Statistical Monitoring Applied to Randomized Trials

Tomasz Burzykowski, PhD
IDDI, Louvain-la-Neuve
I-BioStat, Hasselt University
Belgium

Developing effective quality systems in clinical trials:
An enlightened approach - 13-14 October 2010, Washington DC

Outline

1. Science vs. regulation in clinical trials
2. The continuum from errors to fraud
3. Fraud – prevalence, impact and detection
4. Monitoring strategies
5. SMART project
6. Conclusions

Scientific *vs.* regulatory requirements for a clinical trial

From a *scientific* point of view, a trial must estimate the effect of a treatment without bias.

Randomized trials enable such unbiased inference even in the presence of massive random errors which only cause conservatism (in tests for superiority).

From a *regulatory* point of view, a trial must provide verifiable evidence that it was carried out according to specifications.

Absence of errors must be demonstrated regardless of their consequences.

The continuum from errors to fraud

Type	Typical examples	Intent
Errors	Poorly calibrated equipment	Wholly unintentional
Sloppiness	Data missing or incorrectly copied from source documents	Limited awareness
Fraud	Data fabricated to avoid missing data or create patients	Deliberate
Treatment -related fraud	Data fabricated or falsified to favor treatment	Definite « intention to cheat »

The continuum from errors to fraud

Type	Typical examples	Impact
Errors	Poorly calibrated equipment	Potential (small) loss in power / no bias
Sloppiness	Data missing or incorrectly copied from source documents	Potential (small) loss in power / no bias
Fraud	Data fabricated to avoid missing data or create patients	Unknown effect on power / no bias
Treatment -related fraud	Data fabricated or falsified to favor treatment	Definite bias

The continuum from errors to fraud

Type	Typical examples	Ease of detection
Errors	Poorly calibrated equipment	Difficult to detect
Sloppiness	Data missing or incorrectly copied from source documents	May be hard to detect
Fraud	Data fabricated to avoid missing data or create patients	Detectable through center comparisons
Treatment-related fraud	Data fabricated or falsified to favor treatment	Detectable through treatment by center comparisons

STATISTICS IN MEDICINE
Statist. Med. **18**, 3435–3451 (1999)

THE ROLE OF BIOSTATISTICS IN THE PREVENTION, DETECTION AND TREATMENT OF FRAUD IN CLINICAL TRIALS†

MARC BUYSE[1]*, STEPHEN L. GEORGE[2], STEPHEN EVANS[3], NANCY L. GELLER[4], JONAS RANSTAM[5], BRUNO SCHERRER[6], EMMANUEL LESAFFRE[7], GORDON MURRAY[8], LUTZ EDLER[9], JANE HUTTON[10], THEODORE COLTON[11], PETER LACHENBRUCH[12] AND BABU L. VERMA[13]

for the
ISCB SUBCOMMITTEE ON FRAUD

Fraud

Prevalence

- Treatment-related fraud probably rare

Impact

- No measurable impact on multicenter trial results

Prevention

- Sound design and realistic data collection

Detection

- Statistical approach preferable to site visits

Ref: *Buyse et al, Statistics in Medicine 1999;18:3435*

The effect of scientific misconduct on the results of clinical trials: A Delphi survey

Sanaa Al-Marzouki*, Ian Roberts, Tom Marshall, Stephen Evans

Types of misconduct	Percentage indicating likely or very likely to distort results	Likelihood to occur (%)				
		Very unlikely				Very likely
		1	2	3	4	5
Conduct						
Tampering with treatment packs so as to un-blind allocation	95	17	75	4	4	0
Selective withdrawals on basis of knowledge of allocation	92	8	52	28	12	0
Data falsification	92	64	32	4	0	0
Data fabrication	92	72	24	4	0	0
Treatment recognition in blinded trials	64	4	36	36	24	0
Post-hoc changes in protocol	52	0	20	56	20	4

Refs: *Al Marzouki , Contemp Clin Trials 2005;26:331*

Types of misconduct	Indicating likely or very likely to occur (%)
Over-interpretation of 'significant' findings in small trials	83
Selective reporting based on *p*-values	80
Selective reporting of outcomes in the abstract	76
Subgroup analyses done without interaction tests	75
Negative or detrimental studies not published	68
Putting undue stress on results from subgroup analysis	68
Inappropriate subgroup analyses	64
Selective reporting of (i) subgroups (ii) outcomes (iii) time points	64
Selective reporting of positive results or omission of adverse events data	60
Failure to report results or long delay in reporting	60
Post-hoc analysis not admitted	59
Giving incomplete information about analyses with non significant results	56
Analysis conducted by the sponsor of the trial	54

Refs: *Al Marzouki , Contemp Clin Trials 2005;26:331*

Prevalence of fraud

"Conclusions: According to this expert group, the most important forms of scientific misconduct in clinical trials are selective reporting and the opportunistic use of the play of chance. Data fabrication and falsification were not rated highly because it was considered that these were unlikely to occur."

Refs: *Al Marzouki , Contemp Clin Trials 2005;26:331*

Impact of fraud

Most cases of fraud have little impact on trial results because:

- they introduce random but not systematic errors (*i.e.* noise but no bias) in the analyses
- they affect secondary variables (*e.g.* eligibility criteria)
- their magnitude is too small to have an influence (*e.g.* fraud limited to a single site and/or few patients)

Typical example:

Roger Poisson and the NSABP trial in early breast cancer

Refs: *Altman, Practical Statistics for Medical Research 1991*
Peto et al, Controlled Clin Trials 1997;18:1

Impact of fraud

Some cases of fraud have a devastating impact on trial results because:

- they introduce systematic errors (*i.e.* bias aimed at showing a treatment benefit)
- they affect the primary outcome (*e.g.* survival)
- their magnitude is large (*e.g.* in a single site study)

Typical example:

Werner Bezwoda and high dose chemotherapy for advanced breast cancer

Ref: *Weiss et al, Lancet 2000;355:9999*

Monitoring strategies

Extensive monitoring

- 100% SDV for primary and key secondary outcomes

Reduced monitoring

- Random sampling of centers / patients / outcomes to ensure rate of errors < x%
- Risk-adapted monitoring

Targeted monitoring

- Monitoring based on Key Risk Indicators
- Statistical Monitoring

Monitoring strategies

Extensive monitoring

- 100% SDV for primary and key secondary outcomes

Reduced monitoring

- Random sampling of centers / patients / outcomes to ensure rate of errors < x%
- Risk-adapted monitoring

Targeted monitoring

- Monitoring based on Key Risk Indicators
- Statistical Monitoring

Extensive monitoring

”(...) trial management procedures ensuring validity and reliability of the results are vastly more important than absence of clerical errors.

Yet, it is clerical inconsistencies referred to as ’errors’ that are chased by the growing GCP-departments.”

Refs: *Lörstad, ISCB-27, Geneva, August 28-31, 2006*

Lancet 2005; 366: 172–74

Viewpoint

The Good Clinical Practice guideline: a bronze standard for clinical research

David A Grimes, David Hubacher, Kavita Nanda, Kenneth F Schulz, David Moher, Douglas G Altman

« Monitoring confirms consistency between data collection forms and source documents; if the source documents are wrong because of laboratory, clinical, or clerical errors, then monitoring adds expense without benefit. A common misinterpretation of sponsors is that GCP requires audits of 100% of data; by contrast, random audits might suffice. »

The ICH-GCP guidelines are valuable regarding the technical standards and ethical oversight of clinical trials.[57] However, certain guidelines, such as the one indicating that sponsors should ensure that trials are "adequately monitored," are subject to interpretation and are only as effective as the degree to which they are implemented. The solution is not simple; different types of trials require different monitoring procedures. A rigid set of rules will not suffice and may even impair the quality of the research[23,45,62]; instead, a vast improvement in the quality of clinical research is needed, so that trial procedures match the research goals and societal needs.

Ref: Glickman et al, NEJM 2009;360:816.

Monitoring strategies

Extensive monitoring

- 100% SDV for primary and key secondary outcomes

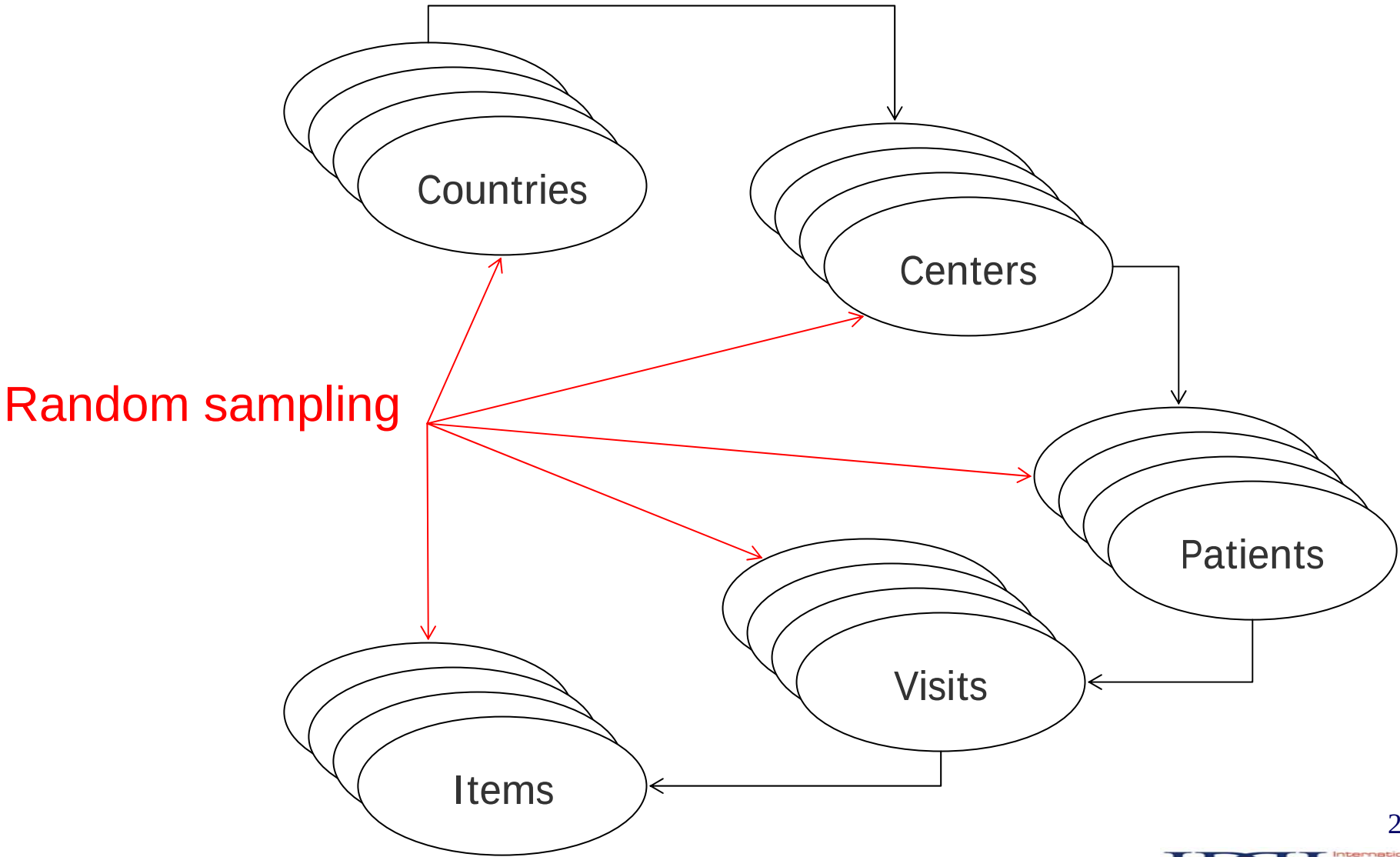
Reduced monitoring

- Random sampling of centers / patients / outcomes to ensure rate of errors < x%
- Risk-adapted monitoring

Targeted monitoring

- Monitoring based on Key Risk Indicators
- Statistical Monitoring

Monitoring with random sampling



Thérapie 2006 Juin-Juillet; 61 (4): 1–7
DOI: 10.2515/therapie:2006053

Adaptation of the Application of Good Clinical Practice Depending on the Features of Specific Research Projects

*Pierre-Henri Bertoye,[1] Soizic Courcier-Duplantier[2], Nicolas Best[3] and participants of Round Table N°2 Giens XXI**

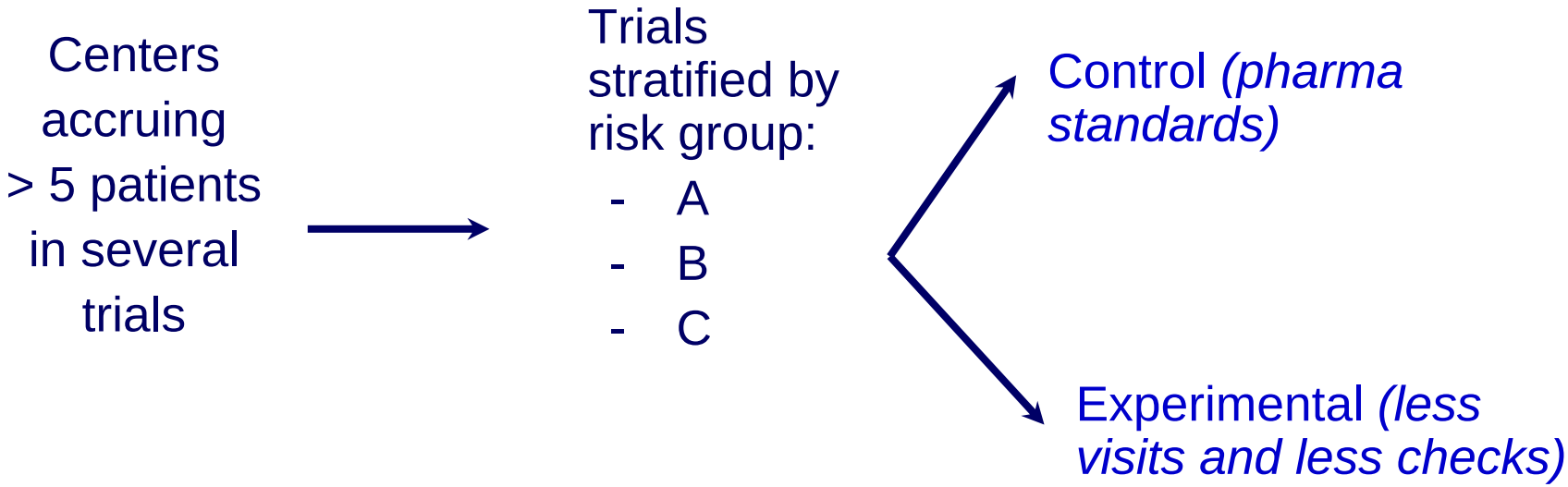
Risk A – Negligible risk (non invasive procedures)

Risk B – Risk similar to that of usual care (trials involving approved drugs)

Risk C – High risk (phase III trials of new agents, new indications or at risk populations)

Risk D – Very high risk (phase I or II trials of new agents)

OPTIMON: OPTimisation of MONitoring for clinical research studies



Goal: non-inferiority of the proportion of patients with at least one severe error in informed consent, suspected unexpected severe adverse events reports, major eligibility criteria, or primary endpoint (expected: 95% with non-inferiority margin of 5%).

Source: Geneviève Chêne, University Teaching Hospital Bordeaux

https://ssl2.isped.u-bordeaux2.fr/optimon/Documents.aspx

Monitoring strategies

Extensive monitoring

- 100% SDV for primary and key secondary outcomes

Reduced monitoring

- Random sampling of centers / patients / outcomes to ensure rate of errors < x%
- Risk-adapted monitoring

Targeted monitoring

- Monitoring based on Key Risk Indicators
- Statistical Monitoring

Targeted monitoring

Trial management committees should consider central statistical monitoring a key aspect of trial monitoring. The systematic application of this approach would be likely to lead to tangible benefits, and resources that are currently wasted on inefficient on-site monitoring could be diverted to increasing sample sizes or conducting more trials.

(…) central statistical monitoring of case report forms is remarkably efficient in the detection of fabricated data, whereas on-site monitoring may fail to uncover fraud.

Ref: Baigent et al, Clinical Trials 2008;5:49.

Potential reductions in clinical trial costs

Assumptions:

- Treatment of chronic disease
- 20,000 patients
- 1,000 sites
- 48 months enrollment (24) + follow-up (24)
- 24 visits per site (every other month)
- 60-page CRF
- 10,000 $ per patient site

Total budget in millions of $:	421
• Coordinating Center	170 (40%)
• Site payments	200 (48%)
• Other costs: travel, meetings, etc	51 (12%)

Ref: Eisenstein et al, Clinical Trials 2008;5:75.

Potential reductions in clinical trial costs

Trial components	Savings (in $ millions)	% of budget
Planning: 6 months → 4 months	1.7	0.4%
Enrollment: 24 months → 18 months	6.7	1.6%
CRF length: 60 pages → 20 pages	14.7	3.5%
Number of sites: 1000 sites → 750 sites	35.6	8.4%
Data collection: paper CRF → EDC	41.3	9.8%
Site management *	89.0	21.1%
All implemented	**149.1**	**35.4%**
All implemented, 50% of site payment	**250.1**	**59.3%**

* Initial evaluation visits: 500 sites → 100 sites
Site visits during trial: 24 per site → 4 per site
Close-out visits: 1000 sites → 0
Source document verification: 100% → 10%

Potential reductions in clinical trial costs

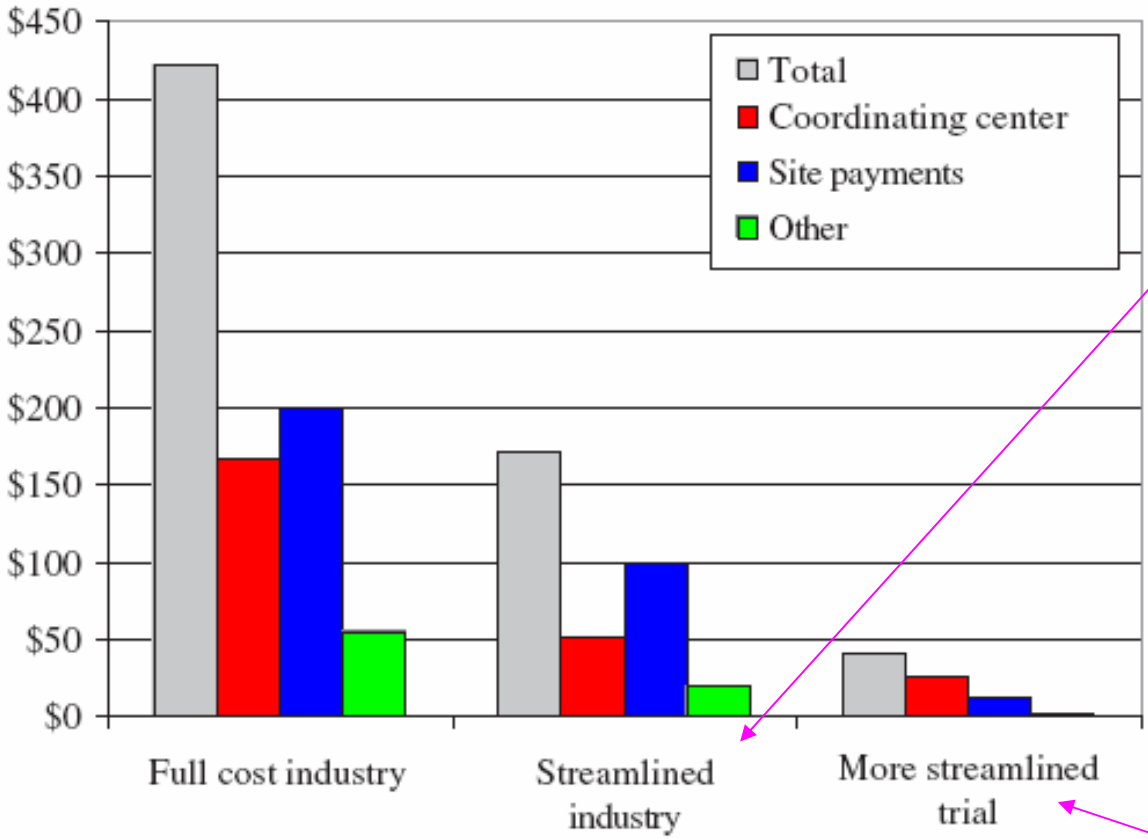


Figure 4 Cost comparison: full cost pharmaceutical industry, streamlined pharmaceutical industry, and more streamlined trial models. $ in US millions

- 4 mths planning
- 18 mths accrual
- 750 sites
- 4 site visits
- 20-page CRF + EDC
- 5,000 $ per site

- 4 mths planning
- 18 mths accrual
- 100 sites
- no site visits
- 5-page CRF + EDC
- 650 $ per site

Ref: Eisenstein et al, Clinical Trials 2008;5:75.

BMJ

RESEARCH

Carbon cost of pragmatic randomised controlled trials: retrospective analysis of sample of trials

Katy Lyle, research fellow,[1] Louise Dent, acting principal research fellow,[1] Sally Bailey, senior programme manager,[1] Lynn Kerridge, chief executive officer,[1] Ian Roberts, professor of epidemiology and public health,[2] Ruairidh Milne, director of strategy and development[1]

Conclusions CO_2 emissions from pragmatic randomised controlled trials are generated in areas where steps could be taken to reduce them. A large proportion of the CO_2 emissions come from travel related to various aspects of a trial.

Examples of “Key Risk Indicators”

Study conduct

- Actual accrual vs. target
- % pts with protocol violations
- % dropouts
- …

Safety

- AE rate
- AE grade 3/4 rate
- SAE rate
- …

Treatment compliance

- % dose reductions
- % dose delays
- Reasons for Rx stops
- …

Data management

- Overdue forms
- Query rate
- Query resolution time
- …

Principles behind statistical checks

- Humans are poor random number generators
  → *check randomness*
  (e.g. Benford's law on first digit, digit preference, etc.)

- Plausible data are hard to fabricate
  → *check plausibility*
  (e.g. correlation structure, outliers, inliers, dates, etc.)

- Clinical trial data are highly structured
  → check *comparability*
  (e.g. between centers, treatment arms, etc.)

fraud detection in biomedical research *Some patterns that may reveal fraud in clinical trial data*

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 randomisation Days of week (Sundays or holidays) Implausible accrual Time trends

Ref: Encyclopaedic Companion to Medical Statistics (Everitt B, Palmer C, Eds..) Arnold Publishers Ltd, London, 2010.

SMART*

A software that utilizes SAS macros to systematically perform a large battery of statistical tests on the values of all variables collected in a clinical trial. These tests generate p-values, ranks and other statistics that are kept in a database for checks of randomness, plausibility and comparability.

* Statistical Monitoring Applied to Randomized Trials

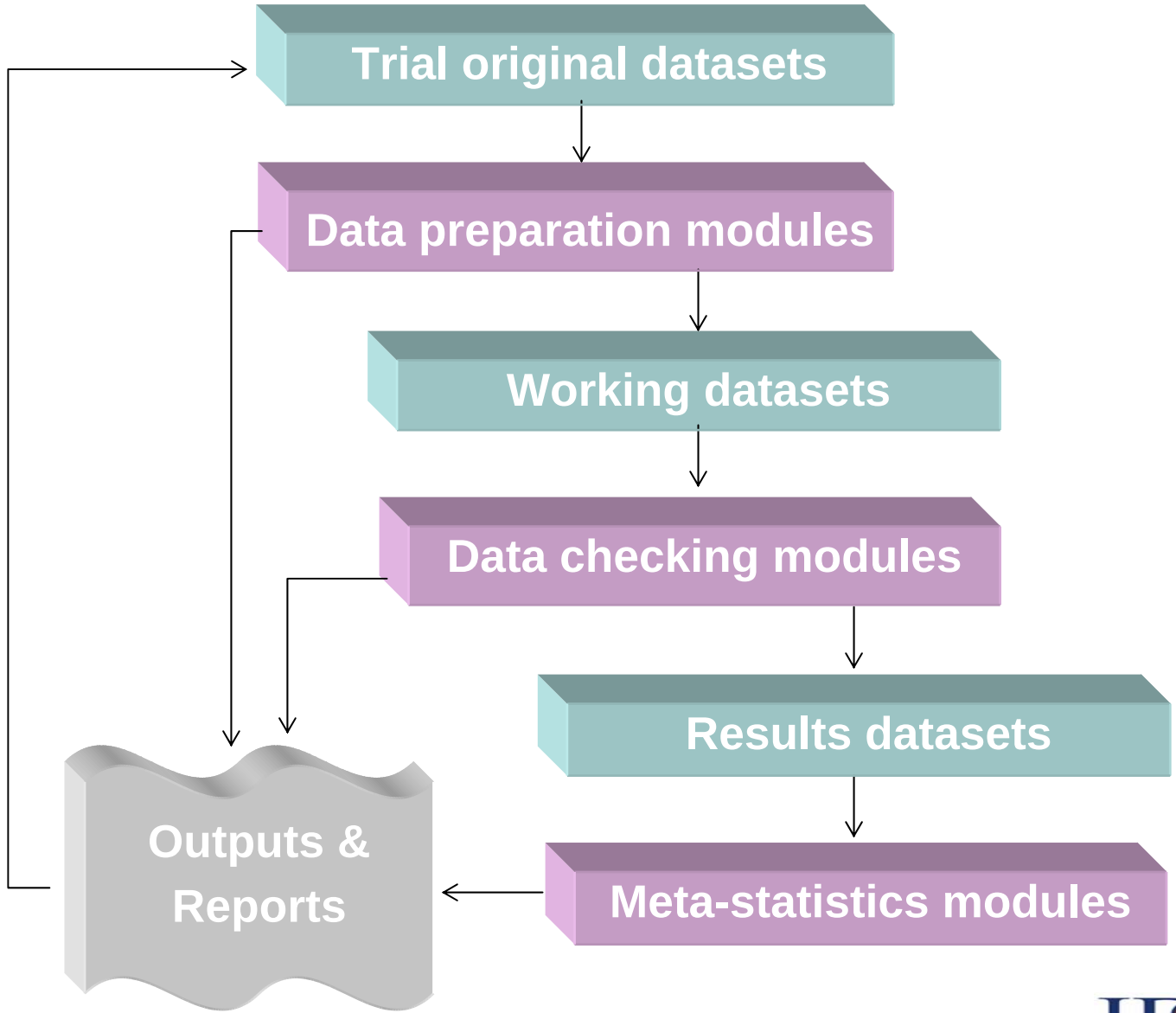
Non-model-based approach

- In multicentric trials, the distribution of all variables can be compared between each center and all others, through
  - χ^2 statistics for discrete variables
  - t-test to compare means of continuous variables
  - F-test to compare variances
  - mutivariate test statistics for more than one variable
  - etc.

Brute force approach

- These tests can be applied automatically, without regard to meaning or plausibility
- They yield very large number of center-specific statistics
- Meta-statistics can be applied to these statistics to identify outlying centers

SMART architecture



Exemplary findings: weight

- Had to be measured at a couple of visits
- Analysis of variance per center:
  - One center reported the same weight for 12/15 pts.
  - Similar, low variability pattern noted for several other centers
  - In total, no variation at all for 11% of all pts.

Exemplary findings: heart rate/blood pressure

- To be taken at each visit, in two positions (supine/standing)
- As for weight: too low / too high variation for several centers
- A particular patient:

VISIT	POS	HR	SYSBP	DIABP
1	1	72	115	75
1	2	70	110	70
2	1	72	115	75
2	2	70	110	70
3	1	70	110	75
3	2	70	110	70
4	1	72	110	75
4	2	70	105	70
5	1	74	115	75
5	2	72	110	70
...	...	...	...	...

Is it worth asking for tedious, but not essential, measurements?

Would not it be better to trim a long CRF?

Conclusions

- Current clinical research practices (such as intensive monitoring and 100% source data verification) are not useful, effective, or sustainable
- A statistical approach to quality assurance could yield huge cost savings and yet increase the reliability of the trial results
- Regulatory constraints should evolve accordingly