Biologicals xxx (2010) 1–9

Contents lists available at ScienceDirect

Biologicals

journal homepage: www.elsevier.com/locate/biologicals

A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials☆

Stephen D. Soroka*, Jarad M. Schiffer, Vera A. Semenova, Han Li, Lydia Foster, Conrad P. Quinn

Microbial Pathogenesis and Immune Response (MPIR) Laboratory, Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC), Atlanta, GA, United States

ARTICLE INFO

Article history:
Received 5 April 2010
Received in revised form
25 August 2010
Accepted 2 September 2010

Keywords:
AVRP
Anthrax vaccine adsorbed
Quality control
ELISA
TNA
Bacillus anthracis

ABSTRACT

A two-stage, multilevel assay quality control (QC) system was designed and implemented for two high stringency QC anthrax serological assays; a quantitative anti-PA IgG enzyme-linked immunosorbent assay (ELISA) and an anthrax lethal toxin neutralization activity (TNA) assay. The QC system and the assays were applied for the congressionally mandated Centers for Disease Control and Prevention (CDC) Phase 4 human clinical trial of anthrax vaccine adsorbed (AVA, BioThrax). A total of 57,284 human serum samples were evaluated by anti-PA enzyme-linked immunosorbent assay (ELISA) and 11,685 samples by anthrax lethal toxin neutralization activity (TNA) assay. The QC system demonstrated overall sample acceptance rates of 86% for ELISA and 90% for the TNA assays respectively. Monitoring of multiple assay and test sample variables showed no significant long term trends or degradation in any of the critical assay reagents or reportable values for both assays. Assay quality control data establish the functionality of the quality control system and demonstrates the reliability of the serological data generated using these assays.

Published by Elsevier Ltd on behalf of The International Association for Biologicals.

1. Introduction

Currently, the Food and Drug Administration (FDA) Office of Regulatory Affairs recommends that laboratories supporting Investigational New Drugs (INDs) and New Drug Applications (NDAs) establish a quality control (QC) program for the detection, correction and prevention of deficiencies or errors in laboratory testing processes [1,2]. Establishing such a QC program is therefore essential for any laboratory implementing a long term regulatory compliant clinical trial of vaccine efficacy. A well designed and executed quality control system for *in vitro* biological assays together with an appropriate Quality Management System (QMS) will monitor assay performance over the course of a clinical study to ensure the consistency and assure the validity of the data generated [3].

A specific example of a long term study is the Centers for Disease Control and Prevention (CDC) Phase 4 human clinical trial of anthrax vaccine adsorbed (AVA, BioThrax). This pivotal study was initiated in 1999 at the request of the US Congress to assess the safety and serological noninferiority of alternate schedules and administration routes of AVA [4]. The study encompassed a 43-month enrollment for 1563 participants, each of whom provided up to 17 serum samples for analyses. The primary endpoints for this study were based on the proportions of vaccinees mounting anti-PA IgG responses to the anthrax toxin protective antigen (PA) component of AVA and the magnitude of those responses [4]. The purpose of the data generated by the Anthrax Vaccine Research Program (AVRP) was to inform significant public health decisions on the use and distribution of the only licensed anthrax vaccine in the US. The wide ranging impact of these decisions on vaccination and emergency preparedness policies, together with the long duration of the study, required that the biological assays used for primary and secondary endpoint determination were precise, accurate, sensitive, specific and validated [5–7]. We constructed a quality control system as part of a comprehensive Quality Assurance activity for the laboratory components of the study. The QC system was able to detect and reject unacceptable assay performance, determine that acceptable assay data had a high level of accuracy and precision, and provide an increased level of confidence in the serological assay data. We report here the monitoring

☆ Required Disclaimer: The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

* Corresponding author. Present address: Microbial Pathogenesis and Immune Response Laboratory, Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Mail Stop D-01, 1600 Clifton Rd., Atlanta, GA 30333, United States. Tel.: +1 404 639 0057; fax: +1 404 639 4139.
E-mail address: ssoroka@cdc.gov (S.D. Soroka).

1045-1056/$36.00 Published by Elsevier Ltd on behalf of The International Association for Biologicals.
doi:10.1016/j.biologicals.2010.09.001

methods and findings that pertain to the AVRP serological assay data.

2. Materials and methods

2.1. Quality control (QC) system

The QC system evaluated and tracked biological assay performance characteristics and ensured appropriate investigations and responses to any out-of-specification behaviors. The QC process was initiated by the output of laboratory data to a secure data folder on a local area network (LAN). Quality control of these data was next applied in a two-stage, multilevel format (Fig. 1).

2.2. Standardized in vitro assays

The analytic QC program was designed specifically to monitor the performance data for two highly standardized and validated serological assays, an enzyme-linked immunosorbent assay (ELISA) for detection of *Bacillus anthracis* PA-specific immunoglobulin (IgG) G antibodies in human sera and a PAN-species *in vitro* anthrax lethal toxin neutralization activity (TNA) assay [8,9]. The anti-PA IgG ELISA was used to generate primary and secondary endpoints for the AVRP Phase 4 human clinical trial [4]. The TNA assay was used to provide functional analyses of serum responses to vaccination with AVA [8]. Both assays employ a customized technology platform comprising standard assay formats, specific protocols and test procedures for sample management, lab testing, data analysis, and database management. It also includes the qualification of equipment and control reagents, as well as the interpretation of performance characteristics [2,8–10]. Ruggedness and robustness of both assays were determined through development studies. Both assays are validated and their respective precision, accuracy, lower limits of detection, lower limits of quantification, and dilutional linearity established. The validation master file is on record with the Food and Drug Administration (FDA; BB MF 12964) [2]. The performance characteristics, expected endpoint values and performance ranges for critical reagents, including reference standards and assay quality control sera, were determined. These data were used to derive, test, and establish multilevel, primary assay QC acceptance criteria and as benchmarks for data generated from AVRP primary and corroborative endpoint testing.

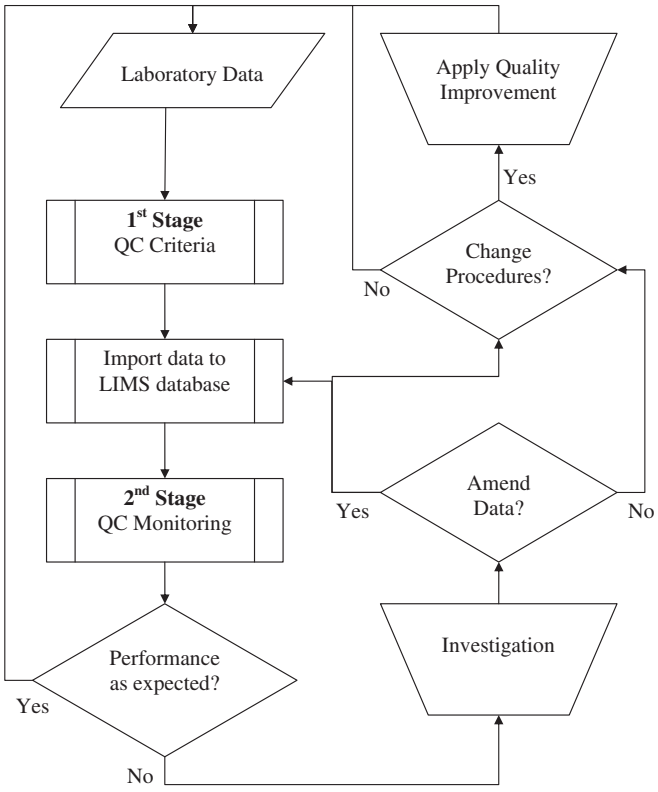

Fig. 1. Flow of serological assay data through a two-stage QC process. The first stage involved analysis of each assay experiment, application of multiple assay QC acceptance criteria from the level of Optical Densities (ODs) to the level of sample test results (i.e., multilevel) to optimize the performance of a single assay run, and review of the data. The second stage of the process applied statistical process controls to monitor assay performance characteristics for multiple assay experiments and to detect and investigate deviations and trends using a customized automated software data monitoring program. The procedure was continually repeated for each set of data until the last sample was tested, analyzed, entered into a laboratory information management system (LIMS) database, and monitored through the quality control program.

2.2.1. Anti-PA IgG ELISA acceptance criteria

The basic components and application of the quantitative anti-PA IgG ELISA have been reported elsewhere [9]. In brief, the assay plate layout consists of a 7-point two-fold, triplicate dilution series of standard reference serum, 8-point two-fold, duplicate dilution series of each test sample (four test samples total per plate), a duplicate single-point negative control, and 3 separate, duplicate single-point positive control sera; plates were typically tested in batches within an experiment. The reportable value of the ELISA is the concentration of anti-PA IgG antibodies expressed in micrograms per milliliter (μg/ml). The anti-PA IgG antibody concentrations of test sera and the positive quality control (QC) sera were calculated by interpolation to a 4-parameter logistic (4-PL) fit of the reference standard data using a calibration factor. The sera AVR414 and AVR801 were used as the reference standards during this study [10]. The analysis was completed using the ELISA for Windows software Version 2.15 [11].

A set of 5 assay plate acceptance criteria and 2 test sample acceptance criteria was applied real-time in sequence to determine data validity. All assay plate acceptance criteria were required to be met, otherwise all test samples on the plate were rejected and sample testing repeated. Specifically these criteria were: 1) the mean Optical Density (OD) value of the negative control must be less than 0.200 OD units; 2) the standard reference serum must have an approximated, weighted r-squared coefficient of determination (r^2) value of $\geq$0.990 to demonstrate "goodness of fit" of the standards data to the 4-PL model; 3) the mean anti-PA IgG concentration for each of three positive quality control sera (high, middle, low concentrations) must have coefficients of variation (CV) <20%, else the concentration was censored; 4) at least 2 of 3 positive control sera must have anti-PA IgG concentrations within 2 standard deviations (SD) of their expected values and 5) all positive control serum anti-PA IgG concentrations must be <3SD from the expected value. The expected percentage of passing plates, based on 2 of 3 controls within 2SD and one within 3SD, was approximately 91%.

For each test sample the overall anti-PA IgG concentration reported was the dilution adjusted arithmetic mean of all calculable serial dilution pair concentrations. All serum specimens tested by ELISA required at least two passing results by independent operators. The criteria specific for each sample tested, to assess both parallelism [12] and intra-assay precision, required that 1) the CV of the overall dilution adjusted mean anti-PA IgG concentration of a test sample was <20% and 2) at least three out of 8 dilution pairs had a calculable concentration within dilution CVs <20%. Exceptions to sample specific criterion #2 included low reactivity samples

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

which are unable to achieve calculable concentrations on three serial dilution pairs. Low reactivity samples were acceptable if all calculable concentrations had within dilution CVs <20%. Samples with a CV between 20 and 50% were assessed for outlier detection and recalculated and acceptance criteria reassessed when necessary. Samples with a CV of the mean anti-PA IgG concentration ≥50% were considered non-parallel to the reference standard curve. Sample concentrations from 2 or more passing results (inter-operator precision) were required to have CV of <30%. Summary calculations (mean, median, CV, and total number of passing results) were automated in a secure laboratory information management system, STARLiMS (STARLiMS Corp, Hollywood, FL).

Additional analyses and computations for multilevel QC monitoring of the data were done in SAS® (SAS Institute Inc. Cary, NC). Variables for these analyses included the dilution factors and OD values of the QC sera, reference standard and test samples curves, the computed concentrations of the positive QC sera, four parameters of the reference standard 4-PL model (upper asymptote infinite concentration response, lower asymptote zero concentration response, symmetric inflection point and slope factor), and QC sera identifiers.

2.2.2. Anthrax lethal toxin neutralization activity (TNA) assay acceptance criteria

A subset of sera was tested by the TNA assay to quantify functional activity of serum antibody responses to AVA. This PAN-species assay is reported elsewhere [8,13]. The assay plate layout consisted of a 7-point two-fold, triplicate dilution series of standard reference serum and three test samples, a negative control, a positive neutralization control, and test serum control. Assay plates were typically tested in batches within an experiment. The standard reference serum used throughout this study in the TNA assay was AVR801. All TNA assay analyses and endpoint calculations were done in SAS® running a customized endpoint calculation algorithm [14]. The program utilized a four-parameter logistic (4-PL) model to fit a dose-response curve to the standard reference serum and test samples [15]. The primary reportable value of this quantitative assay was the effective dilution-50 (ED50), the reciprocal of the dilution corresponding to the inflection point of the 4-PL model that results in 50% neutralization of cytotoxicity activity [16]. Along with the reportable values, additional data captured by the customized software included but were not limited to: ODs of the controls and test samples, the parameters of the 4-PL model, dilutions of the reference standard and test samples, operator ID, testing date, and results for each of the acceptance criteria. Selected variables were then imported into STARLIMS through an automated process that also computed summary calculations of the ED50.

Initial application of QC monitoring for the TNA assay comprised the automation of real-time QC acceptance criteria embedded within the customized interpretive software. There were 12 QC acceptance criteria categorized into three types (primary, secondary, and test sample) and applied in sequence. The multilevel primary and secondary acceptance criteria were used to determine if the assay plate controls and reagents conformed to the validated bioassay model [8]. If these criteria were not met, all test samples on that plate were rejected and testing repeated. The primary acceptance criteria monitored adherent cell acceptability (J774A.1 cell line), anthrax lethal toxin potency, cell density, and the curve formation of the standard reference serum within the 4-PL model. The specific criteria were: 1) the negative control serum (NC) with lethal toxin (≥95% cell lysis) must have a mean OD value ≤0.45; 2) the reference standard's maximum mean triplicate OD value (100% viability) must be ≥0.85 to maintain acceptable curve height; 3) the difference between the mean OD value for the NC and mean triplicate OD for the highest dilution of the standard must be ≤0.25; 4) the difference between the maximum mean triplicate OD and the minimum triplicate OD of the reference standard must be ≥0.55 OD units to demonstrate sufficient curve depth; 5) within dilution CV of OD values for reference standard dilutions must be ≤20% for 6 of the 7 dilutions to demonstrate acceptable intra-assay precision; 6) the mean OD value for the positive neutralization control (PNC) must be ≥0.85 [8]. The secondary acceptance criteria ensured that operator performance, qualified reagent behavior, and 4-PL model fit met expected characteristics. The specific criteria centered around the performance of the reference standard and included: 7) the 4-PL curve of the standard reference serum must have an r^2 value of ≥0.945 to demonstrate acceptable goodness of fit to the 4-PL model fit; 8) at least three dilution points must be on either side of the inflection point to allow for adequate formation and distribution of data for a full sigmoid curve and 9) the ED50 of the reference standard must be within a ±2 SD range from the expected value. Based on assay development data from a 16-month performance period prior to initial study testing the expected passing rate for plates was 92.9%. The test sample acceptance criteria were based on the 4-PL model fit of the test sample data and operator precision. The criteria included: 10) the CV for all triplicate readings for each test sample must be ≤20% for 6 of the 7 dilutions; 11) the test sample dilution curves must have an r^2 of 0.895 in the 4-PL model and 12) the test sample curve must converge using a Gaussian or Marquardt model. If the test sample curve fit demonstrated a lack of curve formation at the higher dilutions, then the test sample was rejected and repeated at a higher starting dilution to ensure formation of a full dose-response curve.

2.3. Monitoring assay performance

Quality control (QC) monitoring of the laboratory assay performance was done using a two-stage format. The first stage included internal QC acceptance criteria that were integrated into the analysis process allowing for real-time acceptance or rejection decisions. The second stage involved an additional QC step to provide statistical process control of multiple plate runs and multiple variables via customized, quality control monitoring (QC-Mon) programs designed using SAS®. The QC-Mon programs monitored the assay performance characteristics and detected potential operational errors and trends within the data. The QC-Mon programs do not automatically correct or reject study data. The programs collated data from SAS® datasets and from STARLIMS using a structured query language (SQL) procedure. The QC-Mon programs comprised 3 monitoring sections, from which data for multiple priority assay performance variables were presented through a series of text outputs and graphs. The monitoring sections were 1) review of manual data entry, 2) review of recent data (RD) and 3) review of long term data (LD). Inspection of RD and LD was in user-defined time ranges. Both QC-Mon programs were fully automated and contained within a SAS macro statement. This design facilitated specific end-user flexibility within the program to adjust date ranges and acceptable passing rates. QC-Mon programs were run weekly by trained analysts.

The review of manual data entry was designed to detect operator data entry errors. Variables of user identifier (ID), date tested, sample ID, QC reagent ID, and dilution entry were checked for inaccuracies. Output of these data was provided when inaccuracies were detected.

The RD portion of the program monitored the most recent laboratory activity based upon the date the program was run (Table 1). The program produced summary statistics, correlations, and graphs; output would only be displayed if there were discrepancies/deficiencies. Shewhart control charts were produced for longitudinal performance of the three positive control

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

Table 1
Listing of the criteria and their description used to monitor serological assay data within the quality control monitoring (QC-Mon) Programs.

Criteria	Description	QCM-ELISA[a]	QCM-TNA	Graph	Output/Table	Recent/Long Term[b]
Model Fit	Observance of the 4-PL fit of the reference standard where the model did not converge, or was improperly fitted to the data	✓	✓		✓	R
Plate Passing Rate	Observance of plate passing percentage by experiment identifying experiments containing > X% failed plates	✓	✓	✓	✓	R
Sample Passing Rate	Observance of passing percentage of samples by experiment on the passing plates of an experiment with >X% failed samples	✓	✓	✓	✓	R
Summary Statistics	Summary statistics of ELISA positive controls concentrations or TNA reference standard ED50	✓	✓		✓	R,L
Positive Control Ranges	Table of ELISA positive control concentrations corresponding to established standard deviation (SD) ranges	✓			✓	R
Correlation Coefficients	Table of correlation coefficients comparing the standard's upper asymptote, lower asymptote, slope, reportable values, and optical density (OD) dilution points	✓	✓		✓	L
Plate Frequency	Display of the frequency and passing rate of plates tested, by operator	✓	✓	✓		R,L
Positive QC Data	Observance of each ELISA positive control concentration by operator; it also displays the recent mean, established mean and SD ranges	✓		✓		R,L
Sample Passing Results	Summary table displaying the sample acceptance criteria results by operator; results are shown as either 'pass' or coded reason for sample failure	✓	✓		✓	R,L
Standard's Reportable Values	Graphical Display for a reference standard's reportable value by operator	✓	✓	✓		R,L
Standard's Parameters	Observance of the reference standard's upper asymptote, lower asymptote, or slope over time with point labels displaying plate QC pass/fail	✓	✓	✓	✓	L
Negative Control	Observance of the OD of the negative control over time with point labels displaying QC pass/fail	✓	✓	✓		L
Standard OD Dilution	Observance of the mean OD associated with each dilution point of the reference standard over time. Dilution points are color-coded	✓	✓	✓		L
Positive Neutralization Control	Observance of the OD associated with the TNA positive neutralization control over time with point labels displaying QC pass/fail		✓	✓		L

[a] Check marks indicate the criteria are being monitored by the TNA (TNAQCM02) and/or ELISA (ELQCM03), if there are graphical data, and if there are output and/or tables.
[b] Monitored via recent (R) and/or long term (L) data.

concentrations in the anti-PA IgG ELISA and for the ED50 of the TNA reference standard. The Shewhart method of process control statistically analyzes variances in the assay to distinguish random variation from error [17]. Control chart target values for the Shewhart method were defined using the expected or assigned endpoint mean value for each control or standard. The corresponding upper and lower control limits were set at 1SD, 2SD, and 3SD from the mean. Assay data were plotted using mean daily results (data points) and were evaluated using eight tests for special causes [18] (Table 2). Investigations were initiated for all detections of special causes – outliers that are assignable but are not representative of the dataset – in the control data.

The LD monitoring portion of the program involved the visual inspection of multiple variables charted longitudinally for the duration of the AVRP study. Some of these variables were utilized in review of recent data (RD) (Table 1). All charts of longitudinal data were automatically saved on a secure network drive. Each time a QC-Mon program was applied, new data were appended to the previous long term data and new files were generated. Any potential long term trends were investigated.

Table 2
Eight tests for special causes within Shewhart control charts.

Test Index	Description
1	1 data point >3 Standard Deviations (SD) from the expected value
2	9 data points in sequence above or below the expected value
3	6 data points in sequence steadily increasing or decreasing
4	14 data points in sequence alternating up and down
5	2 out of 3 data points in sequence >2SD from the expected mean
6	4 out of 5 data points in sequence >1SD from the expected mean
7	15 data points in sequence <1SD from the expected mean
8	8 data points >1SD with no data points within 1SD

The tests were designed and utilized for two serological assays as a method to detect and distinguish random (common cause or expected) variation from assignable (special cause or non-random) variation.

3. Results

3.1. Descriptive statistics

The overall performance of the ELISA Positive control concentrations and the TNA assay reference standard ED50 demonstrated high precision and accuracy for both assays over time (Table 3). The CVs of the ELISA positive controls concentrations ranged from 15.8% to 19.2% with a percent error ranging from 0.2% to 6.5%. The ED50 of the TNA reference standard AVR801 demonstrated a CV of 26.3% and a percent error of 14.0%. These numbers demonstrated similar performance to the precision and accuracy data reported in assay development and validation reports which are on record with the Food and Drug Administration (FDA) as master file BB MF 12964.

3.2. Anti-PA IgG ELISA and TNA assay acceptance criteria

There were 22,221 specimens analyzed for anti-PA IgG by ELISA. Each specimen required a minimum of 2 passing results from 2 different operators. A total of 57,284 test sample aliquots were assayed on 14,713 plates by 8 operators over a 73-month testing period. The overall sample passing rate was 86.2%, with passing rates by operator ranging from 72.5% to 91.9%. The rate for assay plate failures was 9.6%, close to the projected rate of 9% calculated from the QC acceptance criteria of the positive controls. The additional 0.6% of plate failures were attributable to out of range values on the negative control serum and the fit of the standard reference serum to the 4-PL model. Sample failure due to non-parallelism was encountered for only 0.2% of total samples ($n = 124$). All non-passing test samples were retested and data reported.

Of all received specimens, 46.8% ($n = 10,405$) were tested in the TNA assay. In total, there were 11,685 test samples assayed on 3908 plates by 7 different operators over a period of 23 months. The overall sample passing rate was 90.2%. The passing rate for test

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

Table 3
Summary statistics for the ELISA positive quality control (QC) concentrations and the TNA reference standard ED50.

Sample	Reportable Value	*n*	Expected Reportable value	Observed Reportable value	Standard Error	Observed % CV	% Error
AVR216	ELISA Positive Control Concentration (μg/ml)	6362	102.0	99.1	0.2	17.7	2.9
AVR284		6426	100.9	99.4	0.2	19.2	1.5
AVR370		6427	59.8	55.9	0.1	17.8	6.5
AVR1749		8261	446.4	419.7	0.7	15.8	6.0
AVR1750		8271	88.7	88.5	0.2	18.3	0.2
AVR1751		8257	39.4	37.9	0.1	16.6	3.8
AVR801	TNA Reference Standard ED50	3908	656	564	2.4	26.3	14.0

Evaluation of accuracy and precision over time on plates used to generate data for the Anthrax Vaccine Research Program (AVRP). Accuracy was expressed as the percent error (% Error) between the observed reportable values and the expected values. Precision was expressed as the % Coefficient of Variation (CV) of the reportable value. ELISA data were captured over a 73-month testing period; TNA data were captured over a period of 23 months.
n = number of observations.

samples by operator where the number of tests was >16 ranged from 87.8% to 92.2%. There was a 7.4% plate failure frequency ($n = 866$) due to not meeting the assay primary and secondary acceptance criteria. The plate failure rate of 7.4% was comparable to prior studies (7.1%) using the same standard reference serum AVR801 (data not shown). Of the 12 acceptance criteria applied to all test samples, criterion #9 (Standard ED50 being within 2SD of the expected mean) failed to meet its requirement most often (3.4% of all samples tested). Criterion #8 (adequate formation and distribution of data for a full sigmoid standard curve) met the requirement 100% of the time. Seven of the twelve criteria met their requirement >99.0% of the time. Of the 10,539 samples that passed the real-time QC acceptance criteria, 28 samples (0.3%) were determined to have improper curve fits to the 4-PL model. These results were rejected and the samples retested according to protocol.

3.3. Anti-PA IgG ELISA quality control monitoring

In total, there were a total of 3207 experiments containing 1–8 assay plates per experiment, with an average number of plates per experiment between 4 and 5. There were 2457 experiments (76.6%) in which all plates within the experiment passed the plate acceptance criteria. Approximately 94.3% of experiments had over 50% of all plates pass the plate acceptance criteria. Investigation into the 5.7% of experiments that had more than 50% of plates fail plate acceptance criteria revealed the majority of plate failures were due to not passing the acceptance criteria for the positive control concentrations. All test samples on failed plates were subsequently retested. Of the 3121 experiments which contained at least one passing plate, 138 experiments (4.4%) had more than 20% of the test samples on passing plates fail the test sample acceptance criteria. Most of these 138 experiments spanned the 73-month testing period; however, the ELQCM03 program was able to detect a larger than normal percentage of these experiments occurred during a 2-month period. Investigation into this testing period revealed a low frequency of errors and no OD trends or shifts or above average plate failure rates. The above average number of failed test samples decreased in the following month.

Two standard reference sera, AVR414 and AVR801, were used for the AVRP anti-PA IgG serological testing. A total of 5839 assay plates were completed using AVR414, and 8874 using AVR801. Inspection of the 4-PL standards data for AVR414 showed no significant trends with the standard's upper asymptote, lower asymptote, and slope, although there was a slight downward shift down in the optical density (OD) values for the upper asymptote (3.4–2.9 OD units). No corresponding change was evident in the standard's dilution points OD values or in any of the concentrations of the positive control QC sera. Inspection of data from assay plates using reference standard AVR801 showed no significant trends over time for the standard's upper asymptote, lower asymptote, and slope. Examination of the AVR801 OD dilution points revealed a single upward shift at a specific time point in the study. The upward shift in ODs was also evident in the four parameters of the standard curve. No shift or trend was detected in the reportable values (μg/ml) of the three positive controls before or after this OD shift was detected (*T*-test, >0.05). Furthermore, the sample acceptance rate was not affected by the OD shift. This observation makes two key points; firstly, the QCM program provides a high stringency real-time review of assay data that facilitates immediate scrutiny of assay behavior; secondly, that determination of antibody concentrations using a calibrated reference standard is much more robust than use of raw OD values in determining assay endpoints.

Six positive quality control (QC) sera were used in 2 sets of 3 for the AVRP study. Set 1 comprised QC sera AVR216, AVR284, and AVR370; set 2 comprised QC sera AVR1749, AVR1750, and AVR1751. The overall mean anti-PA IgG concentrations and CVs are provided in Table 3. Accuracy of all positive QC sera, determined by percent error compared to the expected values ([(observed − expected)/expected] × 100) ranged from 0.2% to 6.5%. The frequency for each individual positive control concentration that calculated within 2SD of the expected mean ranged from 86.4% to 94.9%. Inspection of positive QC sera mean concentrations over time (months) demonstrated high precision and accuracy for the overall study (Table 3, Fig. 2). An increase in variance was noted in the final stages of the testing periods compared to preceding periods. This was associated with a reduction in the total number of test results in the analysis sample data points compared to the overall test result average ($n < 45$ vs. $n > 200$). There was no associated increase in experiment or test sample fail rates. Inspection of the monthly mean concentrations for all positive control sera demonstrated that 96% had a CV <30% and 85% had a CV <20%, emphasizing good precision of the assay on a month-to-month basis (Fig. 3). Overall, the CVs by month were similar to the established (expected) CVs (~14% on average) for each of the positive controls. Four of the six positive controls had at least one month when the CV exceeded 30%. Investigation of these months with a CV >30% revealed that in each case a set of plates from one experiment contained positive control sera concentrations much different from the expected value of those sera. These assay plates had failed the initial plate QC acceptance criteria and the test samples were repeated. This not only demonstrates the ability of the program to identify these plates, but also exemplifies the ability of our QC acceptance criteria to reject discrepant data. There was only a single month period in the entire duration of the 73-month study when more than one positive serum mean concentration had a CV >30%. Further

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

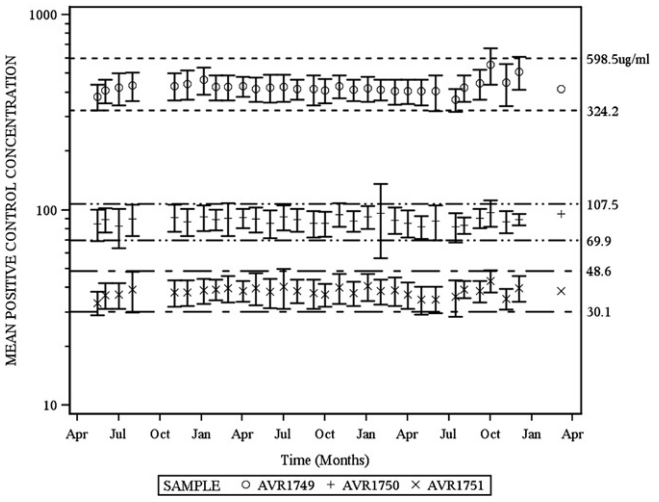

Fig. 2. Accuracy and precision of mean ELISA positive control concentrations. Plot of concentrations (passing and failing results) for AVR1749 (○), AVR1750 (+), and AVR1751 (x) over a 30 month testing period. Standard deviation bars were included for each concentration. Control limits for each positive control, shown with dashed lines, were determined from an established (expected) reportable value. Graphical representation was similar but not shown here for AVR216, AVR284, and AVR370. Increases in accuracy and/or precision were not associated with increases in test sample failure rates.

analysis of these aberrant spikes showed that the sample size for that period was low ($n = 5$) for each QC serum, with all results within 3SD of the expected mean.

Each of the daily test reportable values for each of the positive control sera was averaged and analyzed by inspection using Shewhart control charts. There were 790 discrete test day observations analyzed in this manner. Of these, 358 test days used positive controls AVR216, AVR284, and AVR370, and 432 test days used positive controls AVR1749, AVR1750, and AVR1751. For all positive control sera, the most frequent QC prompt was to have nine data points in sequence on one side of the target value (QC test #2, $n = 56$). Investigation of these indicated that the majority of mean daily concentrations occurred within 1SD of the expected mean value and in many cases overlapped with a QC prompt for 15 points in a sequence within 1SD of the target value (QC test #7, $n = 34$). QC prompts for test #1 (1 data point beyond 3 SD of the target value) were triggered 12 times. The mean daily concentrations before and after each QC test #1 prompt were shown to have concentrations near the target value. Many of the QC prompts for test #1 had a low sample count ($n < 8$) for that particular day resulting from one experiment; of the 3 days with a higher sample count ($8 < n < 21$), investigation showed that one operator's test plates reported a positive control concentration outside of 3SD while the other operators' positive control concentrations were within 3SD, thereby inflating the mean daily concentration. There were no QC prompts for Test #8.

3.4. TNA assay quality control monitoring

There were a total of 931 experiments containing 2–8 plates assayed per experiment over 220 test days. Of these experiments, 873 (93.8%) had 50% or more of its plates pass the primary and secondary acceptance criteria. With multiple experiments tested each day, plate QC was also assessed by day. There were 12 days in which at least 50% or more plates completed on those days failed the plate QCs. The utility of the QCM program was its ability to demonstrate that each day was a sporadic incident and similar rates

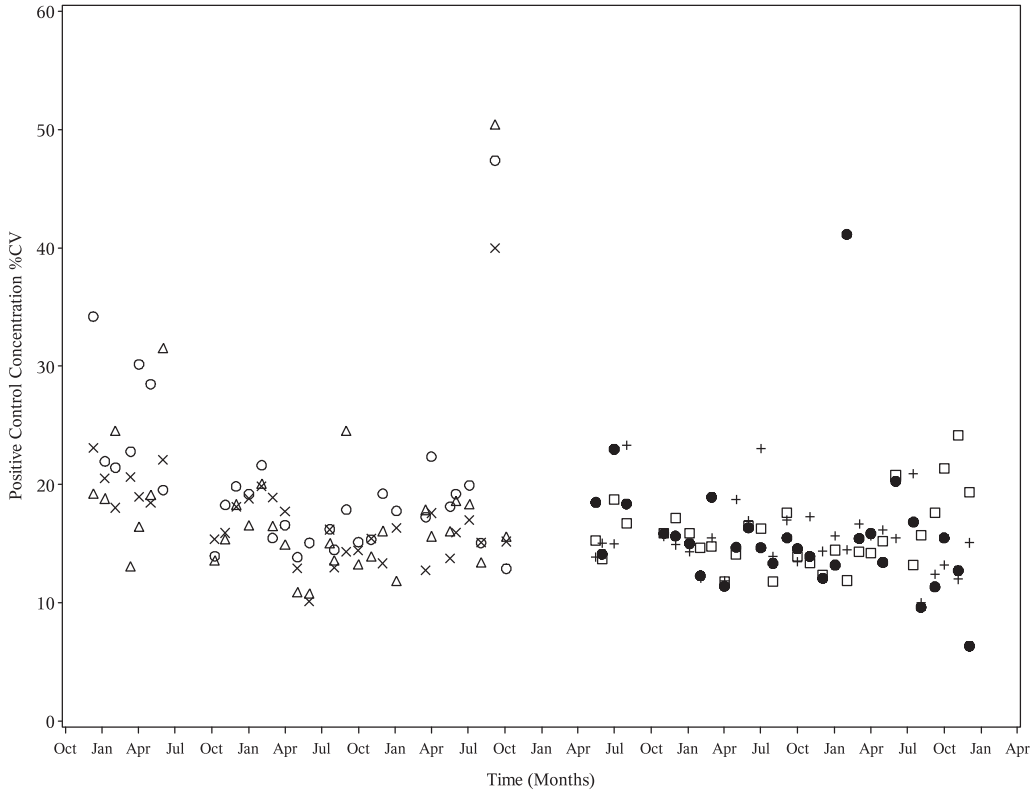

Fig. 3. Percent Coefficient of Variation (%CV) of ELISA positive quality control (QC) concentrations. Plot of %CV for QC concentrations by month for AVR216 (x), AVR284 (○), AVR370 (△), AVR1749 (□), AVR1750 (●), and AVR1751 (+). The monthly CVs were similar to the established CVs (~14% on average) over the course of the study. Overall, 96% of monthly concentrations had a %CV < 30% and 85% had a %CV < 20%, demonstrating good precision over a 73-month testing period.

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

were not seen on the previous testing date or the next testing date. All of the plates passed the plate QCs on approximately 74% of all test dates. Of the test samples on plates that passed the plate QC, there were 25 experiments (2.8%) in which 20% or more of the test samples failed the test sample acceptance criteria. Only 1 experiment (0.11%) had more than 50% of its samples fail the test sample acceptance criteria. In general, daily plate and sample acceptance rates were similar across operators, with indications of potential daily reagent issues but no trends or shifts in the data.

The TNA standard reference serum, AVR801, had an overall mean ED50 of 564 (Standard Error (SE) = 2.4) (Table 3) and a percent error of 14% compared to the established ED50 mean which demonstrated good accuracy in this assay for the duration of the study. Over 23 months of sustained testing the mean ED50 by operator where $n > 16$ ranged from 528 (SE = 4.9) to 632 (SE = 6.6), with a CV range from 22.9% to 27.0%. Inspection of the ED50 values, including results passing and failing QC, by month where $n > 10$ demonstrated good precision and accuracy (Fig. 4). The mean ED50 by month ranged from 469–699 and all mean values for all months were within 1SD of the overall observed mean. The CV of the ED50 ranged from 12.2% to 37.7% in the same period. Of these 91% had a monthly CV <30%.

Inspection of the reportable values and parameters of the standard reference serum AVR801 revealed no significant trends over time. Short term trends seen within the ED50 data were attributable to assay variability and manifested as a wave-like pattern (Fig. 5). The pattern was also evident for the standard's slope, upper asymptote, lower asymptote, and for the OD value of the negative control serum. The pattern was present to a less pronounced extent in the upper and lower regions for the OD range of the standard. The ED50 values for AVR801 were ≤2SD from the mean 94.6% of the time and ≤1SD 75.3% of the time, consistent with a Normal distribution.

Over the 220 test days, the daily mean ED50 values of the reference standard AVR801 were calculated and plotted on Shewhart control charts. In the testing period there were a total of 16 prompts for 8 QC tests. Test #2 was prompted on nine occasions and test #6 and #3 were prompted on five and two occasions, respectively. No two prompts occurred on the same day. All prompts to these tests determined to be part of normal assay variability or minor deviations. There were no prompts for tests #1, 4, 5, 7, and 8 for any test day. Although there were no prompts for special causes, there were two testing days when daily ED50 means were approximately 2SD above the expected mean while the previous and next testing dates showed ED50 values similar to the observed mean. Through investigation it was determined that the cause was related to a reagent used on those particular days. Thus, as is the case for the anti-PA IgG ELISA, these data indicate that the Shewhart control charts for the TNA assay were able to detect both random, expected variations as well as special cause variations. Together with the associated technical investigations these analyses validated the results reported for study primary endpoint analyses.

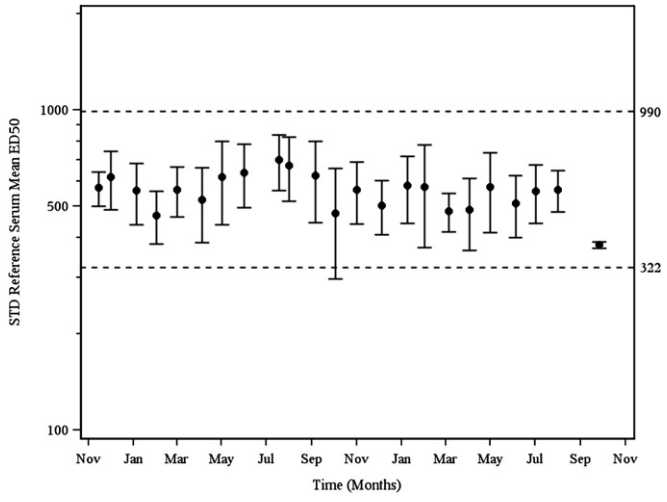

Fig. 4. Accuracy and precision of the mean TNA Effective Dilution-50 (ED50) for standard reference serum AVR801. Plot of mean ED50 values (with standard deviation bars) on all test results (passing and failing) by month. Control limits, displayed for the standard with dashed lines, were determined from an established (expected) reportable value. Data were tested over 23 months. All mean monthly ED50 values were within 1 standard deviation from the overall observed mean (ED50 = 564, SD = 148).

4. Discussion

Clinical trials from which high impact public health decisions are made require supporting data to be valid, reliable and consistent throughout the entire study. The quality control system and its application to AVRP described in this report provide confirmation of the validity and integrity of these data and demonstrate an adaptable method to monitor other biological assays.

The quality control system was designed and applied to the AVRP serological data as a two-stage process. The first stage integrated QC acceptance criteria into the data analysis providing real-time analytic decisions, error detection, and feedback to the test operator. To accommodate the inherent variability of bioassays [19,20] and the use of nonlinear model fits to the data [21,22], we designed multilevel QC criteria that were applied in sequence on multiple variables to encompass different parts of the testing process. This approach improved the assessment of the assays' technical components and ensured reliability of the test results. The acceptance criteria for each assay were derived through development and validation studies. Although some criteria met expectations >99% of the time, particularly for the TNA assay, each criterion evaluated specific, critical markers necessary for the acceptable performance of these bioassays. Equivalent criteria can be applied to other bioassays but the specifics of those criteria must be determined during assay development and validation.

The second stage was designed to detect assay trends across multiple assay plates and multiple testing days. The data were monitored over user-selected short time periods and over the entire course of study testing. Daily and short-term reviews enabled prompt investigation of any assay data or operational discrepancies. Long-term reviews detected the emergence of data 'spikes', assay drift or any unwarranted shift in assay performance due to aberrant behavior in reagents or other materials. The QC-Mon programs standardized and streamlined this process for both assays.

The QC-Mon programs created in SAS® were contained within macros which allowed monitoring of assay data to be fully automated while retaining end-user flexibility for QC test selection. The macro calls were located in a separate program from the macro program which provided protection from inadvertent manipulation of the program. The program and the data outputs were specifically customized for both of the assays reported here and may be customized as required for other assay applications. The extent to which other assay applications can make optimal use of these automated programs depends largely on the type of assay and the type of variables that can be captured from the assay. The use of automated software programs to monitor large quantities of laboratory data is an accepted and essential tool for efficient real-time statistical analyses [22–27]. The customized programs employed here differed compared to other programs in which QC procedures were set up through the use of control charts

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

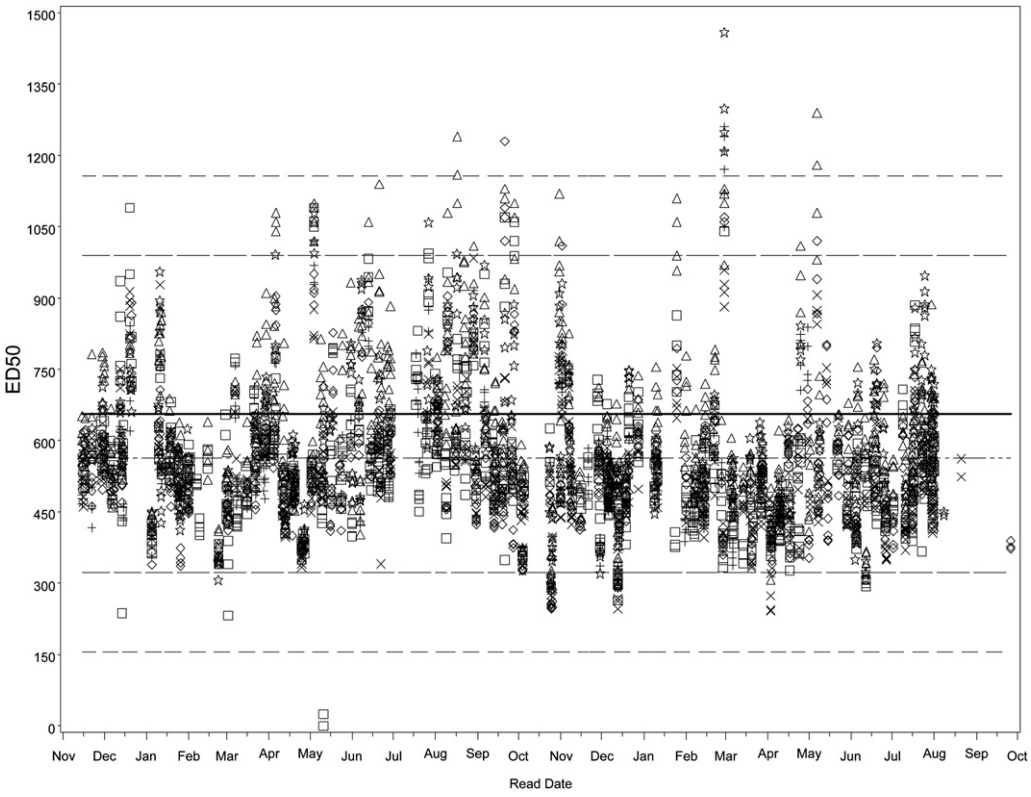

Fig. 5. Individual ED50 data points by date of testing for TNA standard reference serum AVR801. Short term trends attributed to assay variability and displayed as a wave-like pattern. Data points are broken down by operator over a 23 month testing period (operator A: +; operator B: □; operator C: x; operator D: Δ operator E: ◇; operator F: ☆; operator G: ○; expected ED50 mean: solid line; Observed ED50 mean: short dashed line; 2 standard deviations (SD) from the expected ED50 mean: dashed line; 3 standard deviations from the expected ED50 mean: long dashed line). Approximately 95% of all ED50 values were within 2SD of the observed mean.

[25,28,29], sometimes through the use of only one variable [25]. Dependence on single variables in bioassays can make the QC review potentially less reliable. The bioassay variability and performance characteristics for both AVRP serological assays were known and established through development and validation studies and were taken into account during the development of the implemented two-stage QC system. The QC acceptance criteria chosen for these assays relied on the use of multilevel assay-specific variables which were of significant importance in measuring the overall performance of the assay. The second stage QC-Mon programs provided an important tool to monitor the assay performance over time for trends and discrepancies, through the use of control charts, summary tables, reports, and graphical displays of multiple key assay variables. The application of the QCM programs for this study verified that manual data entry errors were few and minimized due to the automated processes set up during assay testing and analysis. These automated processes were specifically employed to reduce the potential for transcription errors. For those errors which did occur, the QC-Mon programs were able to identify them and all errors found were corrected and the data were reanalyzed where necessary.

The Shewhart control charts were set up within the QC-Mon programs for the concentrations of the anti-PA IgG ELISA positive QC sera and the ED50 of the TNA assay's reference standard. This type of control chart was originally designed for manufacturing practices [17], but was later applied to laboratory practices [28]. Rules and tests were then applied to these control charts to facilitate detection of both random, expected variation and assignable, special cause variation. The eight tests for special causes [18] utilized for the assays in this study present similarities to the multi-rule Shewhart procedure by Westgard [29] and were chosen here because their application was already contained within the SAS® Shewhart procedure.

The application of all eight Nelson tests [18] increased the ability to detect errors in the testing process but also increased the chances of detecting random variation. For the AVRP study, it was a calculated decision to utilize all tests and accept the increased detection of random variation. Those tests that gave prompts for our assays detected both special cause variation and random variation. In practice, for the AVRP study, all prompts whether expected or assignable were investigated. Many of the tests that gave prompts within the TNA assay were attributed to the peaks of the sinusoidal wave-like pattern shown over time for the ED50 of the reference standard (i.e., expected assay variation). Establishment of additional assay-specific tests for detecting special causes in addition to select functional tests from above would enhance the ability of these control charts to maximize the detection of non-random variation (assignable causes), while minimizing detection of expected assay variability within our assays. The optimal type and number of control tests need to be carefully considered as the applicability of these tests can vary from assay to assay [30].

The utilization of a two-stage, multilevel QC system was achieved by the use of well characterized reagents, established assay performance (development and validation), operation of a Quality Management System (QMS), and a standardized technology platform for each assay's procedures and analytical methods. The QC system utilized for the AVRP was an essential component of the QMS employed by the study. Outputs and information from the assay QCM can be used to enhance other aspects of the laboratory process that ultimately lead to continuous improvement in laboratory performance.

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001

5. Conclusion

The quality control (QC) system for laboratory data was designed and implemented as a part of a laboratory quality management system (QMS) compliant with the Clinical Laboratories Improvement Amendments (CLIA) and based on Title 21, Code of Federal Regulations (21CFR) Part 58, outlining current Good Laboratory Practices (cGLP), and guidances published by the FDA [31–33]. A multistage, multilevel quality control system was created to monitor the performance of laboratory data from standardized, validated serological assays. Performance of the assays for the AVRP was precise and accurate. The QC-Mon programs detected no significant trends over the course of the study for multiple variables. The system established in our laboratory for the monitoring assay performance provided an increased measure of confidence in the AVRP data obtained from these serological assays. The quality control system established here is an adaptable system for other types of serological assays and has since been applied effectively for use with influenza serological assays as part of the pandemic H1N1 emergency response in 2009 (Hancock, K., personal communication).

Acknowledgements

We would like to thank the Microbial Pathogenesis and Immune Response (MPIR) Laboratory team for their dedication to building and implementing a quality control system within a QMS for the Anthrax Vaccine Research Program: Dababneh H, Walls J, Crenshaw S, Riley P, Noland H, Mahle K, Avery D, Milton A, Shields S (specimen management), Desai R, Abramson D, Cronin L, Caba J, Smith D, Lewis J, Steward-Clark E, Brawner M, Schmidt D, Brown N, Lyde F, Thompson R (laboratory testing), Martin S (quality assurance), and Freeman A, Fox S, Feagins A, Gillis E, Patel N (data analysis and management). We acknowledge the significant contributions of the Anthrax Vaccine Research Program study directors and principal investigators: Wright JG, Plikaytis BD, Centers for Disease Control and Prevention, Atlanta, GA; Parker SD, University of Alabama at Birmingham, Birmingham, AL; Babcock J, Walter Reed Army Institute for Research, Silver Spring, MD; Keitel W, El Sahly H, Department of Molecular Virology & Microbiology and Medicine, Baylor College of Medicine, Houston, TX; Poland G, Jacobson RM, Mayo Clinic, Rochester, MN; Keyserling HL, Emory University School of Medicine, Atlanta, GA. The study was funded through the Centers for Disease Control and Prevention. Foster L, Dababneh H, Crenshaw S, Riley P, Noland H, Mahle K, Avery D, Milton A, Shields S, Cronin L, Caba J, Smith D, Lewis J, Brown N, Lyde F, Thompson R, Fox S, Feagins A, Gillis E, and Patel N were funded by the Atlanta Research and Education Foundation (AREF).

References

[1] Food_and_Drug_Administration. Assuring the quality of test results. Available from, http://www.fda.gov/downloads/ScienceResearch/FieldScience/UCM092168.pdf; November 9, 2007 [cited October 23, 2009].
[2] Food_and_Drug_Administration. Guidance for industry: bioanalytical method validation. Available from, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070107.pdf; 2001 [cited October 23, 2009].
[3] Gray JJ, Wreghitt TG, McKee TA, McIntyre P, Roth CE, Smith DJ, et al. Internal quality assurance in a clinical virology laboratory. II. Internal quality control. Journal of Clinical Pathology 1995 Mar;48(3):198–202.
[4] Marano N, Plikaytis BD, Martin SW, Rose C, Semenova VA, Martin SK, et al. Effects of a reduced dose schedule and intramuscular administration of anthrax vaccine adsorbed on immunogenicity and safety at 7 months: a randomized trial. Jama 2008 Oct 1;300(13):1532–43.
[5] Grabenstein JD. Vaccines: countering anthrax: vaccines and immunoglobulins. Clinical Infectious Diseases 2008 Jan 1;46(1):129–36.
[6] Little SF, Ivins BE. Molecular pathogenesis of Bacillus anthracis infection. Microbes and Infection/Institut Pasteur 1999 Feb;1(2):131–9.
[7] Puziss M, Wright GG. Studies on immunity in anthrax. X. Gel-adsorbed protective antigen for immunization of man. Journal of Bacteriology 1963 Jan;85:230–6.
[8] Li H, Soroka SD, Taylor Jr TH, Stamey KL, Stinson KW, Freeman AE, et al. Standardized, mathematical model-based and validated in vitro analysis of anthrax lethal toxin neutralization. Journal of Immunological Methods 2008 Apr 20;333(1–2):89–106.
[9] Quinn CP, Semenova VA, Elie CM, Romero-Steiner S, Greene C, Li H, et al. Specific, sensitive, and quantitative enzyme-linked immunosorbent assay for human immunoglobulin G antibodies to anthrax toxin protective antigen. Emerging Infectious Diseases 2002 Oct;8(10):1103–10.
[10] Semenova VA, Steward-Clark E, Stamey KL, Taylor Jr TH, Schmidt DS, Martin SK, et al. Mass value assignment of total and subclass immunoglobulin G in a human standard anthrax reference serum. Clinical and Diagnostic Laboratory Immunology 2004 Sep;11(5):919–23.
[11] ELISA for Windows 2.0:. Available from, http://www.cdc.gov/ncidod/dbmd/bimb/ELISA/downloadelisa.htm [cited October 23, 2009].
[12] Plikaytis BD, Holder PF, Pais LB, Maslanka SE, Gheesling LL, Carlone GM. Determination of parallelism and nonparallelism in bioassay dilution curves. Journal of Clinical Microbiology 1994 Oct;32(10):2441–7.
[13] Omland KS, Brys A, Lansky D, Clement K, Lynn F. Interlaboratory comparison of results of an anthrax lethal toxin neutralization assay for assessment of functional antibodies in multiple species. Clinical and Vaccine Immunology 2008 Jun;15(6):946–53.
[14] Taylor TQC, Schmidt D, Freeman A, Li H, Semenova V, et al. Novel mathematical approach to TNA endpoints. Abstracts of the 5th International Meeting on Anthrax; 2003 March 31 to April 4, 2003. Nice, France: 2003.
[15] Findlay JW, Dillard RF. Appropriate calibration curve fitting in ligand binding assays. The AAPS Journal 2007;9(2):E260–7.
[16] Hering D, Thompson W, Hewetson J, Little S, Norris S, Pace-Templeton J. Validation of the anthrax lethal toxin neutralization assay. Biologicals 2004 Mar;32(1):17–27.
[17] Shewhart WA. Economic control of quality of manufactured product. New York: D. Van Nostrand company, inc.; 1931.
[18] Nelson LS. The Shewhart control chart – tests for special causes. Journal of Quality Technology October 1984;16(4):237–9.
[19] Ren S, Frymier PD. Reducing bioassay variability by identifying sources of variation and controlling key parameters in assay protocol. Chemosphere 2004 Oct;57(2):81–90.
[20] Finney DJ. Statistical method in biological assay. 3rd ed. New York: Macmillan; 1978.
[21] Plikaytis BD, Turner SH, Gheesling LL, Carlone GM. Comparisons of standard curve-fitting methods to quantitate Neisseria meningitidis group A polysaccharide antibody levels by enzyme-linked immunosorbent assay. Journal of Clinical Microbiology 1991 Jul;29(7):1439–46.
[22] Rodbard D. Statistical quality control and routine data processing for radioimmunoassays and immunoradiometric assays. Clinical Chemistry 1974 Oct;20(10):1255–70.
[23] Tan IK, Jacob E, Lim SH. Use of computers in quality assurance of laboratory testing. Annals of the Academy of Medicine, Singapore 1990 Sep;19(5):724–30.
[24] Banker CA. Laboratory quality control: use of Shewhart charts and ANOVA. The American Journal of Medical Technology 1980 Apr;46(4):274–9.
[25] Blacksell SD, Cameron AR, Chamnanpood C, Chamnanpood P, Tatong D, Monpolsiri M, et al. Implementation of internal laboratory quality control procedures for the monitoring of ELISA performance at a regional veterinary laboratory. Veterinary Microbiology 1996 Jul;51(1–2):1–9.
[26] Goris N, De Clercq K. Quality assurance/quality control of foot and mouth disease solid phase competition enzyme-linked immunosorbent assay—Part II. Quality control: comparison of two charting methods to monitor assay performance. Revue scientifique et technique (International Office of Epizootics) 2005 Dec;24(3):1005–16.
[27] Westgard JO, Stein B, Westgard SA, Kennedy R. QC Validator 2.0: a computer program for automatic selection of statistical QC procedures for applications in healthcare laboratories. Computer Methods and Programs in Biomedicine 1997 Jul;53(3):175–86.
[28] Levey S, Jennings ER. The use of control charts in the clinical laboratory. American Journal of Clinical Pathology 1950 Nov;20(11):1059–66.
[29] Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. Clinical Chemistry 1981 Mar;27(3):493–501.
[30] Kazmierczak SC. Laboratory quality control: using patient data to assess analytical performance. Clinical Chemistry and Laboratory Medicine 2003 May;41(5):617–27.
[31] Centers_for_Medicare_and_Medicaid_Services. Clinical Laboratory Improvement Amendments (CLIA). Available from, http://www.cms.hhs.gov/clia/ [cited March 16, 2010].
[32] Food_and_Drug_Administration. CFR – Code of Federal Regulations Title 21 Part 58. Available from, http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?CFRPart=58; April 1, 2009 [cited March 16, 2010].
[33] Food_and_Drug_Administration. CFR - Code of Federal Regulations Title 21 Part 11. Available from, http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11; April 1, 2009 [cited March 16, 2010].

Please cite this article in press as: Soroka SD, et al., A two-stage, multilevel quality control system for serological assays in anthrax vaccine clinical trials, Biologicals (2010), doi:10.1016/j.biologicals.2010.09.001