THE ROLE OF BIOSTATISTICS IN DETECTING FRAUD

Sean Chang

Senior Manager, BIOSTATISTICS

PAREXEL

PAREXEL®
YOUR JOURNEY. OUR MISSION.™

AGENDA

- Fraud – Definition, Impact, Examples
- The Role of Biostatistics in Detecting Fraud
- Recommended Approach: First Stage
- Recommended Approach: Second Stage
- Techniques for Detecting Potential Fraud
- References

FRAUD DEFINITION, IMPACT, EXAMPLES

FRAUD DEFINITION

- **Fraud is defined as intentional misconduct in clinical research, mainly data fabrication (making up data values) and data falsification (changing data values), including intentionally selective reporting of outcomes (e.g., not reporting certain adverse events), or manipulating tools and/or equipment.**

- The Food and Drug Administration (FDA) defines research misconduct as “falsification of data in proposing, designing, performing, and recording, supervising or reviewing research or in reporting research results”.

- Research misconduct does not include honest error or differences of opinion.

- It is important to understand whether there is **intent** to cheat and whether errors should be considered **non-random**.

FRAUD VS. UNINTENTIONAL DATA ERRORS

It is important to differentiate fraud from other data errors which may be found during monitoring / review including errors due to:

- "Sloppiness"
- General mistakes
- Misunderstandings
- Transcription errors
- Equipment miscalibration,
- Poor instructions in study protocol, laboratory manuals, etc.

The frequency of such fraud is estimated low in general, typically conducted by single study sites (or even single individuals within a study site) and therefore difficult to detect, in particular by statistical means. In George and Buyse article. the summary of fraud estimates ranged from 0.01% to 0.40% (one to 40 in 10000 active scientists).

Intentional patient level data fabrication can be much more difficult to detect because investigators use medical and scientific knowledge to construct clinically plausible data

THE IMPORTANCE OF DETECTING FRAUD

- **The primary effect of fraud is biased reporting.**
- **Serious secondary effects of fraud include:**
  - Rounding numbers up (e.g. 96 mg Hg) to the border values of inclusion criteria (e.g. 100 mg Hg) so that patients may enter a clinical trial may not impact the overall efficacy of the trial but may cause harm to individual patient's safety.
  - Biased reporting in a non-blinded trial may make the control group look worse and impact the overall efficacy results of the trial.
  - Missing secondary safety endpoint data which is populated with values from a previous visit in order to avoid a query from data management may also lead to incorrect endpoint analysis results.

EXAMPLES OF FRAUD

- Some or even all **subjects in a site do not exist**, they are completely invented.
- The subjects exists but time consuming or expensive **examinations are not performed** and the respective data is invented.
- Subject has been lost to follow-up or requested premature end of study – but investigator continues by **fabricating subsequent data** for this subject.
- Patient-reported outcomes **questionnaires are not completed** by the subject but by the site staff.
- Blood, urine or other types of **samples are not collected** and instead samples from other subjects (internal or external to the study) are used for lab analysis.
- **Study treatment was dispensed to other patients** than the enrolled participants for whom data was later fabricated.

SUMMARY OF PATTERNS THAT MAY INDICATE FRAUD

One variable at a time

Digit preference, Round number preference, Too few or too many outliers, Too little or too much variance, Strange peaks, Data too skewed

Several variables at a time

Multivariate inliers, Multivariate outliers, Leverage, Too weak or too strong, correlation

Repeated measurements

Interpolation, Duplicates, Invented patterns

Calendar time

Breach of randomization, Days of week (Sundays or holidays), Implausible accrual, Time trends

DATA OFTEN AFFECTED BY FRAUD

- **Eligibility:**
  - Data is pushed a little to make a patient eligible for the trial when in fact that patient does not strictly meet the criteria
- **Repeat measurements (e.g. laboratory examination):**
  - When the same measurements are requested repeatedly over time, data may be carried over (aka imputation during data collection/making observations) from the previous visit if the measurements are missing for a particular visit
- **Adverse events:**
  - When adverse events are underreported but not because of different interpretations or lack of compliance by investigators.
- **Compliance:**
  - Subjective measurements, such as the number of pills returned.
- **Patient diaries:**
  - A number of cases of data fabrication have been detected through the color and texture of the pen supposedly used on successive days by the patient, the patient's handwriting, etc.

THE ROLE OF BIOSTATISTICS IN DETECTING FRAUD

PAREXEL

THE ROLE OF BIOSTATISTICS IN DETECTING FRAUD

- Biostatistical procedures for fraud detection may be used as **screening mechanisms or for further investigation** of data that fall under fraud suspicion (for example signaled by the CRA responsible for a study site).

- In general, biostatistical techniques may **support** searching for features, patterns, or trends in the data that would be unlikely to occur in genuine data.

- As the frequency of such fraud is estimated to be low, it is recommended to consider **exploring** features, patterns or trends prior to conducting more formal procedures/test for fraud detection.

- **The process of signaling/detecting fraud is not a one-time effort but rather a continuous effort utilizing biostatistics measures, which can be exploratory or/and confirmatory, that should be started as early as possible. The team will define the time intervals of the process.**

RECOMMENDED APPROACH: FIRST STAGE

PAREXEL

RECOMMENDED APPROACH: FIRST STAGE

Exploring the Data

Using available statistical monitoring review outputs or analytics evaluate:

- ✓ **Visit timing:** Unusual visit timing where distance of actual visit date relative to the target date is either too close or too far away.
- ✓**Visit date plausibility:** Unless indication in emergent, are any visits done during holidays or on Sundays?
- ✓**Adverse event underreporting:** Is a specific site reporting lower rates of adverse events when compared to other sites?
- ✓**Standard deviation distributions:** How standard deviation of specific measures at a site is compared to study-wise standard deviation of the measures?
- ✓**Correlation:** How are measures correlated at a subject, site or study level?
- ✓**Digit preference:** Is there any digit preference observed for specific measures at subject, site or study level?

RECOMMENDED APPROACH: FIRST STAGE

Exploring the Data (continued)

Using available statistical monitoring review outputs or analytics evaluate:

- ✓Number of patients screened vs. randomized vs. treated by site.
- ✓Randomization: Check for double randomization by identical initials, age/year of birth, and gender.
- ✓Timing per the actual clock time at site of randomization.
- ✓Number of protocol violations by site.
- ✓Primary and key secondary endpoints estimates over time, particularly variability.
- ✓Incidence of treatment emergent AE's.
- ✓Incidence of serious AE's.
- ✓Incidence of AE's of special interest.
- ✓Selected lab parameters with special safety or efficacy interest.

The biostatisticians will need to make recommendations whether further efforts such as generating additional outputs or check are required.

RECOMMENDED APPROACH: SECOND STAGE

RECOMMENDED APPROACH: SECOND STAGE

Confirming signals

If any suspicious signals are found while exploring, continue to generate additional outputs and perform corresponding statistical tests to determine whether signals consistently indicate misconduct. Further analysis by study site should be done to look for deviations from expected (normal or log-normal) distribution in the full set of subjects and determine whether (almost) all of these deviations belong to the same study site.

- Stratification by randomized study treatment would usually be informative for fraud detection but is generally not possible (for double blind studies) or desired (for open-label studies with artificial blinding of central sponsor / CRO team). If needed, consider the addition of an unblinded resource to perform this review. The sponsor will have to be the ultimate decision maker if unblinding is needed.

It is strongly recommended not to do unblinded analysis of fraud. If the study has a DMC, then there may be unblinded Stats for the unblinded analysis.

RECOMMENDED APPROACH: SECOND STAGE

- Baseline variable mean and variance comparison. This may be done by t-test and Levene's test, accordingly.
- Continuous fraud data often has a smaller variance than seen with real data – this could be signaled by plots of site-specific standard deviation (around site-specific means).
- If there are repeated measurements over time, then fraud data often tends to show a decreasing variability over time whereas the opposite is expected with genuine data.
- Two variables can be examined for fraud simultaneously by comparing the observed versus the expected correlation (or the correlation observed in other study sites).
- Digit preference at a site can be compared against digit variance at other sites, or tested against Benford's law. A graph that overlays plots of actual proportion and expected proportion along with lower and upper bounds across (first) digit can also be presented.

TECHNIQUES FOR DETECTING POTENTIAL FRAUD

OVERVIEW OF BIOSTATISTICAL TECHNIQUES THAT MAY BE USED TO DETECT FRAUD

One variable at a time:

Descriptive statistics, Box and whisker plot, Frequency histogram, Stem and leaf plots, Tests for slippage

Several variables at a time:

Cross-tabulation/scatterplot, Correlation/regression, Cook's distance, Mahalanobis' distance, Cluster analysis, Discriminant analysis, Chernoff faces, Star (needle, spike) plots, Hotelling's *T* 2, Tests for treatment contrasts

Repeated measurements:

Autocorrelations, Profiles, Polynomial contrasts, Runs tests

Calendar time:

Residual plots, Cusum, Control charts

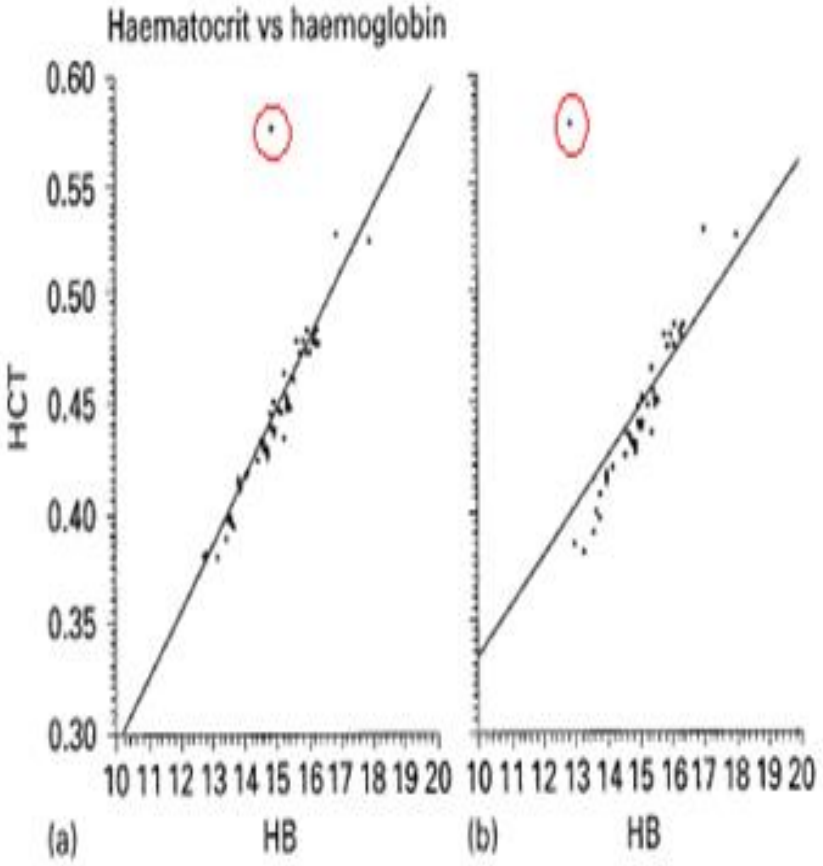
EXAMPLE 1: DETECTING DIGIT/ROUND NUMBER PREFERENCE BY STEM AND LEAF PLOT

Diastolic Blood Pressure (DBP) made to the nearest 2 mmHg shows 21 values ending in zero.

```
Stem Leaf                    #
  10 0                       1
   9
   9 0                       1
   8
   8 000                     3
   7 68                      2
   7 0004                    4
   6
   6 000224                  6
   5 8                       1
     ----+----+----+----+
 Multiply Stem.Leaf by 10**+1
```


EXAMPLE 2: DETECTING INFLUENTIAL OBSERVATION (OUTLIER FROM X-DIRECTION) BY COOK'S DISTANCE

Lock, Wells and Farthing (2001): A bivariate analysis where HCT as dependent and Hb as an independent variable to identify influential observations (in red circle - far distance from x-direction) measured by Cook's distance.

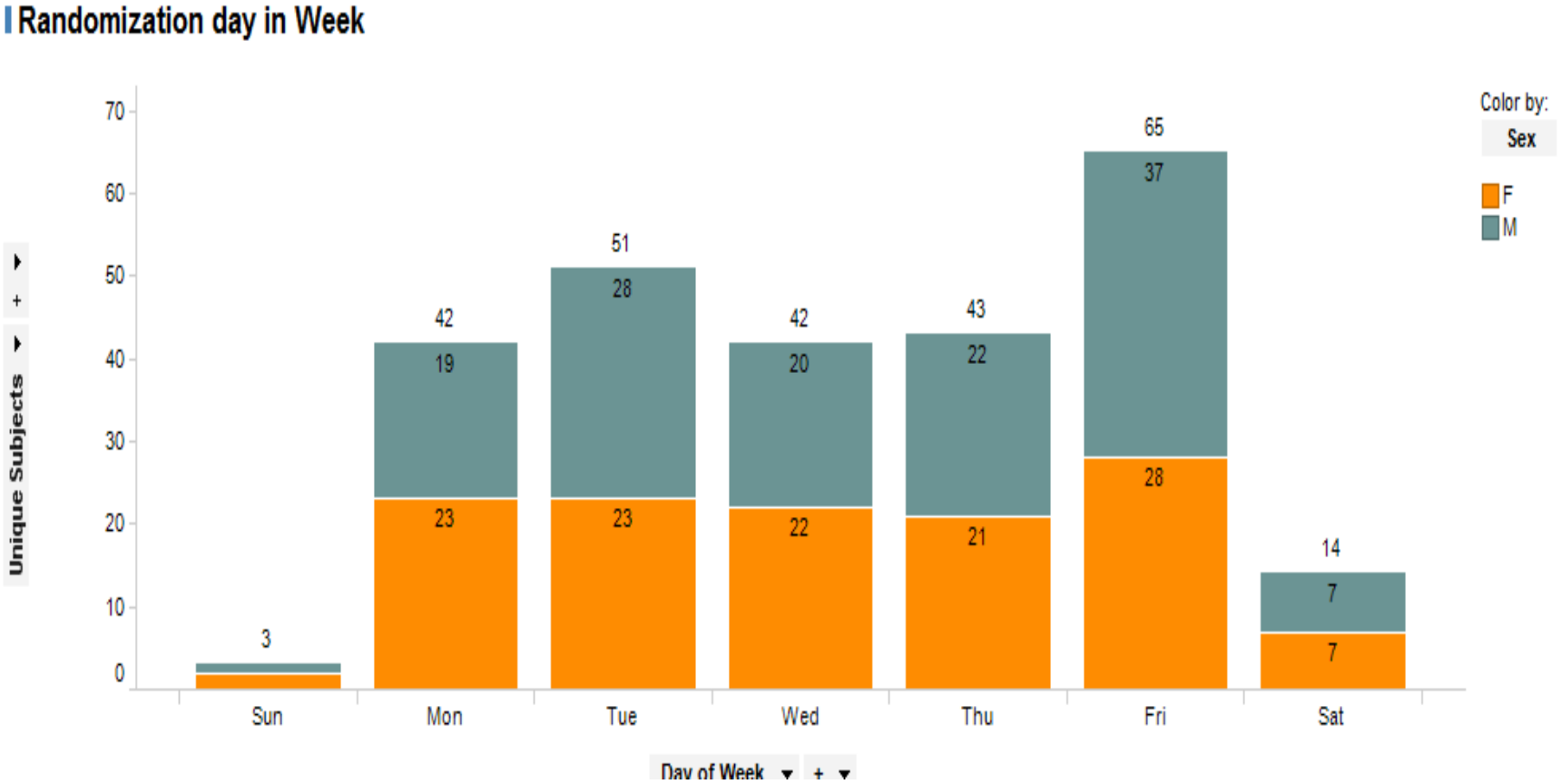


EXAMPLE 3: DETECTING PATTERNS IN REPEATED MEASUREMENT BY RUNS TEST

Test for randomness of observations using data from Example 1 assuming DBPs taken from same patient.

Z	Pr > |Z|
-4.13045	<.0001

EXAMPLE 4: OBSERVING DAYS OF WEEK BY CHART USING PMADS



Randomization on weekends are detected.

TOKYO, JAPAN

REFERENCES

REFERENCES

- Al-Marzouki S., Evans S., Marshal T., and Roberts I. *Are these data real? Statistical methods for the detection of data fabrication in clinical trials*. BMJ 2005; 331; 267-270.

- Buyse, M. et. al. *The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials*. Statistics in Medicine 1999; 18; 3435-3451.

- George, S. and Buyse, M. 2015. *Data fraud in clinical trials*. Clin. Invest. (Lond.) (2015) 5(2), 161–173

- Lock S., Wells F., and Farthing M. Fraud and Misconduct in Biomedical Research. 3rd Edition. 2001. BMJ Books.

- Knepper, D., et. al. *Statistical monitoring in clinical trials: best practices for detecting data anomalies suggestive of fabrication or misconduct*. Therapeutic Innovation & Regulatory science 2016, Vol. 50(2), 144-154.

- Winn, T. Implementing Digital Data Analysis using SAS. 2001 SESUG proceeding.

THANK YOU!

PAREXEL