UNCOVERING FRAUD, MISCONDUCT AND OTHER DATA QUALITY ISSUES IN CLINICAL TRIALS

Biopharmaceutical Applied Statistics Symposium (BASS)

Richard C. Zink, Ph.D.
Principal Research Statistician Developer
JMP Life Sciences
SAS Institute, Inc.

October 2016

SOME DEFINITIONS

- **Fraud**: wrongful or criminal deception intended to result in financial or personal gain
- **Misconduct**: unacceptable or improper behavior
- **Data Quality**: referring to anything and everything that could possibly go wrong with data, including **Fraud** and **Misconduct**

A MOTIVATING EXAMPLE

- National Heart, Lung and Blood Institute (NHLBI) conducted multi-center animal study of two drugs [1]
- Determine whether
  - Drugs could limit infarct size
  - Develop a reproducible animal model for studying myocardial infarction in humans to evaluate new therapies
- There were four centers in the trial
- Several data issues were identified at one center
- Upon investigation, the Medical Fellow was found to have falsified results for this and other studies

A MOTIVATING EXAMPLE

- How was this fraud identified?
  - Inconsistent relationship of left ventricle weight and overall weight of dogs
  - Inconsistent relationship between infarct size and collateral blood flow
  - Lower variability in collateral blood flow
  - Time trends showing notable data differences between pre- and post-discovery

A MOTIVATING EXAMPLE

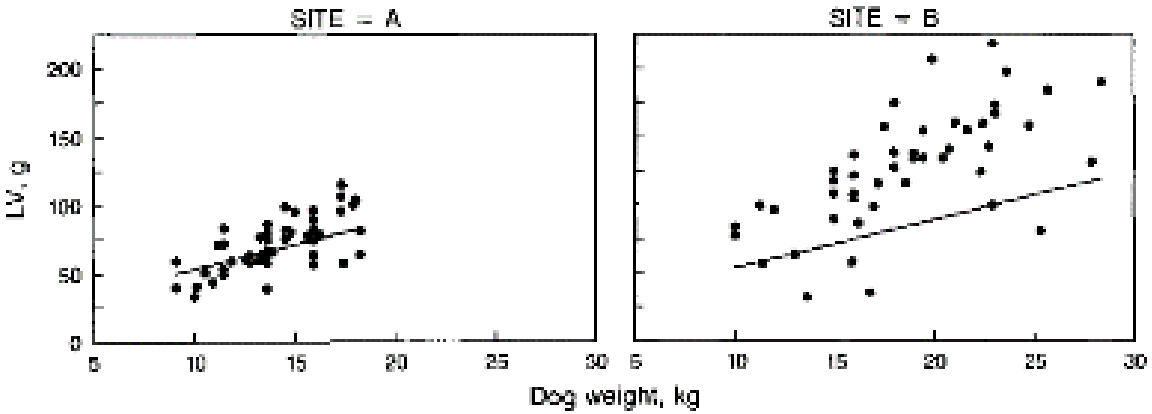
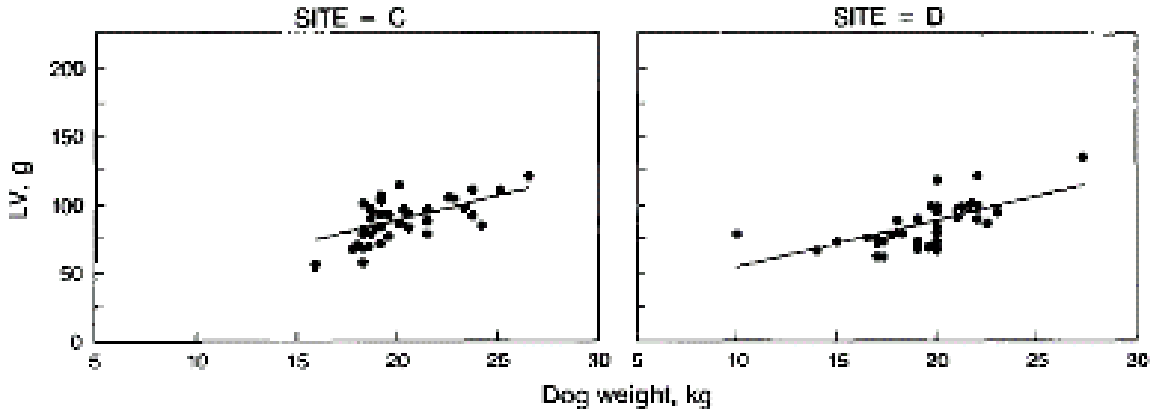


Figure 1A Ventricle weight versus dog weight, sites A and B.



Left ventricle weight versus dog weight by center. Figure from [1]

A MOTIVATING EXAMPLE

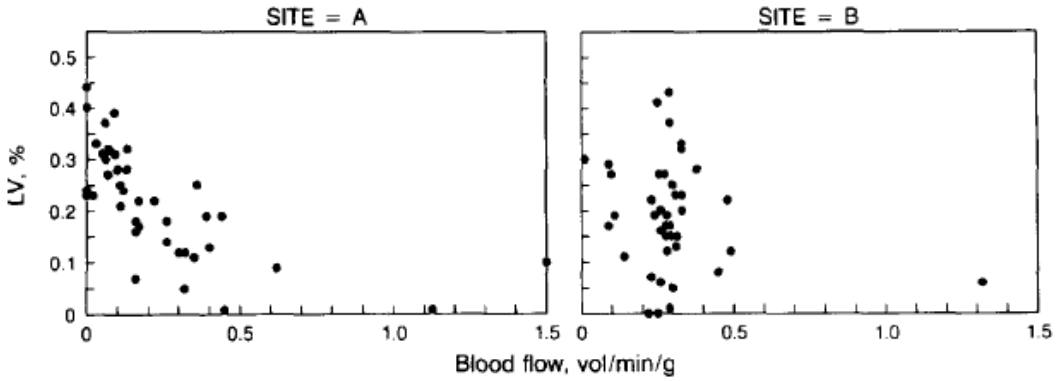


Figure 2A Infarct size versus collateral blood flow.

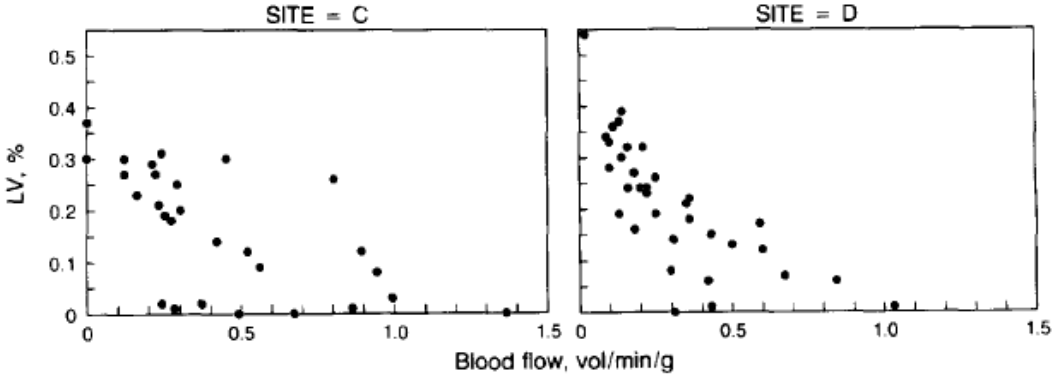


Figure 2B Infarct size versus collateral blood flow.

Infarct size versus collateral blood flow by center. Figure from [1]

A MOTIVATING EXAMPLE

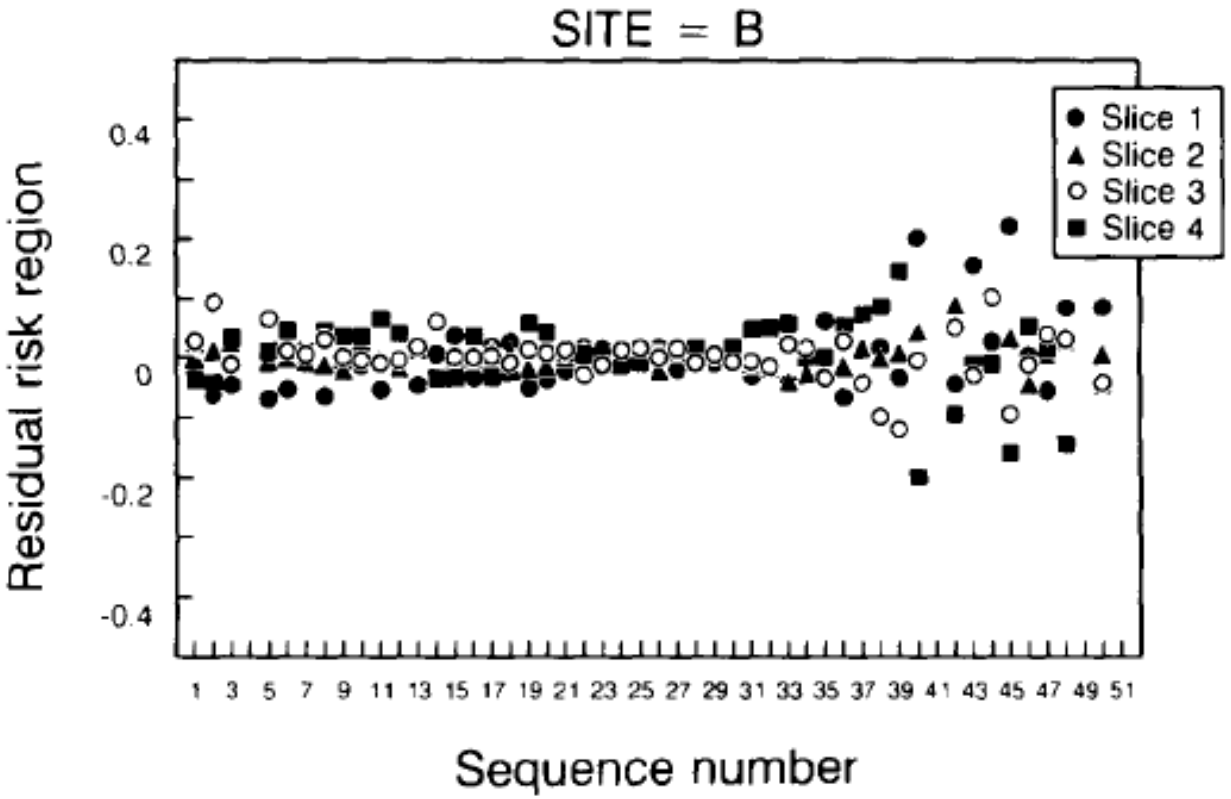


Figure 4 Residual risk region by sequence.

Infarct size by sequence # (dog). Pre-discovery is dog 35 and below. Figure from [1]

A MOTIVATING EXAMPLE

- When first approached, Principal Investigator knew of Medical Fellow's fraud in a trial from previous year
- Hard evidence for current trial
  - A heart was discarded though data was reported from it
  - Pre-discovery dog hearts had no radioactivity
  - Data logs showed discrepancies between dogs reported and analyzed
- Further investigation identified fraud in numerous publications by Medical Fellow

INTRODUCTION

- Fraud and misconduct are important subset of topics involving data quality
  - Is fraud or misconduct more exciting to discuss?
- Quality issues can be due to
  - Carelessness, such as transcription errors
  - Contamination
  - Mechanical failures
  - Poor planning, poor training
  - Fraud or misconduct

INTRODUCTION

- Fraud is the “deliberate attempt to deceive” or the “intention to cheat” [2]
- Fraud/misconduct in clinical trials is difficult to diagnose
  - How to separate from carelessness?
  - Perhaps differences between sites are due to available subjects, or slight variations in technique
  - May identify unusual points indicating a quality problem, but stating that it is explicitly due to fraud may require more evidence [3]

INTRODUCTION

- Many authors agree fraud is uncommon in clinical trials
  - Proportion of investigators committing fraud est. < 1% [2]
  - Other published reports in clinical trials show few or no instances of fraud
- However! Instances may be
  - Undiagnosed
    - Lack of tools
  - Unreported
    - Media firestorm
    - Risk to clinical program

INTRODUCTION

- Why should we bother looking for fraud and misconduct?
  - Ethical to protect the patient
  - Identify problems for correction within the trial
  - Identify problematic sites to avoid in future trials
  - Minimize stress for the study team
  - Reduce risk for a clinical program
- Besides, fraud and misconduct are data quality problems
- Should routinely screen for data quality problems

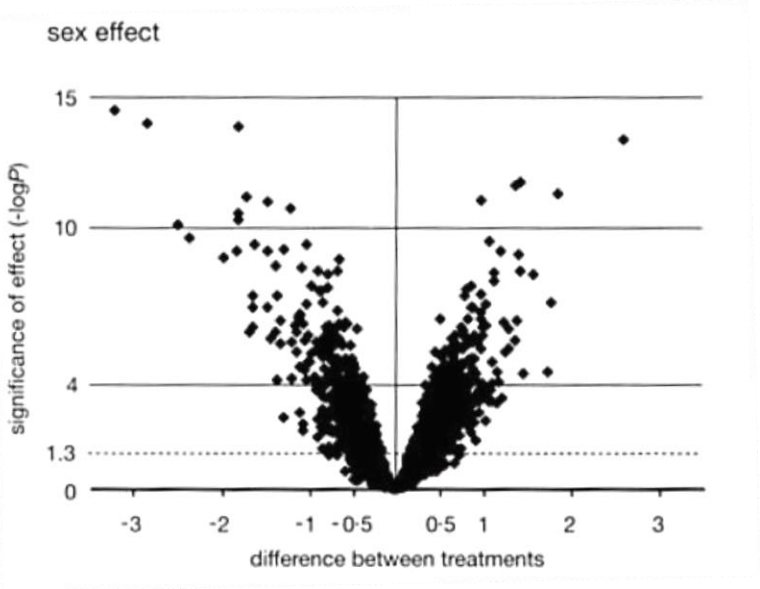
STATISTICAL TESTING

- Fraud/misconduct/quality [2-5]
  - Investigator
  - Patient
  - Lab or CRO
- Statistical testing, pattern matching or clustering
- Graphical displays such as volcano plot to highlight signals and launch follow-up analyses
- Why does this work?
  - Challenging to fabricate data in the many dimensions required for the data to look plausible

STATISTICAL TESTING

- Means, variances, skewness kurtosis per visit
- Identify screening bias
- Frequency of outliers or missing data
- Duplicates or no variation across the trial
- Visits
  - Unusual scheduling (perfect or off schedule)
  - Missing visits
  - Weekends or Holidays
- Clustering for fabricated patients, misuse of samples
- Inliers and outliers
- Unusual trends, autocorrelation

GRAPHICS: VOLCANO PLOTS



- First described in [6]
- X-axis is difference in LS means of $\log_2$ gene expression, a relative measure of RNA abundance
- Y-axis is $-\log_{10}$(p-value)
  - *p*-value of 1 equals 0
  - *p*-value of 0.1 equals 1
  - *p*-value of 0.01 equals 2
  - *p*-value of 0.001 equals 3
  - *p*-value of 0.0001 equals 4
- Diamonds represent one of 3931 genes
- Look for large, significant differences that occur towards upper corners

DIGIT PREFERENCE

- Compare the observed distribution of leading/trailing digits of data collected from clinical site (e.g. blood pressure) [7]
- Alternatively: Benford's Law [8]
  - Digits 1-9 occur with probability $\log_{10}\left(1+\frac{1}{d}\right)$
  - Test goodness of fit with $\sum_{i=1}^{k}\frac{(O_i-E_i)^2}{E_i}\sim\chi^2_{(k-1)}$
- Comparing digits can identify:
  - Rounding issues
  - Miscalibrated equipment
  - Protocol deviations
  - Differences in subjective interpretation
  - Duplications
- Row mean score chi-square tests [9]

DIGIT PREFERENCE

	0	1	2	3	4	5	6	7	8	9
Suspect										
Others										

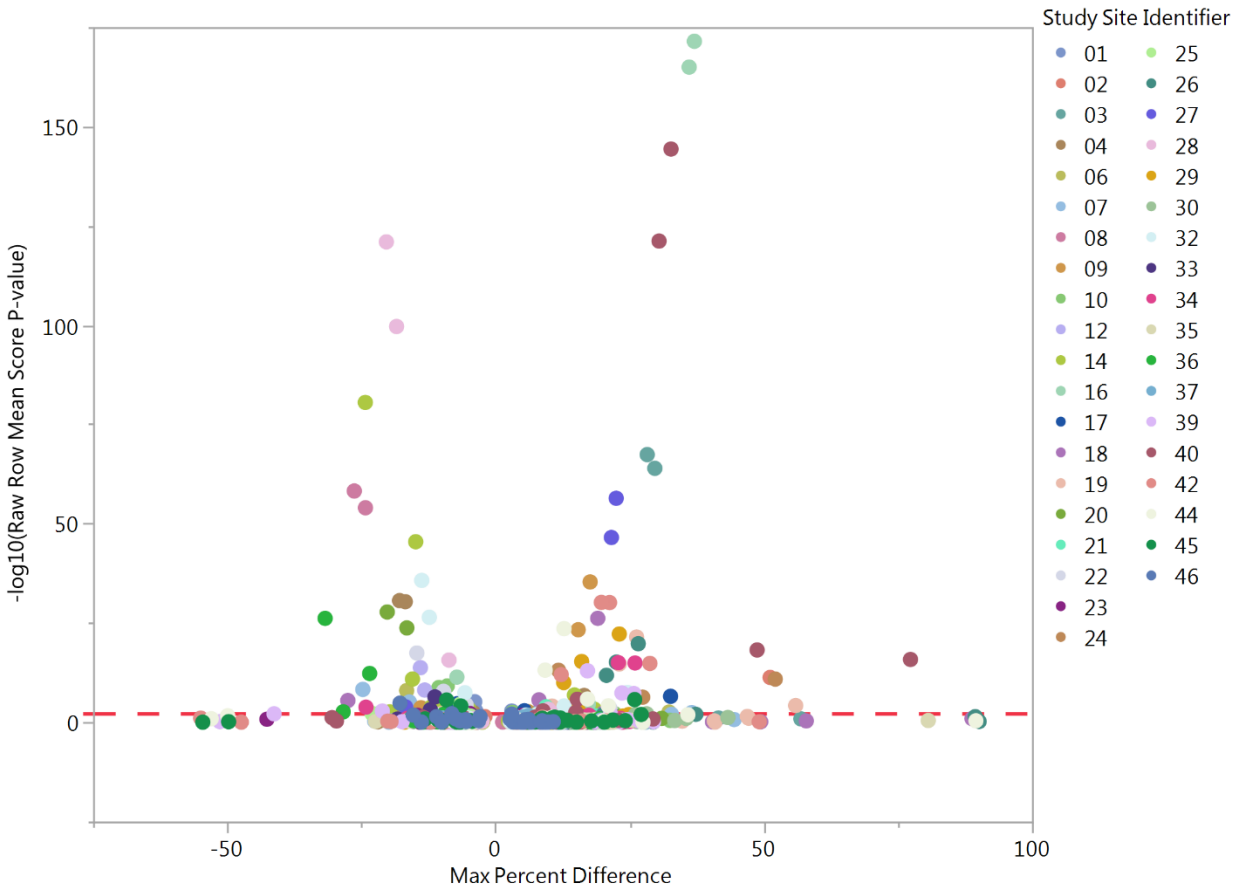
$$Q_s = \frac{(\bar{f}_1 - \bar{f}_2)^2}{\left\{\frac{1}{n_{1+}} + \frac{1}{n_{2+}}\right\}\left\{\frac{nv_a}{n-1}\right\}}$$

$$\mu_a = \sum_{j=1}^{10} \frac{a_j n_{+j}}{n} \text{ and } v_a = \sum_{j=1}^{10} \frac{(a_j - \mu_a)^2 n_{+j}}{n} \text{ and } \bar{f}_i = \sum_{j=1}^{10} \frac{a_j n_{ij}}{n_i}$$

Row mean score test takes advantage of ordinality of digit. $Q_s \sim \chi_1^2$

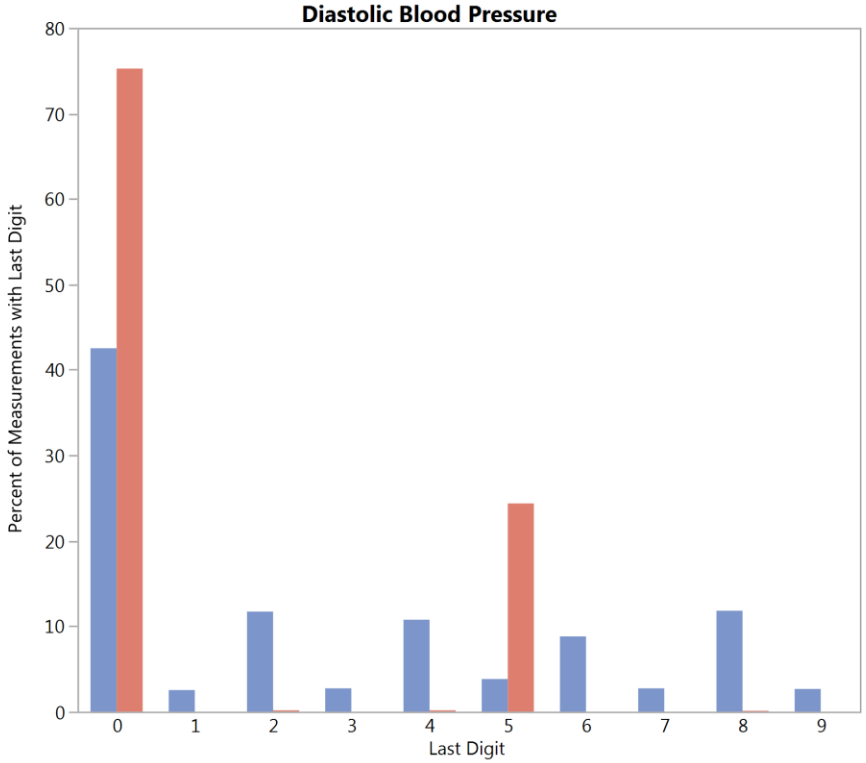
Standardized midranks: $a_j = \frac{2\left[\sum_{k=1}^{j} n_{+k}\right] - n_{+j} + 1}{2(n+1)}$,
when column values not considered equally-spaced.

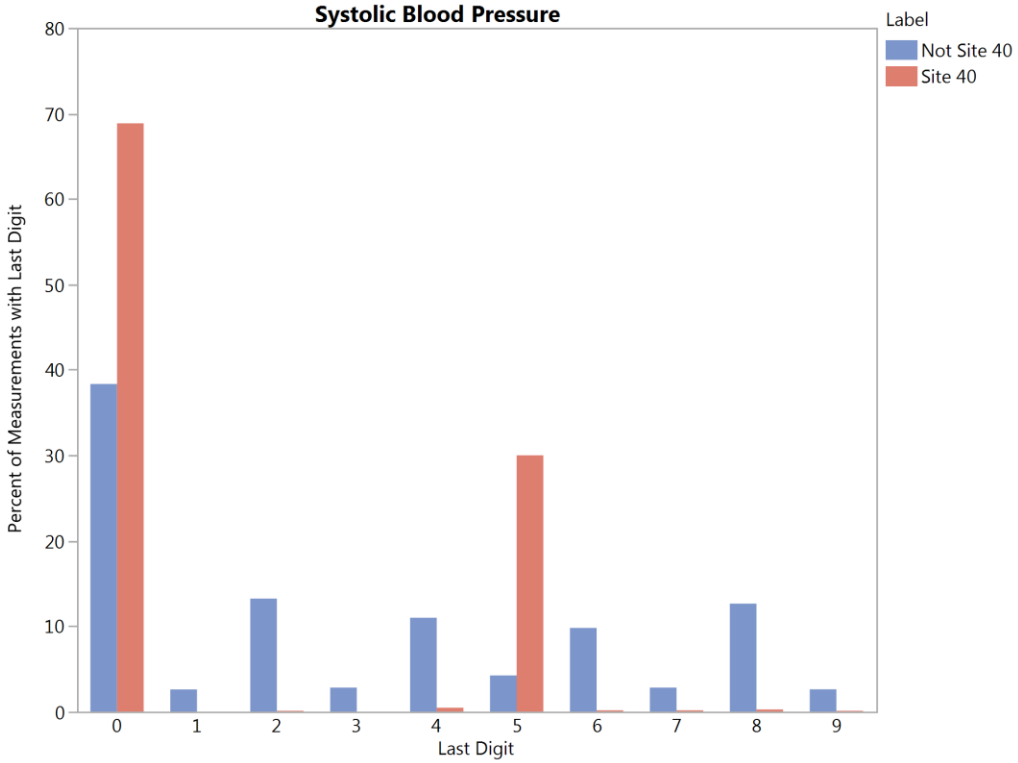
DIGIT PREFERENCE



Analysis of trailing digit preference for ECG, vital signs and laboratory measurements. Each point represents a specific test, such as systolic blood pressure, for a specific site comparing that site to all other sites as a reference. The max percent difference is the largest difference observed among all categories between the suspect site and reference.

DIGIT PREFERENCE





- Site 40 reports a 0 as the trailing digit twice as often as reference
- Rounding to 0 and 5
- Perhaps not following protocol?

PAIRWISE ASSOCIATION

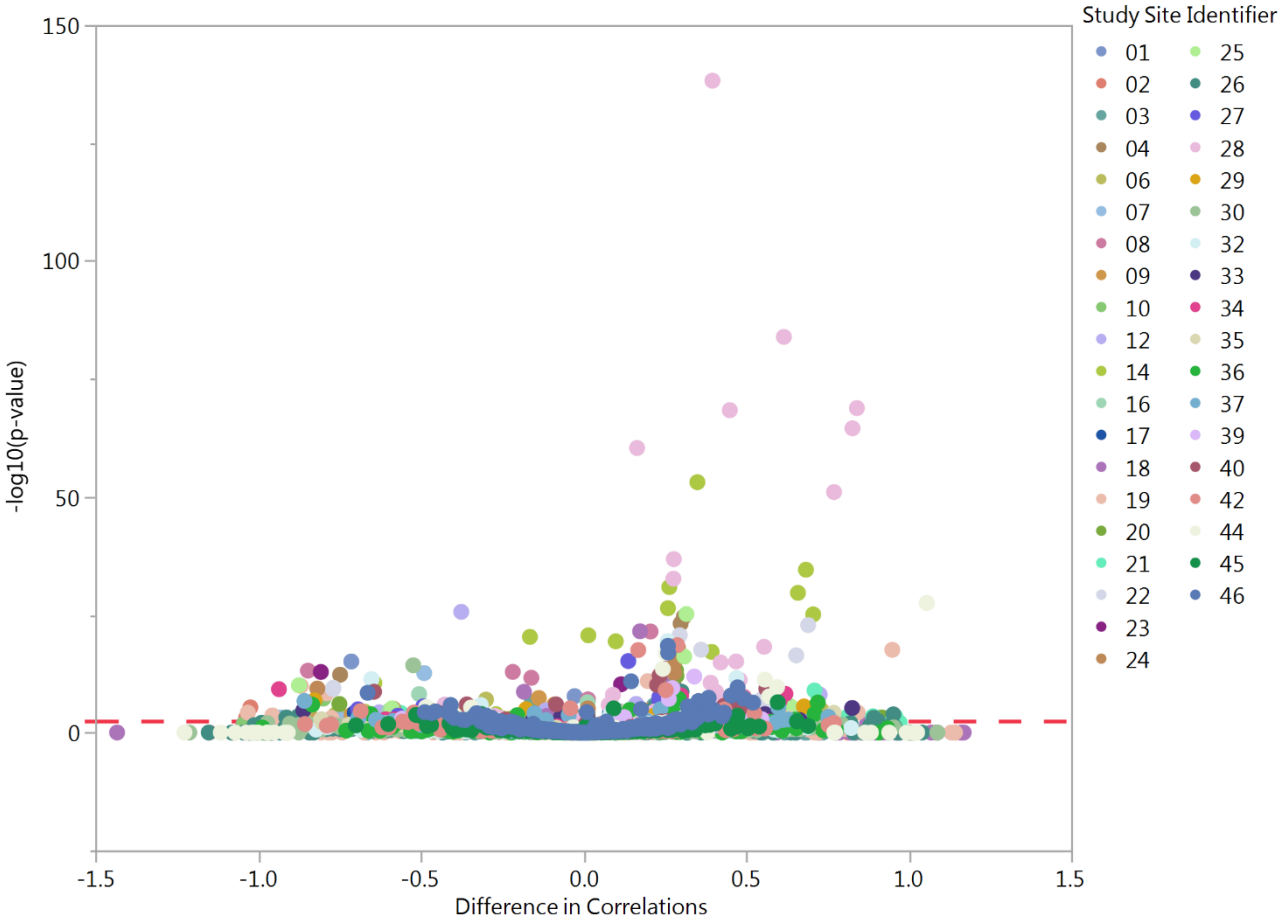
- Use Fisher's transformation of the correlation coefficient [10,11]

$$r_i^* = \frac{1}{2}\log_e\left(\frac{1+r_i}{1-r_i}\right)$$

$$Z = \frac{r_s^* - r_o^*}{\sqrt{\frac{1}{n_s - 3} + \frac{1}{n_o - 3}}}$$

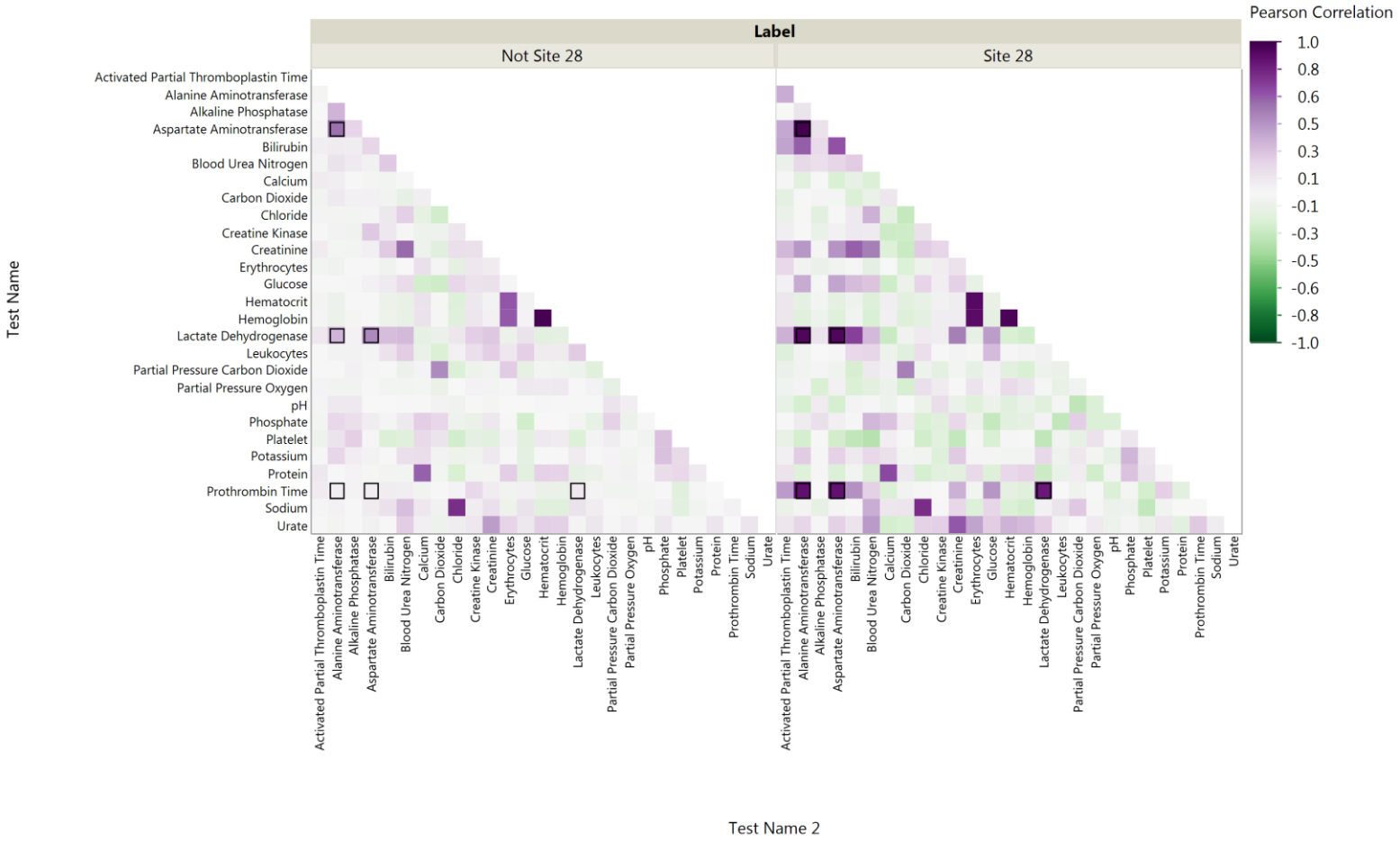
- Resampling-based analyses applied to questionnaire data [12]
- Some authors found correlation higher in fabricated data [13]
- Spearman's correlation based on the ranks
- Evaluate autocorrelation of longitudinal data using differing lags of each variable with itself

PAIRWISE ASSOCIATION



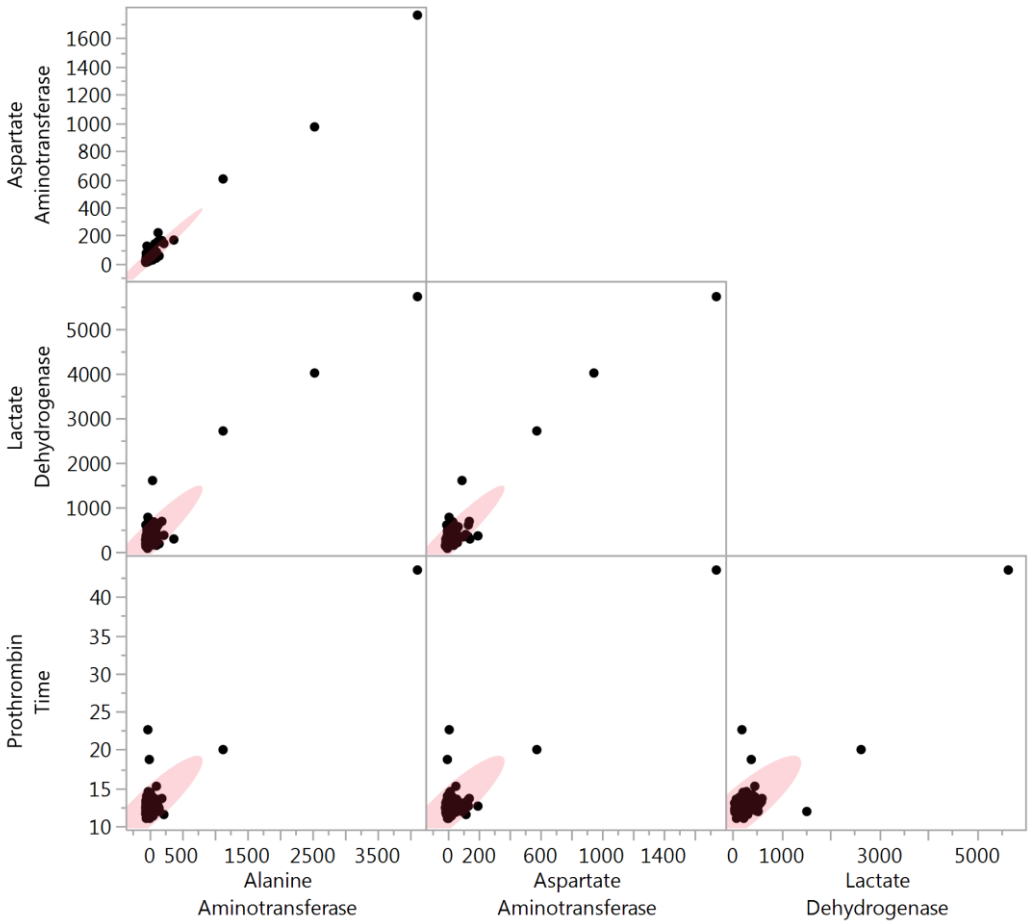
Analysis of correlations for ECG, vital signs and laboratory measurements. Each point represents two specific tests within the same Findings domain, such as systolic and diastolic blood pressure, for a specific site comparing that site to all other sites as a reference.

PAIRWISE ASSOCIATION



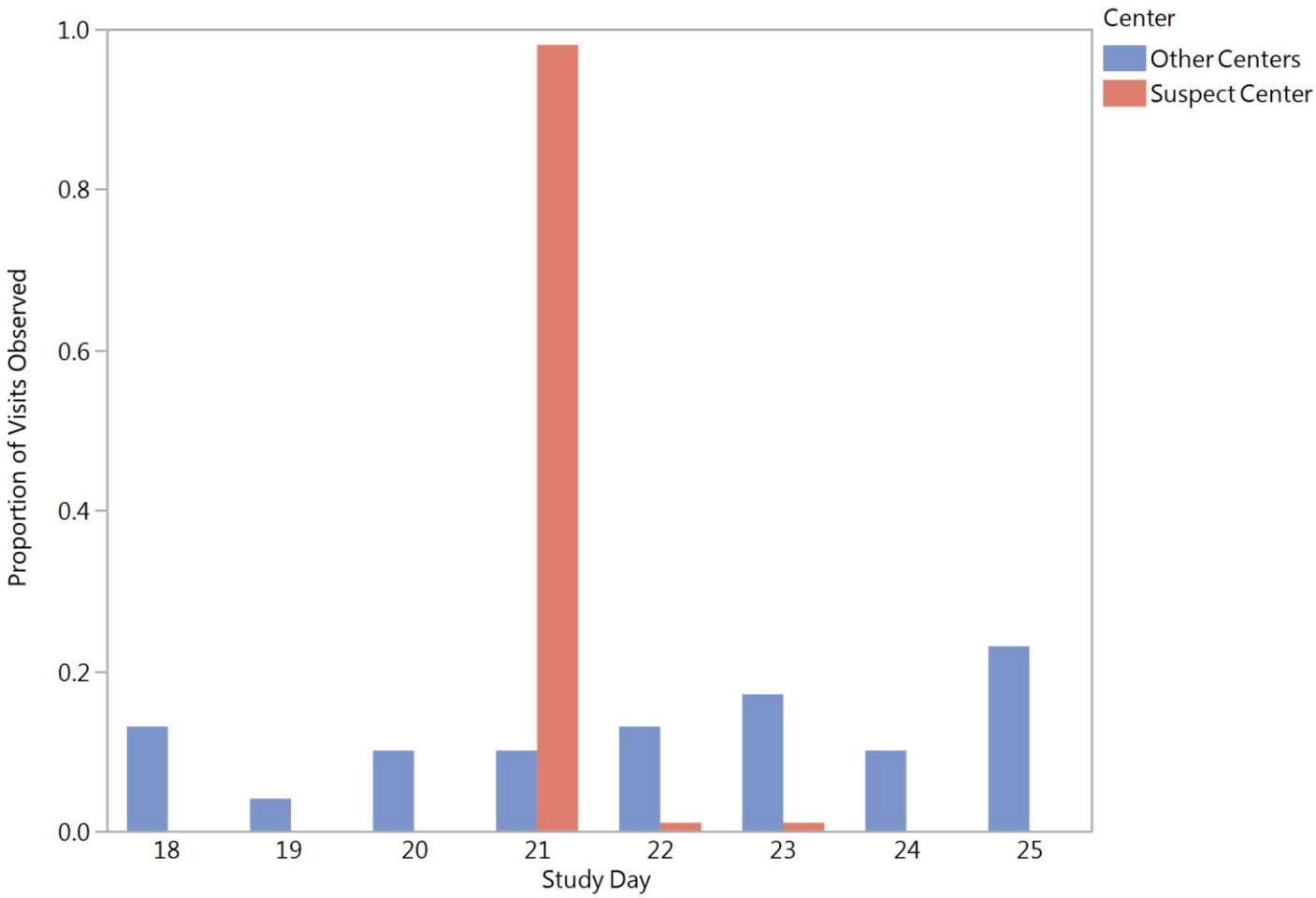
Many unusual differences in correlation patterns at site 28 for labs, particularly alanine aminotransferase (ALT), aspartate aminotransferase (AST), lactate dehydrogenase (LDH) and prothrombin time (PT).

PAIRWISE ASSOCIATION



Outliers force correlation at site 28 to near 1. Outliers all occur within one subject, who can be identified as having potential drug-induced liver injury (DILI) using Hy's law.

CLINIC VISITS – SCHEDULES



Data from [2].

CLINIC VISITS – SCHEDULES

- Scheduling that is too good to be true
- Scheduling where site is falling behind
- Compare study days of visits between centers
  - Randomization is often Day 1
  - study day = date – rand date + 1 where date ≥ rand date
  - study day = date – rand date where date < rand date
- Row mean score chi-square tests as in digit preference

	18	19	20	21	22	23	24	25
Suspect								
Others								

PATTERNS – CONSTANT RESULTS

- Search for results that do not change over the course of study
- Could check for a variance of 0 for numerical data, but need to consider categorical data as well
- Investigator or technician carrying a value forward, or are there legitimate reasons?
  - Hitting a limit of detection
  - Data collection method – consider physical examinations and adverse events
- Frequency of repeated values matters

PATTERNS – CONSTANT RESULTS

System	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5
Eyes	Normal	Normal	Normal	Normal	Normal
ENT	Normal	Normal	Normal	Normal	Normal
Neck	Normal	Normal	Normal	Normal	Normal
Cardio	Normal	Normal	Normal	Normal	Normal
Lungs	Normal	Normal	Normal	Normal	Normal
Skin	Normal	Normal	Normal	**Abnormal: Rash**	Normal
Abdomen	Normal	Normal	Normal	Normal	Normal
Musculo/Skel	Normal	Normal	Normal	Normal	Normal
Genito Urinary	Normal	Normal	Normal	Normal	Normal

Why collect this data? An AE will be entered for the abnormal or worsening body system. Adds a lot of bulk and creates a reconciliation issue

PATTERNS – DUPLICATES

- Similar to constant results, except that sets of measurements are carried forward or occur together
- Problematic within or between patients within the same site
- Data propagation [6]

Date/Time of Measurements	Diastolic Blood Pressure	Heart Rate	Systolic Blood Pressure
1988-08-20T12:30:00	70	90	120
1988-08-20T12:45:00	70	90	120
1988-08-20T13:00:00	70	90	120
1988-08-20T14:00:00	70	90	120
1988-08-20T17:00:00	70	90	120
1988-08-20T16:00:00	70	94	120
1988-08-21T00:30:00	70	80	120
1988-08-22T00:30:00	70	85	110
1988-08-22T06:00:00	70	85	110
1988-08-23T18:00:00	70	86	140

PATTERNS – DUPLICATES

Subject	Date/Time of Specimen Collection	Alkaline Phosphatase	Activated Partial Thromboplastin Time	Aspartate Aminotransferase	Bilirubin	Blood Urea Nitrogen	Calcium	Chloride	Carbon Dioxide
1	1988-01-25T18:36:00	94	23	33	0.0013332	1.071	2.32500005	105	113.048
2	1988-01-25T18:36:00	94	23	33	0.0013332	1.071	2.32500005	105	113.048

Subject	Date/Time of Specimen Collection	Creatinine	Glucose	Hematocrit	Hemoglobin	Potassium	Lactate Dehydrogenase	Partial Pressure Carbon Dioxide	pH
1	1988-01-25T18:36:00	0.0530400018	6.3825	34	11.6	4.1999998	234	5187	7.4400001
2	1988-01-25T18:36:00	0.0530400018	6.3825	34	11.6	4.1999998	234	5187	7.4400001

Subject	Date/Time of Specimen Collection	Platelet	Partial Pressure Oxygen	Protein	Prothrombin Time	Erythrocytes	Sodium	Urate	Leukocytes
1	1988-01-25T18:36:00	290	12103	6.4000001	11.9	3.45	140	0.1475	7.8000002
2	1988-01-25T18:36:00	290	12103	6.4000001	11.9	3.45	140	0.1475	7.8000002

CONCLUSIONS

- Methods to assess fraud, misconduct & data quality using **statistics** and **graphics**
- Why? Good Clinical Practice! [14]
  - Protect the patient
  - Ensure study integrity and validity of final results
- Methods are of great interest given
  - Clinical trial costs [15]
  - Cost and questionable benefit of current data review practices [16-20]
  - Cost of delay: $6-15 million USD per drug, per day [21]

A FINAL EXAMPLE

- In 2013, a scientist at a pharmaceutical services company was convicted of manipulating data for preclinical studies for an anti-cancer therapy [22,23]
- Engaged in these activities since 2003
- Efforts identified in 2009 when quality control procedures identified data irregularities, necessitating the review of hundreds of previously-conducted safety studies

REFERENCES

1. Bailey KR. (1991). Detecting fabrication of data in a multicenter collaborative animal study. *Controlled Clinical Trials* 12: 741-752.
2. Buyse M et al. (1999). The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine*, **18**: 3435-3451.
3. Evans, S. (2001). Statistical aspects of the detection of fraud. In: Lock S & Wells F, eds. *Fraud and Misconduct in Biomedical Research, Third Edition*. London, UK: BMJ Books.
4. Venet D, Doffagne E, Burzykowski T, Beckers F, Tellier Y, Genevois-Marlin E, Becker U, Bee V, Wilson V, Legrand C & Buyse M. (2012). A statistical approach to central monitoring of data quality in clinical trials. *Clinical Trials* 9: 705-713.
5. Zink RC. (2014). *Risk-Based Monitoring and Fraud Detection in Clinical Trials Using JMP® and SAS®*. Cary, NC: SAS Institute.
6. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G & Gibson G. (2001). The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster. *Nature Genetics* 29: 389-395.
7. Kirkwood AA, Cox T & Hackshaw. (2013). Application of methods for central statistical monitoring in clinical trials. *Clinical Trials* 10: 783-806.
8. Hill T.P. (1996). A statistical derivation of the significant-digit law. *Statistical Science* 10: 354-63.
9. Stokes ME, Davis CS, Koch GG. (2012). *Categorical Data Analysis Using SAS, Third Edition*. Cary, NC: SAS Institute, Inc.
10. Fisher, R.A. (1921). On the “probable error” of a coefficient of correlation deduced from a small sample. *Metron* 1: 3-32.
11. Fisher, R.A. (1970). *Statistical Methods for Research Workers, Fourteenth Edition*. Davien, CT: Hafner Publishing Company.

REFERENCES

12. Taylor, R.N., McEntegart, D.J., Stillman, E.C. (2002). Statistical techniques to detect fraud and other data irregularities in clinical questionnaire data. *Drug Information Journal* 36: 115-125.
13. Akhtar-Danesh, N., Dehghan-Ko. M. (2003). How does correlation structure differ between real and fabricated data-sets? BMC Medical Research Methodology. Available at: http://www.biomedcentral.com/1471-2288/3/18.
14. International Conference of Harmonisation. (1996). E6: Guideline for Good Clinical Practice.
15. Tufts Center for the Study of Drug Development. (2014). How the Tufts Center for the Study of Drug Development pegged the cost of a new drug at $2.6 billion. Boston: Tufts Center for the Study of Drug Development.
16. Eisenstein EL, Lemons PW, Tardiff BE, Schulman KA, Jolly MK & Califf RM. (2005). Reducing the costs of phase III cardiovascular clinical trials. *American Heart Journal* 149: 482–488.
17. Funning S, Grahnén A, Eriksson K & Kettis-Linblad A. (2009). Quality assurance within the scope of good clinical practice (GCP): what is the cost of GCP-related activities? A survey with the Swedish association of the pharmaceutical industry (LIF)'s members. *The Quality Assurance Journal* 12: 3-7.
18. Tantsyura V, Grimes I, Mitchel J, Fendt K, Sirichenko S, Waters J, Crowe J & Tardiff B. (2010). Risk-based source data verification approaches: pros and cons. *Drug Information Journal* 44: 745-756.
19. TransCelerate BioPharma Inc. (2013). Position paper: Risk-based monitoring methodology. Available at: http://transceleratebiopharmainc.com/.
20. Bakobaki JM, Rauchenberger M, Joffe N, McCormack S, Stenning S & Meredith S. (2012). The potential for central monitoring techniques to replace on-site monitoring: findings from an international multi centre clinical trial. *Clinical Trials* 9: 257-264.
21. Wells C. (2016). Statistical monitoring – It's just data cleaning, right? JMP Discovery Summit Europe.
22. Benderly BL. (2013, May 3). A prison sentence for altering data. *Science Careers*.
23. Maslen G. (2013, April 25). Scientists sent to prison for fraudulent conduct. *University World News*.