers.

parameter was divided by the curve depth for the reference material, i.e., the distance between the lower and the upper asymptotes for the reference material.

To assess the agreement among the laboratories, the data from each laboratory were compared in terms of the frequency that they reported samples as being nonresponsive (i.e., neutralization not detected). The frequency that the reported $ED_{50}s$ and $NF_{50}s$ were far from the consensus estimates over laboratories and strata were also tabulated. Finally, to depict agreement and variability, the $ED_{50}s$ and $NF_{50}s$ reported from each laboratory were plotted relative to (i) the other laboratories' reported values and (ii) the consensus estimates.

As a prerequisite for interpreting the variance components analysis, the subset of data from each laboratory was first modeled separately, and then the variance components attributable to day-analyst and plate across laboratories were compared. If they were found to be similar, the variance components (laboratory,

day-analyst, and plate) were extracted from the activity model to describe the precision across laboratories. Variability was expressed as the relative standard deviation (R_2), which is equal to $2^s - 1$, where s is the standard deviation attributable to a given level in the hierarchy of random effects. R_2 may be interpreted like a coefficient of variation (26).

RESULTS

Each laboratory reported at least two $ED_{50}s$ and two $NF_{50}s$ for each of the 108 samples. A subset of laboratories (laboratories A, C, E, and G) also reported estimates of the 4PL model curve shape parameters: the lower asymptote, the upper asymptote, and the slope from the CDC ED50.51 SAS data reduction software. Only the data from these four laboratories were used for the characterization of the TNA among species, as laboratories B and F provided that information but not in an electronic format and the software used by laboratory D did not output analogous information. The data analyses included in this study comprised quantitative descriptions of (i) the characteristics of the TNA across species, (ii) the agreement of the results of the TNA among laboratories, and (iii) the precision of the TNA across laboratories.

Characteristics across species. Figure 1 shows the results of the comparison among the species and the human reference sample run on each plate within each of the four laboratories
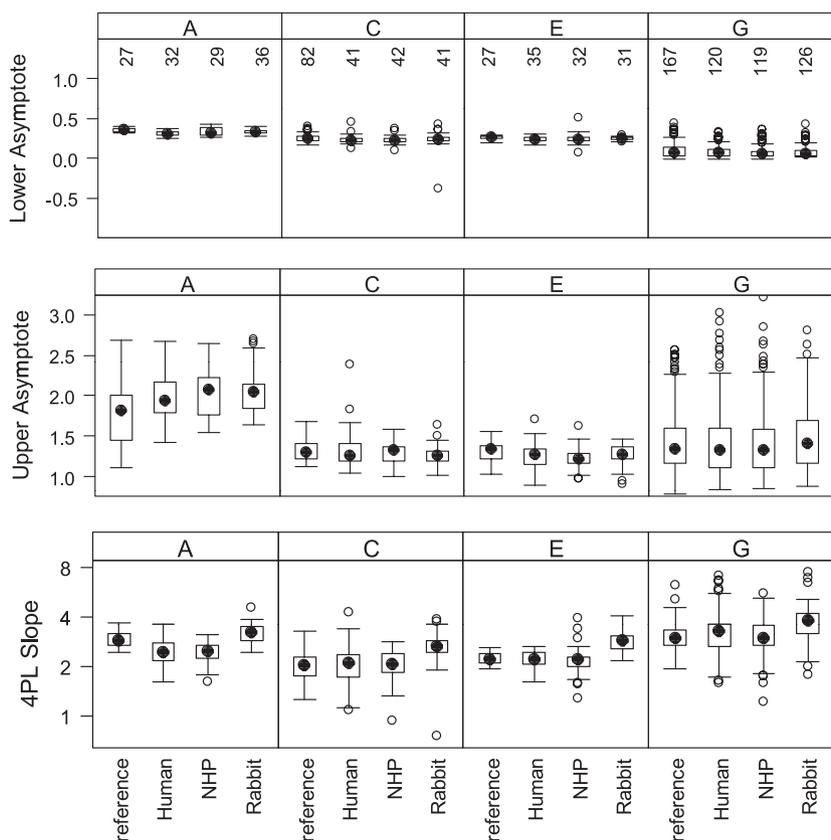

FIG. 1. Box-and-whisker plots of 4PL model curve shape parameters for sera from reference material and three species from four laboratories (laboratories A, C, E, and G); sample sizes are indicated at the top. Only values associated with curves identified as "full" by the data reduction software are shown. Filled dots, medians; boxes, first to third quartiles; whiskers, extensions of 1.5 interquartile ranges past the boxes; open dots, data that were beyond the whiskers. Note the comparable ranges of the y-axis scales for the lower and upper asymptotes. Four estimated upper asymptotes were greater than the upper limit of the graph (all for NHP sera from laboratory G, with values of 3.6, 3.8, 5.5, and 7.0). Note the log scale of the y axis for slope.
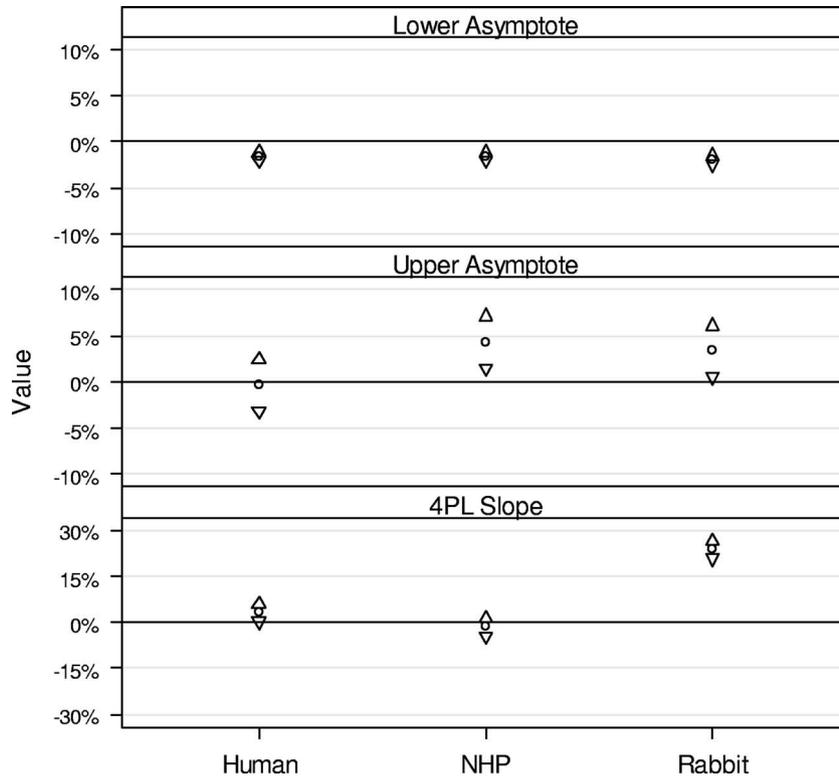
FIG. 2. Estimated differences (dots) with 90% CIs (triangles) for parameter estimates for each species relative to the results for the reference material, expressed as a percentage of the reference curve depth for asymptotes and a percentage of the reference 4PL model slope for that parameter. In equivalence testing, only the farther of the two confidence limits (thus, a one-sided 95% confidence limit) is considered.

that reported the asymptotes and slope data. While each of the parameters might vary across the laboratories shown, the box-and-whisker plots show that the parameters substantially overlap across species within each laboratory. Figure 2 presents the estimated percent differences with 90% confidence intervals (CIs) for parameter estimates for each species relative to the results for the reference material. Much of the variation in these parameters was associated with the laboratory and day-analyst, which was accounted for in the mixed-effects models

(see Table 4). As shown in Fig. 2, all three species had lower asymptotes less than that for the reference material, but when data were expressed as percentages of the reference material curve depth, none was more than 2% smaller. The upper asymptote for human sera was less than that for the reference material, but by no more than 3% of the curve depth; NHP and rabbit sera had upper asymptotes greater than the upper asymptote for the reference material, but by no more than 7%. Finally, the 4PL model slope differences for human and NHP
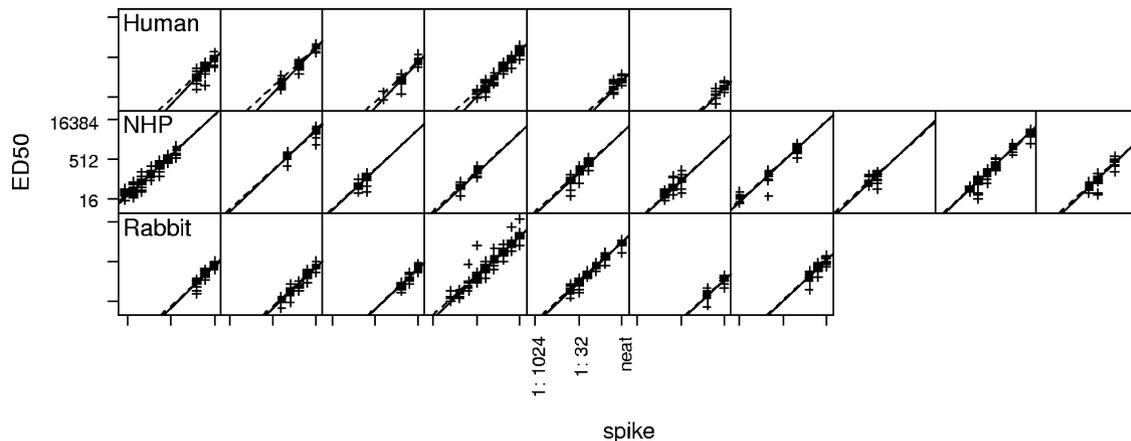

FIG. 3. Dilutional linearity plots for the 23 specimens that were assayed at more than one spiked concentration. Each row of panels depicts the specimens from a particular species. Solid lines have the specimen-specific intercept and species-specific slope estimated in the dilutional linearity model; dashed lines have the same intercept but a slope equal to 1.

TABLE 2. Sample sizes and percentages of samples reported as nonresponsive, organized by stratum and laboratory[a]

Stratum	No. of samples								% of samples nonresponsive							
	All	A	B	C	D	E	F	G	All	A	B	C	D	E	F	G
I	265	28	29	30	28	28	38	84	50	36	93	20	86	29	50	45
II	316	36	41	35	34	34	34	102	3.5		17.1		8.8			1.0
III	364	40	43	41	40	40	40	120								
IV	290	33	33	32	32	32	32	96								
V	290	34	32	32	32	32	32	96								
VI	345	38	40	39	38	38	38	114								
All	1,870	209	218	209	204	204	214	612	7.6	4.8	16	2.9	13	3.9	8.9	6.4

[a] Zeros are not shown.

serum classes were within 6%, while that for rabbit serum was within 27% of the 4PL model slope for the reference material.

The three species were also compared on the basis of dilutional linearity. Figure 3 shows the relationship between known relative spikes and the ED_{50}s measured for each spike. Ideally, the slope for dilutional linearity would be equal to 1. However, the model with the slope constrained to equal 1 fit the data poorly compared to the fit for the model for which the slope was estimated as a free parameter (log likelihoods, -681.34 and -651.25 with 27 and 28 free parameters, respectively; likelihood ratio test, $P \ll 0.001$). The average slope was 1.05 across all three species (95% CI, 1.04 to 1.06). Slope differed among species (log likelihood for the model with species as a predictor, -630.77 with 30 free parameters; likelihood ratio test versus common slope, $P \ll 0.001$). The species-specific slopes were 1.16 (95% CI, 1.13 to 1.19) for humans, 1.03 (95% CI, 1.01 to 1.04) for NHPs, and 1.04 (95% CI, 1.02 for 1.06) for rabbits. While the 95% CIs did not overlap 1.0, none of the upper 95% CIs were over 1.2, indicating overall good dilutional linearity.

Agreement among laboratories. Agreement among laboratories was assessed by comparison of each laboratory's neutralizing activity results for each sample to each other and to the consensus values. In one analysis, the frequency with which each laboratory identified samples as negative was examined. In total, 143 (7.6%) of the reported ED_{50}s yielded a nonresponsive (negative) result, and most of these ($n = 132$) were in stratum I (Tables 2 and 3). The other 11 instances were all in stratum II and were reported predominantly by laboratories B ($n = 7$) and D ($n = 3$). Of the reported ED_{50}s for samples in stratum I, 50% were reported as nonresponsive across all lab-

oratories; the proportions were as high as 93% at laboratory B and 86% at laboratory D and as small as 20% at laboratory C.

The level of agreement for quantitative reported values was generally high (Fig. 4 and 5). Laboratory A tended to report lower ED_{50}s than the other laboratories, while laboratory C tended to report higher values (Fig. 4A). Use of the NF_{50} as the assay readout improved the agreement among the laboratories (Fig. 4B). Half (35 of 71) of the reported values that were far from the consensus values were reported by laboratory D, and nearly all of those (29 of 35) were in strata I to III. Additionally, all but one of the 2-standard-deviation limits above and below the mean for each laboratory were within twofold of the consensus values.

The results for the laboratories involved in the technology transfer program (laboratories A, B, C, E, F, and G) agreed quite well when NF_{50} was used as the readout (Fig. 5), with the results for most samples being well within a twofold difference. The results of all laboratories agreed well for samples with higher titers, again, with the reported values being well within twofold differences.

Precision across laboratories. Variance components analysis revealed that when ED_{50} was used as the assay readout, 46% of the variance was attributable to the laboratory and 35% was residual (unexplained) variance (Table 4). In contrast, when NF_{50} was used, the variance attributable to the laboratory was diminished to 18%. The absolute residual variance was nearly unchanged when NF_{50} was used as the assay readout, but as a percentage of total variance, it increased to 62%. R_2 was reduced when NF_{50} was used as the assay readout (35%) compared to that when ED_{50} was used as the assay readout (45%) (Table 4).

TABLE 3. Percentages of reported ED_{50} and NF_{50} values that were far from the consensus values, organized by stratum and laboratory[a]

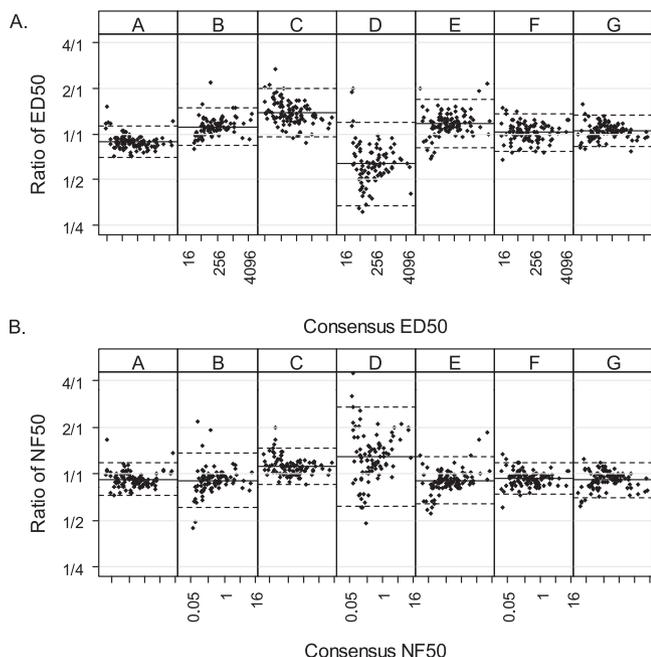| Stratum	% Quantitative reported values far from consensus:															
	ED_{50}								NF_{50}							
	All	A	B	C	D	E	F	G	All	A	B	C	D	E	F	G
I	10	5.6			100	10	5.3	11	6.8	5.6			100	5.0		6.5
II	7.2		8.8	5.7	35	2.9	5.9	3.0	7.5		18	5.7	29	2.9	5.9	3.0
III	6.3		2.3	2.4	35		2.5	5.0	4.7		2.3	2.4	25		2.5	3.3
IV	2.1	3.0	3.0		9.4			1.04	1.7		3.0		13			
V	0.69				6.3				0.34				3.1			
VI	1.4				2.6	7.9		0.88	1.2					7.9		0.88
All	4.1	1.0	2.7	1.5	20	3.1	2.1	2.8	3.4	0.5	4.3	1.5	16	2.6	1.5	1.9

[a] Zeros are not shown.

FIG. 4. Plots of residuals relative to consensus estimates of neutralizing activity. Patterns for both ED_{50} (A) and NF_{50} (B) are shown. Points show the residuals (on the log scale; the y-axis scale is exponentiated for clarity, i.e., $2^{residual}$). Solid line, mean ratio; dashed lines, 2 standard deviations above and below the mean.

FIG. 5. Bland-Altman plots of ratio (y axis) between two laboratories' reported NF_{50}s relative to the geometric mean of the two laboratories' reported NF_{50} (x axis). The y-axis scale is exponentiated for clarity, i.e., $2^{residual}$. Solid line, mean ratio; dashed lines, 2 standard deviations above and below the mean.

DISCUSSION

The appropriate interpretation of the assay data generated in support of nonclinical and clinical studies requires a thorough understanding of assay performance. In the case of application of the anthrax LT TNA to the animal rule for the licensure of anthrax vaccines, the data generated for different species and perhaps in different laboratories will have to be compared. In particular, assays in support of animal studies may be performed in laboratories different from those where the assay was performed in support of clinical trials. If the assay performance is not consistent among species and across laboratories, then data interpretation is more complicated. This interlaboratory study provided a head-to-head comparison of the TNA in three species, in seven different laboratories, and across the range of the assay. The data from this study supplement the species-specific assay validations performed in each key laboratory and provide evidence that the TNA is an appropriately rugged and panspecies assay. This study does not provide information that would support the differences in the protective level of antibody among species. The protective level has been fairly well established in rabbits (10, 16) but not in NHPs. The protective levels will need to be determined for each species and then compared to the antibody levels seen in humans during vaccine clinical trials.

Differences in the titration curves among samples from various species may indicate that the antibodies from each species are dissimilar in avidity or affinity to the LT components. If these qualities are substantially different between species, then extrapolation of the mechanism of protection or the level of antibody required for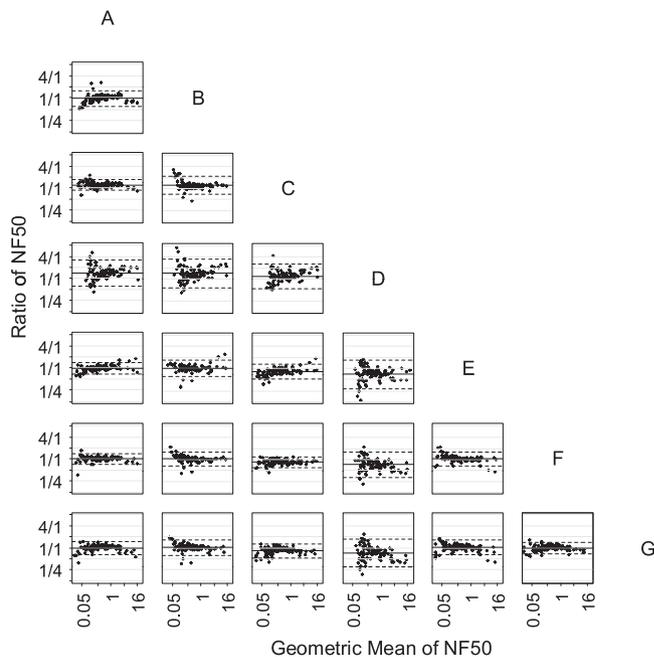 protection among the species becomes difficult. The data in this study suggest that the titration curves for the rabbit samples may be slightly steeper than those for the human and NHP samples. These results imply that the antibody populations produced by rabbits may be slightly different from those produced by humans or NHPs. Given the overall species relatedness, some differences would be anticipated. A difference of less than 30% between the 4PL model slope for the rabbit sera compared with that for the human reference serum was seen in this study. Differences in the curves generated by the human reference material can likely be attributed to normal, unavoidable species variability and not to fundamental differences in the protective action of the antisera. As stated above, however, the level of antibody required for protection cannot be addressed in this study but will be determined by animal challenge studies.

Another determination of interspecies comparability was dilutional linearity modeling. Dilutional linearity assesses rela-

TABLE 4. R_2 and relative contribution to total variance for each response variable associated with laboratory, date-analyst, and plate as well as residual variance and total variance

Source	ED_{50}		NF_{50}	
	R_2 (%)	Contribution (%)	R_2 (%)	Contribution (%)
Laboratory	29	46	13	18
Date-analyst	15	14	11	12
Plate	8.4	4.8	8.8	7.9
Residual	25	35	26	62
Total	45	100	35	100

tive accuracy and is evaluated by spiking known amounts of antibody into normal serum and then performing the assay to determine the ability to recover the expected concentration. Ideally, the slope of the line when the results for the known relative spiked samples are graphed versus the recovered ED_{50} should be near 1.0. The overall dilutional linearity values for all three species are very close to 1.0; however, this study suggests that the assay with human serum may have a steeper dilutional linearity slope than the assays with rabbit or NHP serum. Investigations into the possible causes for this difference indicated that the steeper dilutional linearity slope is not matrix related (data not shown). Given the limited number of samples analyzed for dilutional linearity in this study and the magnitude of the differences (the slopes of ED_{50} relative to spiked concentration equal 1.04, 1.03, and 1.16 for assays with rabbit, NHP, and human sera, respectively), the steeper slope for the assay with human serum probably does not indicate a fundamental difference among species. In the study of Li et al. (9), the acceptance criteria for dilutional linearity in a single laboratory was a slope between 0.7 and 1.3. The results for sera from all species tested in this study were well within that range. Preliminary validation data indicate that the bias introduced by small departures from dilutional linearity can be compensated for by restricting the working range of the assay (data not shown). Additional in-depth dilutional linearity experiments for all three species are being conducted in separate species-specific TNA validation studies.

The level of agreement among the results of the various laboratories, whether the results for the laboratories were compared to a consensus value or directly to each other, was good. Discrepancies were more common for samples with values at the low end of the assay than those with ED_{50}s over 100. We do not have complete information on the lower limit of quantitation for each laboratory but would expect agreement to be poorer at the limits of the assay. For example, ED_{50}s estimated from partial curves would be expected to be more variable as the program tries to extrapolate the best-fit 4PL model from limited data.

Even those laboratories that had participated in an interlaboratory technology transfer program had some methodological differences. One laboratory did not fully participate in the interlaboratory technology transfer and also used a different software program. The results from that laboratory were the furthest from the consensus values. However, even with methodological differences, the comparability among the laboratories for samples that had ED_{50}s above 100 was excellent. The data suggest that the assay can be performed consistently across laboratories and that interlaboratory exchanges of reagents, protocols, and software were successful in improving the agreement among the laboratories. In the future, assessment of proficiency panel samples with well-established values would likely be the most useful way for laboratories to ensure that their assays are performing comparably to others. Avoiding substantive differences in cell preparation, assay protocols, and software will also help to ensure consistency among laboratories.

The precision of the assay was assessed with regard to reproducibility and intermediate precision for date, analyst, and plate. When the ED_{50} was used as the assay readout, the total R_2, which may be interpreted like a coefficient of variation, was only 45%, with 46% of the variance due to laboratory-to-laboratory variability. In the authors' opinion, this level of precision is certainly within the expected performance capabilities of a cell-based neutralization assay. When the NF_{50} readout was used, the total R_2 dropped to 35%, with only 18% of the variance due to laboratory-to-laboratory variability. This improvement in reproducibility suggests that the NF_{50} readout may have a decided advantage for the normalization of data between laboratories and over time and that it will continue to be evaluated as the preferred readout in future studies.

We evaluated the TNA as it has been routinely performed in several laboratories. This study directly demonstrated the consistent accuracy and precision of the assay when it was performed head-to-head in different laboratories. Some of the laboratories analyzed each serum sample only once within an assay; these laboratories' performance was comparable to those of laboratories that assay samples in triplicate by the assay. Thus, the assay has successfully been adapted to a higher-throughput format, which makes it more practical for use in clinical trials with large numbers of samples. The data from this study also established the panspecies nature of the assay. A number of studies have demonstrated that the toxin-neutralizing antibodies correlate with protection (10, 13, 16, 23, 27). On the basis of the performance of the assay demonstrated in this study, the TNA appears to be appropriate from a methodological standpoint for use in bridging the efficacy of the anthrax vaccines in animals to the immunogenicity of those vaccines in humans.

REFERENCES

1. **Burnham, K. P., and D. R. Anderson.** 2002. Model selection and multimodel inference: a practical information-theoretical approach, 2nd. Springer-Verlag, New York, NY.
2. **Campbell, J. D., K. H. Clement, S. S. Wasserman, S. Donegan, L. Chrisley, and K. L. Kotloff.** 2007. Safety, reactogenicity and immunogenicity of a recombinant protective antigen anthrax vaccine given to healthy adults. Hum. Vaccines **3:**205–211.
3. **Friedlander, A. M., R. Bhatnagar, S. H. Leppla, L. Johnson, and Y. Singh.** 1993. Characterization of macrophage sensitivity and resistance to anthrax lethal toxin. Infect. Immun. **61:**245–252.
4. **Gorse, G. J., W. Keitel, H. Keyserling, D. N. Taylor, M. Lock, K. Alves, J. Kenner, L. Deans, and M. Gurwith.** 2006. Immunogenicity and tolerance of ascending doses of a recombinant protective antigen (rPA102) anthrax vaccine: a randomized, double-blinded, controlled, multicenter trial. Vaccine **24:**5950–5959.

5. **Hauck, W. W., R. C. Capen, J. D. Callahan, J. E. De Muth, H. Hsu, D. Lansky, N. C. Sajjadi, S. S. Seaver, R. R. Singer, and D. Weisman.** 2005. Assessing parallelism prior to determining relative potency. PDA J. Pharm. Sci. Technol. **59:**127–137.

6. **Hering, D., W. Thompson, J. Hewetson, S. Little, S. Norris, and J. Pace-Templeton.** 2004. Validation of the anthrax lethal toxin neutralization assay. Biologicals **32:**17–27.

7. **Ivins, B. E., M. L. M. Pitt, P. F. Fellows, J. W. Farchaus, G. E. Benner, D. M. Waag, S. F. Little, G. W. Anderson, P. H. Gibbs, and A. M. Friedlander.** 1998. Comparative efficacy of experimental anthrax vaccine candidates against inhalation anthrax in rhesus macaques. Vaccine **16:**1141–1148.

8. **Lacy, D. B., and R. J. Collier.** 2002. Structure and function of anthrax toxin. Curr. Top. Microbiol. Immunol. **271:**61–85.

9. **Li, H., S. D. Soroka, T. H. Taylor, K. L. Stamey, K. W. Stinson, A. E. Freeman, D. R. Abramson, R. Desai, L. X. Cronin, J. W. Oxford, J. Caba, C. Pleatman, S. Pathak, D. S. Schmidt, V. A. Semenova, S. K. Martin, P. P. Wilkins, and C. P. Quinn.** 2008. Standardized, mathematical model-based and validated *in vitro* analysis of anthrax toxin neutralization. J. Immunol. Methods **333:**89–106.

10. **Little, S. F., B. E. Ivins, P. F. Fellows, M. L. M. Pitt, S. L. W. Norris, and G. P. Andrews.** 2004. Defining a serological correlate of protection in rabbits for a recombinant anthrax vaccine. Vaccine **22:**422–430.

11. **Mosmann, T.** 1983. Rapid colorimetric assay for cellular growth and survival-application to proliferation and cyto-toxicity assays. J. Immunol. Methods **65:**55–63.

12. **National Institute of Allergy and Infectious Diseases.** 2002. NIAID biodefense research agenda for CDC category A agents. NIH Publication No. 03-5308. National Institute of Allergy and Infectious Diseases, Bethesda, MD.

13. **Peachman, K. K., M. Rao, C. R. Alving, R. Burge, S. H. Leppla, V. B. Rao, and G. R. Matyas.** 2006. Correlation between lethal toxin-neutralizing antibody titers and protection from intranasal challenge with *Bacillus anthracis* Ames strain spores in mice after transcutaneous immunization with recombinant anthrax protective antigen. Infect. Immun. **74:**794–797.

14. **Phipps, A. J., C. Premanandan, R. E. Barnewall, and M. D. Lairmore.** 2004. Rabbit and nonhuman primate models of toxin-targeting human anthrax vaccines. Microbiol. Mol. Biol. Rev. **68:**617–629.

15. **Pinheiro, J., D. Bates, S. DebRoy, and D. Sarkar.** 2007. nlme: linear and nonlinear mixed effects models. R package, version 3.1-83. R Foundation for Statistical Computing, Vienna, Austria.

16. **Pitt, M. L. M., S. F. Little, B. E. Ivins, P. Fellows, J. Barth, J. Hewetson, P. Gibbs, M. Dertzbaugh, and A. M. Friedlander.** 2001. In vitro correlate of immunity in a rabbit model of inhalational anthrax. Vaccine **19:**4768–4773.

17. **Pittman, P. R., G. Kim-Ahn, D. Y. Pifat, K. Coonan, P. Gibbs, S. Little, J. G. Pace-Templeton, R. Myers, G. W. Parker, and A. M. Friedlander.** 2002. Anthrax vaccine: immunogenicity and safety of a dose-reduction, route-change comparison study in humans. Vaccine **20:**1412–1420.

18. **Pittman, P. R., S. L. Norris, J. G. Oro, D. Bedwell, T. L. Cannon, and K. T. Mckee.** 2006. Patterns of antibody response in humans to the anthrax vaccine adsorbed (AVA) primary (six-dose) series. Vaccine **24:**3654–3660.

19. **Puziss, M., L. C. Manning, J. W. Lynch, E. Barclaye, I. Abelow, and G. G. Wright.** 1963. Large-scale production of protective antigen of Bacillus anthracis in anaerobic cultures. Appl. Microbiol. **11:**330–334.

20. **Quinn, C. P., P. M. Dull, V. Semenova, H. Li, S. Crotty, T. H. Taylor, E. Steward-Clark, K. L. Stamey, D. S. Schmidt, K. W. Stinson, A. E. Freeman, C. M. Elie, S. K. Martin, C. Greene, R. D. Aubert, J. Glidewell, B. A. Perkins, R. Ahmed, and D. S. Stephens.** 2004. Immune responses to Bacillus anthracis protective antigen in patients with bioterrorism-related cutaneous or inhalation anthrax. J. Infect. Dis. **190:**1228–1236.

21. **Quinn, C. P., V. A. Semenova, C. M. Elie, S. Romero-Steiner, C. Greene, H. Li, K. Stamey, E. Steward-Clark, D. S. Schmidt, E. Mothershed, J. Pruckler, S. Schwartz, R. F. Benson, L. O. Helsel, P. F. Holder, S. E. Johnson, M. Kellum, T. Messmer, W. L. Thacker, L. Besser, B. D. Plikaytis, T. H. Taylor, A. E. Freeman, K. J. Wallace, P. Dull, J. Sejvar, E. Bruce, R. Moreno, A. Schuchat, J. R. Lingappa, N. Marano, S. K. Martin, J. Walls, M. Bronsdon, G. M. Carlone, M. Bajani-Ari, D. A. Ashford, D. S. Stephens, and B. A. Perkins.** 2002. Specific, sensitive, and quantitative enzyme-linked Immunosorbent assay for human immunoglobulin G antibodies to anthrax toxin protective antigen. Emerg. Infect. Dis. **8:**1103–1110.

22. **R Development Core Team.** 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

23. **Reuveny, S., M. D. White, Y. Y. Adar, Y. Kafri, Z. Altboum, Y. Gozes, D. Kobiler, A. Shafferman, and B. Velan.** 2001. Search for correlates of protective immunity conferred by anthrax vaccine. Infect. Immun. **69:**2888–2893.

24. **Selinsky, C. L., V. D. Whitlow, L. R. Smith, D. C. Kaslow, and H. M. Horton.** 2007. Qualification and performance characteristics of a quantitative enzyme-linked immunosorbent assay for human IgG antibodies to anthrax lethal factor antigen. Biologicals **35:**123–129.

25. **Semenova, V. A., E. Steward-Clark, K. L. Stamey, T. H. Taylor, D. S. Schmidt, S. K. Martin, N. Marano, and C. P. Quinn.** 2004. Mass value assignment of total and subclass immunoglobulin G in a human standard anthrax reference serum. Clin. Diagn. Lab. Immunol. **11:**919–923.

26. **Tan, C. Y.** 2005. RSD and other variability measure of the lognormal distribution. Pharmacopeial Forum **31:**653–655.

27. **Weiss, S., D. Kobiler, H. Levy, H. Marcus, A. Pass, N. Rothschild, and Z. Altboum.** 2006. Immunological correlates for protection against intranasal challenge of *Bacillus anthracis* spores conferred by a protective antigen-based vaccine in rabbits. Infect. Immun. **74:**394–398.