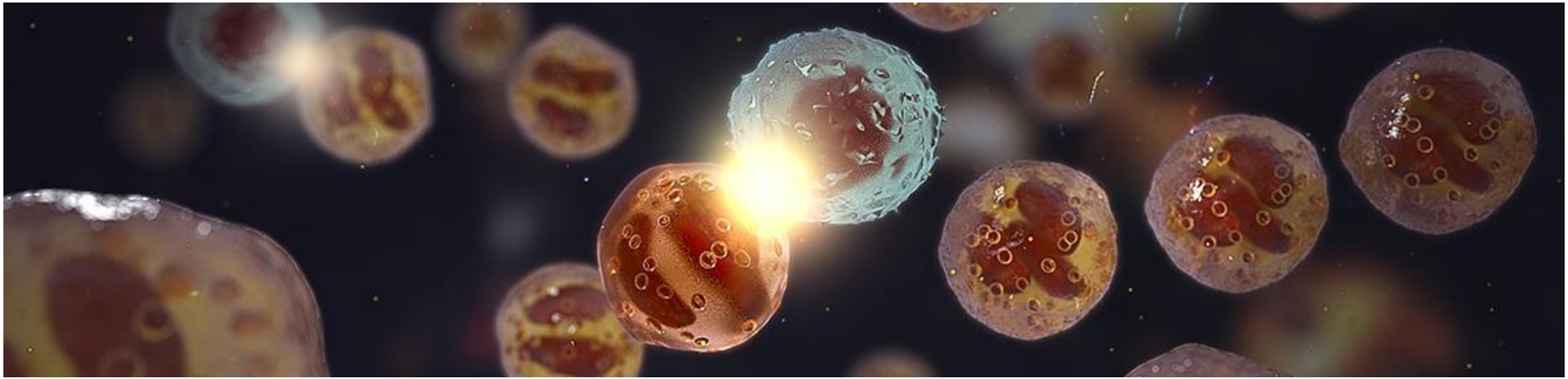
Centralized Statistical Monitoring (CSM) to Detect Data Integrity Issues

Alun Bedding

PhUSE Single Day Event

21st July 2016



Agenda

- What is CSM and why use it?
- Debunking some of the myths around CSM
- What is the regulatory position?
- What sorts of tests are carried out?
- What is going on externally?
- Conclusions

Agenda

- **What is CSM and why use it?**
- Debunking some of the myths around CSM
- What is the regulatory position?
- What sorts of tests are carried out?
- What is going on externally?
- Conclusions

Centralised Statistical Monitoring (CSM)

- An approach for identifying and managing issues affecting data integrity such as non-random errors, GCP misconduct including fraud at investigational sites
- Sophisticated modeling methods to pick up non-random errors or patterns
- CSM should be performed at limited times during a trial, and then prior to any important analysis (interim or final)
  - Why? It requires sufficiently large amounts of data to pick up signals

Why Centralized Statistical Monitoring?

- Much of this has background in fraud and falsification.

- FDA proposed rule for “Reporting Information Regarding Falsification of Data” defines falsification of data as:
  - *creating, altering, recording or omitting data in such a way that the data do not represent what actually occurred*

- Examples of falsification of data include but are not limited to:
  - Creating data that were never obtained
  - Altering data that were obtained by substituting different data
  - Recording or obtaining data from a specimen, sample or test whose origin is not accurately described or in a way that does not accurately reflect the data
  - Omitting data that were obtained and ordinarily would be recorded
- Used to identify strange patterns in the data.
  - Could be mis-calibration of equipment.
  - Sites not properly trained
  - Other less serious issues

Agenda

- What is CSM and why use it?
- **Debunking some of the myths around CSM**
- What is the regulatory position?
- What sorts of tests are carried out?
- What is going on externally?
- Conclusions

Debunking Some Myths Around CSM!!!!

- Isn’t it just data cleaning?
  - NO - It is a statistical analysis of the data.
  - Not comparing treatments but an analysis comparing sites with each other, patients with each other irrespective of treatment.
- What is the difference between CSM and looking at thresholds?
  - No formal statistical analyses are carried out when using simple thresholds.
  - CSM involves univariate and multivariate statistical analysis of all data to compare within and across sites.
- Should be carried out by people trained in statistics?
  - Yes. It involves the use of sophisticated statistical methods, however, the interpretation should be done with the study teams.

How Much Falsification goes on?

- FDA estimates that it will see 73 reports of data falsification across all divisions.
- However, the impact is critical.
- A single investigator involved in data falsification could jeopardize the trial and potentially a drug submission.
- Regulators have the power to terminate a product's development immediately or withdraw approval of an already marketed drug
- The U.S. Code on Crimes and Criminal Procedure makes submission of falsified information to the federal government a criminal act, and convictions can lead to substantial fines and even imprisonment of responsible parties.

BUSINESS

Amgen Finds Data Falsified in Obesity-Diabetes Study Featuring Grizzly Bears

Research published in 2014 retracted by journal Cell Metabolism after scientist is found doctoring results

PHOTO: DANI POZO/AGENCE FRANCE-PRESSE/GETTY IMAGES

By JONATHAN D. ROCKOFF

110 COMMENTS

Agenda

- What is CSM and why use it?
- Debunking some of the myths around CSM
- **What is the regulatory position?**
- What sorts of tests are carried out?
- What is going on externally?
- Conclusions

FDA thoughts?

Centralized monitoring is a remote evaluation carried out by sponsor personnel or representatives (e.g., clinical monitors, data management personnel, or statisticians) at a location other than the sites at which the clinical investigation is being conducted. Centralized monitoring processes can provide many of the capabilities of on-site monitoring as well as additional capabilities.

- Supplement or reduce the frequency and extent of on-site monitoring with monitoring activities that can be done as well or better remotely or with monitoring activities that can be accomplished using centralized processes only. Examples include:
    - Monitor data quality through routine review of submitted data to identify and follow-up on missing data, inconsistent data, data outliers, and potential protocol deviations that may be indicative of systemic or significant errors in data collection and reporting at a site
    - Conduct statistical analyses to identify data trends not easily detected by on-site monitoring, such as
        - Standard checks of range, consistency, and completeness of data
        - Checks for unusual distribution of data within and between study sites, such as too little variance[28]
    - Analyze site characteristics, performance metrics (e.g., high screen failure or withdrawal rates, high frequency of eligibility violations, delays in reporting data), and clinical data to identify trial sites with characteristics correlated with poor performance or noncompliance

Source: FDA Guidance for Industry: Oversight of clinical investigations: a risk based approach to monitoring

ICH E6 R2 Addendum thoughts?

ADDENDUM

The sponsor should develop a systematic, prioritized, risk-based approach to monitoring clinical trials. The flexibility in the extent and nature of monitoring described in this section is intended to permit varied approaches that improve the effectiveness and efficiency of monitoring. A combination of on-site and centralized monitoring activities may be appropriate. The sponsor should document the rationale for the chosen monitoring strategy (e.g., in the monitoring plan).

On-site monitoring is performed at the sites at which the clinical trial is being conducted.

Centralized monitoring is a remote evaluation of ongoing and/or cumulative data collected from trial sites, in a timely manner. Centralized monitoring processes provide additional monitoring capabilities that can complement and reduce the extent and/or frequency of on-site monitoring by such methods as:

(a) Routine review of submitted data.

(b) Identification of missing data, inconsistent data, data outliers or unexpected lack of variability and protocol deviations that may be indicative of systematic or significant errors in data collection and reporting at a site or across sites, or may be indicative of potential data manipulation or data integrity problems.

(c) Using statistical analyses to identify data trends such as the range and consistency of data within and across sites.

(d) Analyzing site characteristics and performance metrics.

(e) Selection of sites and/or processes for targeted on-site monitoring.

Agenda

- What is CSM and why use it?
- Debunking some of the myths around CSM
- What is the regulatory position?
- **What sorts of tests are carried out?**
- What is going on externally?
- Conclusions

Methods Found in Literature (not exhaustive)

Statistical test		Citation
Dates - order checking		Baigent (2008), Kirkwood (2013), Lindblad (2013), Weir (2011)
Duplicates		Lindblad (2013)
Visit timing		Baigent (2008), Kirkwood (2013), O'Kelly (2004)
Outliers	**Univariate using SD**	Kirkwood (2013)
	Univariate Grubb's test - inner quartile range method	Kirkwood (2013)
	Multivariate Euclidean distance	Kirkwood (2013)
	Multivariate Mahalanobis distance	Kirkwood (2013)
		Pogue (2013)
Digit preference - integers (rounding)		Baigent (2008), Kirkwood (2013)
	Digit preference - Benford's law	Baigent (2008), Kirkwood (2013), Lindblad (2013),Walter (2001)
	Comparison with all other sites (either first digit or last)	
	Digit preference - Chi-Square test for uniform distribution	Al-Marzouki (2005), Pogue (2013), Walter (2001)
Comparison of means		
	Using Chernoff faces and Star plots	Kirkwood (2013)
	Using t-test or other pairwise comparisons	Al-Marzouki (2005), Pogue (2013)
Inliers		
	Multivariate Mahalanobis distance	Baigent (2008), Kirkwood (2013), Weir (2011)
Adverse event rates		Kirkwood (2013), Lindblad (2013)
Correlations		Kirkwood (2013), Weir (2011)
		O'Kelly (2004)
Variability		Al-Marzouki (2005), Kirkwood (2013), Pogue (2013), Weir (2011)
		Lindblad (2013)
		Pogue (2013)
Categorical variables		Kirkwood (2013)
Kurtosis		Baigent (2008), Weir (2011)
Cluster analysis		Weir (2011)
Frequency analysis		Pogue (2013)
Multidimensional approach		Pogue (2013), Weir (2011)

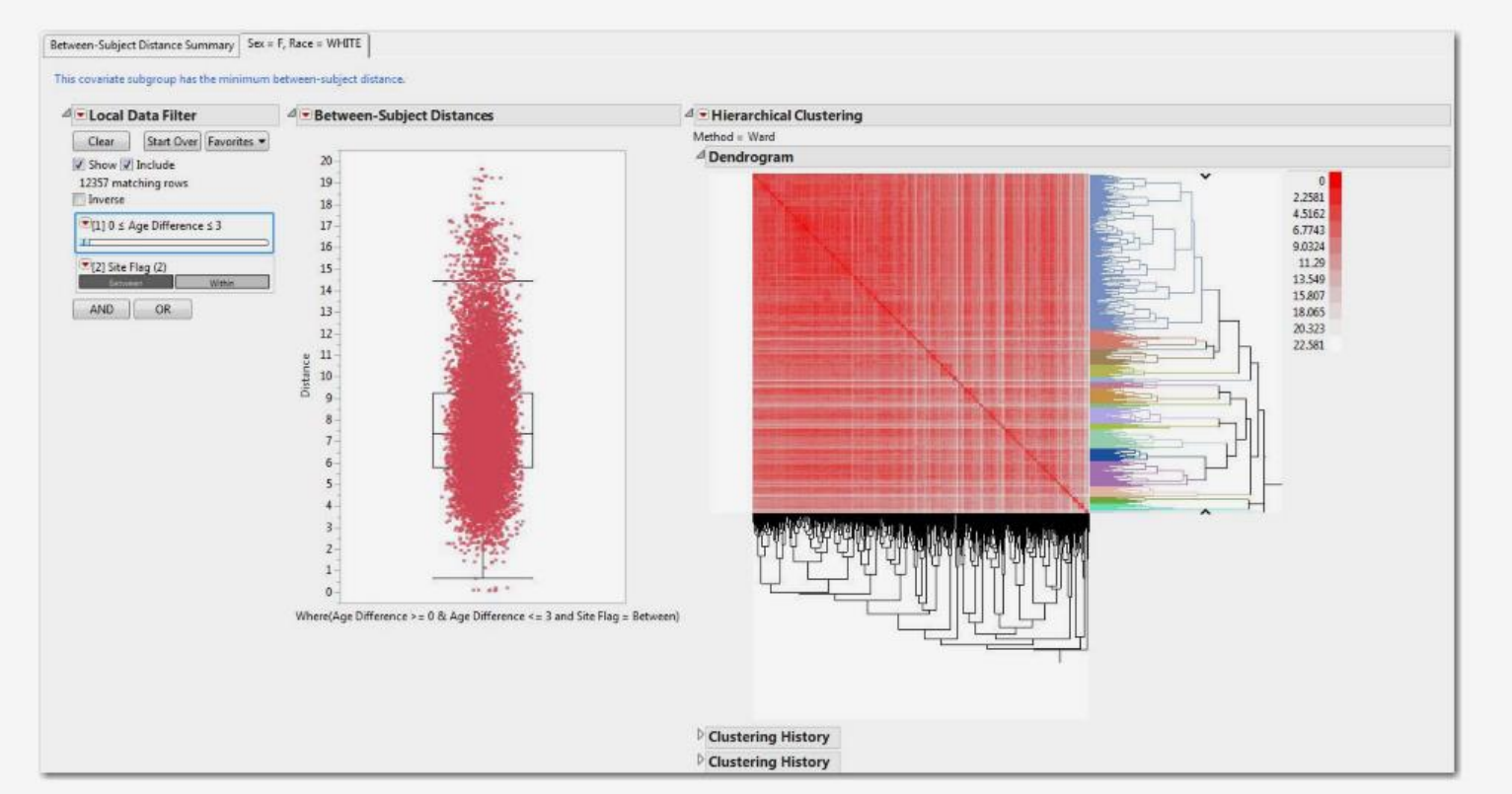
CSM Methods

- Cluster Analysis
  - Cluster Subjects Across Sites – Patients going to multiple sites or possible fabrication across sites
  - Cluster Subjects Within Study Sites – Patients within a site to identify possible fabricated data
- Digit Preference (leading and trailing)
- Multivariate Inliers and Outliers – to close to the mean or too far might indicate fabricated data

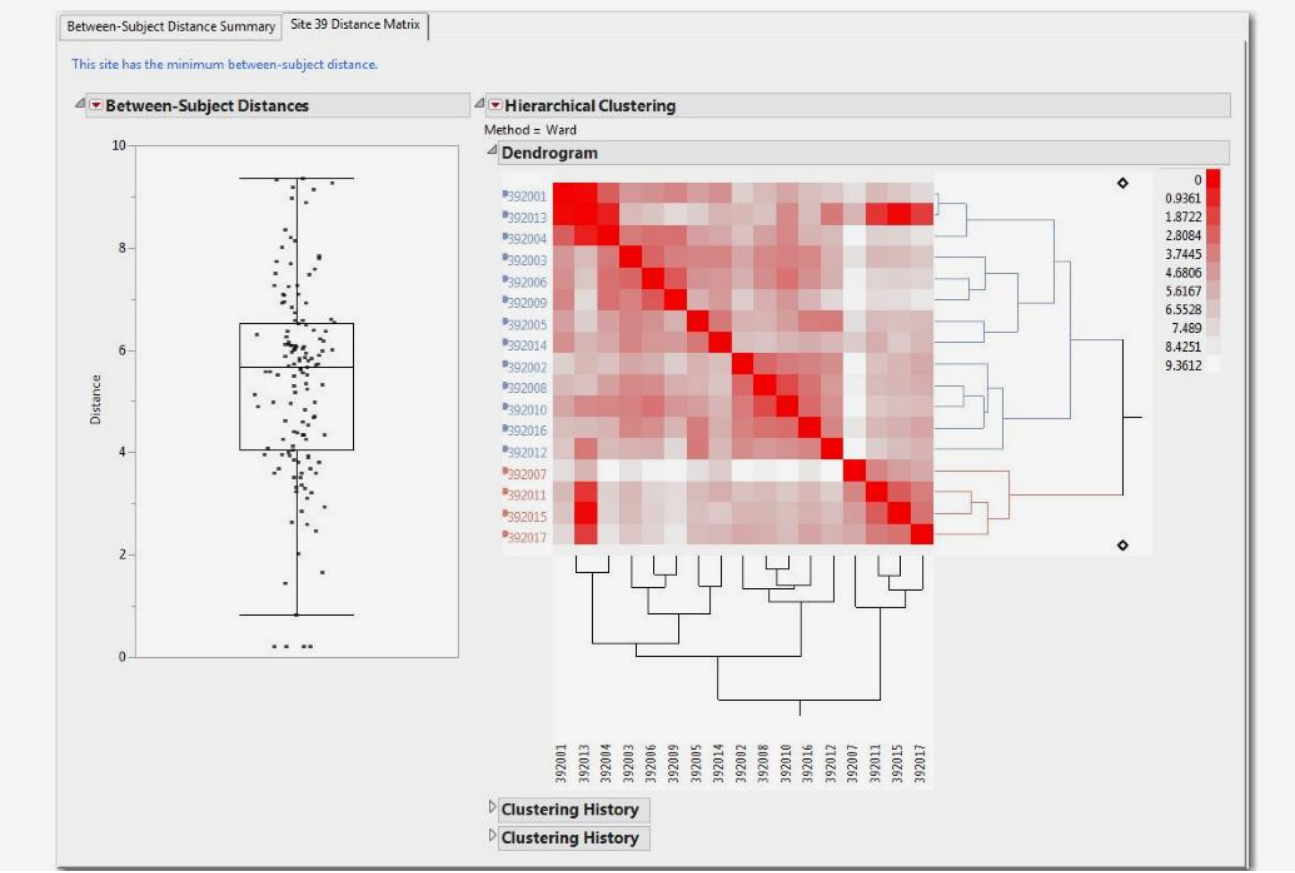
Clustering Subjects Across Sites and Within Sites

- Both use similar methods
- Euclidian distance is calculated between subjects for a number of covariates (age, sex, vital signs, ECG, lab data)
- If the distance is small there is similarity and maybe they are the same subject.
- Subjects twice in the same study or in more than one site.
- Why clustering?
- It can help identify sets of similar subjects

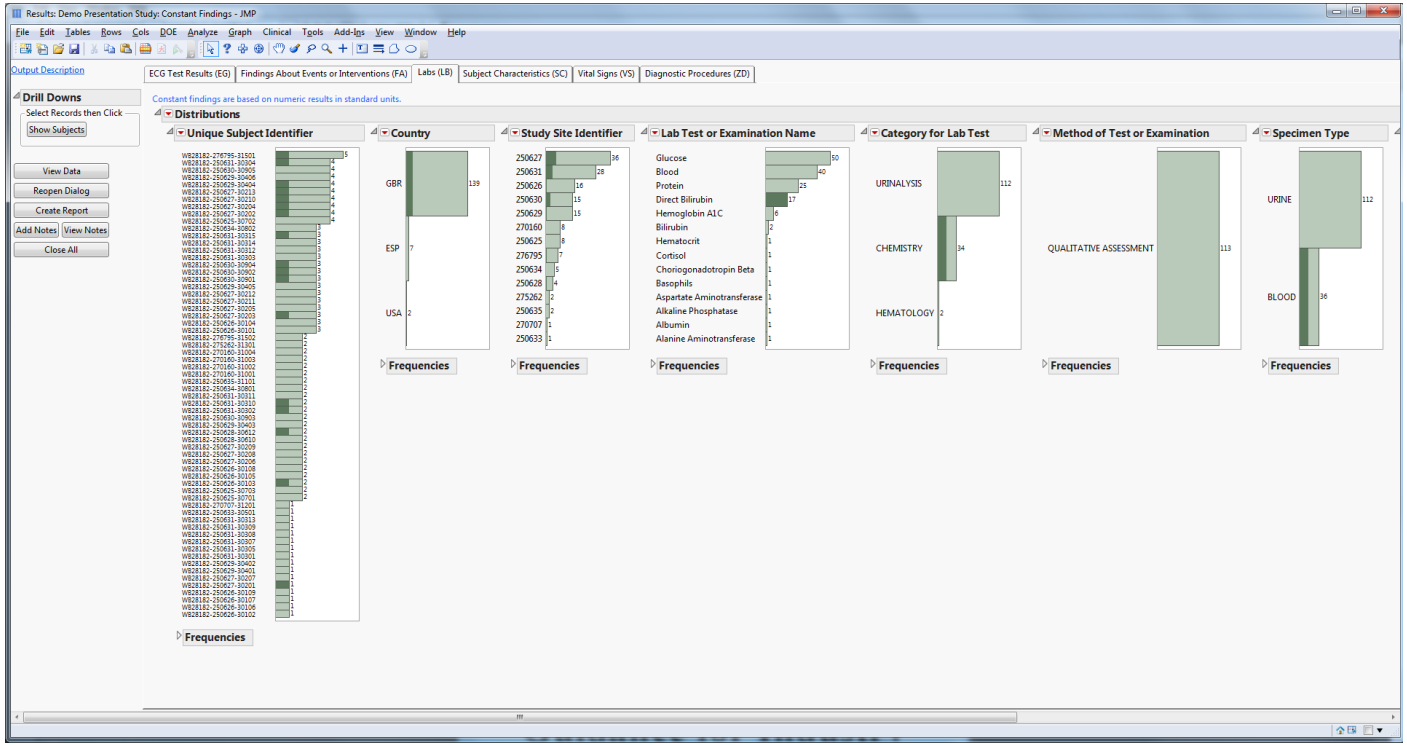
Clustering Subjects Across Sites



Clustering Subjects Within a Site



Drill Downs


Results: Demo Presentation Study: Constant Findings - JMP
File Edit Tables Rows Cols DOE Analyze Graph Clinical Tools Add-Ins View Window Help
Output Description
ECG Test Results (EG)
Findings About Events or Interventions (FA)
Labs (LB)
Subject Characteristics (SC)
Vital Signs (VS)
Diagnostic Procedures (ZD)
Drill Downs
Select Records then Click
Show Subjects
View Data
Reopen Dialog
Create Report
Add Notes
View Notes
Close All
Constant findings are based on numeric results in standard units.
Distributions
Unique Subject Identifier
Country
GBR 139
ESP 7
USA 2
Study Site Identifier
250627 36
250631 28
250626 16
250630 15
250629 15
270160 8
250625 8
276795 7
250634 5
250628 4
275262 2
250635 2
270707 1
250633 1
Lab Test or Examination Name
Glucose 50
Blood 40
Protein 25
Direct Bilirubin 17
Hemoglobin A1C 6
Bilirubin 2
Hematocrit 1
Cortisol 1
Choriogonadotropin Beta 1
Basophils 1
Aspartate Aminotransferase 1
Alkaline Phosphatase 1
Albumin 1
Alanine Aminotransferase 1
Category for Lab Test
URINALYSIS 112
CHEMISTRY 34
HEMATOLOGY 2
Method of Test or Examination
QUALITATIVE ASSESSMENT 113
Specimen Type
URINE 112
BLOOD 36
Frequencies

Digit Preference

- Identify potential issues such as:
  - Investigators doing manual data entry
  - Miscalibrated equipment
  - The use of methods different from the protocol or problems with subjective measurements.
- The consistency of the trailing or leading digit for variables at a site compared to that of all the other sites combined.
- Compared using a Cochran-Mantel Haenszel test.
- This assumes that the data is ordinal, which the trailing/leading digit should be, ranging from 0 to 9.

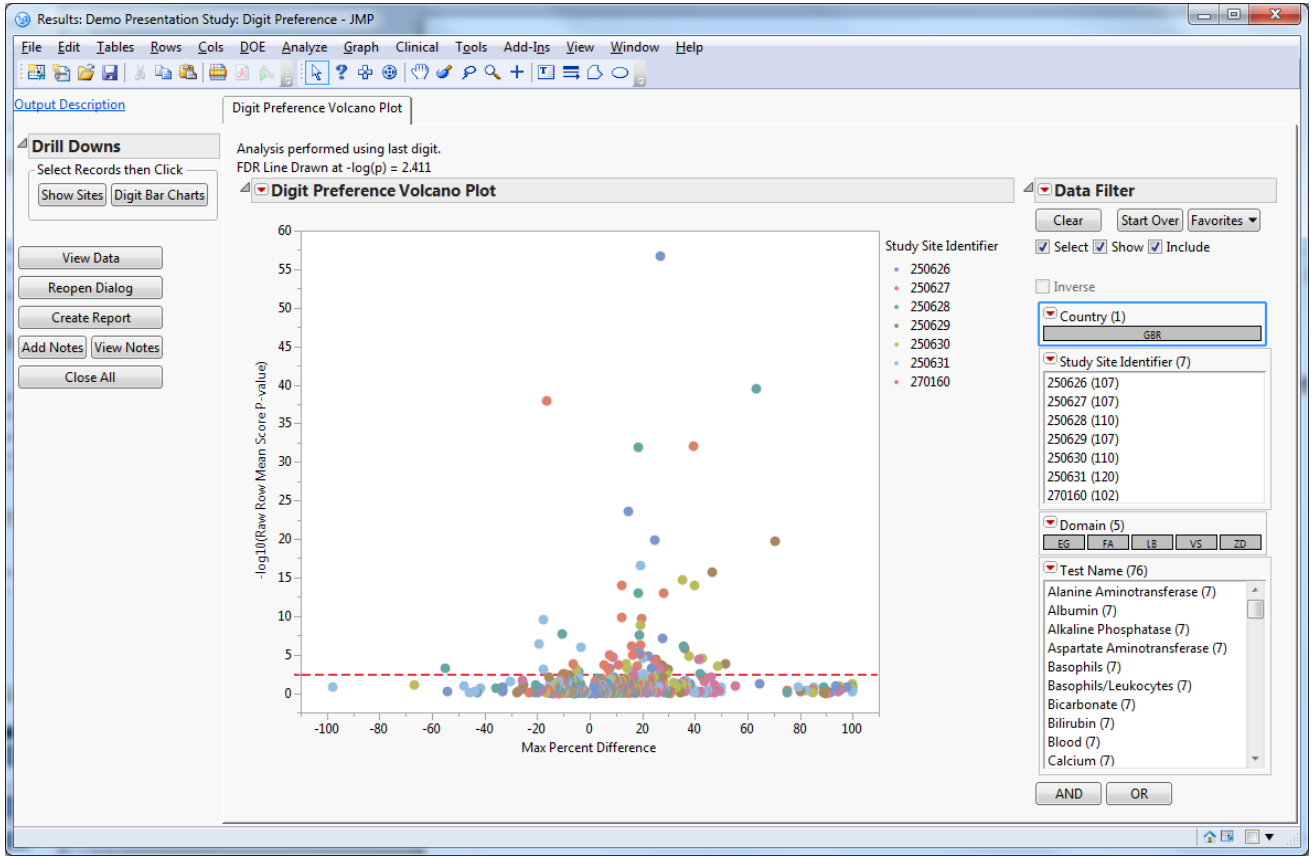
Digit Preference

- For each variable k, let $F_{k,i}$ be the distribution of the trailing digit of the kth variable at the ith site. Then want to test:

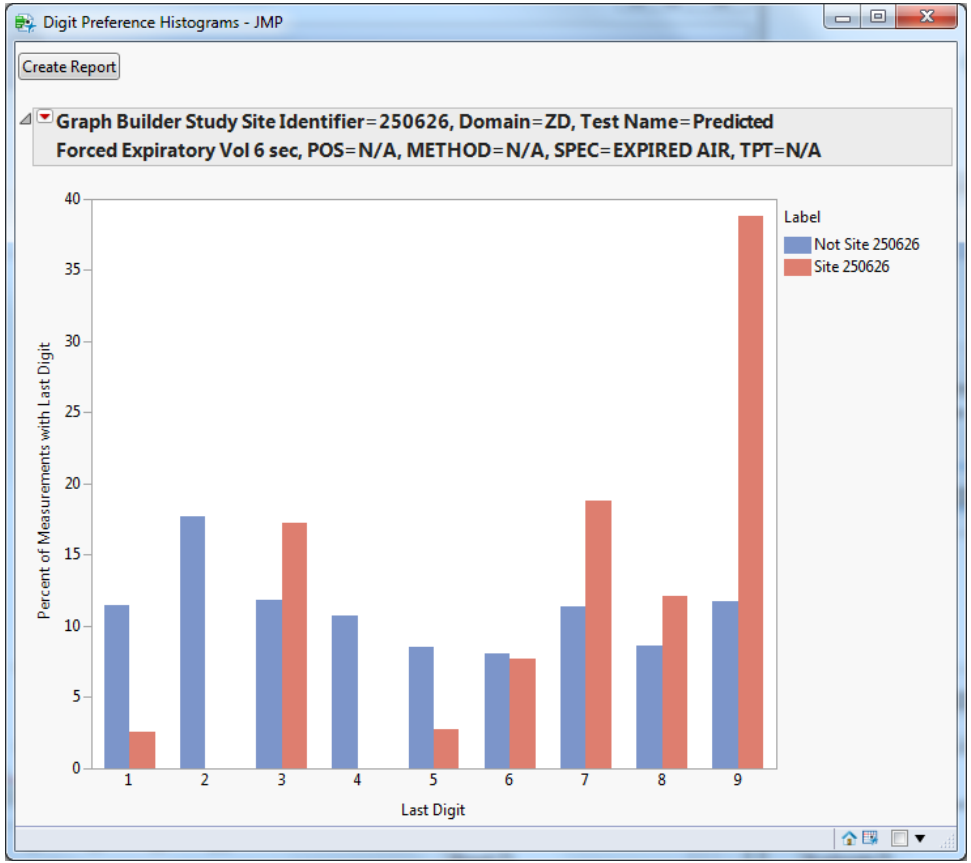
  $H_0 : F_{k,1} = \Lambda = F_{k,s}$ where there are s sites involved in the study

- Displayed using:
  - Volcano plots of –log10(p-value) versus max percent difference
- Multiplicity adjustments are easily applied in the volcano plot
  - Any of the raw p-values at or above the reference line are considered significant when accounting for the False Discovery Rate (FDR) method of Benjamini and Hochberg.
  - Among the rejected null hypotheses from a family or set of multiple tests, the expected proportion of erroneous rejections is defined as the FDR (typically set at 0.05).

Volcano Plot



Digit Preference Histogram

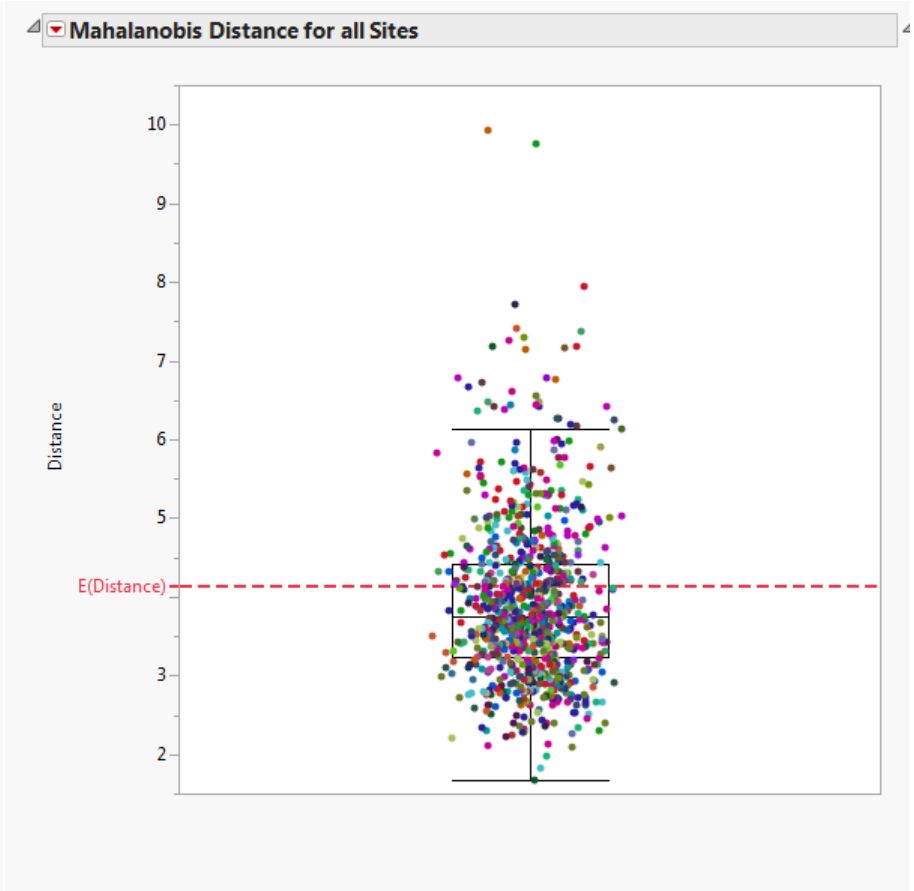


Multivariate Inliers and Outliers

- We can all look at univariate outliers (values different from the rest of the sample) and inliers (values close to the mean).
- But it does not take into account the relationships between variables/covariates.
- For this we need a multivariate procedure.
- Mahalanobis distance can be used to calculate the distance between two vectors of data (to assess similarity) or from a vector to a particular point in multivariate space (typically the multivariate mean or centroid).
- Let $x_i^T = \left(x_{i1}, \Lambda\ , x_{ip}\right)$ be the ith observation, with sample mean vector $\bar{x}$ and covariance matrix s. For each observation, calculate $D_i = \left(x_i - \bar{x}\right)^T S^{-1}\left(x_i - \bar{x}\right)$
- The test statistic $\frac{(n-p)n}{\left(n^2-1\right)p} D_i$ has an approximate F distribution with p & $n-p$

  degrees freedom.

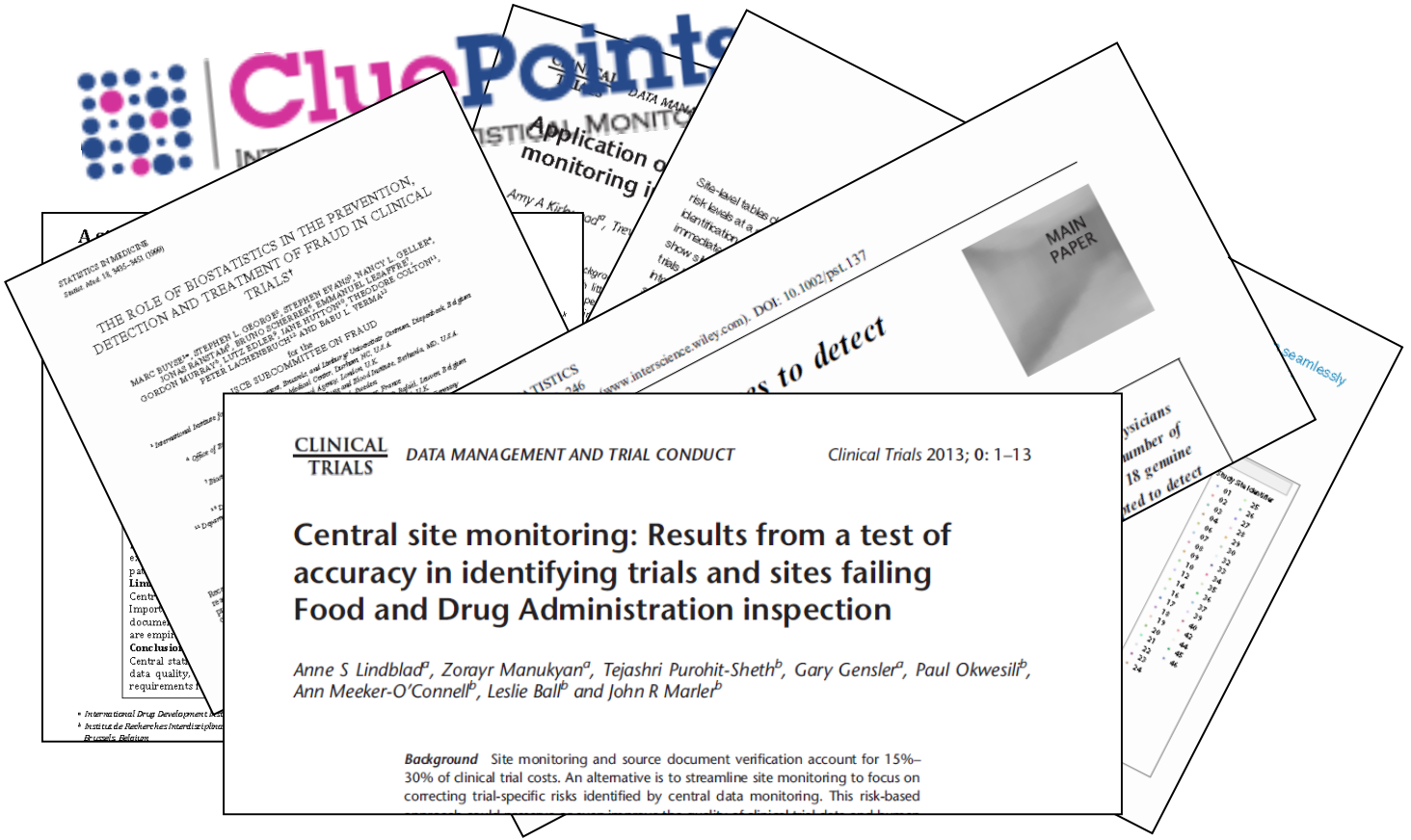
Multivariate Inliers and Outliers



Agenda

- What is CSM and why use it?
- Debunking some of the myths around CSM
- What is the regulatory position?
- What sorts of tests are carried out?
- **What is going on externally?**
- Conclusions

What is Being Done Externally?



Survey by TransCelerate

Sponsored Special Section by TransCelerate BioPharma: Original Research

DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
2016, Vol. 50(1) 15-21

DOI: 10.1177/2168479015620248
tirs.sagepub.com

Detecting Data Quality Issues in Clinical Trials: Current Practices and Recommendations

David Knepper, MS, MBA[1], Christian Fenske, MPH[2], Patrick Nadolny, MS[3], Alun Bedding, PhD[4], Elena Gribkova, MSc[5], John Polzer, DVM, MS, MS[6], Jennifer Neumann, CCDM[7], Brett Wilson, BSP[8], Joanne Benedict, MS[9], and Andy Lawton, CSTAT[10]

Abstract

Background: Data quality issues in clinical trials can be caused by a variety of behaviors including fraud, misconduct, intentional or unintentional noncompliance, and significant carelessness. Regardless of how these behaviors are defined, they may compromise the validity of the study results. Reliable study results and quality data are needed to evaluate products for marketing approval and for decisions that are made on the use of medicine. This article focuses on detecting data quality issues, irrespective of origin or motive. Early detection of data quality issues are important so that corrective actions taken can be implemented during the conduct of the trial, recurrence can be prevented, and data quality can be preserved. *Methods:* A survey was distributed to TransCelerate member companies to assess current strategies for detecting and mitigating risks involving fraud and misconduct in clinical trials. A review of literature across many industries from 1985 to 2014 was conducted using multiple platforms. *Results:* Eighteen TransCelerate member companies anonymously responded to the survey. All of the respondents had one or more existing strategies for fraud and misconduct detection. The literature search identified current practices and methodologies across many industries. *Conclusions:* TransCelerate recommends the creation of an integrated, multifaceted approach to proactively detect data quality issues. Detection methods should include a strategy tailored to the characteristics of the study. Some sponsors are taking advantage of more advanced methods and integrated processes and systems to proactively detect and address issues, relying on advances in technology to more efficiently review data in real time. Further research is underway to assess statistical data quality detection methodology in clinical trials.

Survey by TransCelerate

- TransCelerate member companies were sent a survey to collect high level information specific to fraud and misconduct.
- 18 companies replied

Strategies	Total Responses	Strategy Used	Not Used	All Studies
Site Monitoring	18	17	1	16
Auditing	17	14	3	14
Data Review Activities	17	14	3	13
Central Monitoring	17	14	3	4
Advanced Statistical Analysis	17	8	9	3

- The three most common tactics used to detect fraud and misconduct are detecting exists, and detecting duplicate results.
- Other tactics used include variation not as expected, site performance measures, that patient is unique, and data quality indicators.

Does it work? Further TransCelerate Paper

Sponsored Special Section by TransCelerate BioPharma: Original Research

DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
1-11

DOI: 10.1177/2168479016630576
tirs.sagepub.com

Statistical Monitoring in Clinical Trials: Best Practices for Detecting Data Anomalies Suggestive of Fabrication or Misconduct

David Knepper, MS, MBA[1], Anne S. Lindblad, PhD[2], Gaurav Sharma, PhD[2], Gary R. Gensler, MS[2], Zorayr Manukyan, PhD[3], Abigail G. Matthews, PhD[2], and Yodit Seifu, PhD[4]

Abstract

Background: Traditional site-monitoring techniques are not optimal in finding data fabrication and other nonrandom data distributions with the greatest potential for jeopardizing the validity of study results. TransCelerate BioPharma conducted an experiment testing the utility of statistical methods for detecting implanted fabricated data and other signals of noncompliance. *Methods:* TransCelerate tested statistical monitoring on a data set from a chronic obstructive pulmonary disease (COPD) clinical study with 178 sites and 1554 subjects. Fabricated data were selectively implanted in 7 sites and 43 subjects by expert clinicians in COPD. The data set was partitioned to simulate studies of different sizes. Analyses of vital signs, spirometry, visit dates, and adverse events included distributions of standard deviations, correlations, repeated values, digit preference, and outlier/inlier detection. An interpretation team, including clinicians, statisticians, site monitoring, and data management, reviewed the results and created an algorithm to flag sites for fabricated data. *Results:* The algorithm identified 11 sites (19%), 19 sites (31%), 28 sites (16%), and 45 sites (25%) as having potentially fabricated data for studies 2A, 2, 1A, and 1, respectively. For study 2A, 3 of 7 sites with fabricated data were detected, 5 of 7 were detected for studies 2 and 1A, and 6 of 7 for study 1. Except for study 2A, the algorithm had good sensitivity and specificity (>70%) for identifying sites with fabricated data. *Conclusions:* We recommend a cross-functional, collaborative approach to statistical monitoring that can adapt to study design and data source and use a combination of statistical screening techniques and confirmatory graphics.

Preparing the Database

- De-identified a large completed COPD database
  - 178 sites, 1154 subjects
  - Objective data including spirometry data and vital signs

- Replaced real data with fabricated data
  - Selected spirometry and vital sign data for fabrication
  - Selected 43 subjects at 7 sites for implantation of fabricated data
  - Asked independent physicians knowledgeable in COPD to replace missing real data with plausible fabricated data in a manner likely to evade detection

Four Test Conditions to Mimic Four Studies

- Created four data sets mimicking four study conditions
  - All sites and all subjects … Large completed study
    (178 sites and 1154 subjects)
  - Some sites and all subjects … Small completed study
    (61 sites and 824 subjects)
  - All sites and some subjects … Large study mid-study
    (178 sites and 627 subjects)
  - Some sites and some subjects … Small study mid-study
    (61 sites and 338 subjects)

Statistical Analysis

- Sent four data sets to a blinded independent data analysis center
- Asked data analysis center to perform battery of statistical tests to:
  - Find all sites with fabricated data
  - Identify which data are fabricated
  - Evaluate which statistical tests are more or less effective in the four test conditions
- Domains included
  - Enrollment, randomization, visit timing
  - Clinical outcomes (spirometry data: FEV1, FVC, VS)
  - Safety (Adverse Events)

Results

- Results yielded >70% sensitivity and specificity for all but the smallest study (small study, mid-study condition)
  - >70% of sites were detected in three of the four test conditions
  - Only 3 of 7 sites were detected in the smallest study
  - >79% of subjects were detected across all studies
  - Correctly identified >74% of sites without fabricated data
  - Correctly identified >70% of subjects without fabricated data
  - Overall, >92% of sites and >97% of subjects were correctly flagged

Agenda

- What is CSM and why use it?
- Debunking some of the myths around CSM
- What is the regulatory position?
- What sorts of tests are carried out?
- What is going on externally?
- **Conclusions**

Conclusions

- CSM is an integral part of a risk based approach to the over-sight of a study, however, the utility of Statistical Monitoring can be challenged by small, fast enrolling studies
- The timing and cadence of Statistical Monitoring activities is dependent on the volume and rate of change of the dataset
- The Data-Science model, using a multi-disciplinary interpretation team, is optimal to determine the clinical relevance of the statistical findings
- The field is evolving and still experimental, although the methods have been used in other industries
- Statistical Monitoring is useful to augment traditional and central monitoring approaches to detect data quality issues