universiteit hasselt

I-BioStat Interuniversity Institute for Biostatistics and statistical Bioinformatics

Central Statistical Monitoring of Clinical Trials

Marc Buyse, ScD

Intternational Hexa-Symposium
November 14-15, 2013
Diepenbeek, Belgium

Science
AAAS

Embedding statistics in science and society will pave the route to a data informed future, and statisticians must lead this charge.

Marie Davidian and Thomas A. Louis

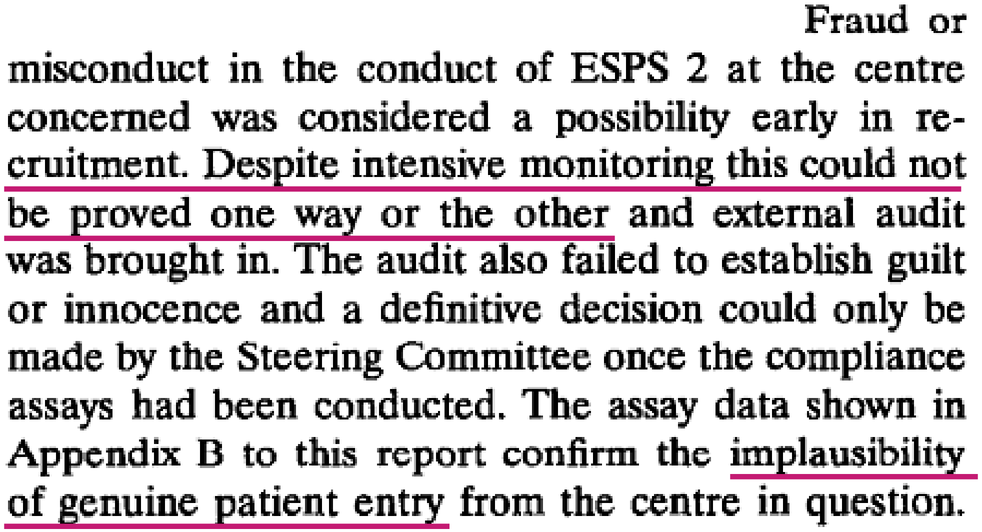
6 APRIL 2012 VOL 336 SCIENCE www.sciencemag.org

Statistical detection of fraud

The second European Stroke Prevention Study (ESPS2, 1997) accrued 7,040 patients, of which 438 (!) were fabricated using historical data at one center

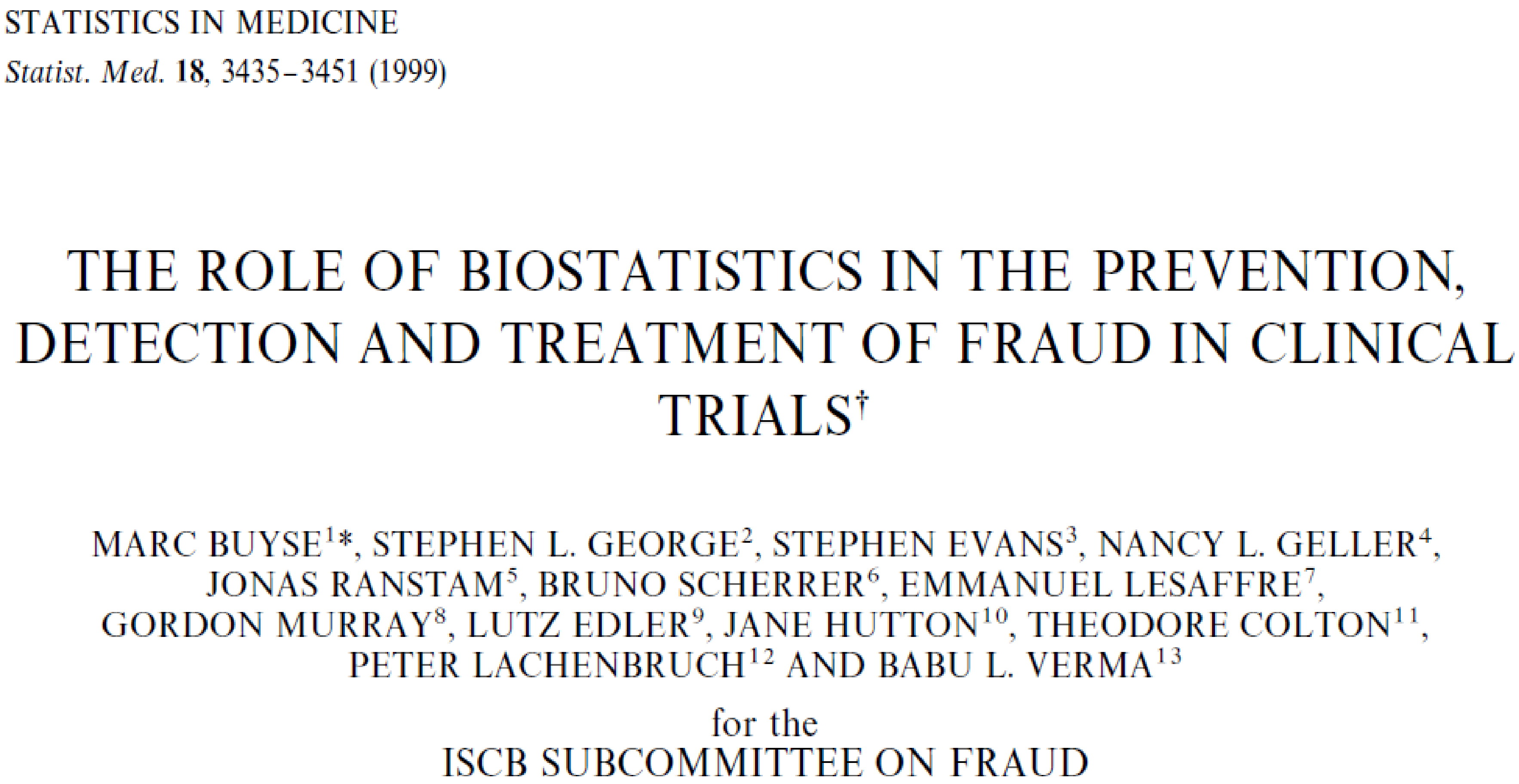
ELSEVIER

JOURNAL OF THE NEUROLOGICAL SCIENCES



Fraud or misconduct in the conduct of ESPS 2 at the centre concerned was considered a possibility early in recruitment. Despite intensive monitoring this could not be proved one way or the other and external audit was brought in. The audit also failed to establish guilt or innocence and a definitive decision could only be made by the Steering Committee once the compliance assays had been conducted. The assay data shown in Appendix B to this report confirm the implausibility of genuine patient entry from the centre in question.

Statistical detection of fraud



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

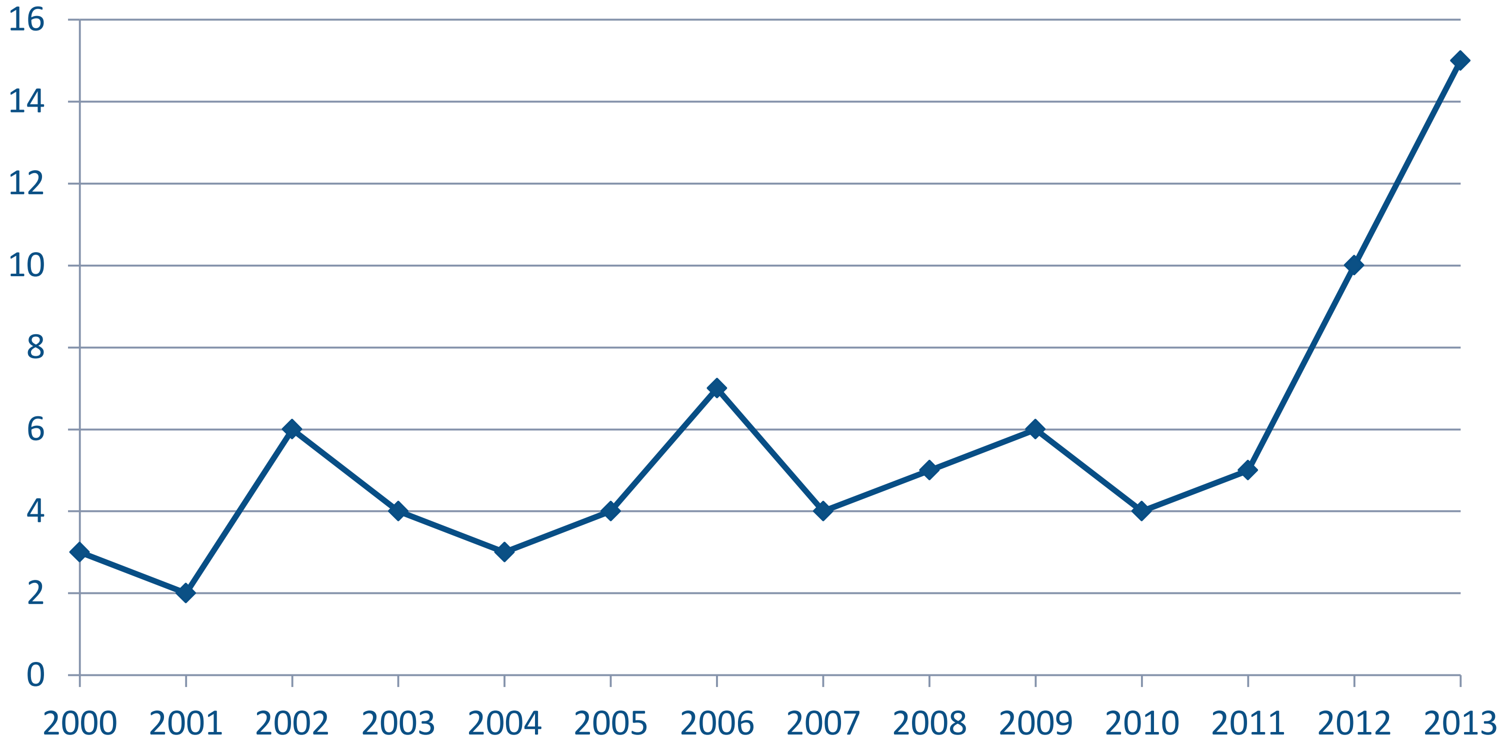
THE ROLE OF BIOSTATISTICS IN THE PREVENTION, DETECTION AND TREATMENT OF FRAUD IN CLINICAL TRIALS[†]

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

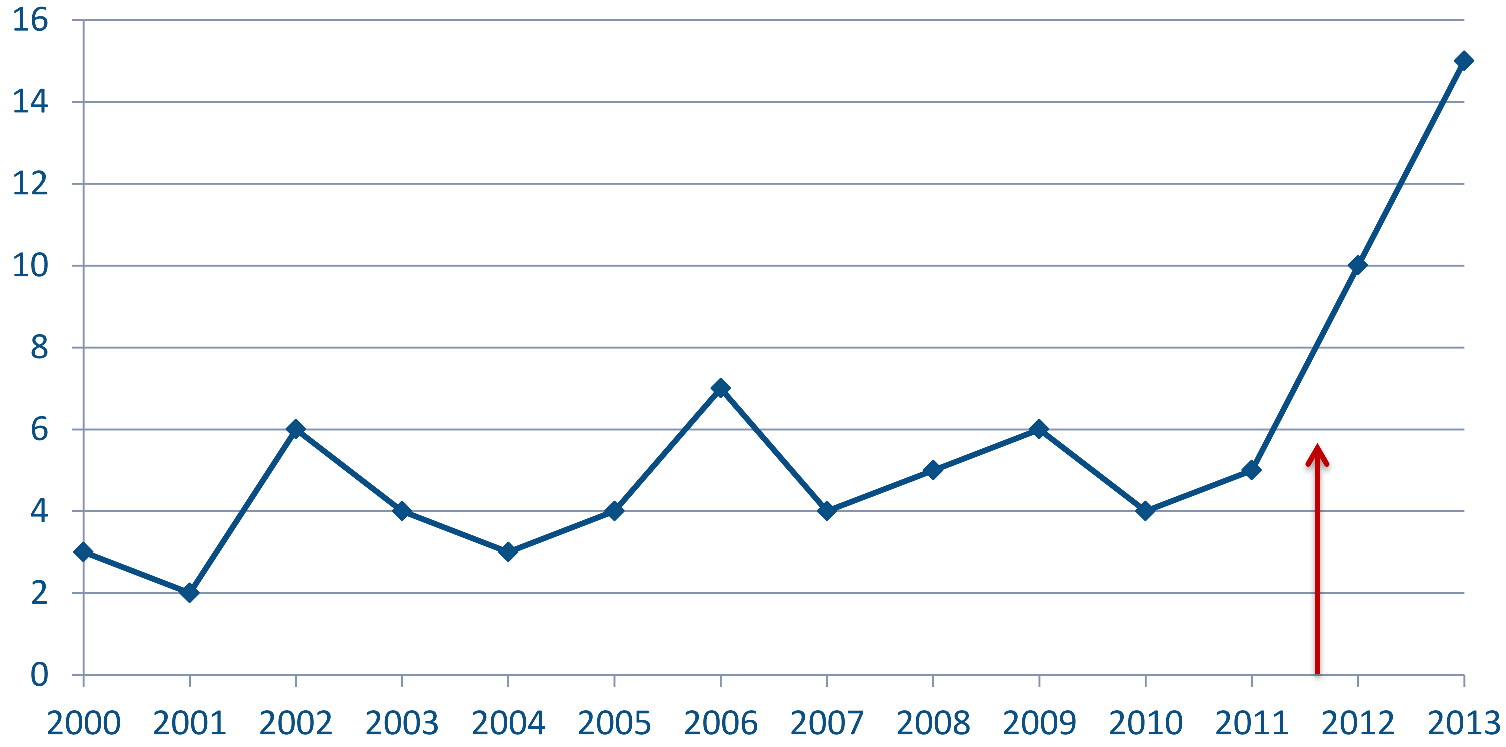
Statistical detection of fraud

Number of citations of paper on fraud over time

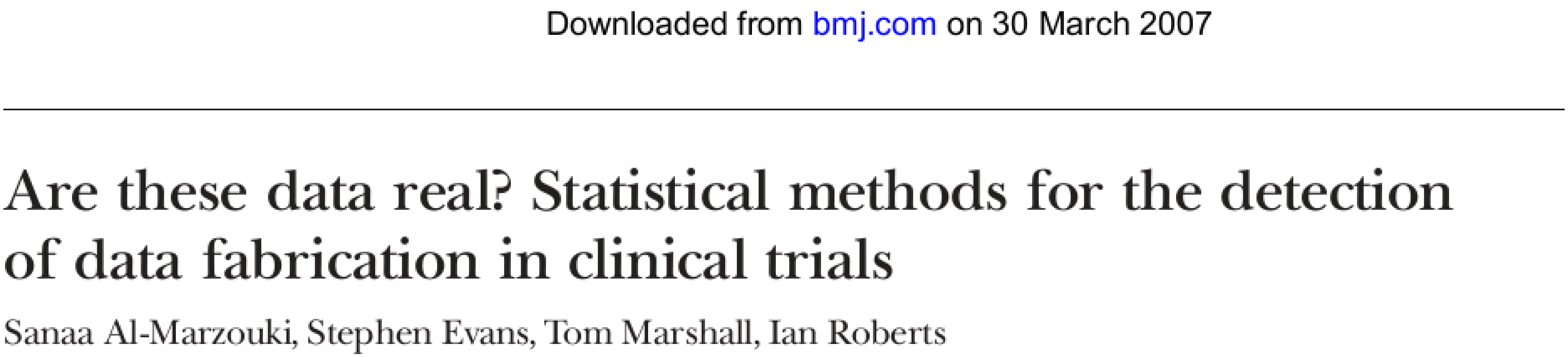


Statistical detection of fraud

Number of citations of paper on fraud over time



Statistical detection of fraud

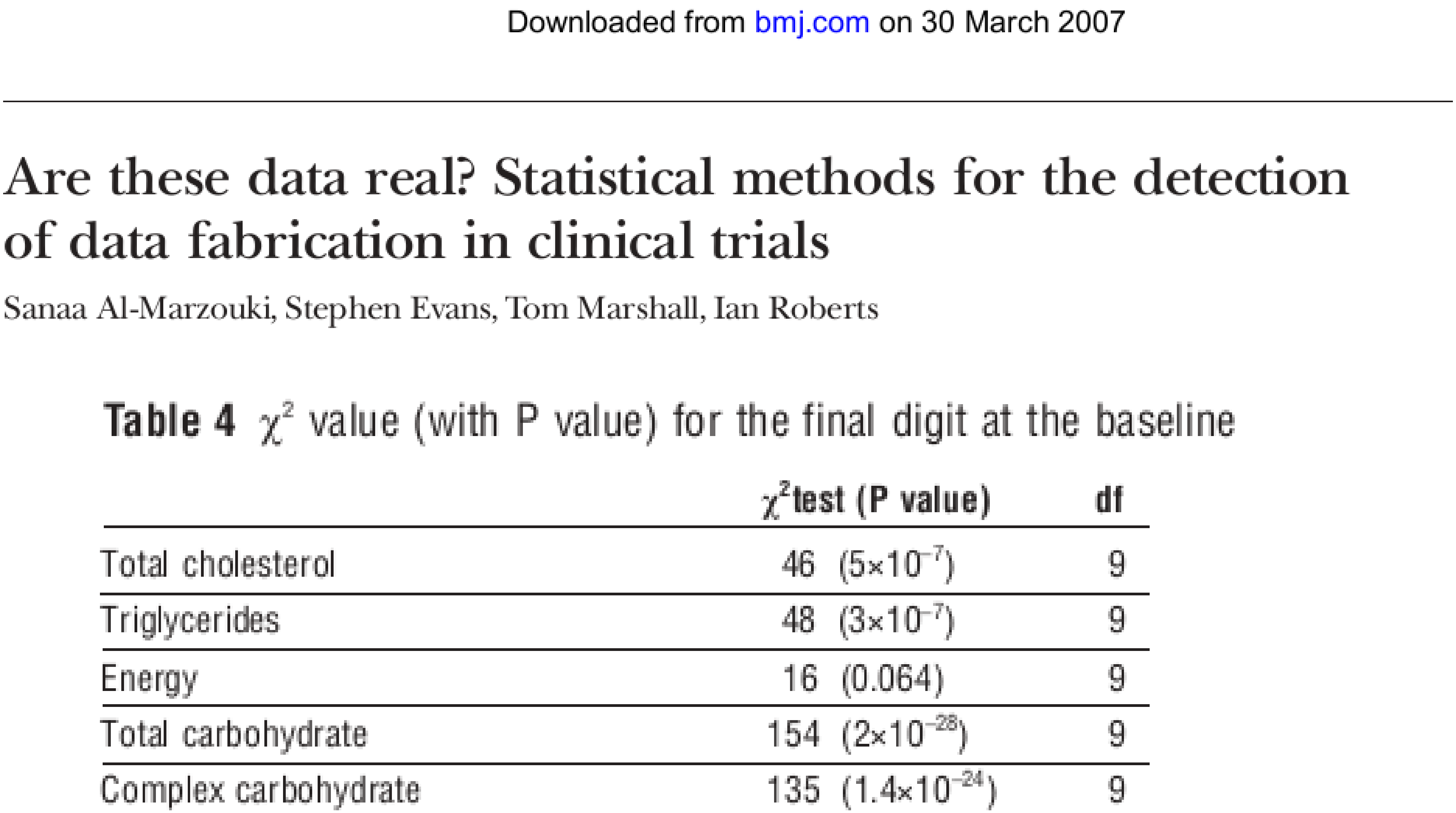


Downloaded from bmj.com on 30 March 2007

Are these data real? Statistical methods for the detection of data fabrication in clinical trials

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

Statistical detection of fraud



Downloaded from bmj.com on 30 March 2007

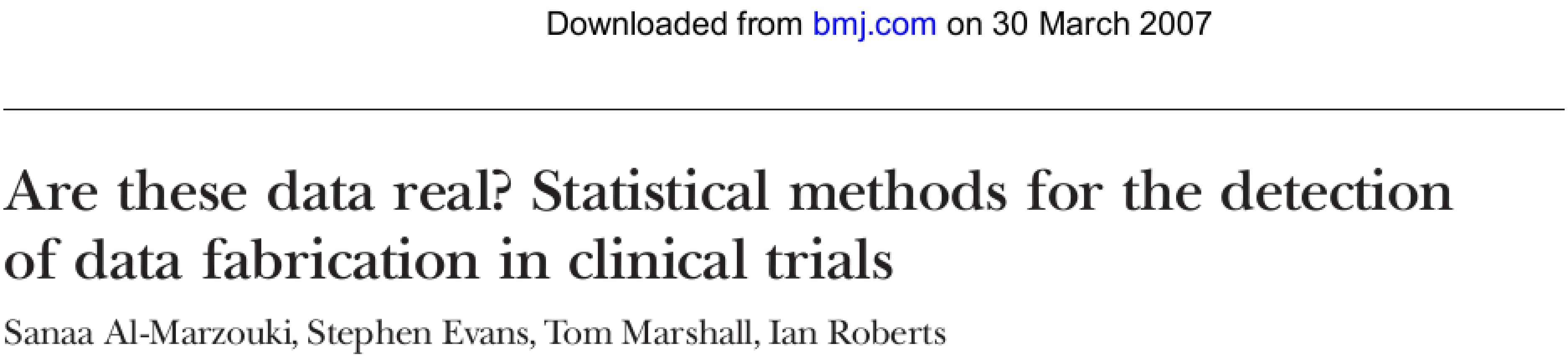
Are these data real? Statistical methods for the detection of data fabrication in clinical trials

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

Table 4 χ^2 value (with P value) for the final digit at the baseline

	χ^2 test (P value)	df
Total cholesterol	46 (5×10^{-7})	9
Triglycerides	48 (3×10^{-7})	9
Energy	16 (0.064)	9
Total carbohydrate	154 (2×10^{-28})	9
Complex carbohydrate	135 (1.4×10^{-24})	9

Statistical detection of fraud



Downloaded from bmj.com on 30 March 2007

Are these data real? Statistical methods for the detection of data fabrication in clinical trials

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

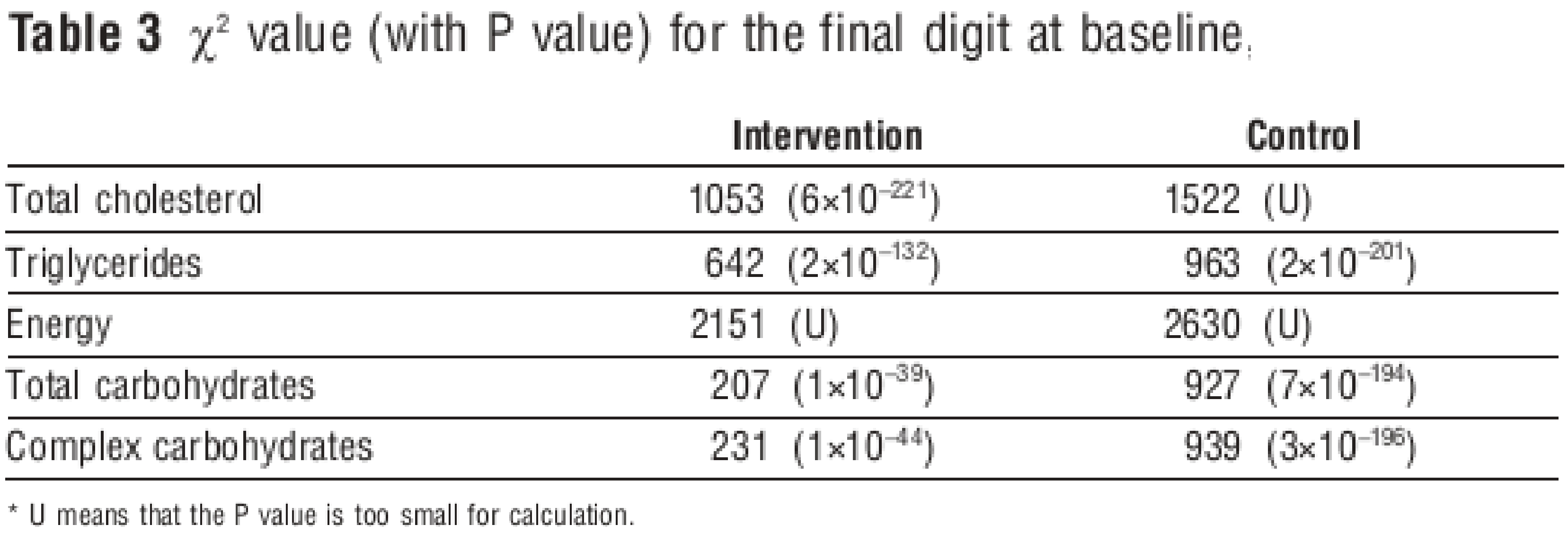
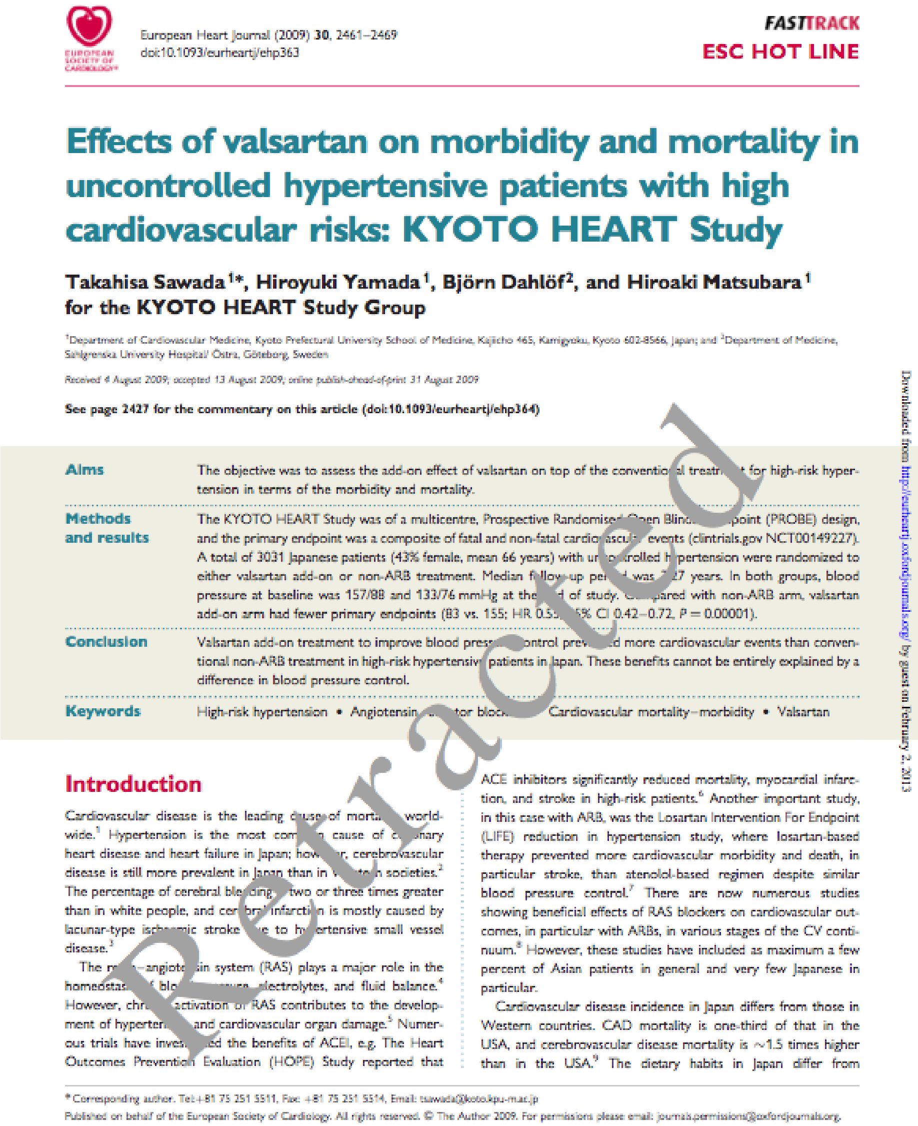


Table 3 χ^2 value (with P value) for the final digit at baseline.

	Intervention	Control
Total cholesterol	1053 (6×10^{-221})	1522 (U)
Triglycerides	642 (2×10^{-132})	963 (2×10^{-201})
Energy	2151 (U)	2630 (U)
Total carbohydrates	207 (1×10^{-39})	927 (7×10^{-194})
Complex carbohydrates	231 (1×10^{-44})	939 (3×10^{-196})

* U means that the P value is too small for calculation.

Importance of fraud



European Heart Journal (2009) 30, 2461–2469
doi:10.1093/eurheartj/ehp363

FASTTRACK
ESC HOT LINE

Effects of valsartan on morbidity and mortality in uncontrolled hypertensive patients with high cardiovascular risks: KYOTO HEART Study

Takahisa Sawada[1]*, Hiroyuki Yamada[1], Björn Dahlöf[2], and Hiroaki Matsubara[1] for the KYOTO HEART Study Group

[1]Department of Cardiovascular Medicine, Kyoto Prefectural University School of Medicine, Kajiicho 465, Kamigyoku, Kyoto 602-8566, Japan; and [2]Department of Medicine, Sahlgrenska University Hospital/ Östra, Göteborg, Sweden

Received 4 August 2009; accepted 13 August 2009; online publish-ahead-of-print 31 August 2009

See page 2427 for the commentary on this article (doi:10.1093/eurheartj/ehp364)

Aims The objective was to assess the add-on effect of valsartan on top of the conventional treatment for high-risk hypertension in terms of the morbidity and mortality.

Methods and results The KYOTO HEART Study was of a multicentre, Prospective Randomised Open Blinded Endpoint (PROBE) design, and the primary endpoint was a composite of fatal and non-fatal cardiovascular events (clintrials.gov NCT00149227). A total of 3031 Japanese patients (43% female, mean 66 years) with uncontrolled hypertension were randomized to either valsartan add-on or non-ARB treatment. Median follow-up period was [illegible]27 years. In both groups, blood pressure at baseline was 157/88 and 133/76 mmHg at the end of study. Compared with non-ARB arm, valsartan add-on arm had fewer primary endpoints (83 vs. 155; HR 0.55, 95% CI 0.42–0.72, $P = 0.00001$).

Conclusion Valsartan add-on treatment to improve blood pressure control prevented more cardiovascular events than conventional non-ARB treatment in high-risk hypertensive patients in Japan. These benefits cannot be entirely explained by a difference in blood pressure control.

Keywords High-risk hypertension • Angiotensin receptor blocker • Cardiovascular mortality–morbidity • Valsartan

Introduction

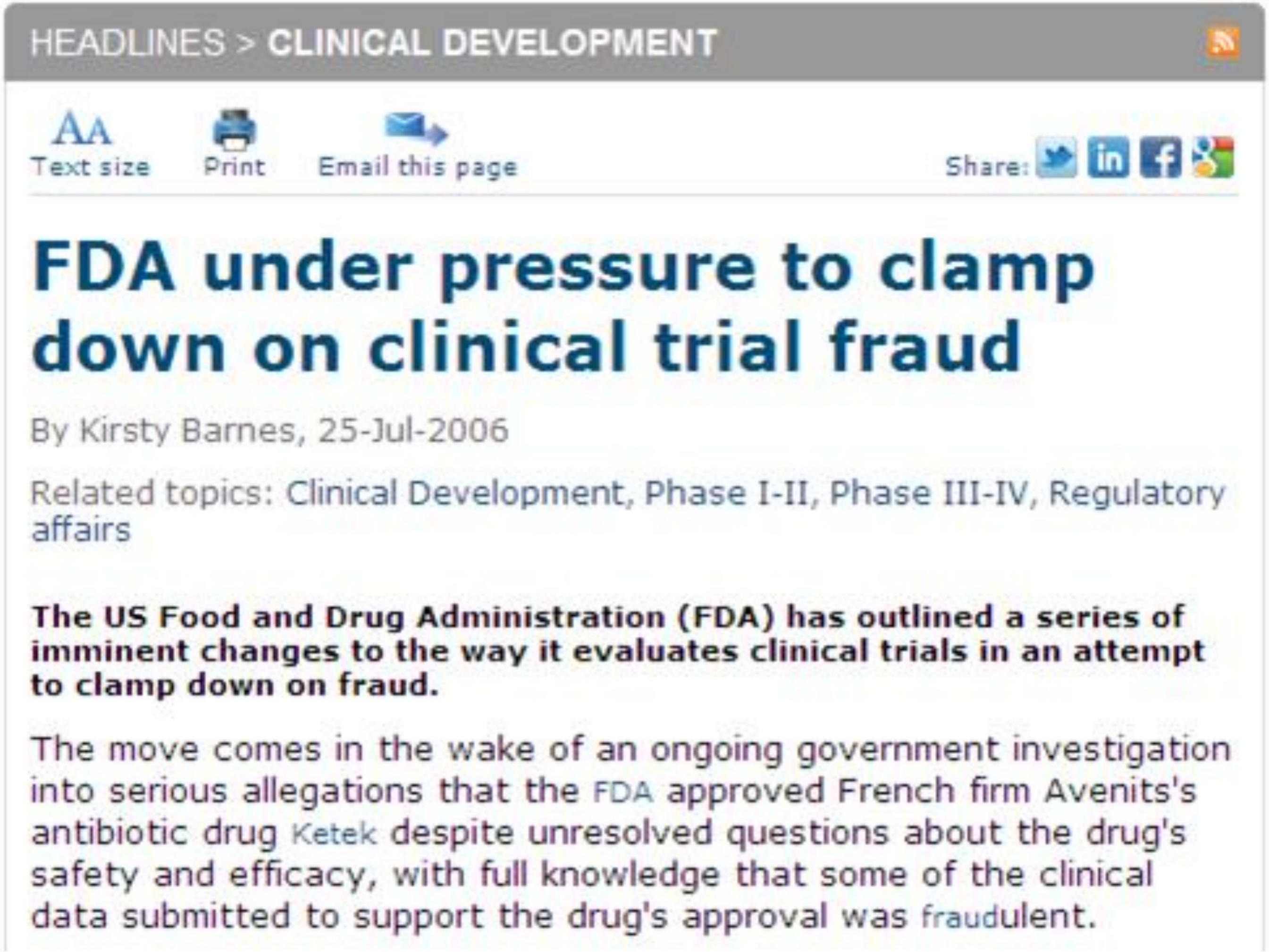
Cardiovascular disease is the leading cause of mortality worldwide.[1] Hypertension is the most common cause of coronary heart disease and heart failure in Japan; however, cerebrovascular disease is still more prevalent in Japan than in Western societies.[2] The percentage of cerebral bleeding is two or three times greater than in white people, and cerebral infarction is mostly caused by lacunar-type ischaemic stroke due to hypertensive small vessel disease.[3]

The renin–angiotensin system (RAS) plays a major role in the homeostasis of blood pressure, electrolytes, and fluid balance.[4] However, chronic activation of RAS contributes to the development of hypertension and cardiovascular organ damage.[5] Numerous trials have investigated the benefits of ACEI, e.g. The Heart Outcomes Prevention Evaluation (HOPE) Study reported that ACE inhibitors significantly reduced mortality, myocardial infarction, and stroke in high-risk patients.[6] Another important study, in this case with ARB, was the Losartan Intervention For Endpoint (LIFE) reduction in hypertension study, where losartan-based therapy prevented more cardiovascular morbidity and death, in particular stroke, than atenolol-based regimen despite similar blood pressure control.[7] There are now numerous studies showing beneficial effects of RAS blockers on cardiovascular outcomes, in particular with ARBs, in various stages of the CV continuum.[8] However, these studies have included as maximum a few percent of Asian patients in general and very few Japanese in particular.

Cardiovascular disease incidence in Japan differs from those in Western countries. CAD mortality is one-third of that in the USA, and cerebrovascular disease mortality is ~1.5 times higher than in the USA.[9] The dietary habits in Japan differ from

* Corresponding author. Tel:+81 75 251 5511, Fax: +81 75 251 5514, Email: tsawada@koto.kpu-m.ac.jp

Importance of fraud



FDA under pressure to clamp down on clinical trial fraud

By Kirsty Barnes, 25-Jul-2006

Related topics: Clinical Development, Phase I-II, Phase III-IV, Regulatory affairs

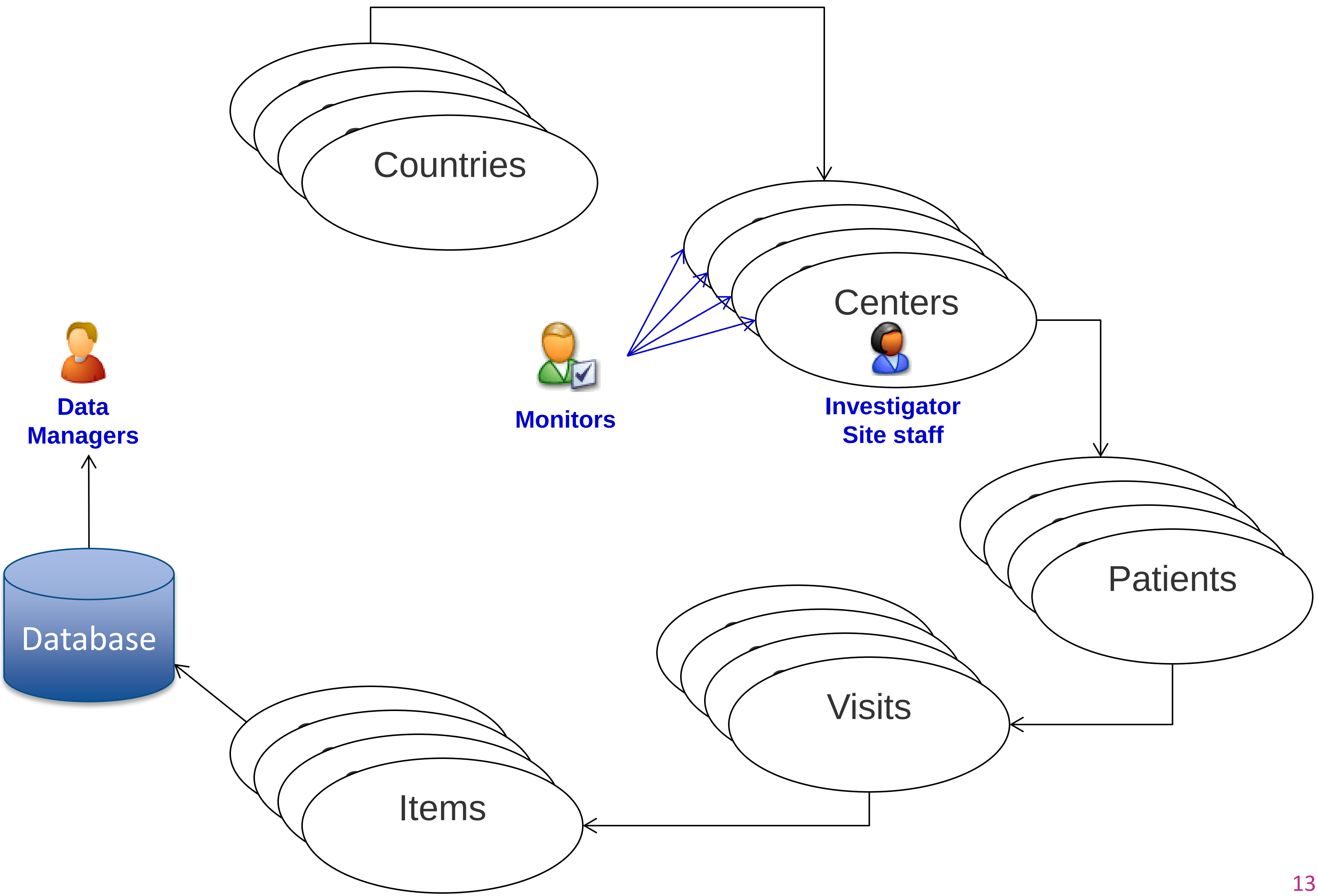
The US Food and Drug Administration (FDA) has outlined a series of imminent changes to the way it evaluates clinical trials in an attempt to clamp down on fraud.

The move comes in the wake of an ongoing government investigation into serious allegations that the FDA approved French firm Avenits's antibiotic drug Ketek despite unresolved questions about the drug's safety and efficacy, with full knowledge that some of the clinical data submitted to support the drug's approval was fraudulent.

FDA’s requirement to ensure data quality

100% (manual) source data verification !

Central Data Management / On-site Monitoring



FDA’s requirement to ensure data quality

100% (manual) source data verification !

Typical phase III clinical trial

- 100 centers
- 10 patients / center
- 10 visits / patient
- 100 data items / visit

→ 10^6 data items to check (hospital files *vs.* case report form)

The cost of 100% source data verification

What proportion of the total budget of a clinical trial is spent on “100% source data verification”?

- < 1%
- 1 – 5%
- 5 – 10%
- 10 – 20%
- > 20%

Ref: Funning et al, Quality Assurance J (2008)

The cost of 100% source data verification

What proportion of the total budget of a clinical trial is spent on “100% source data verification”?

- < 1%
- 1 – 5%
- 5 – 10%
- **> 15%**
- > 20%

Cost of a typical phase III clinical trial : 100 M$

Cost of source data verification : 15 M$

Ref: Funning et al, Quality Assurance J (2008)

Errors discovered in clinical data

What proportion of clinical data items are corrected during the course of a trial?

- < 1%
- 1 – 5%
- 5 – 10%
- 10 – 20%
- > 20%

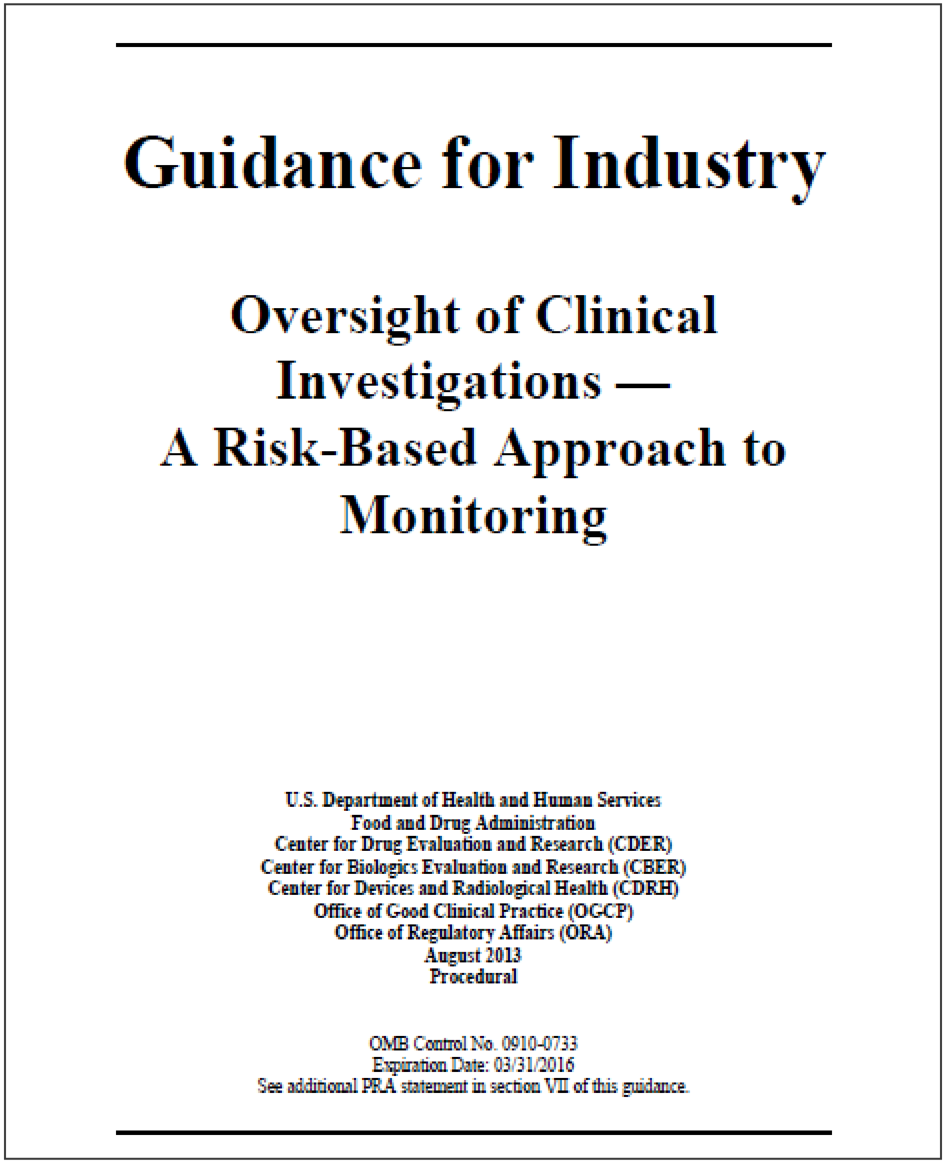
Errors discovered in clinical data

What proportion of clinical data items that are corrected during the course of a trial?

- < 1%
- **< 5%**
- 5 – 10%
- 10 – 20%
- > 20%

Source: Medidata (J. Pines)

Risk-Based Monitoring

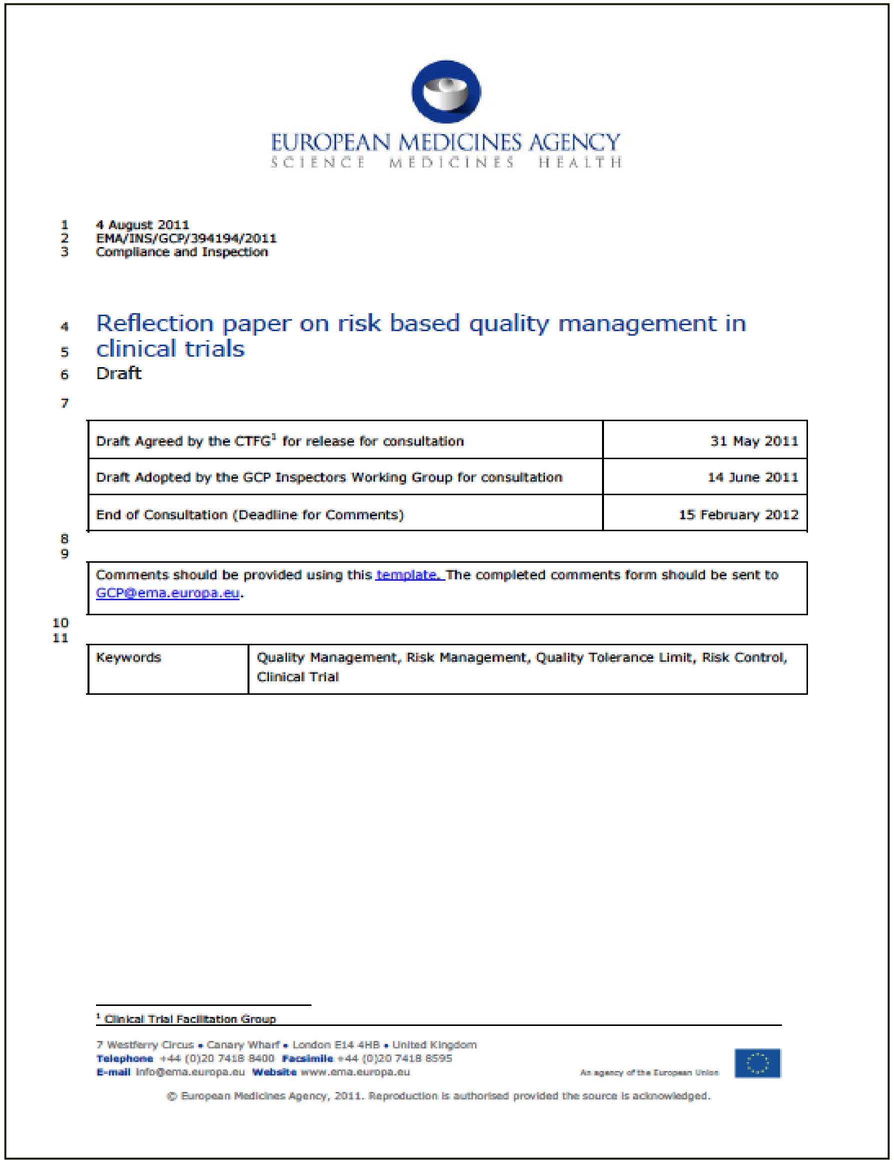


Guidance for Industry

Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring

U.S. 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)
Office of Good Clinical Practice (OGCP)
Office of Regulatory Affairs (ORA)
August 2013
Procedural

OMB Control No. 0910-0733
Expiration Date: 03/31/2016
See additional PRA statement in section VII of this guidance.



EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

4 August 2011
EMA/INS/GCP/394194/2011
Compliance and Inspection

Reflection paper on risk based quality management in clinical trials

Draft

Draft Agreed by the CTFG[1] for release for consultation	31 May 2011
Draft Adopted by the GCP Inspectors Working Group for consultation	14 June 2011
End of Consultation (Deadline for Comments)	15 February 2012

Comments should be provided using this template. The completed comments form should be sent to GCP@ema.europa.eu.

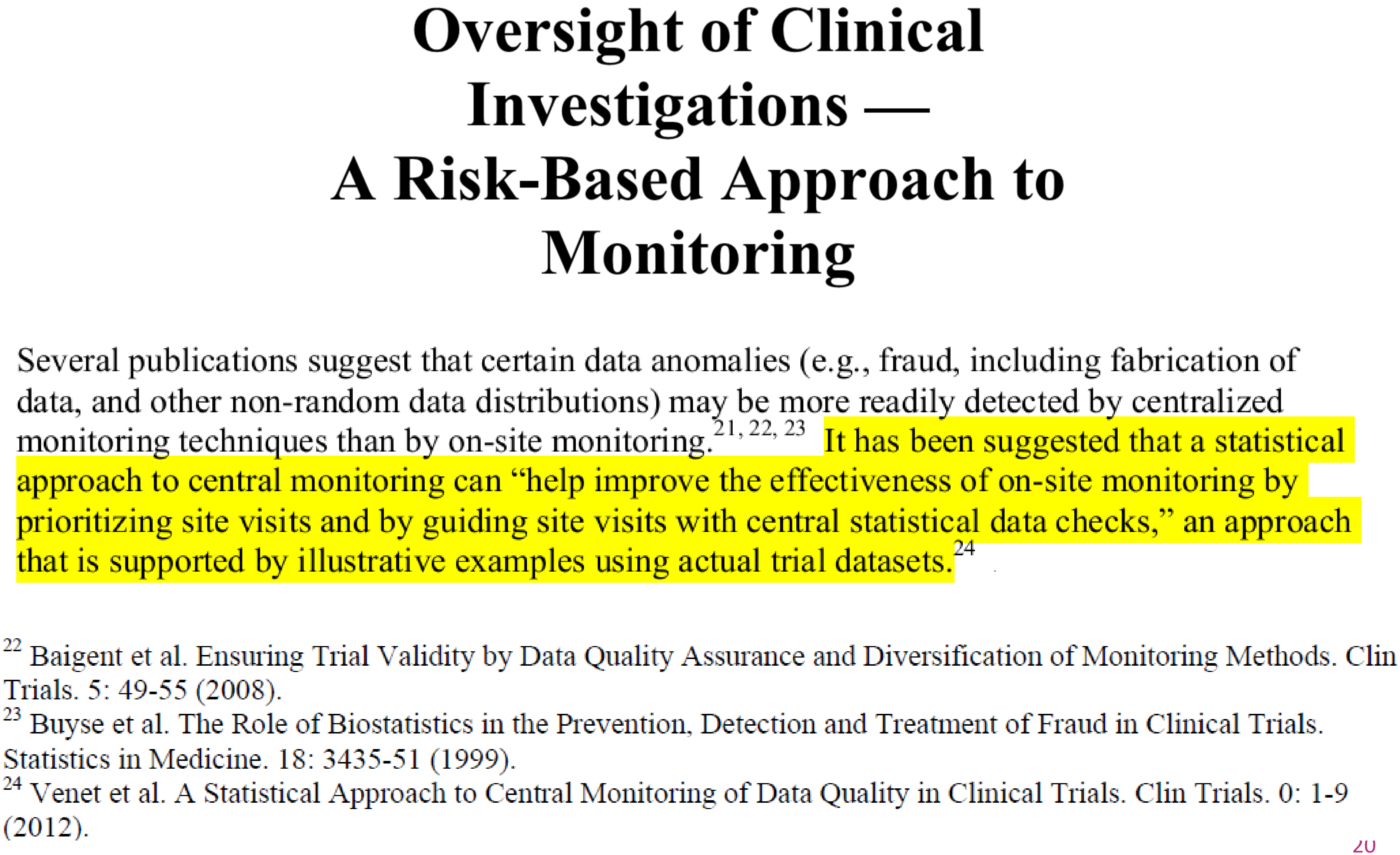
Keywords	Quality Management, Risk Management, Quality Tolerance Limit, Risk Control, Clinical Trial

[1] Clinical Trial Facilitation Group

7 Westferry Circus • Canary Wharf • London E14 4HB • United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8595
E-mail info@ema.europa.eu **Website** www.ema.europa.eu

An agency of the European Union

FDA Guidance for Industry



Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring

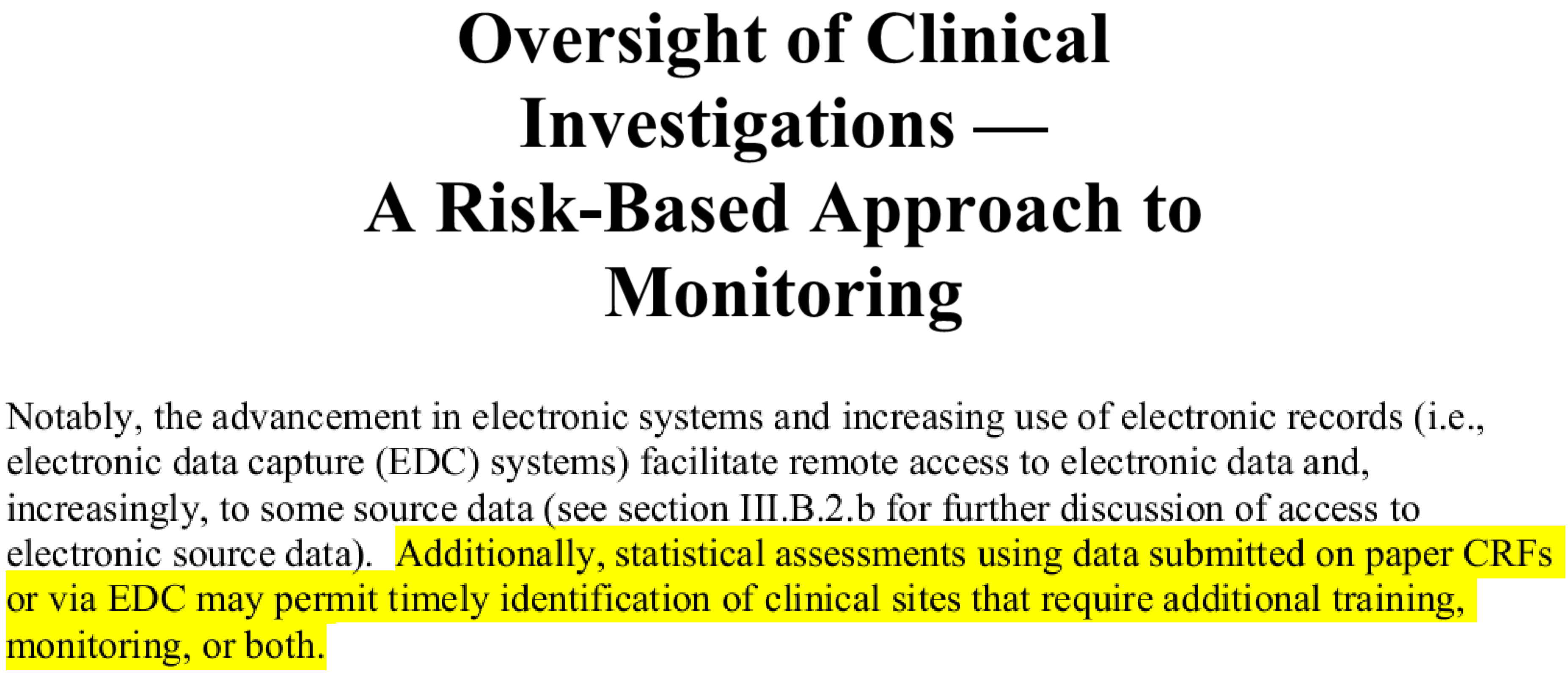
Several publications suggest that certain data anomalies (e.g., fraud, including fabrication of data, and other non-random data distributions) may be more readily detected by centralized monitoring techniques than by on-site monitoring.[21, 22, 23] It has been suggested that a statistical approach to central monitoring can "help improve the effectiveness of on-site monitoring by prioritizing site visits and by guiding site visits with central statistical data checks," an approach that is supported by illustrative examples using actual trial datasets.[24]

[22] Baigent et al. Ensuring Trial Validity by Data Quality Assurance and Diversification of Monitoring Methods. Clin Trials. 5: 49-55 (2008).

[23] Buyse et al. The Role of Biostatistics in the Prevention, Detection and Treatment of Fraud in Clinical Trials. Statistics in Medicine. 18: 3435-51 (1999).

[24] Venet et al. A Statistical Approach to Central Monitoring of Data Quality in Clinical Trials. Clin Trials. 0: 1-9 (2012).

FDA Guidance for Industry

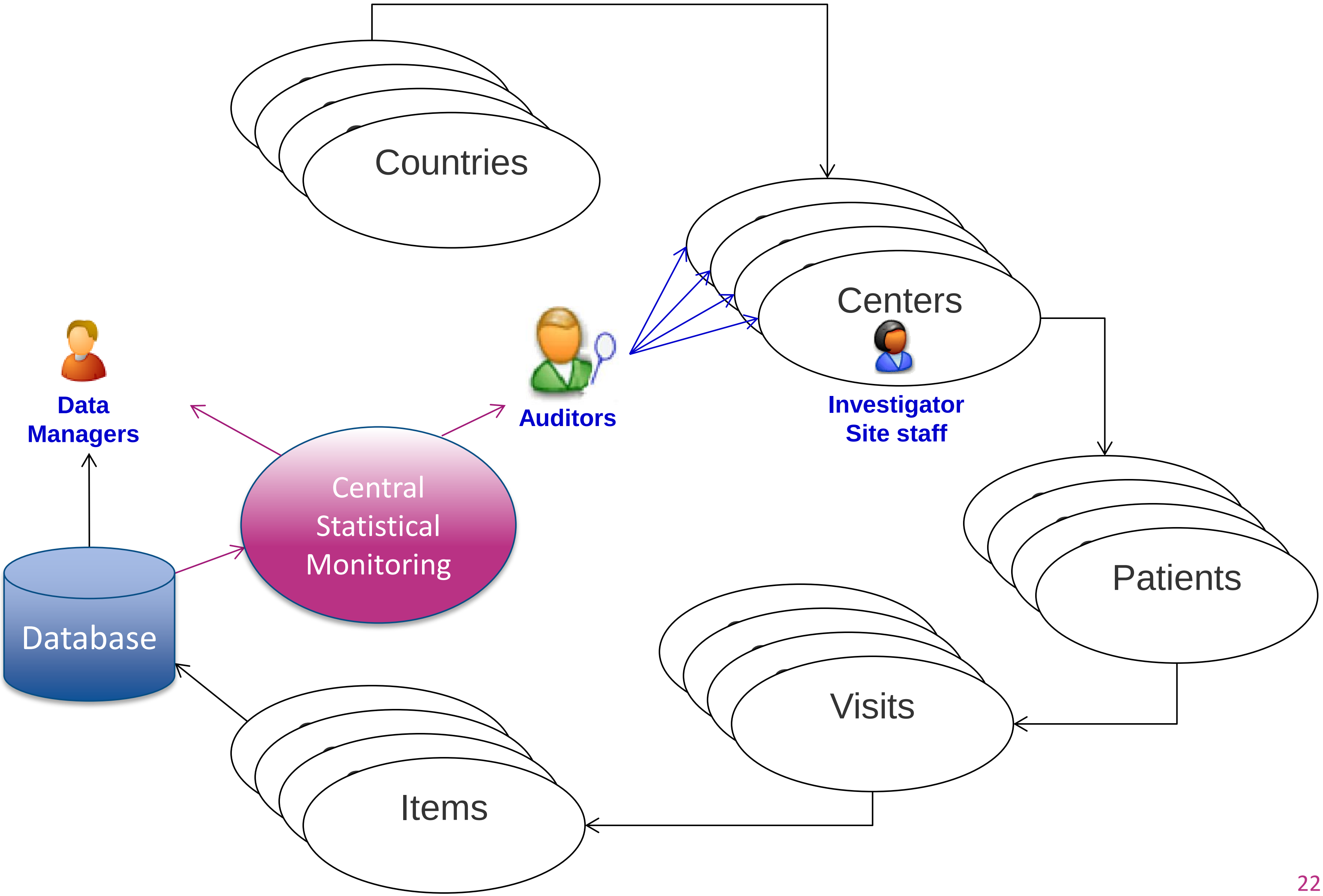


Oversight of Clinical Investigations — A Risk-Based Approach to Monitoring

Notably, the advancement in electronic systems and increasing use of electronic records (i.e., electronic data capture (EDC) systems) facilitate remote access to electronic data and, increasingly, to some source data (see section III.B.2.b for further discussion of access to electronic source data). Additionally, statistical assessments using data submitted on paper CRFs or via EDC may permit timely identification of clinical sites that require additional training, monitoring, or both.

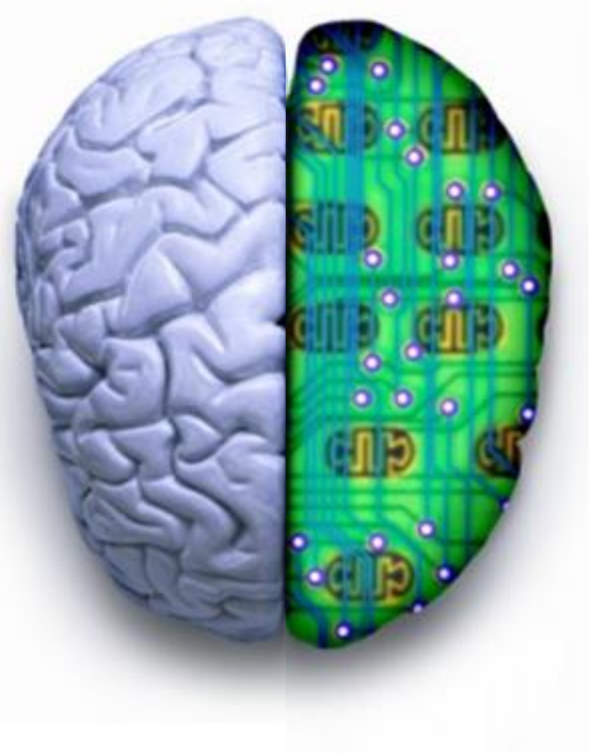
Ref: www.fda.gov/downloads/Drugs/…/Guidances/UCM269919.pdf

Central Statistical Monitoring



Humans *vs.* Computers

Humans are not good at fabricating data, nor at detecting erroneous data patterns



Computer algorithms are very good at detecting data patterns (they are also reliable and cheap)

Data types

Variable types are automatically determined

- Center, subject, visit identifiers
- Binary
- Categorical
- Numerical (continuous if ≥ 10 distinct values)
- Dates

Data preparation

Uninformative variables are removed

- Tables without patient identifiers
- Auxiliary variables (database management)
- Variables with too many missing values
- Variables with no variability
- Variables with too much variability (*e.g.* mixed units)

Statistical tests

- For each variable, all relevant statistical tests are selected based on the type of the variable
- The tests compare each center against all other centers
- One *P*-value is generated per center per test

Tests for numerical variables

- Variables are transformed to have approximate normal distribution

- Non repeated measures
  - Means and variances, using linear mixed effects model to account for between-center variability
  - Outliers

- Repeated measures
  - Within-patient variance
  - Sequence outliers (*e.g.* 30,32,33,32,55,32)
  - Propagation of values (*e.g.* 32,32,32,32)

Tests for binary variables

- Non-repeated measures
  - Beta-binomial model to account for between-center variability
  - Binomial model if little between-center variability

- Repeated measures
  - Markov model with two different states (0 and 1)
    - Start 1: State 1 at the beginning of the sequence
    - 1 -> 0: The transition from state 1 to state 0
    - 0 -> 1: The transition from state 0 to state 1

Tests for categorical variables

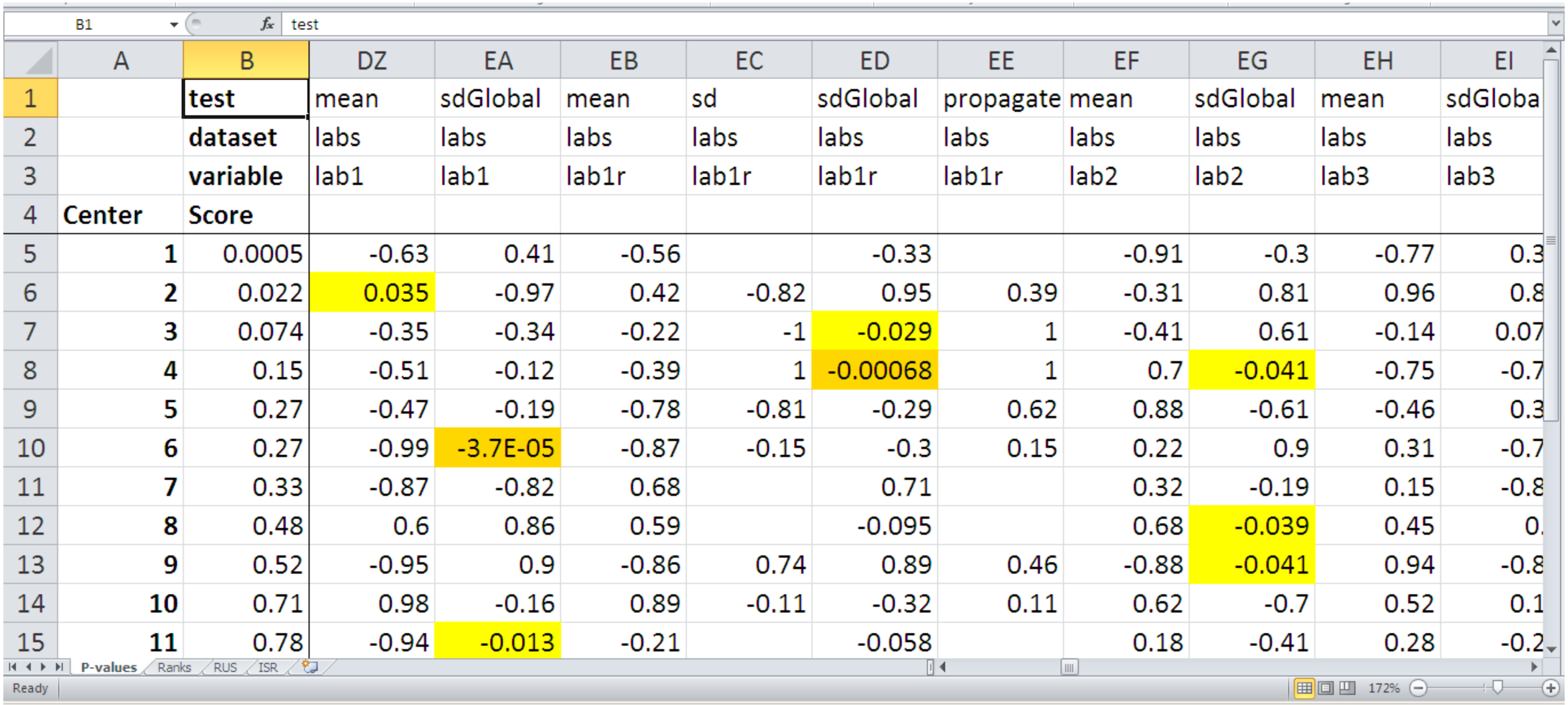
- Categorical variables are dichotomized
  - e.g. x having possible values A, B and C:
    3 variables are created
    $y_1 = 1$ if $x = A$, $y_1 = 0$ otherwise
    $y_2 = 1$ if $x = B$, $y_2 = 0$ otherwise
    $y_3 = 1$ if $x = C$, $y_3 = 0$ otherwise
- Tests as for binary variables
  - single *P*-value is calculated as the minimum of the *P*-values of all binary tests
  - simple correction for multiplicity (Bonferroni)
  - *e.g.* in the example: $p = \min(p_A, p_B, p_C) \times 3$

Statistical tests

- *P*-values (p_{ij}) form a matrix with as many rows as centers and as many columns as individual tests
- Typical phase III clinical trial
  - 100 sites
  - 1000 items to test
  - 10 statistical tests per item

  $\rightarrow$ 10^6 *P*-values

P-values



	A	B	DZ	EA	EB	EC	ED	EE	EF	EG	EH	EI
1		test	mean	sdGlobal	mean	sd	sdGlobal	propagate	mean	sdGlobal	mean	sdGloba
2		dataset	labs	labs	labs	labs	labs	labs	labs	labs	labs	labs
3		variable	lab1	lab1	lab1r	lab1r	lab1r	lab1r	lab2	lab2	lab3	lab3
4	Center	Score										
5	1	0.0005	-0.63	0.41	-0.56		-0.33		-0.91	-0.3	-0.77	0.3
6	2	0.022	0.035	-0.97	0.42	-0.82	0.95	0.39	-0.31	0.81	0.96	0.8
7	3	0.074	-0.35	-0.34	-0.22	-1	-0.029	1	-0.41	0.61	-0.14	0.07
8	4	0.15	-0.51	-0.12	-0.39	1	-0.00068	1	0.7	-0.041	-0.75	-0.7
9	5	0.27	-0.47	-0.19	-0.78	-0.81	-0.29	0.62	0.88	-0.61	-0.46	0.3
10	6	0.27	-0.99	-3.7E-05	-0.87	-0.15	-0.3	0.15	0.22	0.9	0.31	-0.7
11	7	0.33	-0.87	-0.82	0.68		0.71		0.32	-0.19	0.15	-0.8
12	8	0.48	0.6	0.86	0.59		-0.095		0.68	-0.039	0.45	0.
13	9	0.52	-0.95	0.9	-0.86	0.74	0.89	0.46	-0.88	-0.041	0.94	-0.8
14	10	0.71	0.98	-0.16	0.89	-0.11	-0.32	0.11	0.62	-0.7	0.52	0.1
15	11	0.78	-0.94	-0.013	-0.21		-0.058		0.18	-0.41	0.28	-0.2

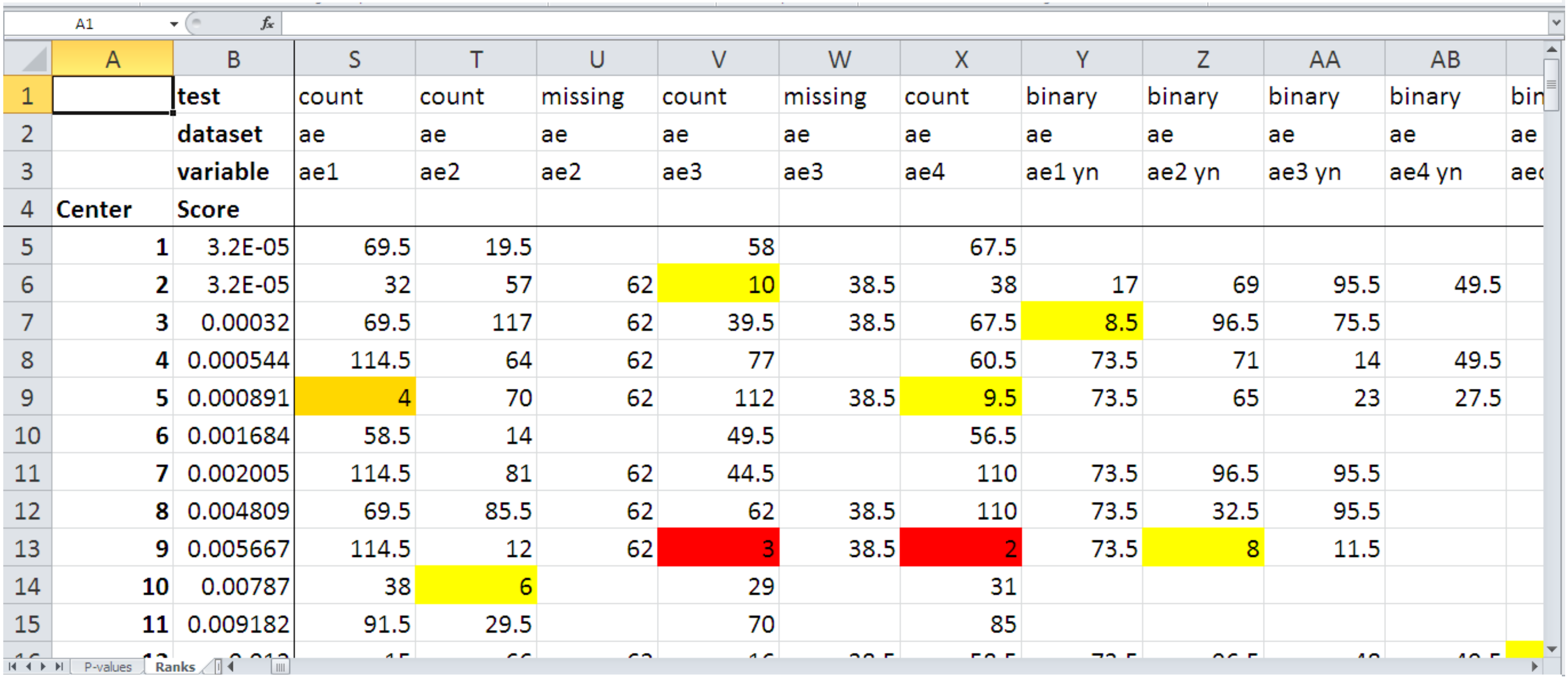
P-values

- Color conventions:
  - Red : $0 < p < 10^{-5}$
  - Orange : $10^{-5} < p < 10^{-3}$
  - Yellow: $10^{-3} < p < 5\ 10^{-2}$
  - No color : $5\ 10^{-2} < p < 1$
- *P*-values are signed for directional tests

Ranking

- For each test, centers are ranked from most extreme to least extreme *P*-value (*e.g.* if there are 100 centers, the rank will range between 1 and 100)
- Ranks form a matrix with as many rows as centers and as many columns as tests

Ranking



	A	B	S	T	U	V	W	X	Y	Z	AA	AB	
1		**test**	count	count	missing	count	missing	count	binary	binary	binary	binary	bin
2		**dataset**	ae	ae	ae	ae	ae	ae	ae	ae	ae	ae	ae
3		**variable**	ae1	ae2	ae2	ae3	ae3	ae4	ae1 yn	ae2 yn	ae3 yn	ae4 yn	aed
4	**Center**	**Score**											
5	**1**	3.2E-05	69.5	19.5		58		67.5					
6	**2**	3.2E-05	32	57	62	10	38.5	38	17	69	95.5	49.5	
7	**3**	0.00032	69.5	117	62	39.5	38.5	67.5	8.5	96.5	75.5		
8	**4**	0.000544	114.5	64	62	77		60.5	73.5	71	14	49.5	
9	**5**	0.000891	4	70	62	112	38.5	9.5	73.5	65	23	27.5	
10	**6**	0.001684	58.5	14		49.5		56.5					
11	**7**	0.002005	114.5	81	62	44.5		110	73.5	96.5	95.5		
12	**8**	0.004809	69.5	85.5	62	62	38.5	110	73.5	32.5	95.5		
13	**9**	0.005667	114.5	12	62	3	38.5	2	73.5	8	11.5		
14	**10**	0.00787	38	6		29		31					
15	**11**	0.009182	91.5	29.5		70		85					

Ranking

- Color conventions:
  - Red : Rank $\leq$ 3
  - Orange : 3 < Rank $\leq$ 5
  - Yellow: 5 < Rank $\leq$ 10
  - No color : 10 < Rank
- Convention for tied ranks:
  - Mid-ranks used for tied ranks

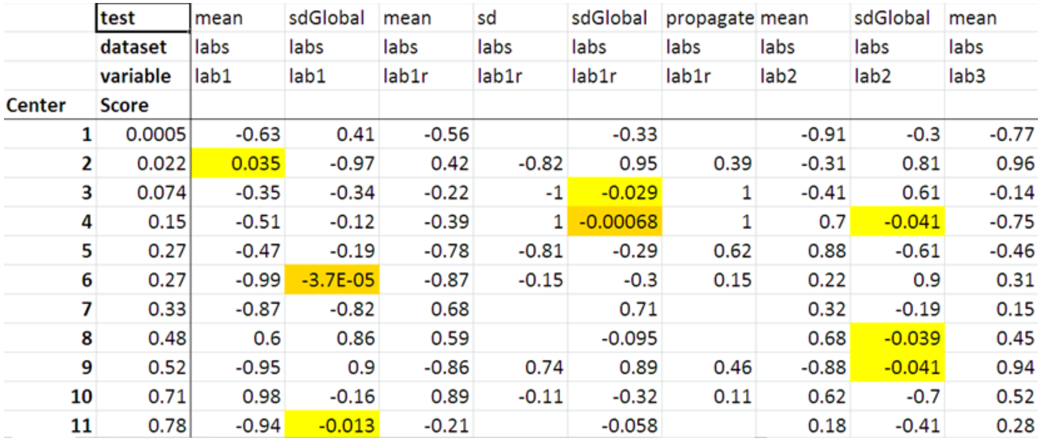
Center scoring

$$score_i = \exp\left(\frac{1}{N}\sum_{j=1}^{N}\log p_{ij}\right)$$

- Some tweaking…
  - Tests with extreme *P*-values are eliminated
  - Uninformative tests are eliminated
  - *P*-values are weighted to account for correlation between tests
- Statistical significance of center scores
  - Estimated using resampling

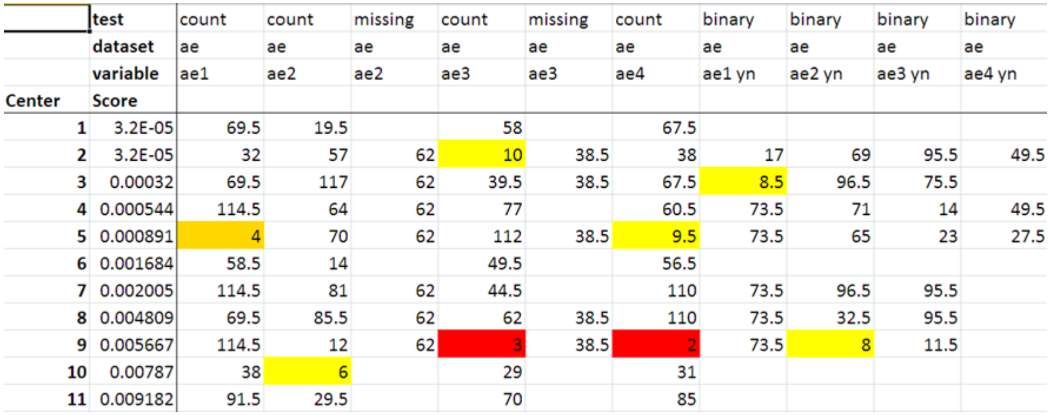
In summary

Central statistical testing engine



	test	mean	sdGlobal	mean	sd	sdGlobal	propagate	mean	sdGlobal	mean
	dataset	labs	labs	labs	labs	labs	labs	labs	labs	labs
	variable	lab1	lab1	lab1r	lab1r	lab1r	lab1r	lab2	lab2	lab3
Center	Score									
1	0.0005	-0.63	0.41	-0.56		-0.33		-0.91	-0.3	-0.77
2	0.022	0.035	-0.97	0.42	-0.82	0.95	0.39	-0.31	0.81	0.96
3	0.074	-0.35	-0.34	-0.22	-1	-0.029	1	-0.41	0.61	-0.14
4	0.15	-0.51	-0.12	-0.39	1	-0.00068	1	0.7	-0.041	-0.75
5	0.27	-0.47	-0.19	-0.78	-0.81	-0.29	0.62	0.88	-0.61	-0.46
6	0.27	-0.99	-3.7E-05	-0.87	-0.15	-0.3	0.15	0.22	0.9	0.31
7	0.33	-0.87	-0.82	0.68		0.71		0.32	-0.19	0.15
8	0.48	0.6	0.86	0.59		-0.095		0.68	-0.039	0.45
9	0.52	-0.95	0.9	-0.86	0.74	0.89	0.46	-0.88	-0.041	0.94
10	0.71	0.98	-0.16	0.89	-0.11	-0.32	0.11	0.62	-0.7	0.52
11	0.78	-0.94	-0.013	-0.21		-0.058		0.18	-0.41	0.28

Matrix of *P*-values p_{ij}



	test	count	count	missing	count	missing	count	binary	binary	binary	binary
	dataset	ae	ae	ae	ae	ae	ae	ae	ae	ae	ae
	variable	ae1	ae2	ae2	ae3	ae3	ae4	ae1 yn	ae2 yn	ae3 yn	ae4 yn
Center	Score										
1	3.2E-05	69.5	19.5		58		67.5				
2	3.2E-05	32	57	62	10	38.5	38	17	69	95.5	49.5
3	0.00032	69.5	117	62	39.5	38.5	67.5	8.5	96.5	75.5	
4	0.000544	114.5	64	62	77		60.5	73.5	71	14	49.5
5	0.000891	4	70	62	112	38.5	9.5	73.5	65	23	27.5
6	0.001684	58.5	14		49.5		56.5				
7	0.002005	114.5	81	62	44.5		110	73.5	96.5	95.5	
8	0.004809	69.5	85.5	62	62	38.5	110	73.5	32.5	95.5	
9	0.005667	114.5	12	62	3	38.5	2	73.5	8	11.5	
10	0.00787	38	6		29		31				
11	0.009182	91.5	29.5		70		85				

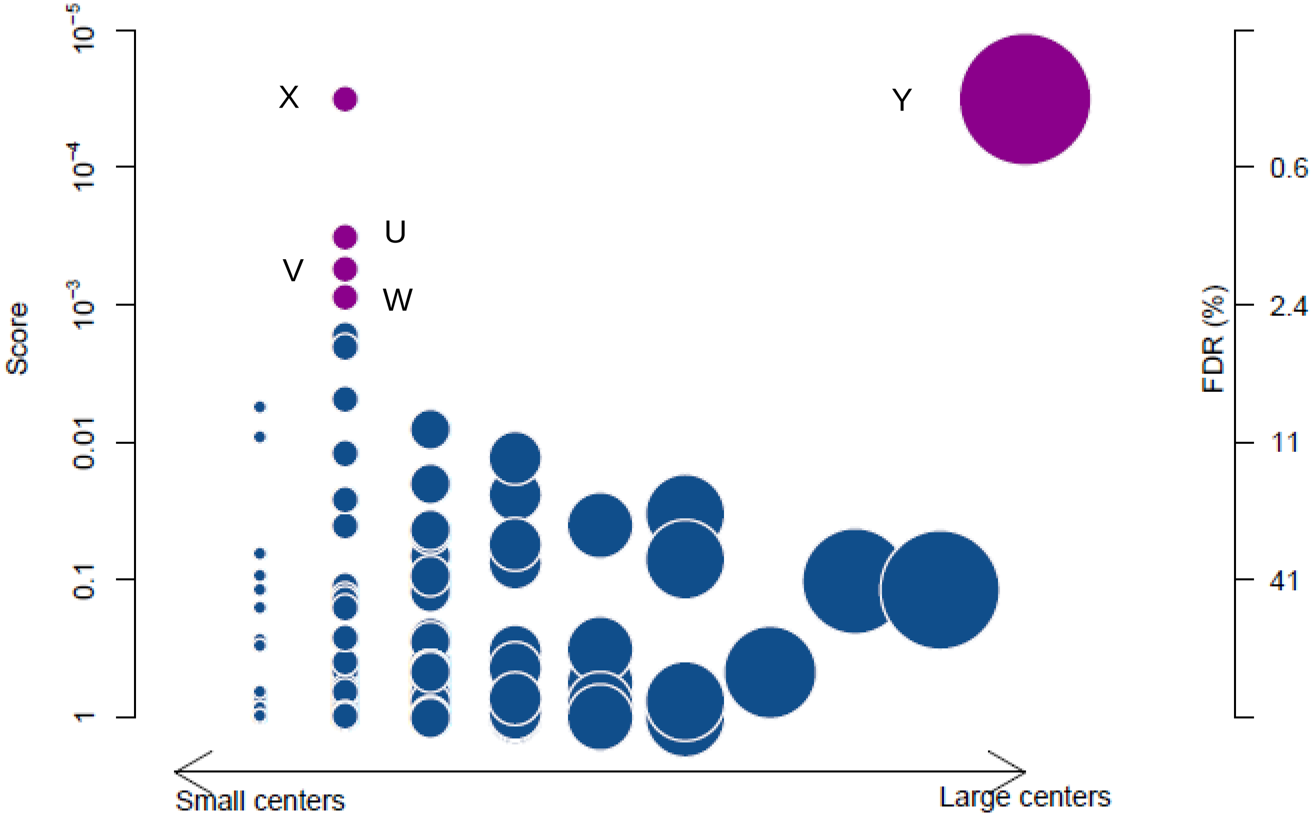
Matrix of ranks r_{ij}

$$score_i = \exp\left(\frac{1}{N}\sum_{j=1}^{N}\log p_{ij}\right)$$

Score$_i$

Visual displays

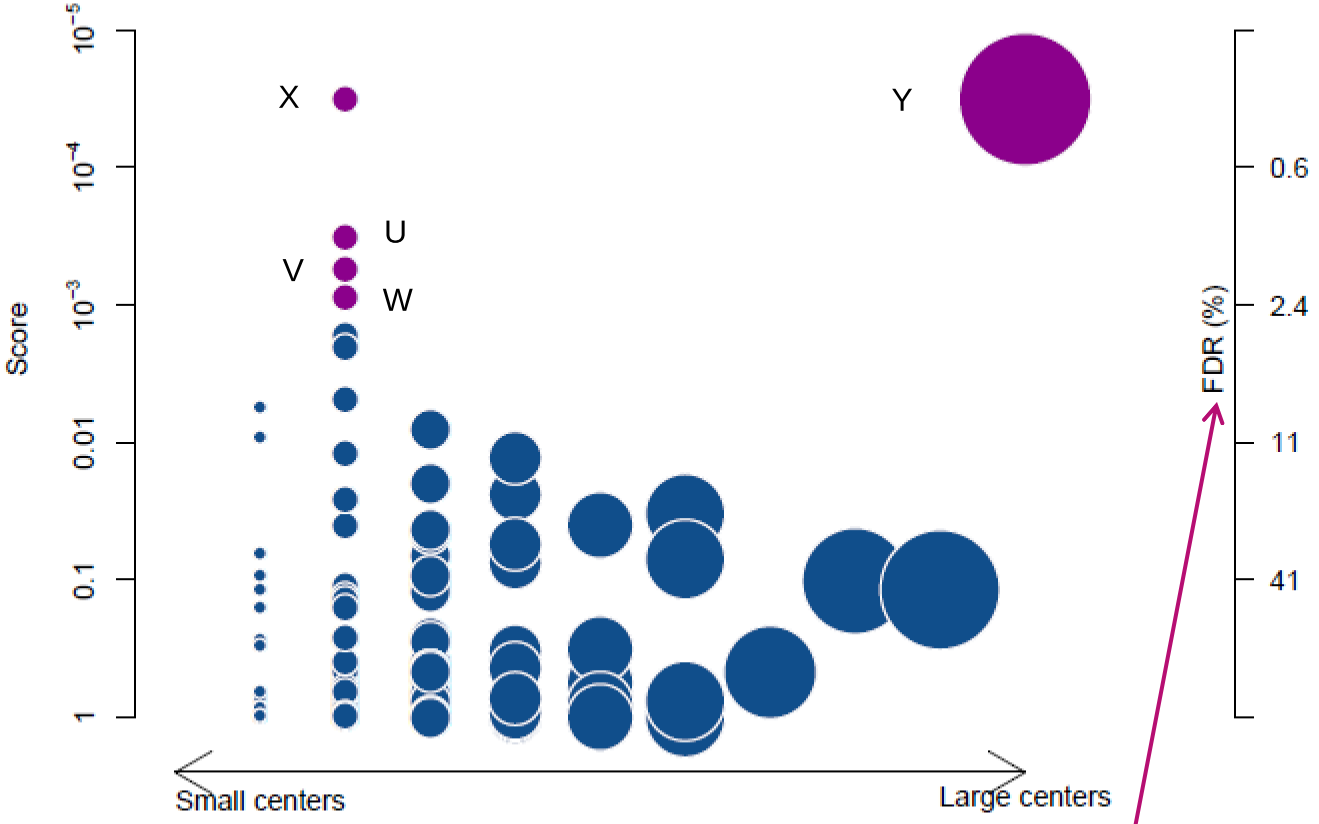
Bubble plot



Circles are proportional to the center size

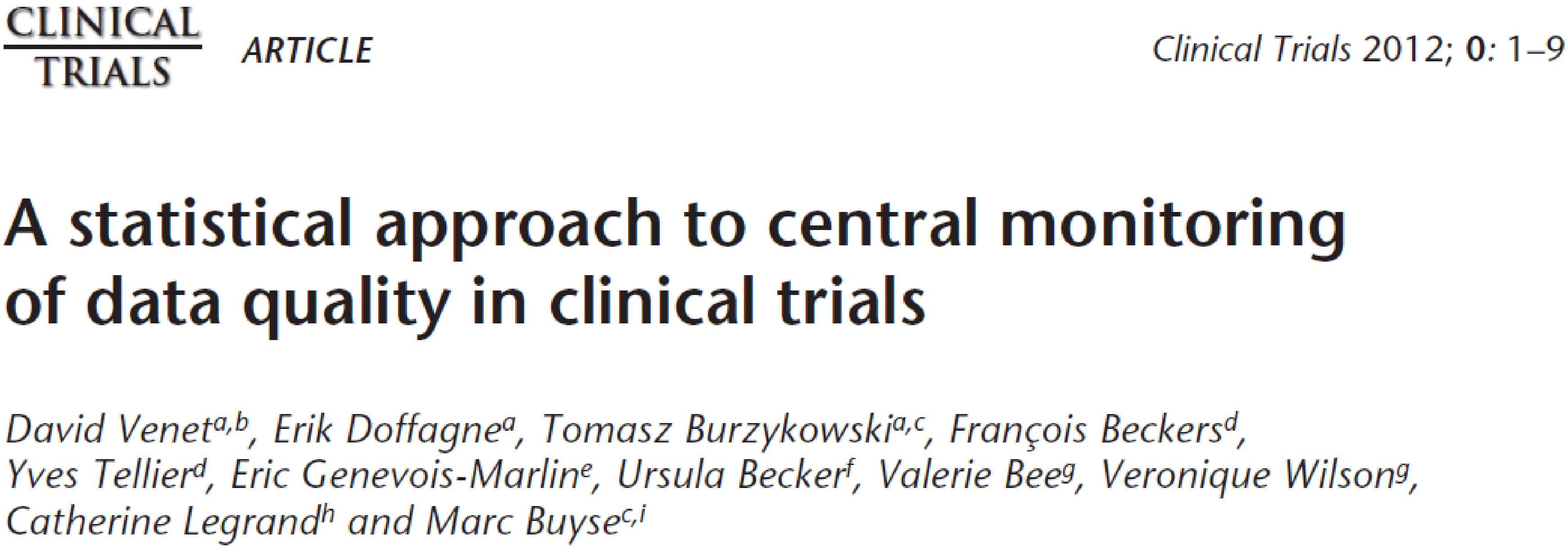
Visual displays

Bubble plot



False Discovery Rate

Central Statistical Monitoring



CLINICAL TRIALS

ARTICLE

Clinical Trials 2012; **0**: 1–9

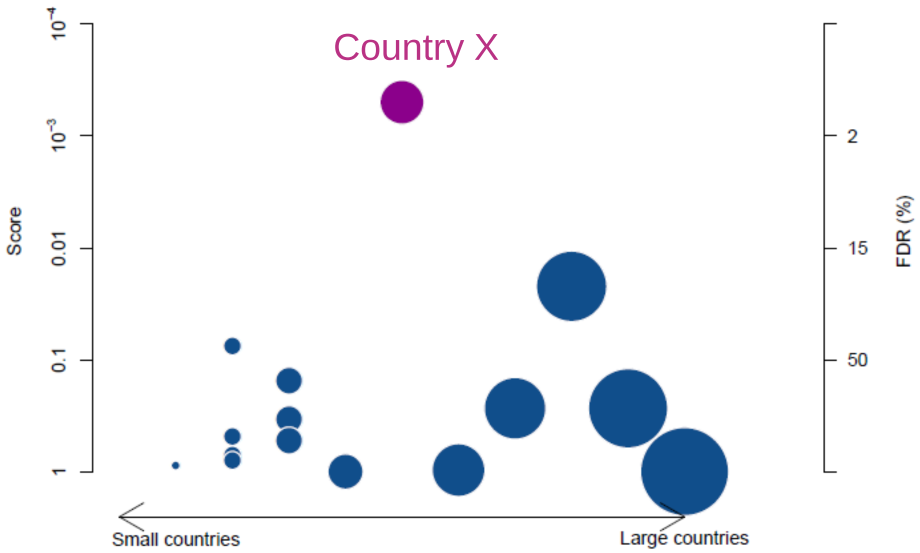
A statistical approach to central monitoring of data quality in clinical trials

David Venet[a,b], Erik Doffagne[a], Tomasz Burzykowski[a,c], François Beckers[d], Yves Tellier[d], Eric Genevois-Marlin[e], Ursula Becker[f], Valerie Bee[g], Veronique Wilson[g], Catherine Legrand[h] and Marc Buyse[c,i]

An example of fraud

Major Depression Trial

- 800 patient trial
- 70 centers
- After run-in period, MADRAS score < 12 for patient to be eligible



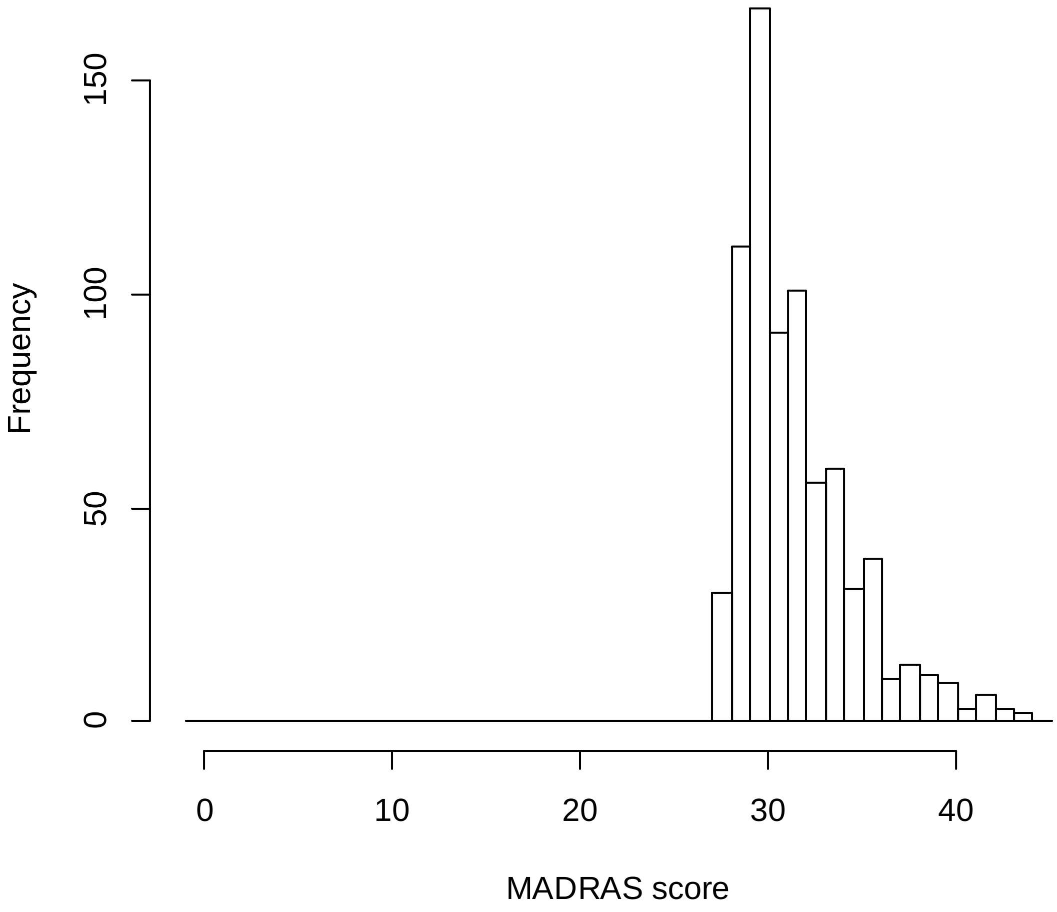
Circles are proportional to the country size

Abnormal Pattern:
No ineligible patients (out of 35 patients in 3 centers) in country X

Interpretation:
The MADRAS score was « pushed » down to make patients eligible

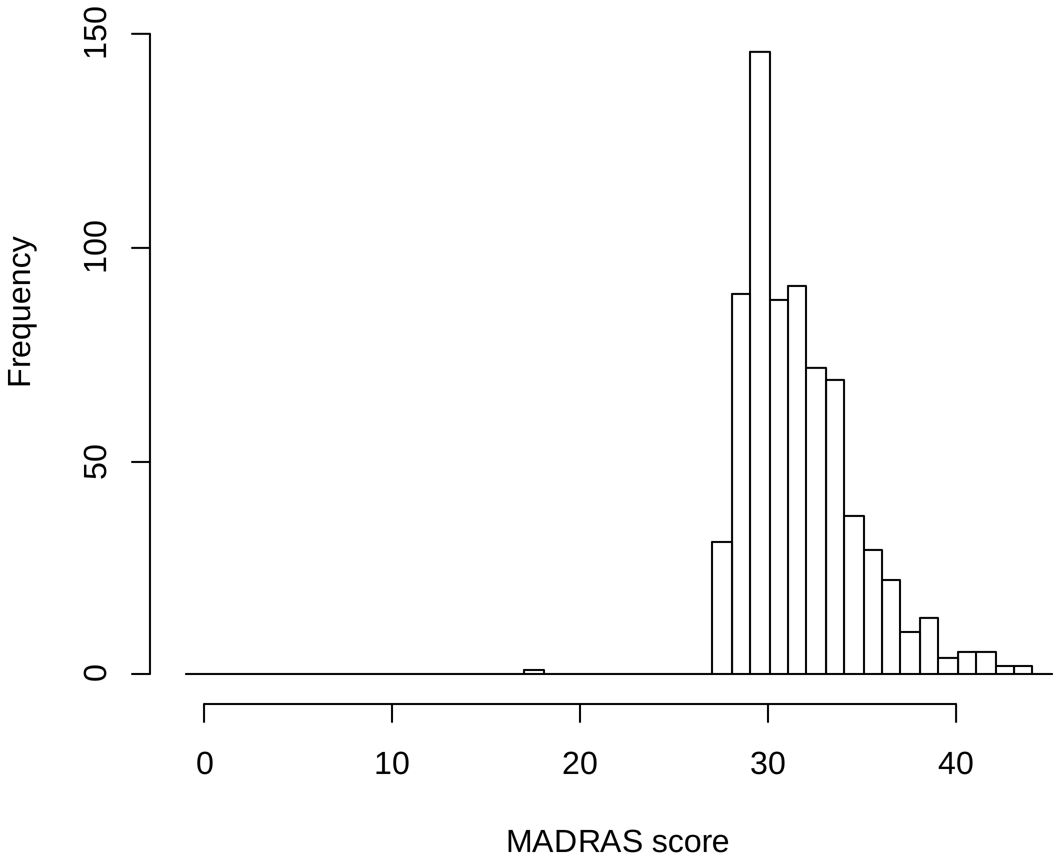
An example of fraud

Visit 1 (baseline)



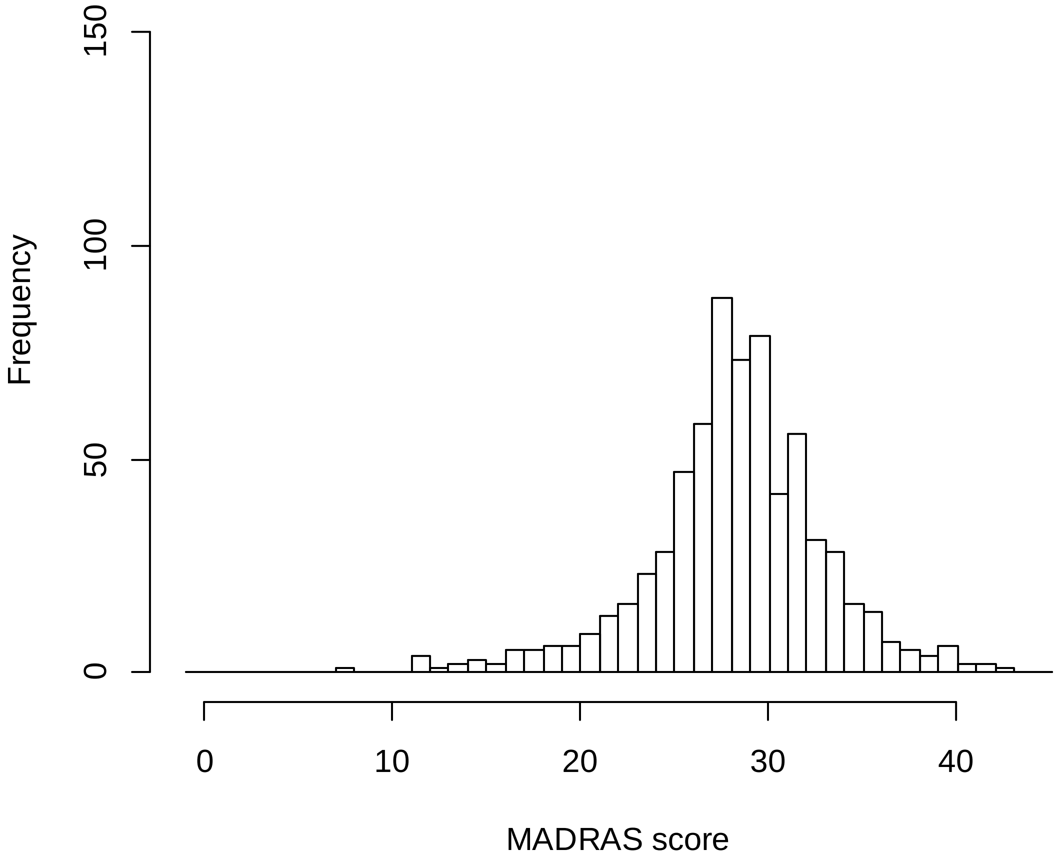
An example of fraud

Visit 2 (run-in)



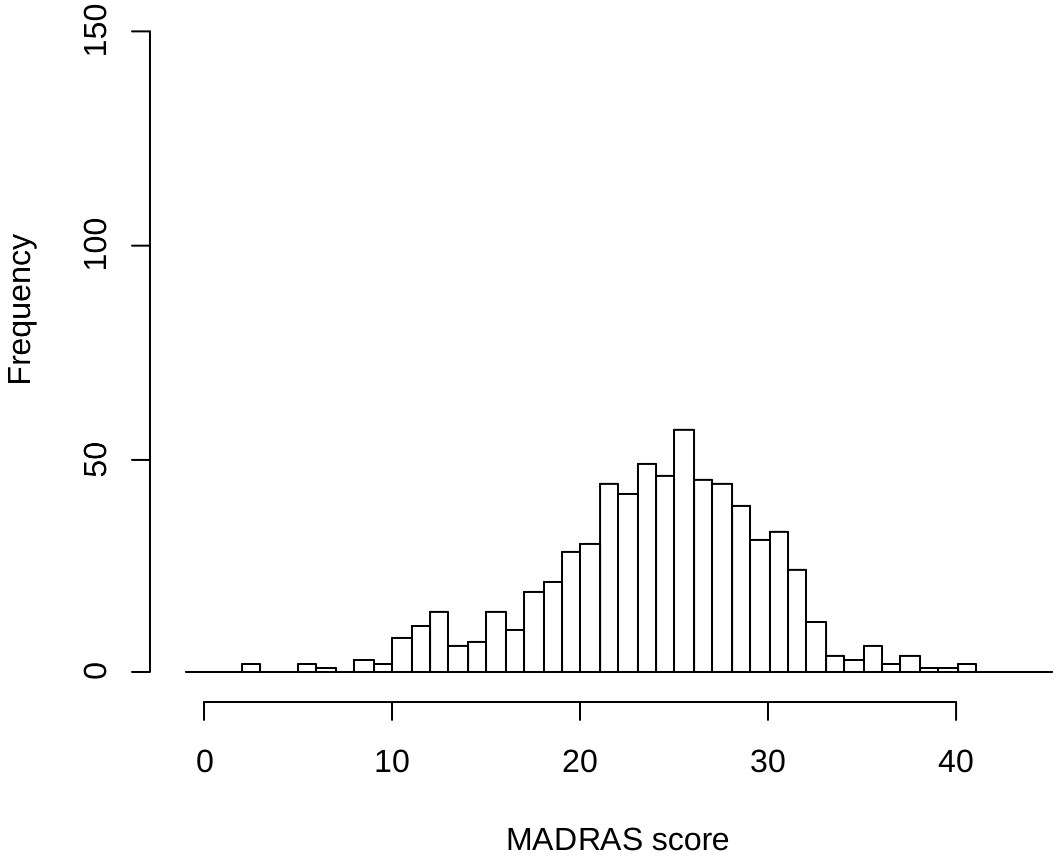
An example of fraud

Visit 3 (run-in)

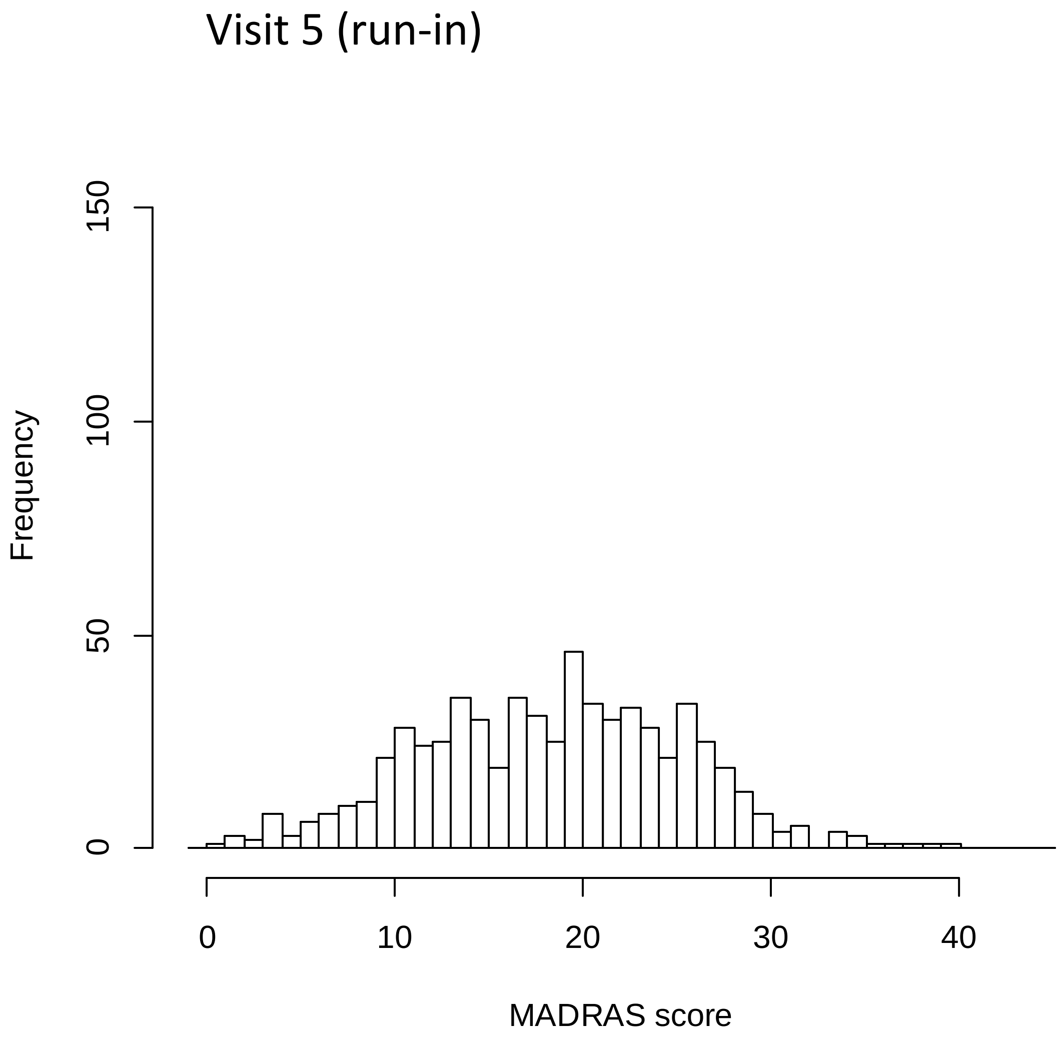


An example of fraud

Visit 4 (run-in)

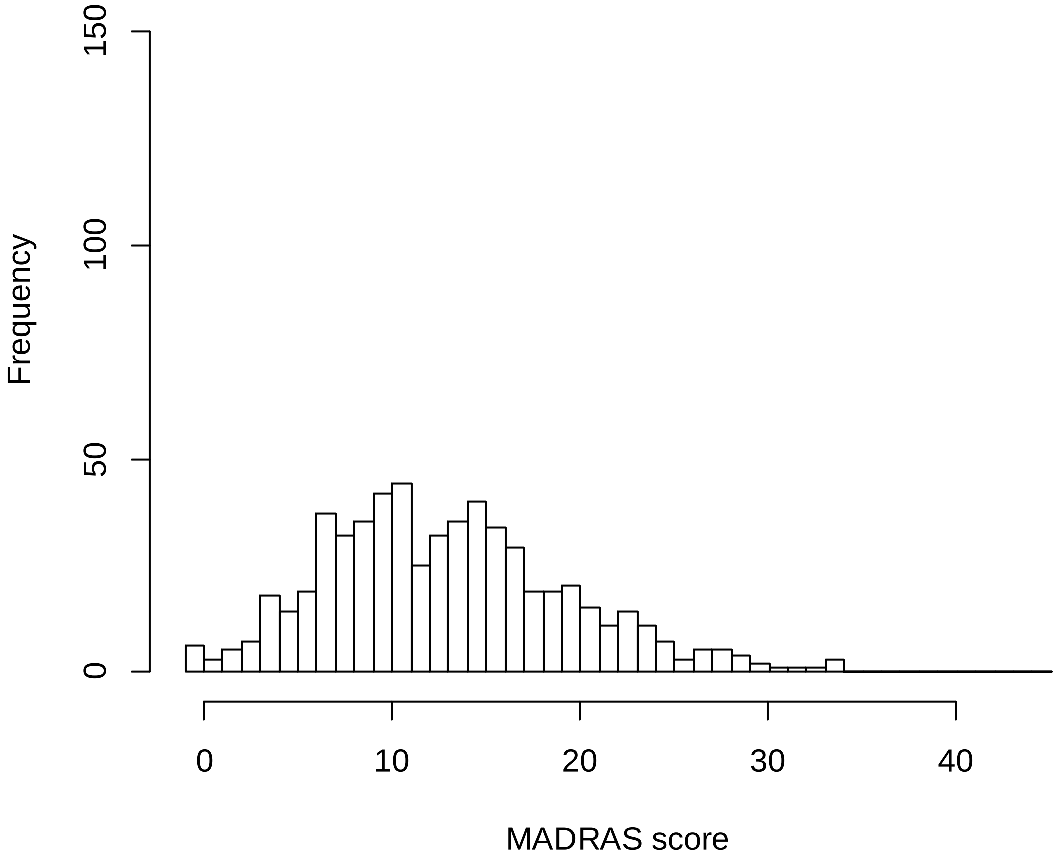


An example of fraud



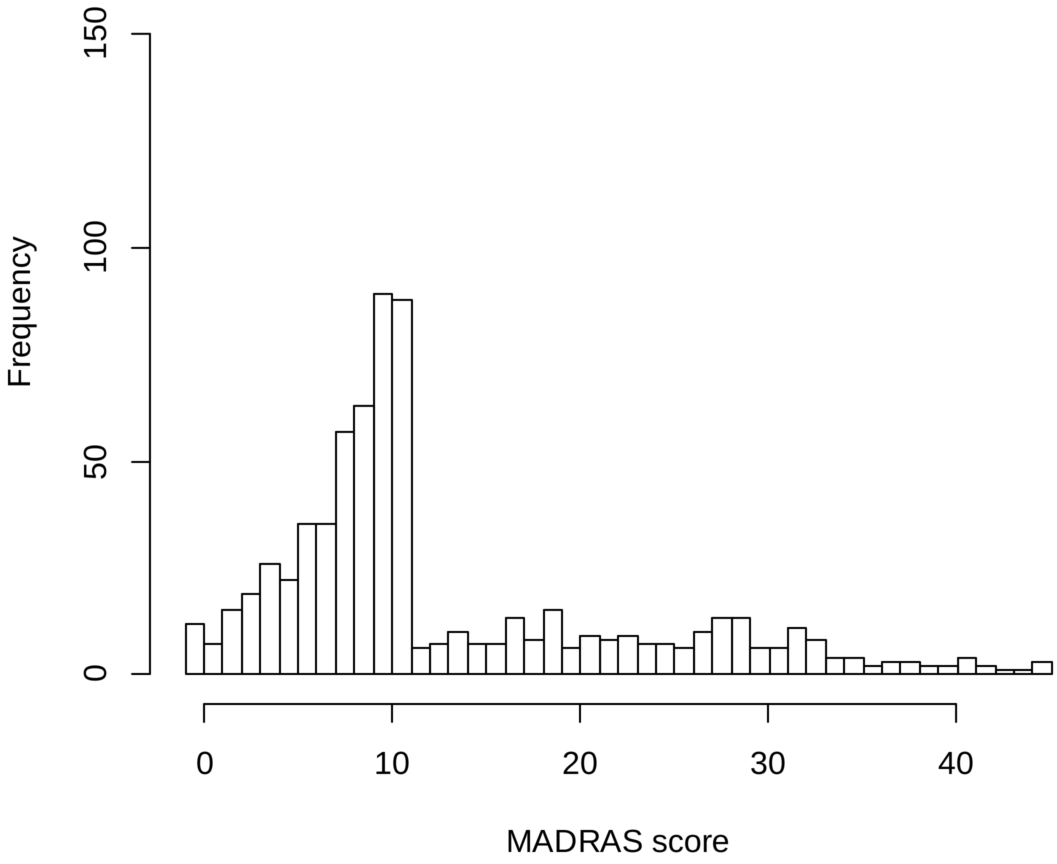
An example of fraud

Visit 6 (run-in)



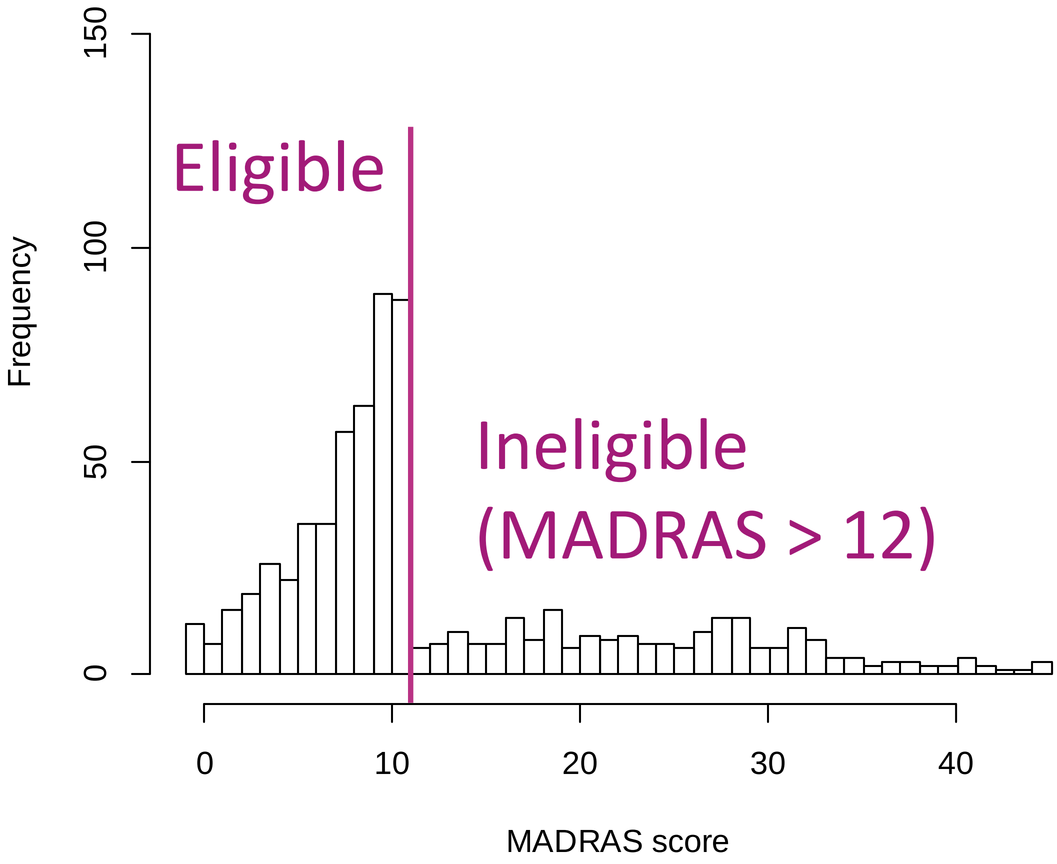
An example of fraud

Visit 7 (eligibility)



An example of fraud

Visit 7 (eligibility)



Continuum from errors to fraud

Type	Typical examples	Intent
Errors	Technical problems *(e.g. miscalibrated thermometers)*	Unintentional

Continuum from errors to fraud

Type	Typical examples	Intent
Errors	Technical problems *(e.g. miscalibrated thermometers)*	Unintentional
Sloppiness	Incorrect reporting *(e.g. under-reporting of AEs)*	Limited awareness

Continuum from errors to fraud

Type	Typical examples	Intent
Errors	Technical problems *(e.g. miscalibrated thermometers)*	Unintentional
Sloppiness	Incorrect reporting *(e.g. under-reporting of AEs)*	Limited awareness
Tampering	Fabricated data *(e.g. propagation of blood pressure)*	Deliberate

Continuum from errors to fraud

Type	Typical examples	Intent
Errors	Technical problems *(e.g. miscalibrated thermometers)*	Unintentional
Sloppiness	Incorrect reporting *(e.g. under-reporting of AEs)*	Limited awareness
Tampering	Fabricated data *(e.g. propagation of blood pressure)*	Deliberate
Fraud	Falsified data *(e.g. modification of eligibility criteria)*	Intention to cheat

10 years of R&D

IDDI
ADDING VALUE TO CLINICAL DATA

gsk GlaxoSmithKline Vaccines

A.E.R.O.

universiteit hasselt

sanofi aventis

TRIO Translational Research In Oncology

INSTITUT DE STATISTIQUE
CORE
UNIVERSITE CATHOLIQUE DE LOUVAIN

Bristol-Myers Squibb
Together we can prevail.®

IRIDIA
Institut de Recherches Interdisciplinaires
et de Développements en Intelligence Artificielle

Roche

10 years of R&D

IDDI
ADDING VALUE TO CLINICAL DATA

CluePoints
INTELLIGENT STATISTICAL MONITORING

gsk
GlaxoSmithKline
Vaccines

A.E.R.O.

universiteit
hasselt

sanofi aventis

TRIO
Translational
Research
In Oncology

INSTITUT DE STATISTIQUE
CORE
UNIVERSITE CATHOLIQUE DE LOUVAIN

Bristol-Myers Squibb
Together we can prevail.®

IRIDIA
Institut de Recherches Interdisciplinaires
et de Développements
en Intelligence Artificielle

Roche

Tomasz Burzykowski, PhD

Marc Buyse, ScD

Catherine Legrand, PhD

David Venet, PhD

Lieven Desmet, PhD

Catherine Timmermans, PhD

Erik Doffagne, MSc

Pierre-Andre Marchand, MSc

Sabrina El Bachiri, MSc

Suzanne Hackett, MSc

Marjolein Crabbe, PhD