Central site monitoring: Results from a test of accuracy in identifying trials and sites failing Food and Drug Administration inspection

Anne S Lindblad[a], Zorayr Manukyan[a], Tejashri Purohit-Sheth[b], Gary Gensler[a], Paul Okwesili[b], Ann Meeker-O'Connell[b], Leslie Ball[b] and John R Marler[b]

Background   Site monitoring and source document verification account for 15%–30% of clinical trial costs. An alternative is to streamline site monitoring to focus on correcting trial-specific risks identified by central data monitoring. This risk-based approach could preserve or even improve the quality of clinical trial data and human subject protection compared to site monitoring focused primarily on source document verification.

Purpose   To determine whether a central review by statisticians using data submitted to the Food and Drug Administration (FDA) by clinical trial sponsors can identify problem sites and trials that failed FDA site inspections.

Methods   An independent Analysis Center (AC) analyzed data from four anonymous new drug applications (NDAs) where FDA had performed site inspections overseen by FDA's Office of Scientific Investigations (OSI). FDA team members in the OSI chose the four NDAs from among all NDAs with data in Study Data Tabulation Model (SDTM) format. Two of the NDAs had data that OSI had deemed unreliable in support of the application after FDA site inspections identified serious data integrity problems. The other two NDAs had clinical data that OSI deemed reliable after site inspections. At the outset, the AC knew only that the experimental design specified two NDAs with significant problems. FDA gave the AC no information about which NDAs had problems, how many sites were inspected, or how many were found to have problems until after the AC analysis was complete. The AC evaluated randomization balance, enrollment patterns, study visit scheduling, variability of reported data, and last digit reference. The AC classified sites as 'High Concern', 'Moderate Concern', 'Mild Concern', or 'No Concern'.

Results   The AC correctly identified the two NDAs with data deemed unreliable by OSI. In addition, central data analysis correctly identified 5 of 6 (83%) sites for which FDA recommended rejection of data and 13 of 15 sites (87%) for which any regulatory deviations were identified during inspection. Of the six sites for which OSI reviewed inspections and found no deviations, the central process flagged four at the lowest level of concern, one at a moderate level, and one was not flagged.

Limitations   Central data monitoring during the conduct of a trial while data checking was in progress was not evaluated.

Conclusion   Systematic central monitoring of clinical trial data can identify problems at the same trials and sites identified during FDA site inspections. Central data monitoring in conjunction with an overall monitoring process that adapts to identify

---

[a]The EMMES Corporation, Rockville, MD, USA, [b]US Food and Drug Administration, Silver Spring, MD, USA
Author for correspondence: John R Marler, Division of Neurology Products, US Food and Drug Administration, 10903 New Hampshire Avenue, Building 22 Room 4340, Silver Spring, MD 20993-0002, USA.
Email: john.marler@fda.hhs.gov

risks as a trial progresses has the potential to reduce the frequency of site visits while increasing data integrity and decreasing trial costs compared to processes that are dependent primarily on source documentation. *Clinical Trials* 2014; **11**: 205–217. http://ctj.sagepub.com

Introduction

On-site clinical data quality monitoring contributes significantly to the cost of clinical trials [1–3]. The authors' personal experience is that the sponsors of trials are reluctant to try new quality control methods because they might increase the risk of failing Food and Drug Administration (FDA) data audits. There are limited data to support or allay their concern.

A number of reports recommend reducing on-site monitoring and increasing the extent and sophistication of central statistical analysis to evaluate clinical trial data quality [4–10]. In 1999, a subcommittee of leading statisticians and clinical trial methodologists from the International Society for Clinical Biostatistics [4] suggested that many changes made in clinical trial monitoring practices were an overreaction to reports of fraud in the National Surgical Adjuvant Breast and Bowel Project (NSABP) [11] and were unlikely to improve the quality of clinical trials. They concluded that data quality problems 'can largely be prevented through design of the trial protocol and case report form, and detected by statistical procedures and computerized checks that make use of the unique structure of clinical trial data'. The International Council on Harmonization [9] guidelines recommends more careful monitoring early in the data collection process. Woodin [6] wrote that 'maybe it's time for us all to stand back and think about what makes sense to do'. She asked,

> What is really important for verifying the subjects actually exist and that the data we are collecting are valid? We are expending enormous amounts of time and energy (which equal money, of course) on an activity that may not need to be done. What's really important to review and verify?

She concludes that we need to reassess the purpose and need for 100% source document review.

Weir and Murray [5] tested central monitoring methods that were dependent on visual inspection of graphs to identify sites with data quality problems. Tantsyura *et al*. [8] compared advantages and disadvantages of several source data verification approaches. They advocate a mixed risk-based and random approach where, for example, screening, baseline, adverse events (AEs), drug doses, and trial endpoints require 100% source data verification,

and other data elements undergo central monitoring. They estimate that this approach could reduce source verification by 50%–70%.

Others have tested central monitoring using clinical trial data. In an experiment reported by O'Kelly [12], he placed data fabricated by three physicians among the data from 3 of 18 sites in a trial. His central data monitoring approach correctly identified one of the three sites and incorrectly identified a site that had no fabricated data. Lienard *et al*. [13] evaluated the impact of site visits on trial recruitment and data quality using data from an ongoing trial. They found no effect of site visits on quality or recruitment and concluded that their preliminary results warranted further studies. They comment that 'the lack of systematic investigations on the actual returns of on-site monitoring is surprising in view of the high labor intensity, and therefore the high cost of this activity'. Bakobaki *et al*. [10] report that over 90% of the problems identified from the review of site monitoring reports in a 9385-participant 6-site trial performed in Africa could also have been identified by central monitoring processes. They suggest that 'a change in focus of site visits to one of education and training rather than data and procedure checking'.

The increasing complexity and rising cost of clinical trials hinders drug development. Several groups estimate that 15%–30% of trial cost is attributed to site monitoring, particularly, source document verification [1–3]. Eisenstein *et al*. [14] estimate cost savings for refinements in the typical pharmaceutical industry clinical trial. They estimate that 'implementing a modified site management strategy that largely replaces on-site with remote monitoring could in itself reduce clinical trial costs in our pharmaceutical industry simulation by more than 20%'.

A grant from the 'Regulatory Science and Review Enhancement' (RSR) Program of the FDA Center for Drug Evaluation and Research (CDER) allowed us to investigate the potential of central monitoring in a setting in which traditional site monitoring already had occurred. Our purpose was not to develop an algorithm that would guarantee successful clinical site audits by FDA; instead, we sought to evaluate whether a risk management process that relied on central analysis of trial data alone could play a significant role in protecting clinical data quality. This

report describes the results of collaboration between a team from CDER and an independent statistical coordinating center to determine whether central review of data submitted to the FDA by clinical trial sponsors could have identified problem sites and trials before they failed FDA on-site inspections.

Methods

Analysis plan

Prior to receipt of any data, personnel (A.S.L., Z.M., and G.G.) at the Analysis Center (AC) at The EMMES Corporation, Rockville, Maryland, prepared a statistical analysis plan. A consultant and FDA members of the collaboration reviewed the plan before any data were transferred to the AC. The plan included statistical tests for global agreement between pooled distributions versus a single site. Frequency of rejection of the test of no difference was summarized by site. The test ignored the correlated nature of observations contributed from the same participant. Because there were fewer than 10 participants at most sites, and a large number of variables, the AC team determined that these tests alone were unlikely to flag problem sites accurately and added the visual inspection methods described below. The AC combined data from small sites to create larger units for analysis but concluded that the selection of problem sites should rely mainly on inspection of visual representations of data with consideration given to the type and number of problems identified.

Test data

To test their plan, the AC analyzed data from four new drug applications (NDAs) to attempt to identify two that had significant problems identified during FDA site inspections. FDA team members in the Office of Scientific Investigations (OSI) (T.P.-S., P.O., A.M.-O., and L.B.) chose the four NDAs from among all NDAs with data in Study Data Tabulation Model (SDTM) format. OSI first chose two NDAs for which FDA site inspections had found significant problems with data integrity, and OSI had deemed the data unreliable (i.e., not usable) in support of the application. Then OSI chose two additional applications with a similar amount of clinical data that OSI, after inspection, deemed reliable. The selected NDAs included trials with about 200–1000 participants. The number of sites in each trial ranged from about 20 to 100; across the trials, 10 or more participants had enrolled at fewer than 25 sites. The trials included from two to four treatment arms using one to one, two to one, or three to one randomization ratios. For each NDA, FDA had already conducted inspections at clinical sites selected based on the

application, trial, and clinical site–related attributes after preliminary examination of submitted data. Some clinical sites had participated in more than one of the trials that comprised an NDA submission. From the outset, the AC knew that the experimental design specified two NDAs with significant problems but did not know which two. FDA gave the AC no information about how many sites or which sites were inspected or how many were found to have problems until after the AC reported the results of their analysis. Data fields that contained specific information that could identify sponsors, the specific NDA, trial participants, sites, or investigators were deleted before the data were transferred from the FDA to the AC for analysis and reporting.

FDA classification of sites

After completing a site inspection, FDA makes one of three determinations for each site: (a) No Action Indicated (NAI), (b) Voluntary Action Indicated (VAI), or (c) Official Action Indicated (OAI). NAI indicates that no significant deviations from regulations were noted and the data are acceptable. VAI indicates that deviations from regulations were noted but they do not significantly affect data reliability; thus, the data in general are acceptable. OAI indicates that there were significant deviations from regulations that significantly affect data reliability; the data are considered unreliable. In this report, these regulatory categories that summarize the site visit results, NAI, VAI, and OAI, are termed 'No Problems', 'Minor Problems', and 'Serious Problems', respectively. The AC did not know the categories FDA used to describe the outcome of site audits until after they completed their analysis.

Data analysis

Raw data tables in SAS transfer format were assembled in analysis files using SAS 9.2. Analyses were coded using a function-based approach in R [15] to ensure reproducibility and application to other data sets. The AC verified data generated by the programs against samples from the raw data tables as an additional quality control check of programming.

Data quality indicators

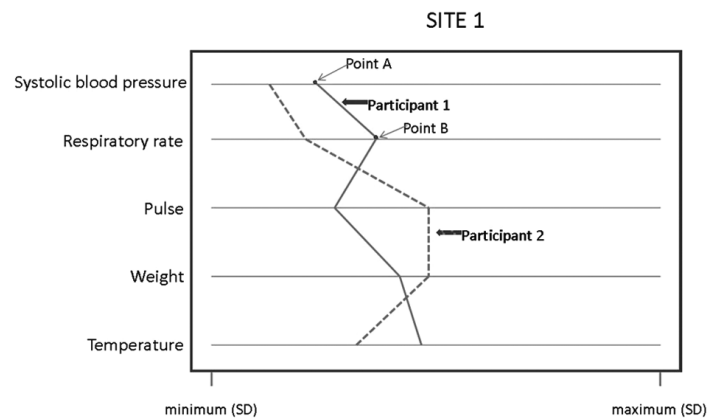Below, we describe the eight indicators the AC chose and give the rationale for choosing them.

1. *Enrollment.* Unusually high enrollment in a short period of time may result in difficulties in submitting accurate data.

2. *Randomization balance.* Participant allocation should be reasonably balanced across treatment arms according to the randomization plan; it can be unequal when planned. For example, a 2-to-1 randomization plan for treatment A versus B should result in about twice as many participants assigned to treatment A as to treatment B.

3. *Visit timing.* Visits to collect study data typically are performed close to a target date. For example, for a weekly visit schedule, we expected target dates of 7, 14, 21, and so on days from randomization. We calculated the actual days from randomization to the visits for each participant within a site and subtracted that number from the target date for that visit. This difference was summarized in box plots, one for each site. We considered it noteworthy when we observed that a site always conducted visits exactly on the target date or conducted visits on days or times medical clinics were unlikely to be open.

4. *Variability in outcome measurements.* Measurement error is expected to be similar across sites. Too much or too little variability in an outcome measurement compared to other sites would raise concerns. For continuous measurements repeated over time, the AC calculated the variance of the measure for each participant and calculated the mean and standard deviation of these variances. For each participant, the standard deviation was plotted relative to the minimum and maximum standard deviation across all participants at all sites for each measure. The AC used the technique of parallel coordinate plots [16] to display the variability on the x-axis for multiple outcomes on the y-axis for all the participants at a site. Figure 1 is an example of a parallel coordinate plot for a single site with two participants for the variables blood pressure, respiratory rate, pulse, weight, and temperature. The standard deviation for each vital sign is calculated for each participant and plotted between the minimum and maximum standard deviations for that vital sign for all participants at all sites. Then a line for each participant is drawn to connect the plotted values for each vital sign. In the example in Figure 1, the standard deviation of all blood pressure measurements for a participant is the first point on the line (shown as point A for participant 1). The standard deviation of all respiratory rates for participant 1 is the second point on the line (shown as point B), and participant 2 is represented by a separate line. Any additional participants at this site would have an individual 'line' on the plot. Parallel coordinate plots that show heavy repetition of points near the minimum mark of the x-axis

indicate low variability of that measure across all participants at that site. When the pattern of variability at one site differs from the pattern at other sites for one or more variables, there may be problems with data quality at that site.

5. *Carry-over.* Most outcome measurements are expected to change over time because of disease state, testing conditions, or measurement error. Lack of change, termed 'carry-over' may indicate problems in data quality. We calculated frequencies of identical consecutive measurements of continuous and categorical variables and compared the frequencies across sites.

6. *Missing data.* Sites with a large quantity of missing data may prompt major concern about data quality. We compared the average number of missing observations per participant for different outcomes at each site in order to identify any sites with unusually high or low frequencies of missing data.

7. *Digit preference.* Some continuous or semicontinuous outcome measurements, on average, are equally likely to end in 1 of the 10 digits 0–9. For those outcome variables, we used bar charts to examine the overall frequency distribution of each last digit at each site. Unusual patterns at a site could indicate improper technique or equipment calibration issues.

8. *Adverse events.* Some sites may overreport or underreport adverse events or severity of adverse events. For each participant, we determined the maximum severity of any adverse event reported. For each degree of severity, we calculated the proportion of participants within a site for whom that maximum degree was reported.

Site review

For each trial, the AC examined all quality indicators for selected variables within the data sets. Statisticians from the AC identified sites with data patterns that differed from the data patterns at other sites from the same trial within the NDA application. We refer to these unusual data patterns below as 'discrepancies'. A site could have discrepancies in any of the quality measures the AC evaluated. Statistical significance of the magnitude of the discrepancy was not a requirement for a site to be flagged as discrepant.

After the AC reviewed all the sites for all of the quality indicators, it considered all the discrepancies at each site and assigned one of four levels of concern to summarize its findings. The AC defined these four categories to describe, from a coordinating center's perspective, the seriousness of the aggregate of data anomalies identified within a site based on their potential impact on the credibility of the trial.

SITE 1

Figure 1. Example of a parallel coordinate plot for vital signs with accompanying data.
SD: standard deviation.

Measurement	Baseline	Year 1	Year 2	Year 3	SD	Study minimum SD	Study maximum SD
Participant 1							
Systolic blood pressure	110	110	120	110	5	3.3	20.3
Respiratory rate	15	18	15	16	1.4	0.4	3.2
Pulse	70	72	74	74	1.9	0.0	4.9
Weight	155	158	162	159	2.9	1.4	5.2
Temperature	97.9	98.6	98.7	98.9	0.4	0.1	0.6
Participant 2							
Systolic blood pressure	110	115	120	115	4.1	3.3	20.3
Respiratory rate	16	17	18	17	0.8	0.4	3.2
Pulse	70	76	72	72	2.5	0.0	4.9
Weight	180	185	188	186	3.4	1.4	5.2
Temperature	98.4	98.6	98.0	98.8	0.3	0.1	0.6

For each participant, the standard deviation for each different vital sign is calculated and plotted between the minimum and maximum standard deviations for that vital sign for all participants at all sites. A line connects the plotted values for each vital sign for a given participant. The plotted standard deviation of all blood pressure measurements for participant 1 is shown as point A. Its position on the x-axis is plotted relative to the minimum and maximum standard deviations of blood pressure across all measurements for all participants at all sites. The standard deviation of all respiratory rates for participant 1 is point B, and so on. Participant 2 is represented by a separate line. Lines are added for each additional participant at this site.

Discrepancies that involved randomization or a primary outcome were rated 'Very Serious'. When a discrepancy was not rated 'Very Serious', it was rated 'Serious' unless the discrepancy resulted from rounding or instrumentation problems or the issue was not wide spread across the other centers, that is, four or fewer sites with similar discrepancies. Otherwise, a discrepancy was deemed 'Non-Serious'. The AC assigned one of four levels of concern to each sites based on the severity of all the discrepancies at the site:

High Concern. One or more Very Serious discrepancies or more than three Serious discrepancies;
Moderate Concern. One to three Serious discrepancies and no Very Serious discrepancies;
Mild Concern. One or more Non-Serious discrepancies and no Serious or Very Serious discrepancies;
No Concern. No discrepancy.

In addition, in order to simulate central data monitoring during an ongoing clinical trial, the AC evaluated sites using data from the first 33% and 50% of the participants enrolled at the site. The AC then compared the discrepancies identified from analysis of 33%, 50%, and 100% of the data for each site.

Trial review

Any trial with at least one site rated as High Concern was flagged as a trial with significant data quality problems. The AC presented its results for each trial, and the reasons sites were flagged as High Concern in a written report to the FDA team members before FDA team members revealed the two NDAs that contained unacceptable data.

This research was approved by the FDA Research Involving Human Subjects Committee. Each member of the AC signed a confidentiality agreement. In all results presented below, specific site references and types of measures have been coded to protect

Table 1. Comparison of FDA inspection results with the Analysis Center (AC) central review result

AC central review result		FDA findings (number of sites)		
		No problems	Minor problems	Serious problems
NDA 1	No Concern			
	Mild Concern	1	1	
	Moderate Concern		1	
	High Concern			
NDA 2	No Concern			1
	Mild Concern		1	
	Moderate Concern		1	
	High Concern		1	
NDA 3	No Concern			
	Mild Concern	3		
	Moderate Concern			
	High Concern			
NDA 4	No Concern		1	
	Mild Concern	1	3	
	Moderate Concern	1		
	High Concern			5
All NDAs	No Concern		1	1
	Mild Concern	5	5	
	Moderate Concern	1	2	
	High Concern		1	5
	Total sites	6	9	6

NDA: new drug application; FDA: Food and Drug Administration.

the identity of the NDAs, trial participants, and trial sponsors.

Results

The AC evaluated data from 413 sites for the four NDAs and flagged 25 sites (6%) as High Concern, 54 sites (13%) as Moderate Concern, 112 (27%) as Mild Concern, and 223 (54%) as No Concern. Using only the analysis of data submitted with the NDA application, the AC correctly identified the two NDAs that failed FDA inspection (NDAs 2 and 4). NDAs 2 and 4 had 17 sites with High Concern. NDAs 1 and 3 had no sites with High Concern.

Table 1 provides a summary of FDA and AC findings by NDA. In all, FDA inspected 21 sites from the four NDAs. Among these sites, FDA had found Serious Problems at 6 sites (29%), and Minor Problems at 9 (43%). The AC identified as High Concern 5 of the 6 (83%) sites where FDA had found Serious Problems and identified some level of concern at 8 of the 9 sites (89%) where FDA had found Minor Problems. The AC indicated some level of concern for 13 of the 15 (87%) sites where FDA inspection had identified any problem, whether serious or minor.

FDA had found problems at two sites for which the AC had no concerns. For NDA 4, the FDA had found Minor Problems at a site that had fewer than five participants; the AC did not find problems at this site other than minor last digit preference in a vital sign. This same site participated in one of the other trials that comprised the same NDA; for this trial, the AC flagged this site for concerns about vital sign reporting, one of the issues identified during the FDA inspection of this site. For NDA 2, the FDA inspection had found Serious Problems with drug accountability record keeping and reporting at a site; the AC did not have drug accountability data for any of the NDAs.

There was only one site that the AC flagged as High Concern for which FDA inspections had identified only Minor Problems with no significant data integrity concerns. The AC found that compared to all other sites, this site had a high proportion of within-participant repetition of the same response for an outcome from one time point to the next as well as other instances where this site's data profile appeared to differ from other sites. Of the six sites for which FDA inspections found no problems, the central process had rated 4 at 'Mild Concern', 1 at 'Moderate Concern', and 1 at 'No Concern'.

For NDA 4, the AC correctly flagged as High Concern each of the five sites identified by FDA inspections as having significant data integrity problems. The AC had flagged these sites due to (1) small

standard deviations for specific groups of variables such as laboratory values, vital signs, and primary outcome variables; (2) reporting of too few adverse events compared to other sites; and (3) treatment imbalance. FDA findings for these same sites included multiple pertinent discrepancies between reported data and source documents, failure to follow blinding procedures, absence of laboratory test reports, failure to follow protocol procedures, and inappropriate changes to source documents. Of note is that FDA inspections also had found discrepancies in drug delivery or drug disposition documentation or procedures, which were factors the AC did not review.

Figures 2–4 show examples of discrepancies that led the AC to flag two particular sites as High Concern. In Figure 2, within-participant variability stands out as low relative to participants at other sites for two components of the primary outcome, measures V3 and V4, for sites 898 and 889. For these two sites, the lines in the parallel coordinate plot for these two variables are clustered near the minimum of the standard deviations for all participants, unlike the lines for sites 827 and 985. A similar observation applies to the variable V1; however, there are multiple sites with low variability in this outcome measure. Figure 3, which displays the distribution of outcome V3 values, confirms the concerns raised in Figure 2 by showing a smaller mean and standard deviation for outcome V3 values for sites 898 and 889 compared to other sites (sites 827 and 883 as an example) whose values for the mean and standard deviation are nearly triple those of sites 898 and 889. In Figure 4, the overall mean and standard deviation of within-participant variance for laboratory values are also smaller at these two sites relative to participants at other sites.

Across all four NDAs and within each trial, the AC found data discrepancies that could have been caused by problems in practice or instrumentation rather than data reporting. In all the trials, the AC found that sites measured vital signs with different precision. This discrepancy in precision is illustrated in Figure 5 that shows for some sites the last digit of blood pressure values was almost always rounded to zero or five (for example, sites 839 and 843) although other sites within that same trial reported the full range of digits (for example, sites 885 and 886). The sites may have reported the data exactly as in their source documentation, but the AC findings suggest there were underlying equipment or procedural problems.

Figure 6 shows why another site was flagged as High Concern. Site 885 had a randomization pattern different from other sites. This study had a 2-to-1 randomization ratio in favor of treatment A. Site 839 shows the expected pattern; by the time 35 participants had enrolled, nearly 15 more participants had been assigned to treatment A compared to treatment B. Site 885 demonstrates that by the time the 20th participant enrolled there, 6 more participants had been assigned to treatment B than to treatment A. For this site, one of the significant problems that FDA found was failure to follow blinding procedures.

The AC performed post hoc tests to determine whether interim analyses could identify future problems. Partial data sets using 33% or 50% of the data showed problems such as small standard deviations and digit preference that were present in the full data sets.

Discussion

FDA has acknowledged for many years that monitoring to confirm 100% source documentation is not needed to ensure credible results from clinical trials. In fact, the FDA actively encourages sponsors to develop new quality control paradigms that improve or maintain data quality and protect trial participants while streamlining processes and reducing clinical trial costs. On April 2011, the FDA withdrew a 1988 guidance on clinical monitoring that emphasized on-site monitoring visits by sponsor or CRO personnel because the 1988 guidance no longer reflected the current FDA opinion. A clinical monitoring guidance published August 2013 [17] makes it clear that sponsors can use a variety of approaches to meet their trial oversight responsibilities.

We have shown that central review of data by a trial sponsor could have identified data quality problems prior to FDA on-site inspection without extensive checking for source documentation. Our results using data when only one-third and one-half of the complete data were available suggest that central review of data during a trial could guide a risk management system for optimizing the efficiency of data quality control. For example, central data review may determine early in a trial that some sites require additional training, equipment calibration, or increased efforts on-site to ensure data quality. Other sites may require additional effort to prevent dropout and missing data. Our experience leads us to doubt that a fixed algorithm will work for all trials or even for all phases of a single trial from start-up to final follow-up. We expect that successful central monitoring will require creative and flexible responses to the differences between sites and trial designs that adapt to solve data quality problems as soon as they become apparent so that any further damage to trial integrity can be minimized.

For the trials in the four NDAs that we used for our experiment, we found that checking randomization balance and inspecting standard deviations
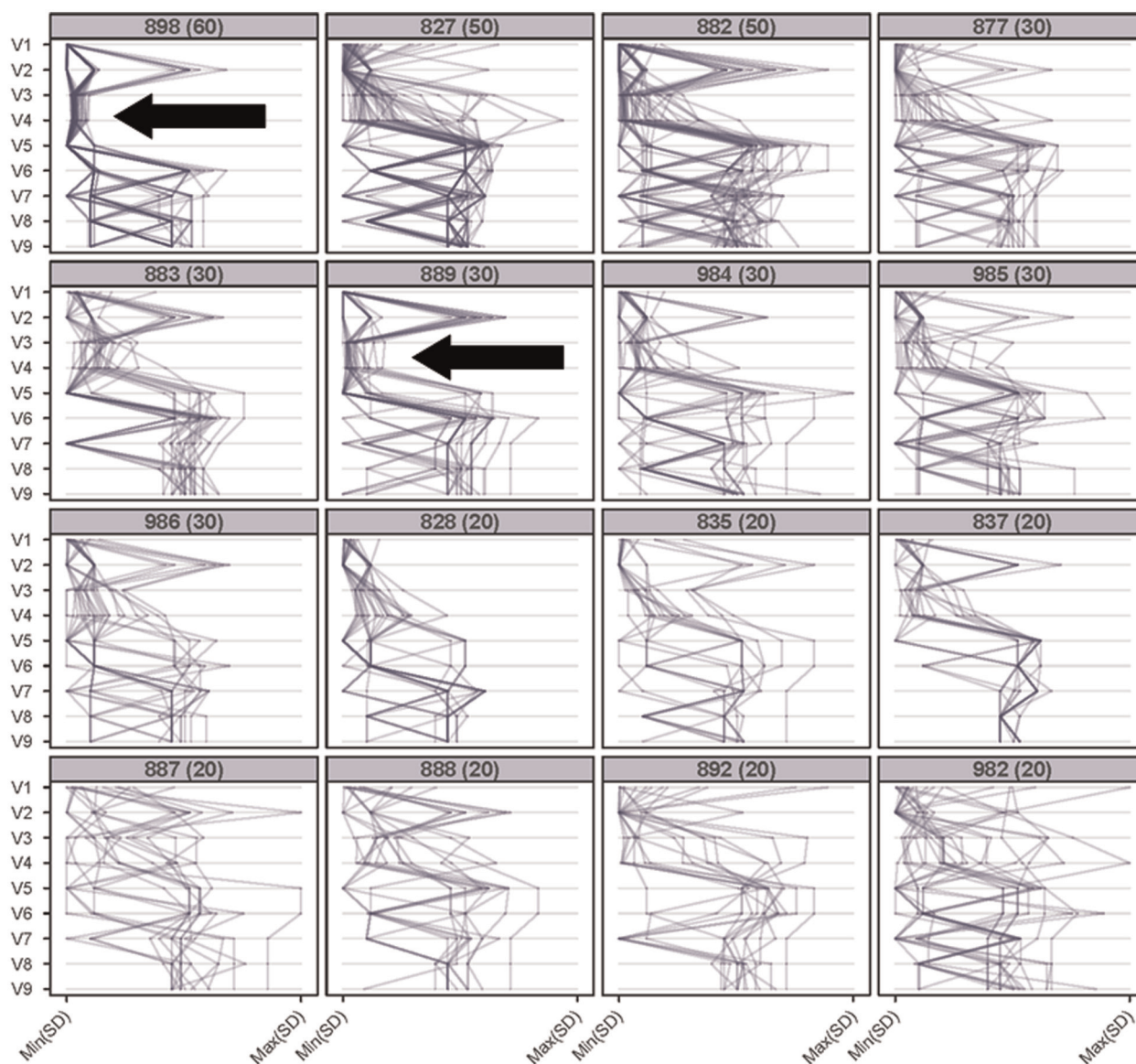
Figure 2. Parallel coordinate plot showing small within-participant variability for outcome measures V3 and V4 for sites 898 and 889.
AC: Analysis Center; FDA: Food and Drug Administration.
This shows 16 parallel coordinate plots that display the standard deviations for each patient for variables V1–V9 at 16 sites. In the trial, each variable was measured at different scheduled times. The AC calculated the standard deviation of the measurements V1–V9 for one patient at different times. Then as illustrated in Figure 1, a line connects nine points, each point representing the standard deviation for a different variable on the gridlines labeled V1–V9. The 'min' and 'max' labels on the x-axis represent the smallest and largest standard deviations for each of the variables V1–V9 across an entire trial and all sites. We used the 'min' and 'max' because each variable has a different range of values. The value in parentheses is the approximate number of patients at each site.
FDA site inspections found Serious Problems at sites 889 and 898. For these two sites, the figure shows small standard deviations for most patients for outcome measures V3 and V4. The arrows indicate where lines for all the patients are concentrated near the minimum for these measures. At all other sites, the lines intersect at V3 and V4 across a wider range of values along the x-axis, implying the within-participant variability of V3 and V4 for sites 898 and 889 is different from other sites.

identified the most serious problems. Parallel coordinate plots allowed us to simultaneously visualize differences in the variability of multiple, potentially related, variables. Calculated means and standard deviations supplemented the graphical displays with quantitative values that could be tested for statistically significant differences. Digit preference was useful for identifying potential training or
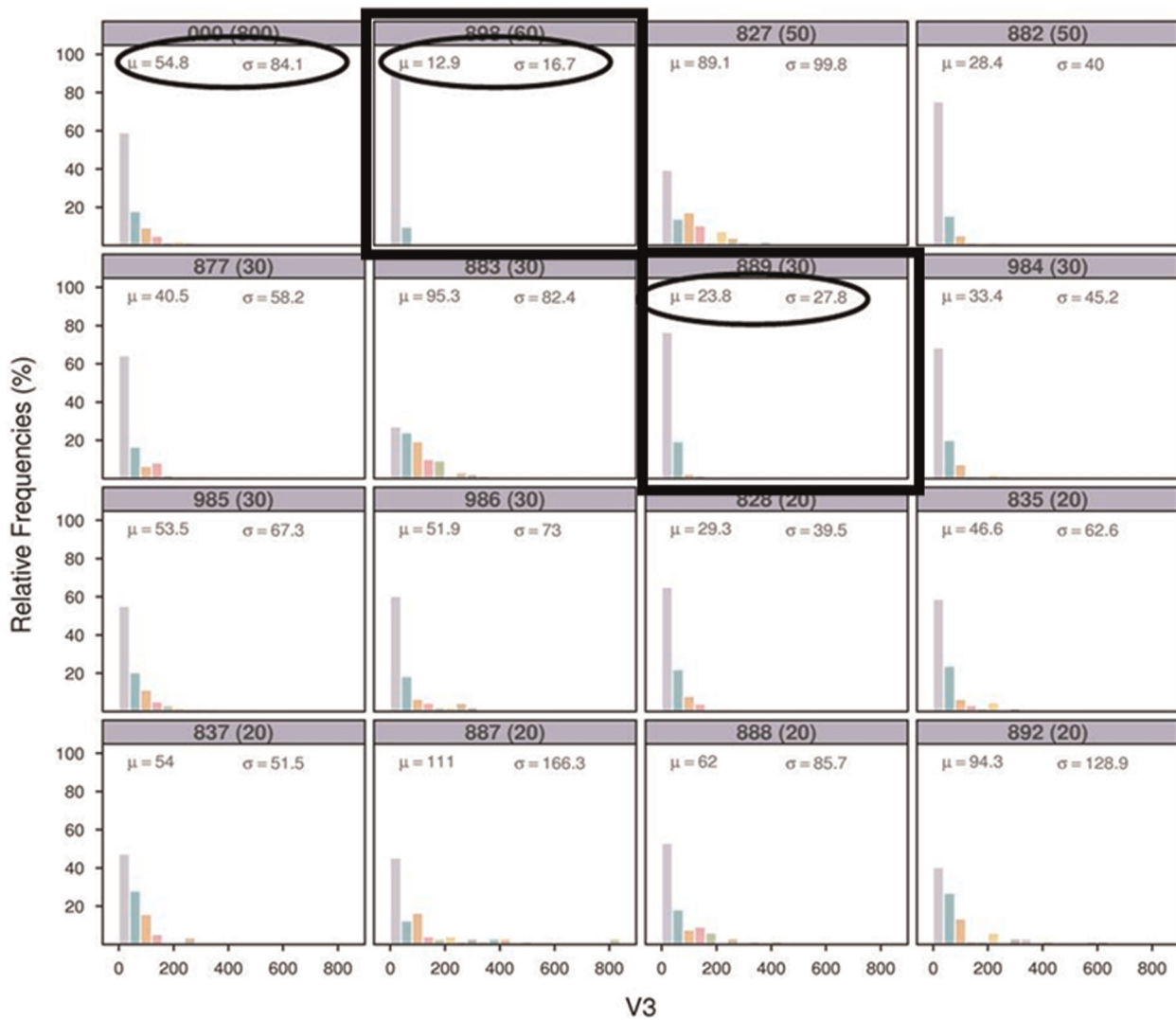
Figure 3. Distribution of outcome V3 values by site.

FDA: Food and Drug Administration.

This is a companion plot to Figure 2. It shows the frequency of each of the values for the outcome measure V3 for all participants for all data points at each site. Site 000 is a reference plot which shows the distribution of all measures of V3 across all sites. The value in parentheses is approximate sample size for each site. As noted in Figure 2, FDA found Serious Problems at sites 889 and 898. This figure shows a mean (μ) and standard deviation (σ) of the outcome measure V3 at these two sites (sites highlighted by boxes and mean and standard deviation circled) that is much less than the mean and standard deviation in the 000 reference plot (circled mean and standard deviation) and smaller than each of the other 14 sites.

equipment issues and perhaps may indicate a lack of attention to detail.

The two approaches, site inspections by the FDA and central monitoring, must, by necessity, focus on different symptoms of data quality problems; yet, the approaches led to similar conclusions. It is important to note that all of these studies had been monitored by their respective sponsors and that databases had been locked. Thus, aggregate data inspection found serious problems that clinical monitors who traveled to individual sites and looked at individual records did not find. We hope our retrospective analyses will stimulate prospective use of aggregate data inspections as a critical part of a comprehensive data quality risk management process.

Limitations

The knowledge by the AC that two of the four NDAs included data deemed unreliable may have increased the likelihood of correctly identifying the two NDAs but would have had a smaller impact on specific site identification. Our analyses were based
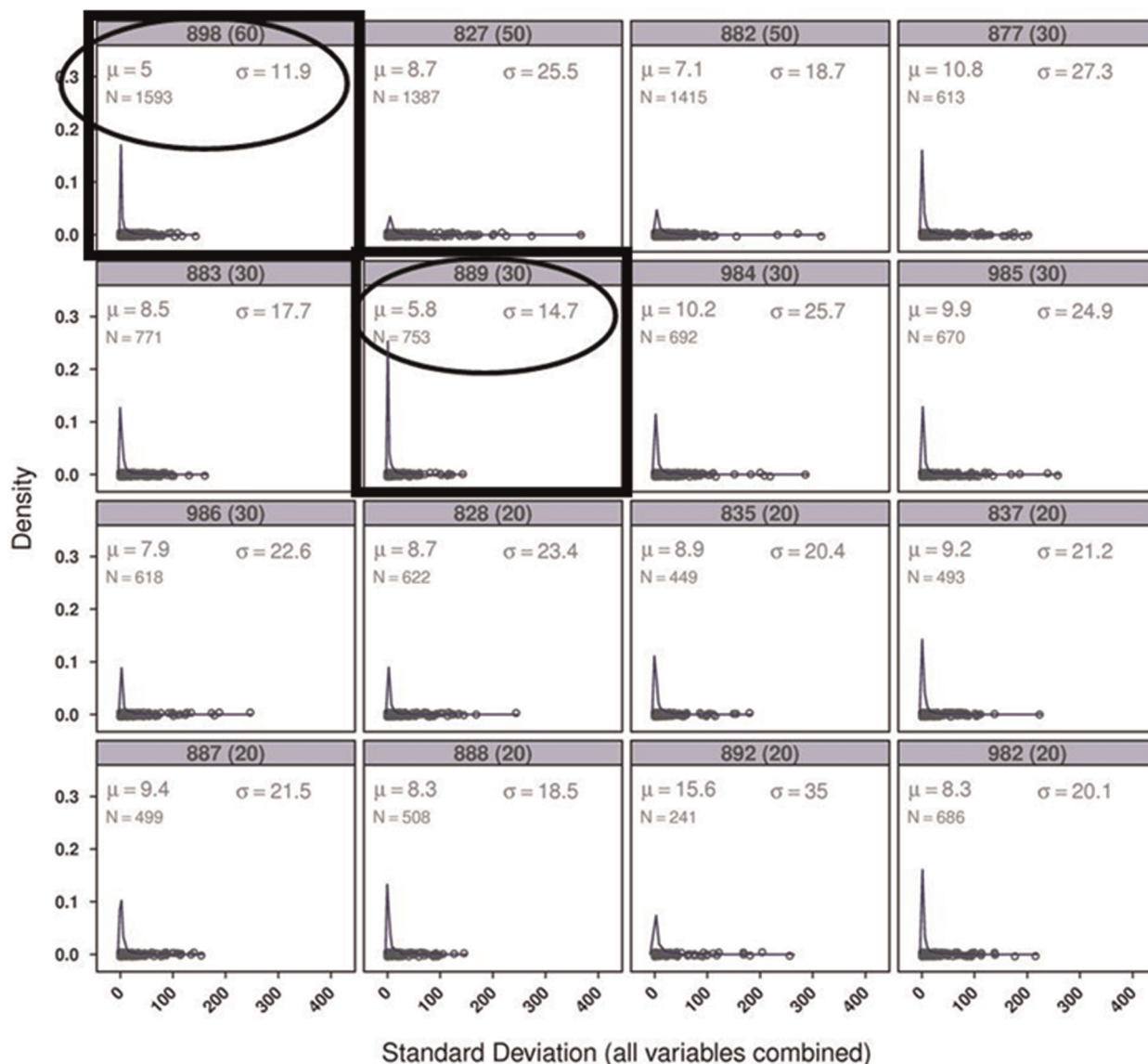
Figure 4. Distribution of within-participant variability among all laboratory values.

AC: Analysis Center; FDA: Food and Drug Administration.

This shows the distribution of the standard deviations for each laboratory value that was repeated multiple times for each participant in a trial. A larger 'peak' close to zero indicates a smaller standard deviation and less variability. The mean (μ) and standard deviation (σ) of all the standard deviations summarize the variability at each site. The value in parentheses is approximate number of participants at each site. FDA site inspections found serious problems at sites 889 and 898. The AC found less variability across all laboratory values at these two sites (circled values with centers highlighted by boxes) compared to the 14 other sites.

on data locked for FDA submission, suggesting data queries had been resolved. It is possible that the increased variability in unedited data could obscure some anomalous data patterns. The potential for successful implementation of central monitoring during an ongoing trial may not be well represented by the observations from retrospective inspection of locked data sets in this analysis. Initial programming of a centralized monitoring approach tailored to a

given trial and ongoing implementation and data inspection requires resources. Whether this expense can be offset by fewer, or more efficient, on-site monitoring visits requires further study.

Conclusion

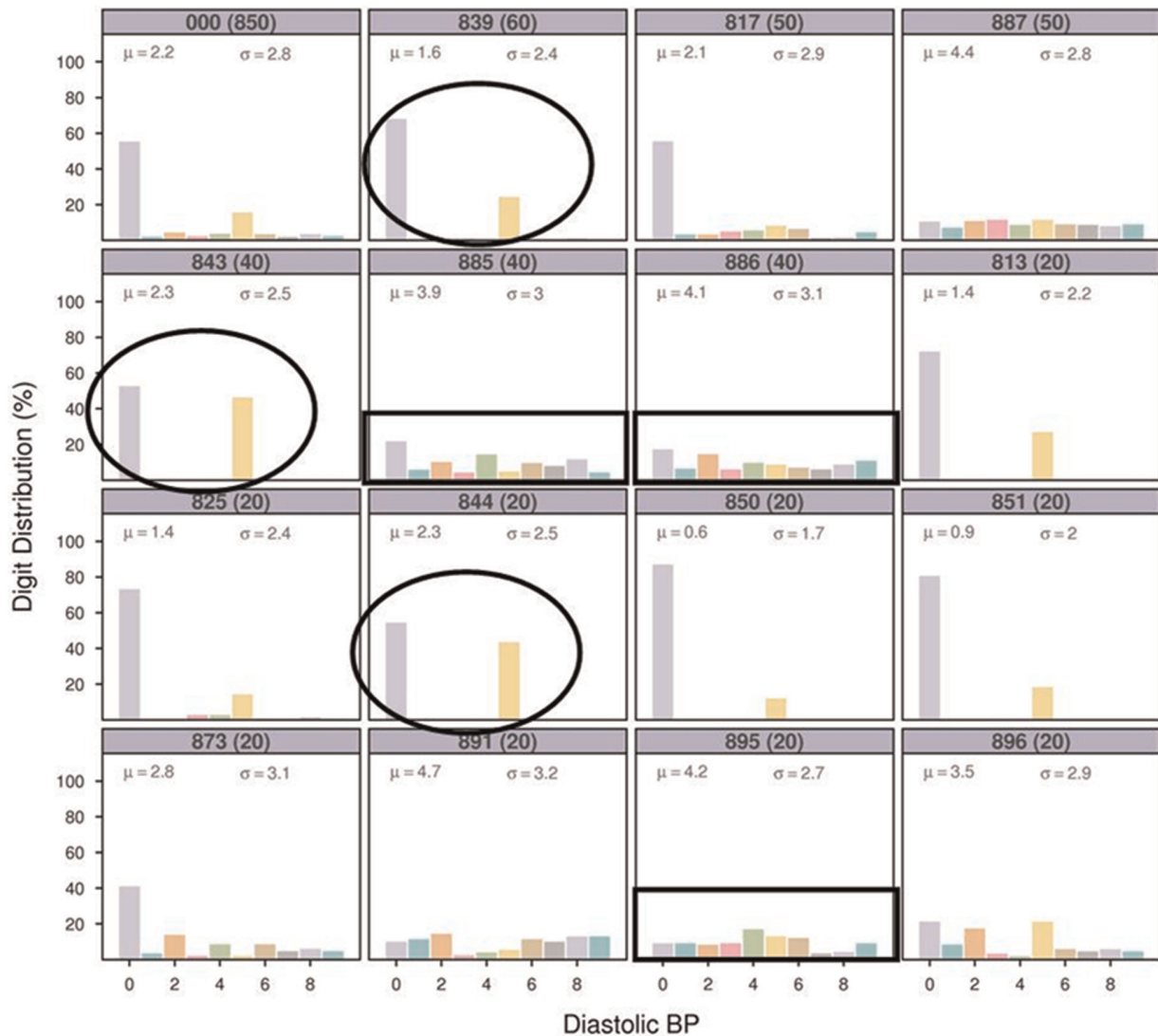Our results suggest that central data monitoring used in conjunction with an overall monitoring

Figure 5. Distribution of third digit of diastolic blood pressure.
NDA: new drug application.
This shows the distribution of the last digit (ones column) of all diastolic blood pressure measures over time. The value in parentheses is approximate sample size for each site. In this NDA, sites show a wide range of measurement bias for the last digit of diastolic blood pressure. Some sites round to the nearest multiple of 5 (e.g., sites 839, 843, and 844 are highlighted with circles), whereas others are far more precise (e.g., sites 885, 886, and 895 are highlighted with rectangles). This issue is considered a Mild Concern and, if detected while a trial was ongoing, could be addressed over the phone with additional coordinator training.

plan that adapts to identified risks as a trial progresses has the potential to decrease the need for frequent on-site monitoring visits, to increase the likelihood of detecting discrepant trends in data before trial closure, and to improve data integrity and participant safety while decreasing trial costs compared to processes that are dependent primarily on source documentation.
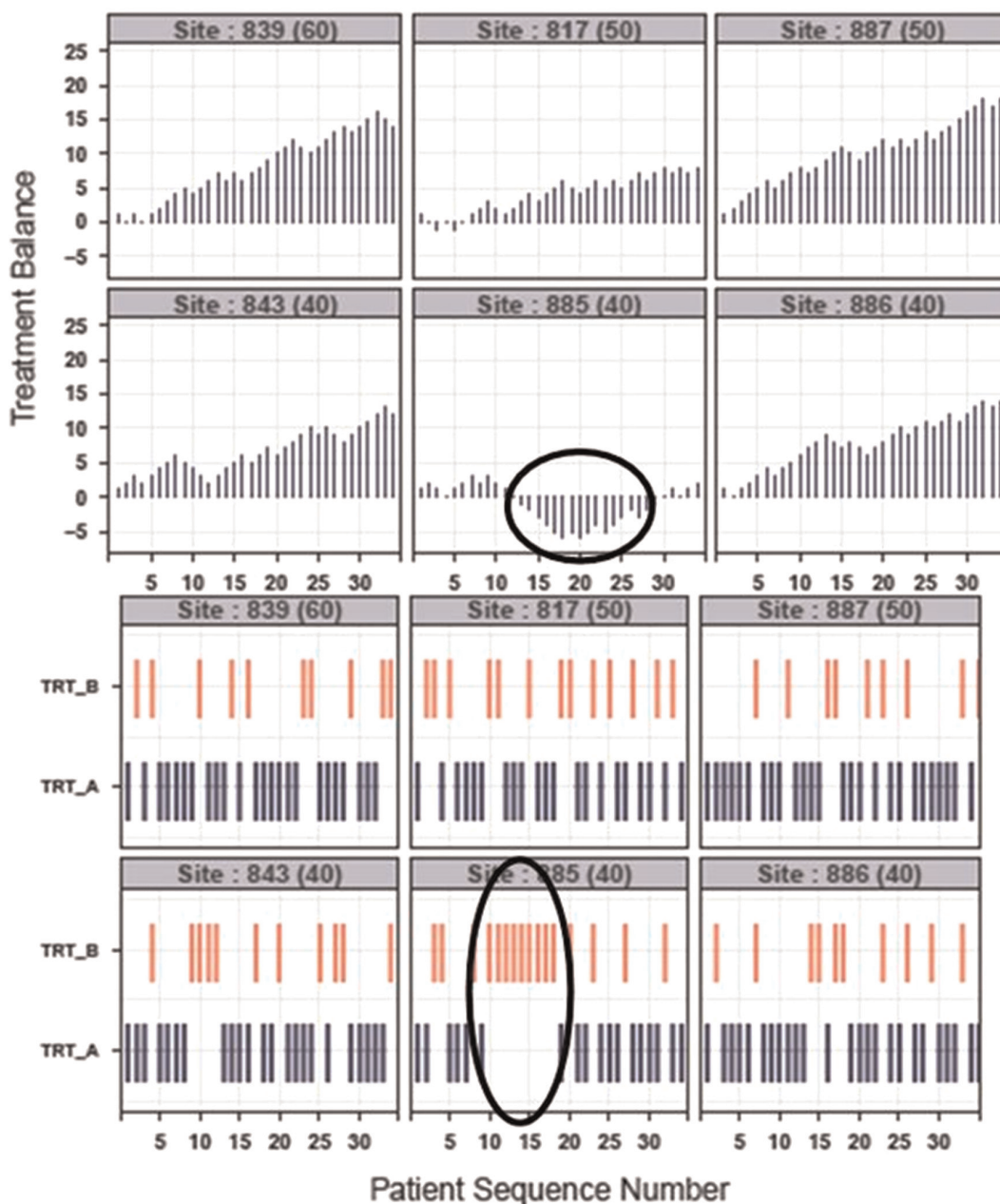
Figure 6. Randomization assignment imbalance (top) and treatment assignment sequence (bottom).
FDA: Food and Drug Administration.
This displays treatment balance over time (top six panels) and sequential treatment assignment at six selected sites (bottom six panels). In this trial, the randomization schema was 2:1 in favor of TRT_A. The value in parentheses is approximate sample size. FDA site inspections identified Serious Problems at site 885, including improper blinding procedures. All other sites show an expected, progressive imbalance of assignments to TRT_A. The long sequence of assignments to only TRT_B for participants 10 through 18 at site 885 suggests a deviation from the randomization plan (anomaly is circled).

Conflict of interest

The authors declare that there are no conflicts of interest.

References

1. **Eisenstein EL, Lemons II PW, Tardiff BE**, *et al*. Reducing the costs of phase III cardiovascular clinical trials. *Am Heart J* 2005; **149**: 482–88.

2. **National Research Council**. *Assuring Data Quality and Validity in Clinical Trials for Regulatory Decision Making: Workshop Report*. The National Academies Press, Washington, DC, 1999.

3. **Eisenstein EL, Collins R, Cracknell BS**, *et al*. Sensible approaches for reducing clinical trial costs. *Clin Trials* 2008; **5**: 75–84.

4. **Buyse M**. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Stat Med* 1999; **18**: 3435–51.

5. **Weir C, Murray G**. Fraud in clinical trials; detecting and preventing it. *Significance* 2011; **8**: 164–68.

6. **Woodin KE**. How much is enough? The high cost of source document review (SDR). CenterWatch NEWSON-LINE. Available at: http://www.centerwatch.com/news-online/article/578/how-much-is-enough–the-high-cost-of-source-document-review-sdr (2006, accessed 1 March 2012).

7. **Mitchel JT, Kim YJ, Choi J**, *et al*. Evaluation of data entry errors and data changes to an electronic data capture clinical trial database. *Drug Inf J* 2011; **45**: 421–30.

8. **Tantsyura V, Grimes I, Mitchel J**, *et al*. Risk-based data verification approaches: Pros and cons. *Drug Inf J* 2010; **44**: 745–56.

9. **International Conference on Harmonisation (ICH)**. ICH topic E9: Statistical principles for clinical trials. Step 5 version. International Conference on Harmonisation (ICH), Geneva, 16 September 1998. Available at: http://www.gpo.gov/fdsys/granule/FR-1998-09-16/98-24754/content-detail.html

10. **Bakobaki JM, Rauchenberger M, Joffe N**, *et al*. The potential for central monitoring techniques to replace on-site monitoring: Findings from an international multi-centre clinical trial. *Clin Trials* **9**: 257–64.

11. **NCI issues information on falsified data in NSABP trials**. *J Natl Cancer Inst* 1994; **86**: 487–89.

12. **O'Kelly M**. Using statistical techniques to detect fraud: A test case. *Pharm Stat* 2004; **3**: 237–46.

13. **Lienard JL, Quinaux E, Fabre-Guillevin E**, *et al*. Impact of on-site initiation visits on patient recruitment and data quality in a randomized trial of adjuvant chemotherapy for breast cancer. *Clin Trials* 2006; **3**: 486–92.

14. **Duley L, Antman K, Arena J**, *et al*. Specific barriers to the conduct of randomized trials. *Clin Trials* 2008; **5**: 40–48.

15. **Murrel P**. *R Graphics*. Chapman and Hall/CRC, London, 2006.

16. **Wegman EJ**. Hyperdimensional data analysis using parallel coordinates. *J Am Stat Assoc* 1990; **85**: 664–75.

17. **US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH)**. *Guidance for industry oversight of clinical investigations – A risk-based approach to monitoring*. OMB Control No. 0910-0733. Food and Drug Administration, Silver Spring, MD, August 2013.