東京大学大学院 医学系研究科 公共健康医学専攻 生物統計学分野
Department of Biostatistics, School of Public Health, Graduate School of Medicine, The University of Tokyo

東京大学 医学部 健康総合科学科 疫学・生物統計学教室
Department of Epidemiology and Biostatistics, School of Integrated Health Sciences, Faculty of Medicine, The University of Tokyo

Biostatistician - The best job of the 21st century?

Marc Buyse, ScD
San Francisco, CA

March 14, 2018

IDDI TRIAL DESIGN · RANDOMIZATION · DATA MANAGEMENT · BIOSTATISTICS
ADDING VALUE TO CLINICAL DATA

universiteit hasselt

Menu

Statistician… one of the « top jobs of 2018 »

1. Statistician… or biostatistician?
2. Statistician… or data scientist?
3. Statistician… or simply researcher?

I will illustrate my talk with three examples that I am familiar with. They are not meant to be representative or exhaustive…

BEST JOBS U.S.News & WORLD REPORT RANKINGS

U.S. News Top 10 Jobs of 2018

1. Software Developer
2. Dentist
3. Physician Assistant
4. Nurse Practitioner
5. Orthodontist
6. Statistician
7. Pediatrician
8. Obstetrician and Gynecologist (tie)
9. Oral and Maxillofacial Surgeon (tie)
10. Physician (tie)

https://money.usnews.com/careers/best-jobs/rankings/the-100-best-jobs

BEST JOBS U.S.News & WORLD REPORT RANKINGS

U.S. News Top 10 Jobs of 2018

	# jobs
1. Software Developer	250,000
2. Dentist	23,000
3. Physician Assistant	40,000
4. Nurse Practitioner	56,000
5. Orthodontist	1,100
6. Statistician	**12,400**
7. Pediatrician	5,300
8. Obstetrician and Gynecologist (tie)	3,900
9. Oral and Maxillofacial Surgeon (tie)	1,200
10. Physician (tie)	8,400

https://money.usnews.com/careers/best-jobs/rankings/the-100-best-jobs

BEST JOBS U.S.News & WORLD REPORT RANKINGS

U.S. News Top 10 Jobs of 2018

		# jobs	salary ($)
1.	Software Developer	250,000	100,000
2.	Dentist	23,000	153,000
3.	Physician Assistant	40,000	102,000
4.	Nurse Practitioner	56,000	101,000
5.	Orthodontist	1,100	208,000
6.	Statistician	12,400	85,000
7.	Pediatrician	5,300	167,000
8.	Obstetrician and Gynecologist (tie)	3,900	208,000
9.	Oral and Maxillofacial Surgeon (tie)	1,200	208,000
10.	Physician (tie)	8,400	197,000

https://money.usnews.com/careers/best-jobs/rankings/the-100-best-jobs

The New York Times

TECHNOLOGY

For Today's Graduate, Just One Word: Statistics

By STEVE LOHR AUG. 5, 2009

DATA

Data Scientist: The Sexiest Job of the 21st Century

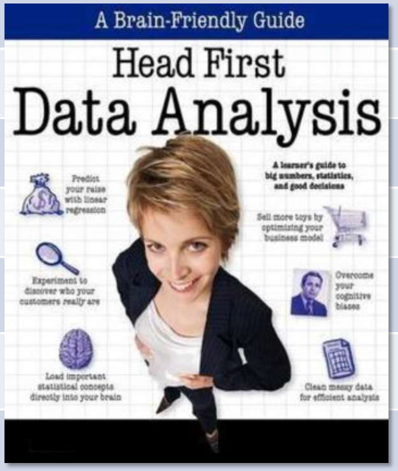
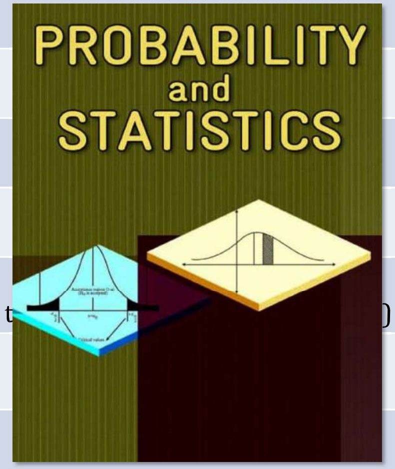
by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

Someone seeks help to analyze data…

Data Scientists…	Statisticians…
say the data look interesting	say there was no proper design
make interesting findings	reject null hypotheses
make interesting findings	fail to reject null hypotheses
believe that more data means less errors	believe that more data means more errors
do not pretend they understand what they do	pretend they understand what they do (but you don't)
generate statements that look really interesting but are probably untrue	"generate statements that are probably true and definitely useless" *

Statisticians have a « marketing » problem

Data Scientists…	Statisticians…
	

* *Stephen Senn, http://www.senns.demon.co.uk/wdict.html*

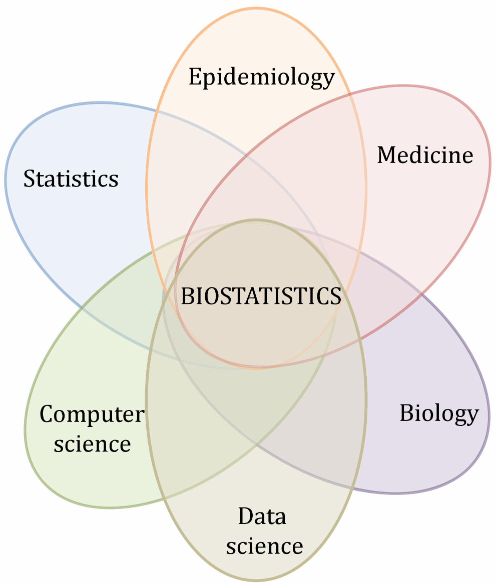
Two complementary professions

Data Scientists needed for...	Statisticians needed for...
Discovery (finding the unexpected)	Testing (confirming the anticipated)
Exploring big, poorly structured, messy data	Designing controlled experiments to generate reliable data
Correcting errors using future data	Controlling errors using current data
Implementing efficient algorithms for machine learning	Generating reliable evidence for human learning

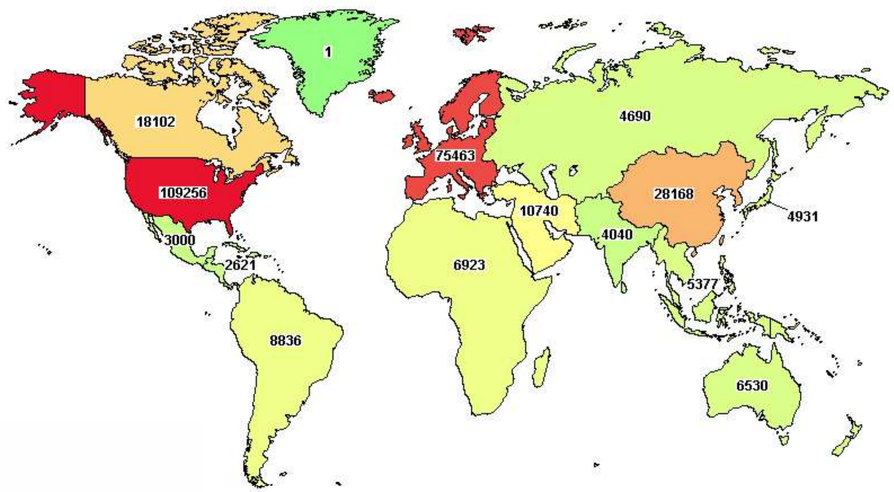
Can the two cooperate?

Statistician… or biostatistician?

A multidisciplinary profession



> 235,000 on-going clinical trials worldwide



Source: clinicaltrials.gov

Development cost per new drug > 1 BN $

Ref: Scanning et al, Nat Rev Drug Discov 11: 191 (2012).

Development cost per new drug > 1 BN $

Clinical development

- too lengthy
- too costly
- too risky
- inadequate for precision medicine
- inadequate for personalized medicine

Surrogate Endpoints (too lengthy)

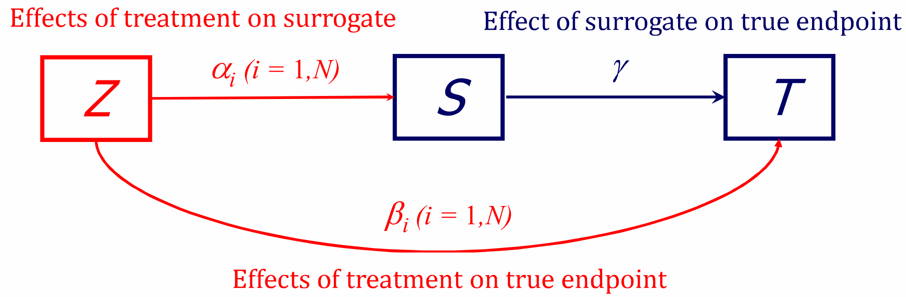
Central Statistical Monitoring (too costly, too risky)

Generalized Pairwise Comparisons (inadequate for personalized medicine)

Surrogate endpoints

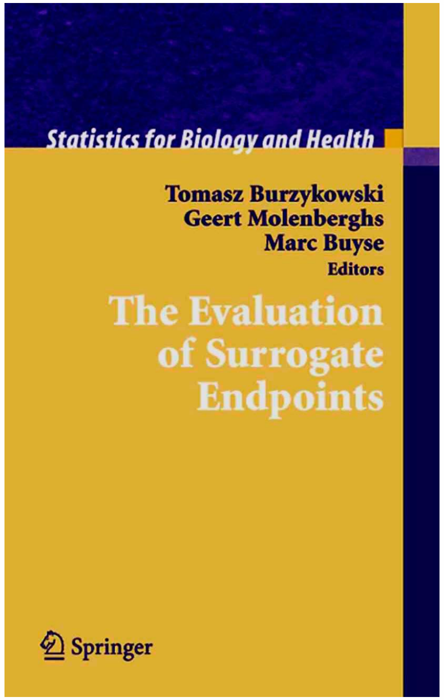
- **Clinical context**: can new treatments be assessed using earlier endpoints (or biomarkers) instead of later clinical endpoints?
- **Potential**: months or years of development time gained
- **Statistical challenges**:
  - Reliable predictions are hard!
  - Complex modeling
  - Association *vs.* causation

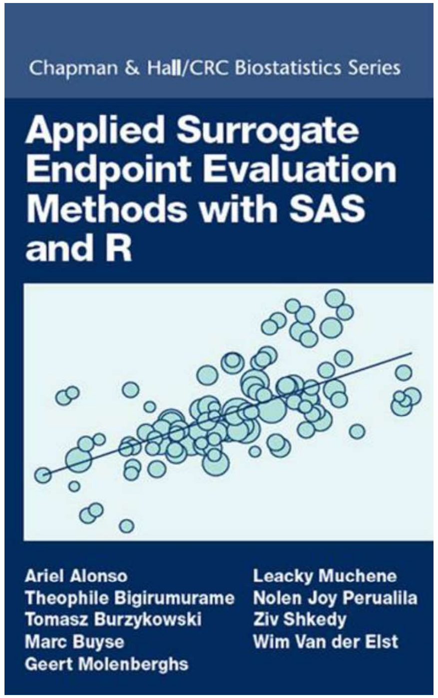
Evaluation of surrogate endpoints



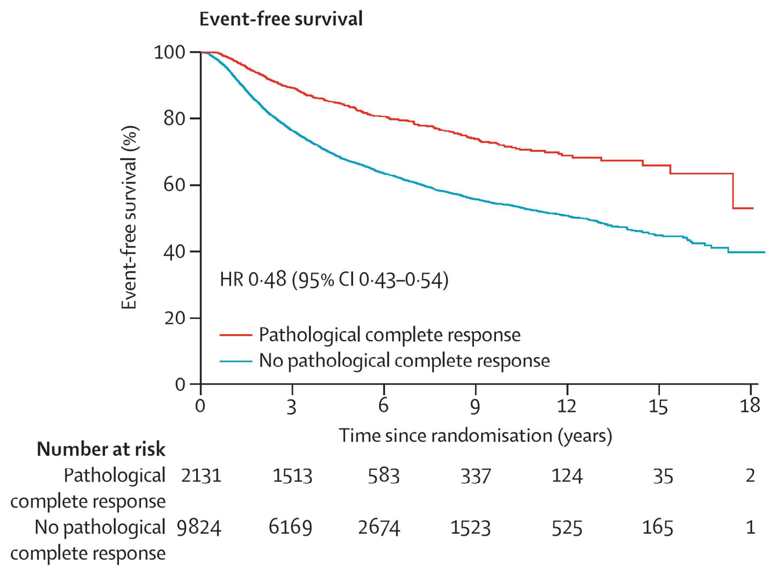
S and *T* must be correlated ("individual- level surrogacy")

α and β must be correlated ("trial-level surrogacy")



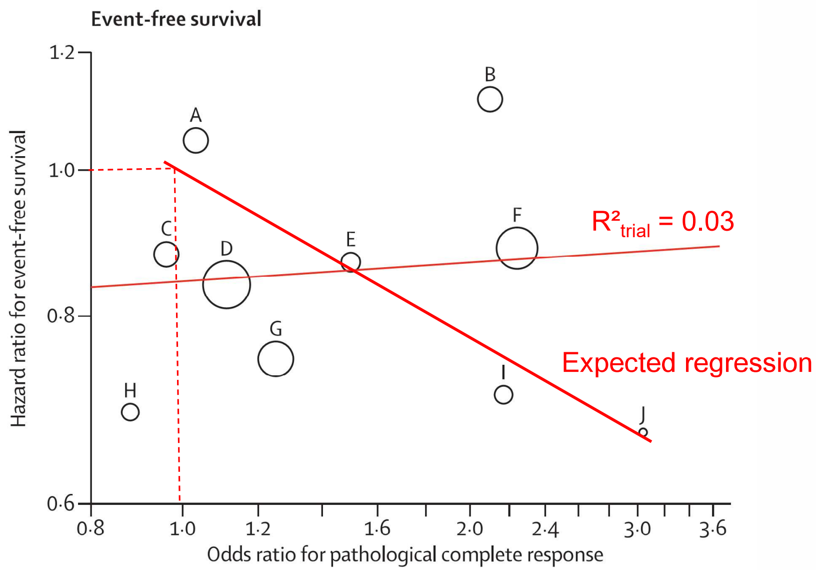


Is pathological response a surrogate for survival?



Ref: Cortazar et al, Lancet 2014.

Is pathological response a surrogate for survival?



Ref: Cortazar et al, Lancet 2014.

Surrogate endpoints

Depth

Hierarchical models
(G Molenberghs)

Errors-in-variables models
(T Burzykowski)

Copulas
(T Burzykowski)

Information theory
(A Alonso)

Bayesian models
(Z Shkedy)

Causal inference
(A Alonso)

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Surrogate endpoints

Breadth

Initial datasets	Cooperative Groups	Pharma	Agencies
Oncology	MAGIC – colorectal (*P Piedbois*)	BMS – Lung	FDA
Ophthalmology	GASTRIC – stomach (*K Oba, X Paoletti*)	Roche – Breast	IQWiG
Schizophrenia	ARCAD – colorectal (*D Sargent*)	Novartis – AML	
	ICECaP – prostate (*C Sweeney*)	Boehringer – Mesothelioma	
	EORTC – melanoma (*S Suciu*)		

IQWiG Institute for Quality and Efficiency in Health Care

Surrogate endpoints

OXFORD ACADEMIC

JNCI JOURNAL of the NATIONAL CANCER INSTITUTE

REVIEW

Disease-Free Survival as a Surrogate for Overall Survival in Adjuvant Trials of Gastric Cancer: A Meta-Analysis

Koji Oba, Xavier Paoletti, Steven Alberts, Yung-Jue Bang, Jacqueline Benedetti, Harry Bleiberg, Paul Catalano, Florian Lordick, Stefan Michiels, Satoshi Morita, Yasuo Ohashi, Jean-pierre Pignon, Philippe Rougier, Mitsuru Sasako, Junichi Sakamoto, Daniel Sargent, Kohei Shitara, Eric Van Cutsem, Marc Buyse, Tomasz Burzykowski; on behalf of the GASTRIC group

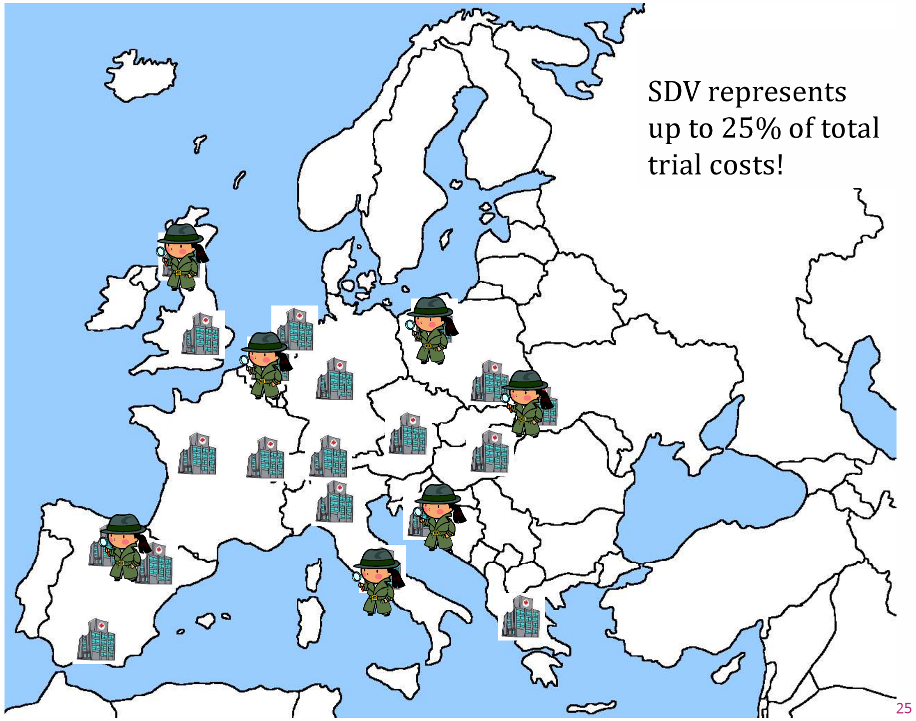
Manuscript received February 12, 2013; revised July 25, 2013; accepted July 25, 2013.

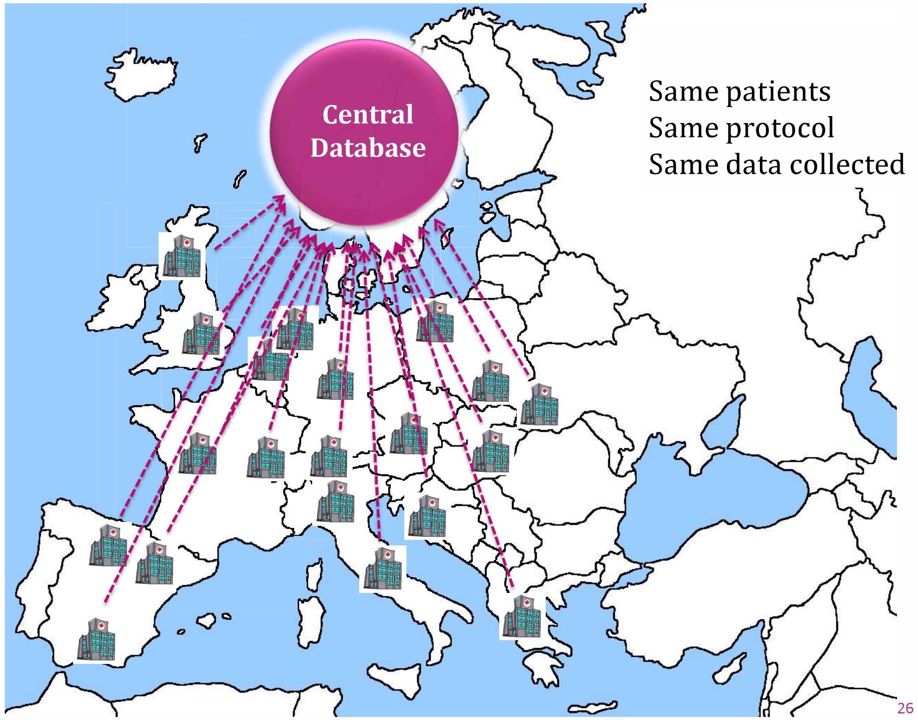
Correspondence to: Koji Oba, PhD, Translational Research and Clinical Trial Center, Hokkaido University Hospital, Kita 14, Nishi 5, Kita-ku, Sapporo, Hokkaido 0608648, Japan (e-mail: k.oba@huhp.hokudai.ac.jp).

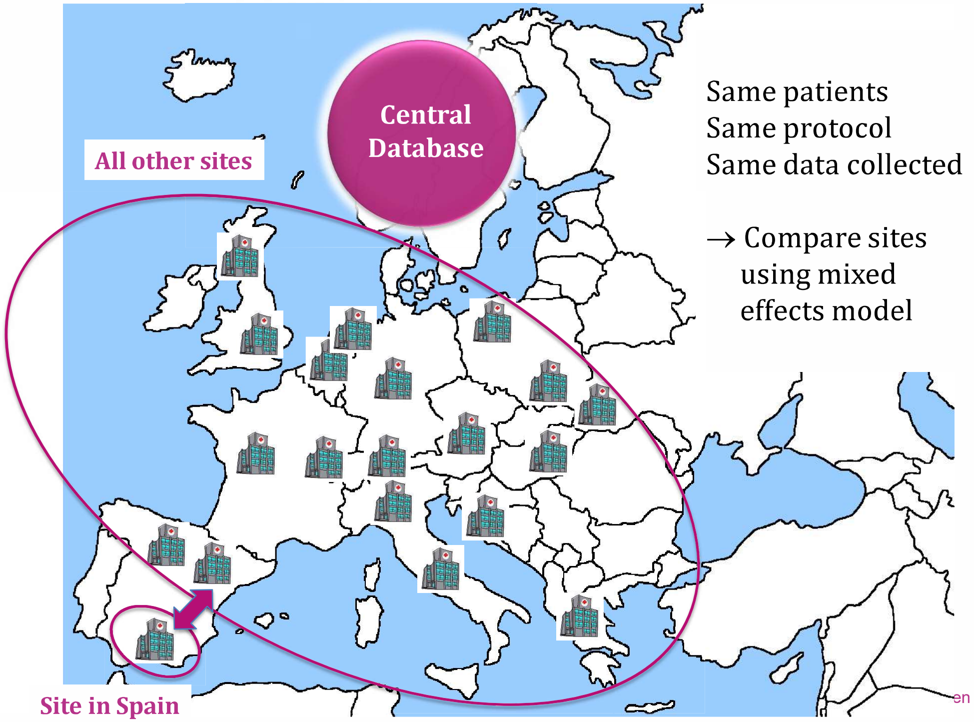
Statistician… or data scientist?

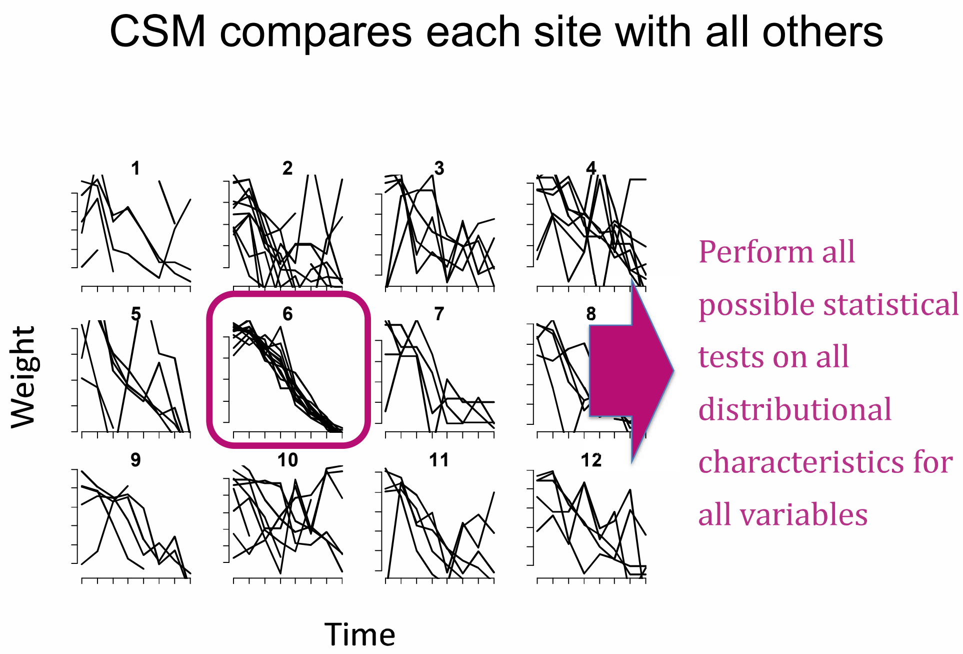
Central statistical monitoring

- **Clinical context**: can central statistical monitoring help eliminate source data verification (SDV) and target on-site monitoring visits?
- **Potential**: cut clinical trial budgets by up to 25%
- **Statistical challenges**:
  - Use data consistency across sites as proxy for quality
  - Allow for natural / expected variability
  - Translate statistical findings into actionable signals


SDV represents
up to 25% of total
trial costs!
25


Central
Database
Same patients
Same protocol
Same data collected
26


Central
Database
All other sites
Same patients
Same protocol
Same data collected
→ Compare sites
using mixed
effects model
Site in Spain


CSM compares each site with all others
Weight
Time
1
2
3
4
5
6
7
8
9
10
11
12
Perform all
possible statistical
tests on all
distributional
characteristics for
all variables

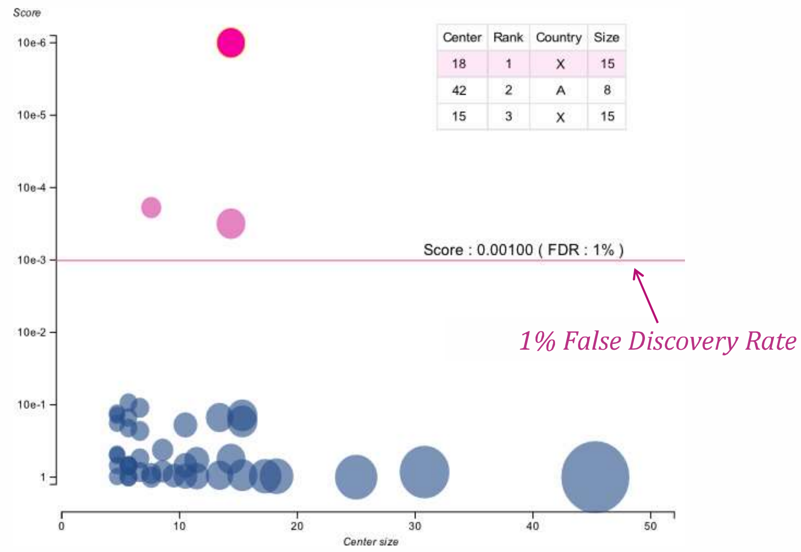
$(S \times T)$ *P*-value matrix

	Var$_1$			Var$_2$			…	Var$_V$
Site	Test$_a$	…	Test$_c$	Test$_d$	…	Test$_f$	…	…
1	p_{11}	p_{12}	…				…	p_{1T}
2	p_{21}	…						…
…	…							…
S	p_{S1}	…					…	p_{ST}

Score sites $\tilde{p}_k = [p_{11} \cdot p_{12} \cdots p_{ST}]^{1/T}$

Resampling $s_k = P[x_k \leq \tilde{p}_k]$

Example: 3 centers with highly atypical data



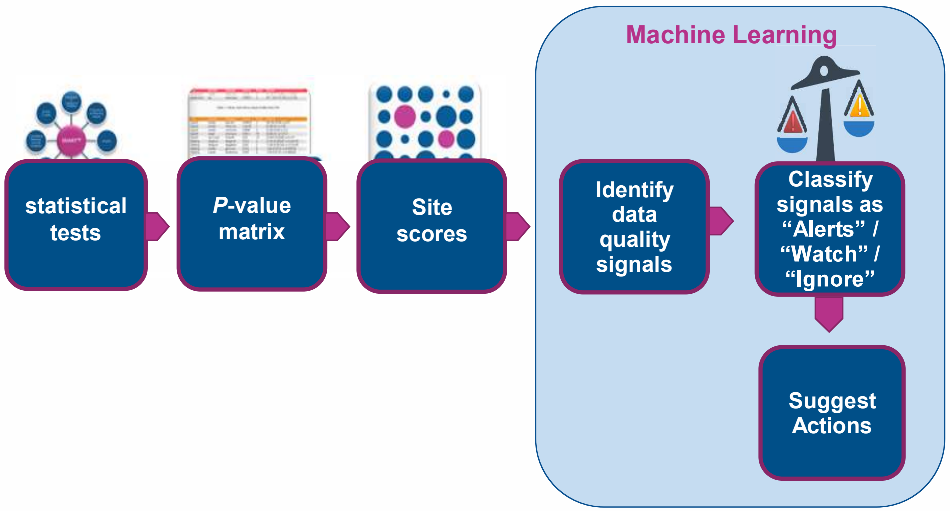
Operating characteristics

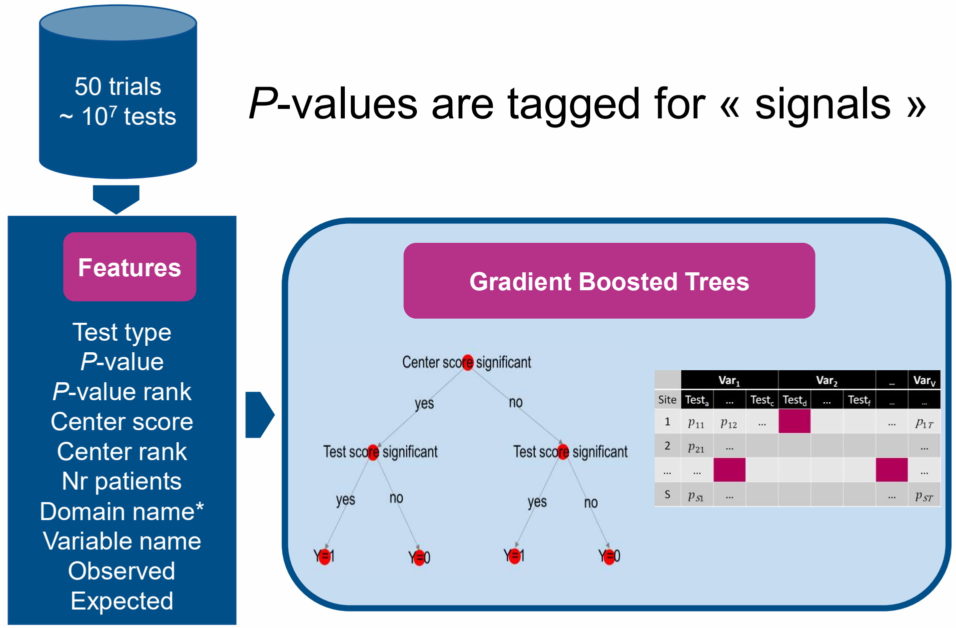
Unsupervised statistical monitoring for the detection of atypical data in multicenter clinical trials

Journal Title
XX(X):1–6
©The Author(s) 2018
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

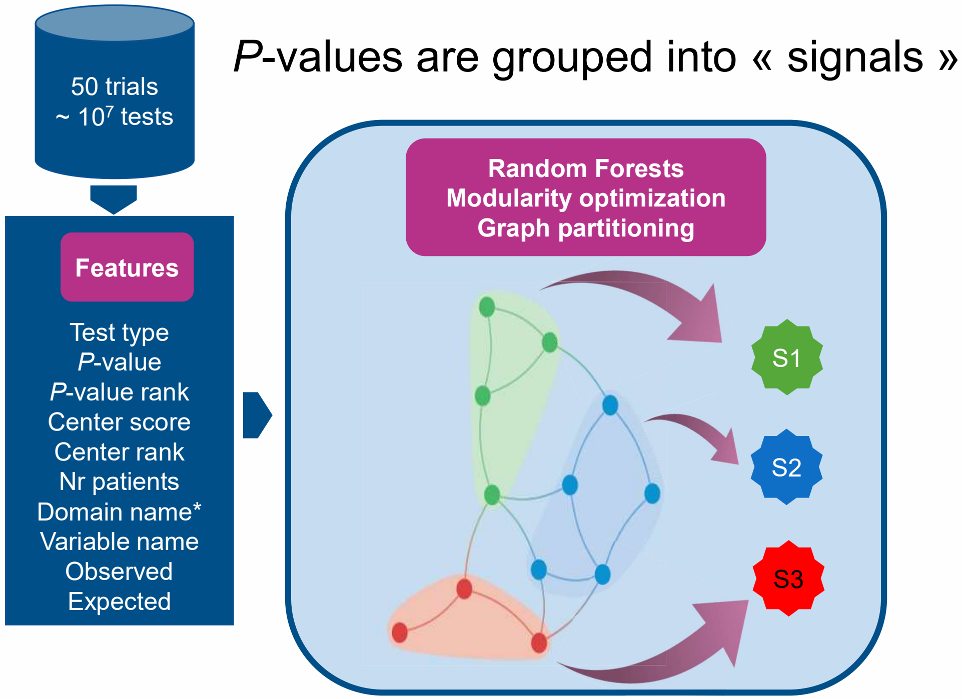
Laura Trotta, PhD[1,*], Yuusuke Kabeya, MSc[2,*], Marc Buyse, ScD[3,4], Erik Doffagne, MSc[1], David Venet, PhD[5], Lieven Desmet, PhD[6], Tomasz Burzykowski, PhD[7,8], Akira Tsuburaya, MD[9], Kazuhiro Yoshida, MD[10], Yumi Miyashita[11], Satoshi Morita, PhD[12], Junichi Sakamoto, MD[11,13], Paurush Praveen, PhD[1,*] and Koji Oba, PhD[2,*].

Machine learning helps create « signals »





Domain = Demography, Physical Examination, Exposure, Adverse Events, Outcome, …



CluePoints for quality control

THE LANCET Oncology

" The CluePoints® statistical monitoring software (CluePoints Inc., Cambridge, USA) was applied to check the quality and consistency of the clinical data across all participating centres. ClluePoints® did not detect atypical data patterns at some of the participating centres that could have had a significant impact on the efficacy and safety analyses of the trial. "

Ref: Tsuburaya et al. Lancet Oncology 2014.

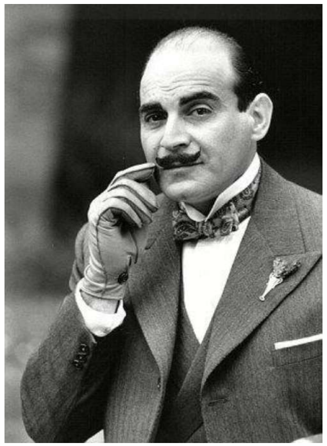
CluePoints for « detective » work

Gastric Cancer (2016) 19:21–23
DOI 10.1007/s10120-015-0555-3

EDITORIAL

A Hercule Poirot of clinical research

Junichi Sakamoto[1]



Ref: Sakamoto. Gastric Cancer 2016.

Statistician… or simply researcher?

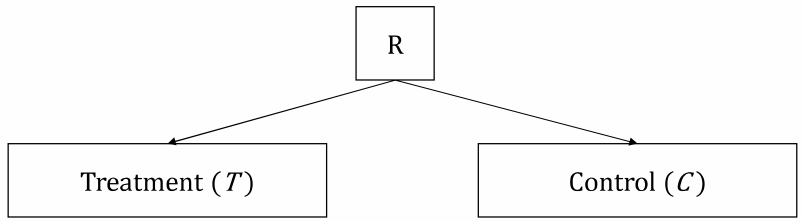
Generalized pairwise comparisons

- **Clinical context:** can all outcomes measured in randomized clinical trials be used in a single, patient-relevant, measure of treatment effect?
- **Potential:** pave the way to personalized medicine
- **Statistical challenges:**
  - Paradigm shift away from population parameters
  - Only tractable analytically in simplest cases
  - Interpretational difficulties *e.g.* with censoring

Current analyses of randomized clinical trials

- A single (« primary ») endpoint drives decision-making
- Composite endpoints consider time to *first* event, instead of time to *most relevant* endpoint
- Other (« secondary ») endpoints are analyzed descriptively
- Safety endpoints / adverse are informally balanced against efficacy, resulting in debatable risk / benefit analyses
- Patient preferences are not formally taken into account

Randomized clinical trial

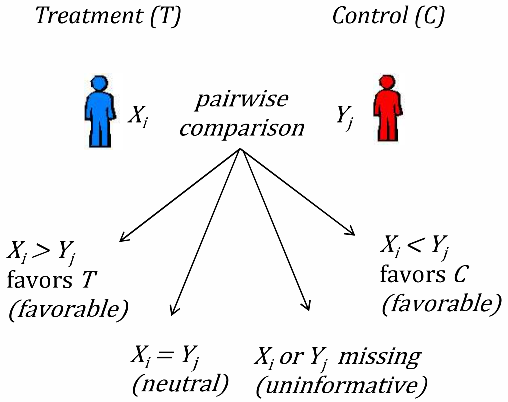


Let X_i be the outcome of i^{th} subject in T $(i = 1, \ldots, n)$

Let Y_j be the outcome of j^{th} subject in C $(j = 1, \ldots, m)$

Pairwise comparisons

Assume a continuous outcome measure



Ref: Buyse, Stat Med 29:3245, 2010.

Mann-Whitney form of the Wilcoxon test

The Wilcoxon test statistic can be derived from consideration of all possible pairs of subjects, one from each treatment group.

Let

$$U_{ij} = \begin{cases} +1 & \text{if } X_i > Y_j \\ -1 & \text{if } X_i < Y_j \\ 0 & \text{otherwise} \end{cases}$$

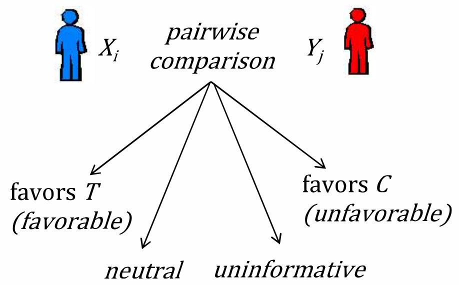
$$U = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}$$

The Wilcoxon-Mann-Whitney test statistic W can be written as

$$W = m \cdot n \cdot (1 - U)/2$$

Any single outcome measure

Now let X_i and Y_j be observed outcomes for any outcome measure (continuous, time to event, binary, categorical, …)



Binary outcome measure

Pairwise comparison	Pair is
$X_i = 1,\ Y_j = 0$	favorable
$X_i = 1,\ Y_j = 1$ or $X_i = 0,\ Y_j = 0$	neutral
$X_i = 0,\ Y_j = 1$	unfavorable
X_i or Y_j missing	uninformative

GPC test is equivalent to χ^2 test

Continuous outcome measure

Pairwise comparison	Pair is
$X_i - Y_j > \tau$	favorable
$\lvert X_i - Y_j \rvert \leq \tau$	neutral
$X_i - Y_j < -\tau$	unfavorable
X_i or Y_j missing	uninformative

$\tau = 0$ is Wilcoxon test

τ can be chosen to reflect clinical relevance

Time to event outcome measure

Pairwise comparison	Pair is
$X_i - Y_j > \tau$ or $X_i' - Y_j > \tau$	favorable
$\lvert X_i - Y_j \rvert \leq \tau$	neutral
$X_i - Y_j < -\tau$ or $X_i - Y_j' < -\tau$	unfavorable
otherwise	uninformative

$\tau = 0$ is Gehan test

τ can be chosen to reflect clinical relevance

Several prioritized outcome measures

Outcome with higher priority	Outcome with lower priority	Pair is
favorable		favorable
unfavorable		unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

Ref: Buyse, Stat Med 29:3245, 2010.

Thresholds of clinical relevance

Survival difference > 12 months	Survival difference ≤ 12 months	Pair is
favorable		favorable
unfavorable		unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

Ref: Buyse, Stat Med 29:3245, 2010.

Benefit / risk analyses

Survival	Serious toxicity (e.g. CTC grade 3/4)	Pair is
favorable		favorable
unfavorable		unfavorable
neutral or ?	favorable	favorable
neutral or ?	unfavorable	unfavorable
neutral or ?	neutral	neutral
?	?	?

Ref: Buyse, Stat Med 29:3245, 2010.

The net treatment benefit Δ

$$U_{ij} = \begin{cases} +1 & \text{if } (X_i, Y_j) \text{ pair is favorable} \\ -1 & \text{if } (X_i, Y_j) \text{ pair is unfavorable} \\ 0 & \text{otherwise} \end{cases}$$

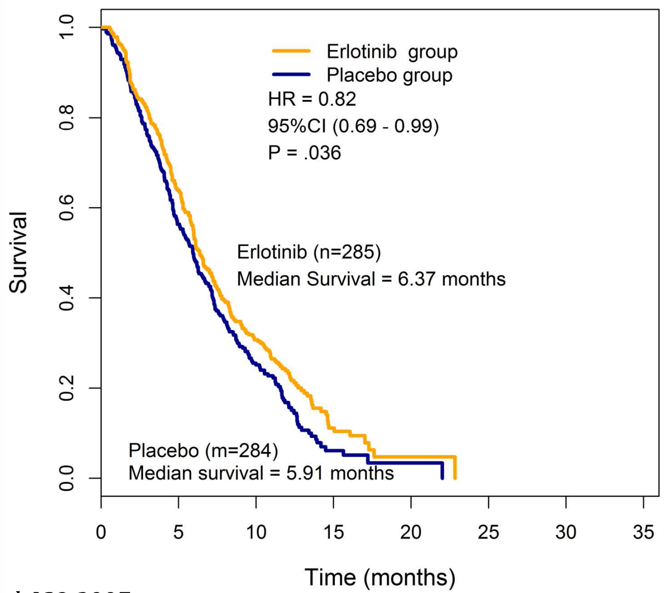
$$U = \frac{1}{m \cdot n} \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}$$

U is the difference between the proportion of favorable pairs and the proportion of unfavorable pairs. It is the « net treatment benefit », denoted Δ.

This measure is analogous to Pocock's « win ratio » (Δ is the « win difference »).

Ref: Pocock et al. Eur Heart J 33: 176, 2012.

Survival benefit of erlotinib

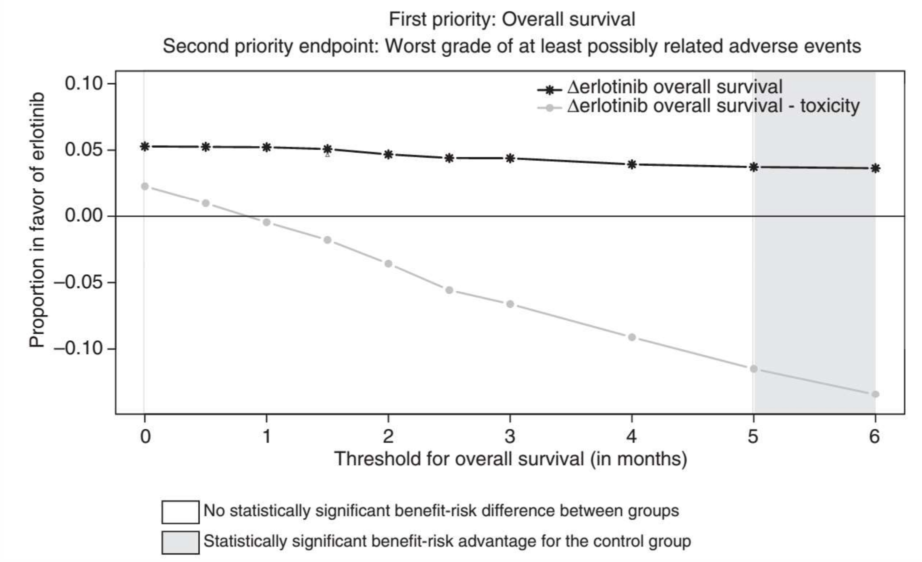


Ref: Moore et al. JCO 2007.

Toxicities of erlotinib

Worst grade related AE	Erlotinib group (n=282)	Placebo group (n=280)
Grade 1	48 (17%)	69 (24.6%)
Grade 2	118 (41.8%)	89 (31.8%)
Grade 3	29%	19%
Grade 4		
Grade 5	4 (1.4%)	3 (1.1%)

Prioritized outcomes: OS and worst toxicity



Ref: Peron et al. BJC 2015.

Personalized medicine

BENEFIT - Biostatistical Estimation of Net Effects For Individualization of Therapy

IDDI
ADDING VALUE TO CLINICAL DATA

EORTC
European Organisation for Research and Treatment of Cancer
AISBL International Non-Profit Association under Belgian law IVZW

Bristol-Myers Squibb

ISBA

Hospices Civils de Lyon

Lyon 1

UHASSELT I-BioStat KU LEUVEN
Interuniversity Institute for Biostatistics and statistical Bioinformatics

Personalized medicine

The quiet statisticians have changed our world - not by discovering new facts or technical developments but by changing the ways that we reason, experiment and form our opinions...

Ian Hacking

Acknowledgments

I gratefully acknowledge help and inspiration from the individuals mentioned in this talk, and many others. The job of a biostatistician is by nature collaborative. Which makes it one of the best jobs of the 21st century.