UNIVERSITÄTS KLINIKUM FREIBURG

Lecture 5
Critique of the quality of medical science - reporting and initiatives for improvement

Willi Sauerbrei, Edwin Kipruto

Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center -University of Freiburg, Germany

WS 2025/26

Improve research in the health sciences

STRATOS (Statistical Analysis)

EQUATOR (Reporting)

PROGRESS (Prognosis Research)

EBM – Systematic reviews and meta-analyses required

Study registration

Data sharing

All trials campaign

Critique of the quality of medical science -

Bad reporting is a key problem

How should medical science change?

The scandal of poor medical research

DOUGLAS G ALTMAN
BMJ, 1994

We need less research, better research, and research done for the right reasons

***The Lancet* Research: Increasing Value, Reducing Waste Series, 2014**

In 2009, we published a Viewpoint by Iain Chalmers and Paul Glasziou called "Avoidable waste in the production and reporting of research evidence", which made the extraordinary claim that as much as 85% of research investment was wasted.

Our belief is that research funders, scientific societies, school and university teachers, professional medical associations, and scientific publishers (and their editors) can use this Series as an opportunity to examine more forensically why they are doing what they do—the purpose of science and science communication—and whether they are getting the most value for the time and money invested in science.

Kleinert and Horton 2014

Of 1575 reports about cancer prognostic markers published in 2005, 1509 (96%) detailed at least one significant prognostic variable. However, few identified biomarkers have been confirmed by subsequent research and few have entered routine clinical practice. This Pattern — initially promising findings not leading to improvements in health care — has been recorded across biomedical research. So why is research that might transform health care and reduce health problems not being successfully produced?

Global biomedical and public health research involves billions of dollars and millions of people. In 2010, expenditure on life sciences (mostly biomedical) research was US$240 billion. The USA is the largest funder, with about $70 billion in commercial and $40 billion in governmental and non-profit funding annually, representing slightly more than 5% of US health-care expenditure. Although this vast enterprise has led to substantial health improvements, many more gains are possible if the waste and inefficiency in the ways that biomedical research is chosen, designed, done, analysed, regulated, managed, disseminated, and reported can be addressed

Macleod et al. 2014

Research: increasing value, reducing waste 2

Increasing value and reducing waste in research design, conduct, and analysis

John P A Ioannidis, Sander Greenland, Mark A Hlatky, Muin J Khoury, Malcolm R Macleod, David Moher, Kenneth F Schulz, Robert Tibshirani

Therapy studies are often too small

The Continuing Unethical Conduct of Underpowered Clinical Trials

Halpern et al, JAMA 2002: 288: 358-362

Why Most Published Research Findings Are False

John P.A. Ioannidis

PLoS Medicine 2005, 8: 696-701.

Confusion Caused by Bad Quality of Design, Analysis and Reporting of Many Studies

Willi Sauerbrei

Bier H(ed): Current Research in Head and Neck Cancer.
Adv Otorhinolaryngol. Basel, Karger, 2005, vol 62,pp 184-200

Almost all articles on cancer prognostic markers report statistically significant results

Panayiotis A. Kyzas[a], Despina Denaxa-Kyza[a], John P.A. Ioannidis[a,b,c,]*

[a]*Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece*
[b]*Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece*
[c]*Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts-New England Medical Center, Boston, USA*

European Journal of Cancer 2007; 43: 2559-2579

Reporting research is as important a part of a study as its design or analysis

Jordan, K.P. & Lewis, M. (2009) Improving the quality of reporting of research studies. *Musculoskeletal Care*, **7**, 137-142

Publish houses of brick, not mansions of straw

Papers need to include fewer claims and more proof to make the scientific literature more reliable, warns **William G. Kaelin Jr.**

Nature (2017) Vol 545, 387

I worry about sloppiness in biomedical research: too many published results are true only under narrow conditions, or cannot be reproduced at all. The causes are diverse […].

Earlier statements about poor quality of research

- Methodology

„…less than 1% of research workers clearly apprehend the rationale of the statistical techniques they commonly invoke“

Hogben L., 1950

„…almost any volume of a medical journal contains faults that can be detected by first-year students after only three or four hours‘ guidance in the scrutiny of reports.“

Mainland D., 1952

- Reporting

„…incompleteness of evidence is not merely a failure to satisfy a few highly critical readers. It not infrequently makes the data that are presented of little or no value.“

Mainland D., 1938

„…the idea is to give all of the information to help others to judge the value of your contribution; not just the information that leads to judgement in one particular direction or another.“

Feynman R., 1974

For further references see Altman and Simera, 2016

Deficiencies in medical research … and the result

A clinical example - **Question in 2011**

Is it important to measure Ki-67 in breast cancer?
Is it a prognostic and/or a predictive marker?

Ki-67 is a proliferation marker

Biological arguments for its prognostic/predictive value

Question is discussed for a long time:

More than 25 studies BEFORE 2000

Systematic Reviews 1

1. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12 155 patients (de Azambuja, E. et al. BJC 2007)

   - 68 studies were identified and 46 studies including 12,155 patients were evaluable for our meta-analysis
   - Our meta-analysis suggests that Ki-67/MIB-I positivity confers a higher risk of relapse and a worse survival in patients with early breast cancer

   … further comments later

Systematic Reviews 2

2. Proliferation markers and survival in early breast cancer: A systematic review and meta-analysis of 85 studies in 32,825 patients (Stuart-Harris, R. et al. The Breast 2008)

   - Ki-67: 43 studies, 15,790 patients

   - Some evidence for publication bias
   - Whether these proliferation markers provide additional prognostic information to commonly used prognostic indices remains unclear (Stuart-Harris et al. 2008)

Systematic Reviews 3

3. Ki-67 in breast cancer: prognostic and predictive potential (Yerushalmi, R. et al. Lancet Oncol 2010)

   - The use of Ki67 as a predictive and prognostic marker in breast cancer has been widely investigated. ..issues of quality testing …makes comparisons difficult. .. An accurate analysis … develop a standard methodology …
   - Although some intriguing preliminary findings exist … that high Ki67 levels might predict … indicate that Ki67 might have a valuable role

Certainly, research can and has to improve

Other markers – situation better?

Systematic Review for 260 studies, 130 different markers

Neuroblastoma (Riley et al. 2003)

Published reports for 3 prognostic factors

Marker name	Papers	OS and DFS reports	Total successful estimates	Estimates univar./multivariate	Different cut-off groups	Different stage groups	Different age groups
MYC-N	151	194	94	77/17	9	9	4
CD44	8	8	3	3/0	1	1	2
MDR	16	30	16	13/3	8	3	3

Result of bad primary research

Bladder cancer

"After 10 years of research, evidence is not sufficient to conclude whether changes in P53 act as markers of outcome … decade of research … is frustrating"

Coronary disease

"Multiple types of reporting bias, and publication bias, … association between CRP and prognosis sufficiently uncertain that no clinical practice recommendations can be made."

Osteosarcoma

"93 papers were studied ….Only 7 papers were of sufficient quality to analyze. .. Because of heterogeneity of the studies, pooling results is hardly possible. There is a need for standardization of studies and reports"

General

"As a consequence of the poor quality of research, prognostic markers may remain under investigation for many years after initial studies without any resolution of the uncertainty. Multiple separate and uncoordinated studies may actually delay the process of defining the role of prognostic markers".

Methodological issues as one relevant component to improve research

Design

Data quality and completeness

Statistical Analysis Plan

Analysis of single studies

Reporting

Prospectively planned (meta-)analyses

Collaboration

- between disciplines
- between study groups

Guidance for analysis

Study registries

Design and analysis of single studies

Which phase?

→ PROGRESS

Write a protocol and a SAP!

Sufficient sample size?

Continuous variables –

Categorization by using cutpoints introduce severe problems

Models based on cut-points: problems!

Cut-points are still popular in clinical and epidemiological research

Use of cut-points in a model gives a step function

How many cut-points?

Where should the cut-points be put?

Poor approximation to the true relationship

Almost always fits the data less well than a suitable continuous function

A very bad combination

Analysis with the optional cutpoint approach and selective reporting

…fishing for significance

On Fishing for Significance and Statistician‘s Degree of Freedom in the Era of Big Molecular Data

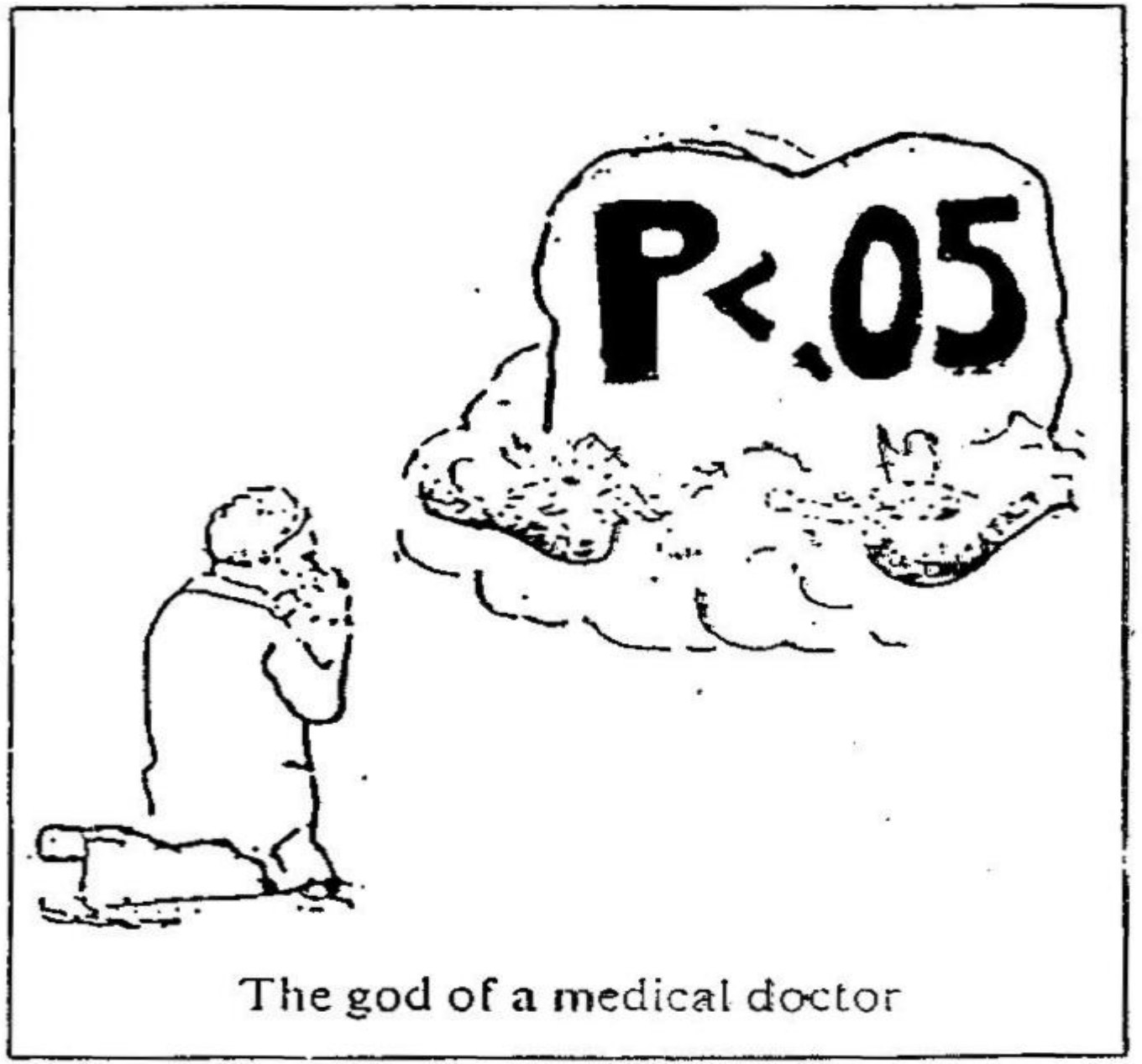
Anne-Laure Boulesteix[1], Roman Hornung[2], Willi Sauerbrei[3]

In
Ott, Max; Pietsch, Wolfgang; Wernecke, Jörg. Berechenbarkeit der Welt? Philosophie und Wissenschaft im Zeitalter von Big Data. Springer Fachmedien Wiesbaden, 2017, 155-170

Almost all articles on cancer prognostic markers report statistically significant results

- Kyzas et al. (2007) review 1915 prognostic factor articles in cancer
- Nearly all articles present significant findings
- < 1.5% were fully ‘negative’ in that they did not present statistically significant prognostic results & did not elaborate on non-significant trends

Biased reporting !



The god of a medical doctor

…but the p-value depends on sample size

Does an exposition to a substance influence the status of a disease?

Simple situation: 2x2 table

		D^+	D^-	
Exposition	E^+	30	270	300
	E^-	6	144	150
		36	414	450

$$RR = \frac{30}{300} / \frac{6}{150} = 2.5$$

- 95%-confidence interval for RR: [1.06 ; 5.87]
- p-value of χ^2 - Test: 0.027 ("significant")

The effect does not change

…but the p-value depends on sample size

	sample size		
	halving 225	450	doubling 900
RR	2.5	2.5	2.5
95%- CI	[0.75 ; 8.37]	[1.06 ; 5.87]	[1.37 ; 4.57]
p-value	0.118	0.027	0.002

Example extended - there is a second stratum

Stratum I

	D+	D−	
E+	30	270	300
E−	6	144	150
	36	414	450

RR = 2.5

Stratum II

	D+	D−	
E+	20	20	40
E−	40	70	110
	60	90	150

RR = 1.38

Stratum (Confounder) ignored in the analysis…

Combined

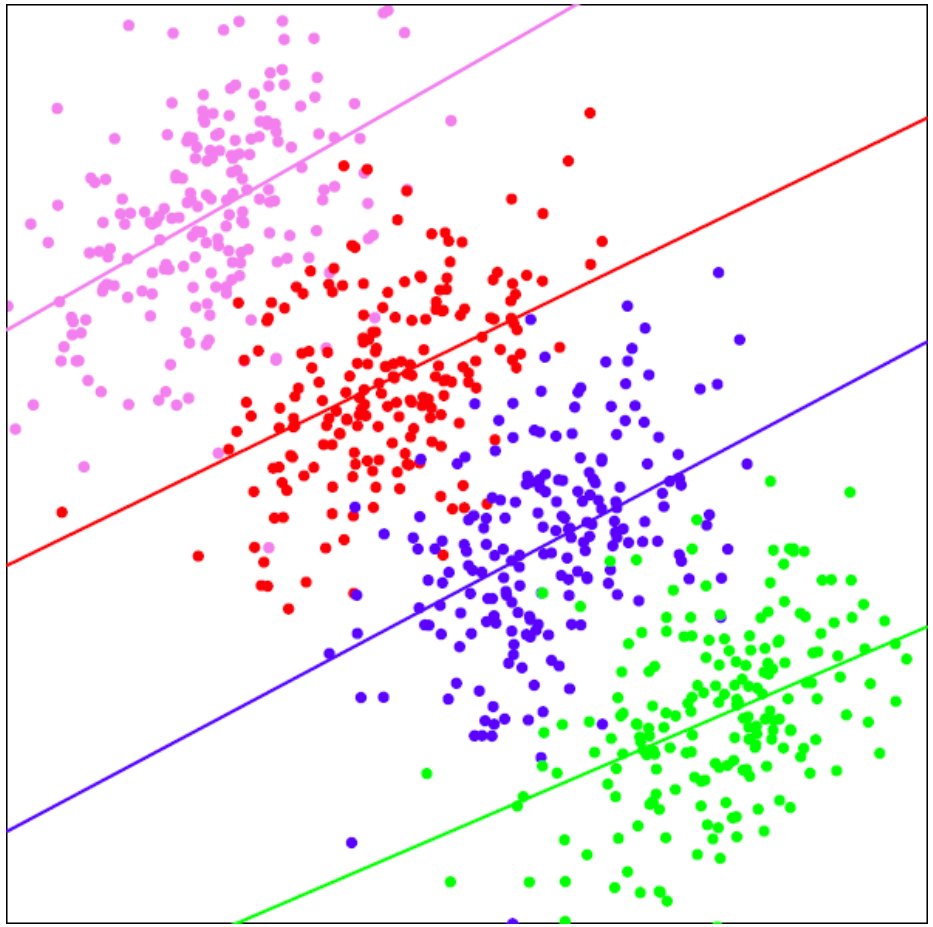
	D+	D−	
E+	50	290	340
E−	46	214	260
	96	504	600

RR = 0.83
95%-CI: [0.58; 1.20]
P = 0.63
(→„not significant“)

…can lead to wrong conclusions!

→ Stratified or multivariable analysis necessary!

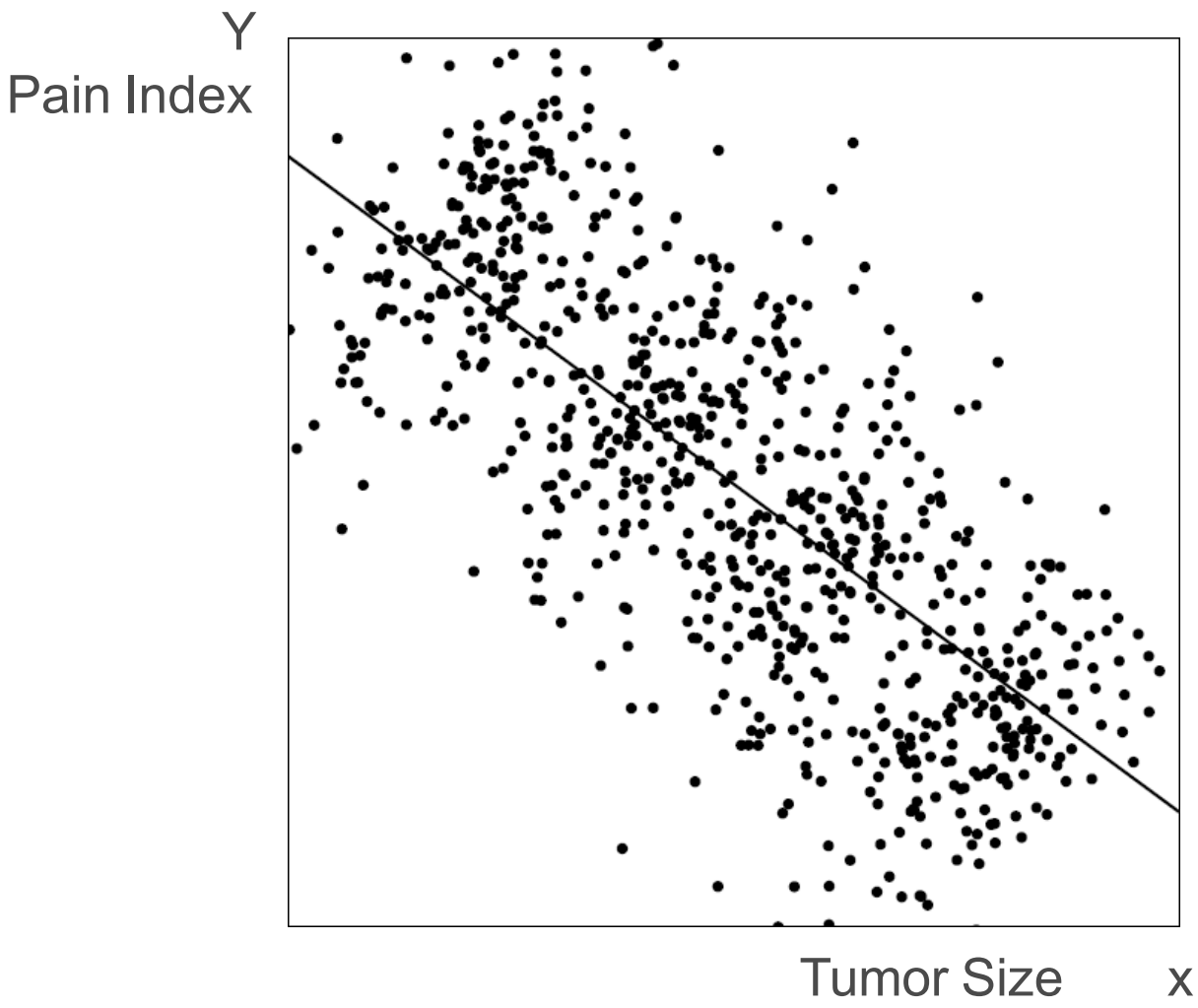
Confounder problem for continuous variables - influence of tumor size on pain index



Each Color signifies a different type of tumor

→ Positive correlation

Influence of tumor size on pain index?



Ignore type of tumor in analysis

Ohhh!
Negative correlation!

→Wrong conclusion!!!

Simpson‘s Paradoxon

Multivariable analysis required

Problem caused by ignoring one strong confounder is obvious.
However, in real analyses many (weak) confounders may be available.

Which confounders (variables) to include in a model?

Systematic reviews and meta-analysis

We live in the time of Evidence Based Medicine

Systematic reviews and meta-analysis relevant for the

selection of treatment

but also for prognostic factors, risk factors,
diagnostic methods, …

Evidence-Based Assessment and Application of Prognostic Markers: The Long Way from Single Studies to Meta-Analysis

WILLI SAUERBREI[1], NORBERT HOLLÄNDER[1],
RICHARD D. RILEY[2], AND DOUGLAS G. ALTMAN[3]

Communications in Statistics—Theory and Methods, 35: 1333–1342, 2006

… back to meta-analysis of Ki-67 in Breast Cancer

Azambuja et al. BJC 2007

- 65 studies identified
- 46 studies evaluable
- 38 studies evaluable for aggregation of results for DFS

Unadjusted hazard ratios of 38 studies

- 20 different cut-points and
- 5 different methods of measurement

It is possible to calculate a pooled estimate for Ki-67 positivity but …

publication bias, 38 studies selected, univariate analysis, different cutpoints, …

… does such an estimate tell us anything?

Older examples … did practise improve?

BAG-1 as a biomarker in early breast cancer prognosis: a systematic review with meta-analyses

E S Papadakis[1], T Reeves[*,1], N H Robson[1], T Maishman[3], G Packham[1] and R I Cutress[1,2]

British Journal of Cancer (2017) 116, 1585–1594

Papadakis et al (2017)

Identified 18 papers, providing results from 20 studies

Assessed quality of reporting by REMARK criteria

Performed ‚meta-analysis'

However, we identified severe weaknesses

(Sauerbrei & Haeussler (2018), British Journal of Cancer)

„This study illustrates key steps required for an evidence-based biomarker assessment; however, we have identified several major weaknesses in the assessment of the quality of reporting and the meta-analyses. We concluded **that results and inferences from this study are not justified by the assessments and analyses presented**."

Reply of Papadakis et al:

„We felt that this was **important**, particularly since BAG-1 is already included in multi-gene assays widely used as **part of routine clinical practice**…"

Comment on Papadakis et al (2017)

1. Assessment of the quality of reporting according to REMARK
   - Overly positive assessment of reporting, strongly contradicting a recent review on the topic (Sekula et al. 2017), prompted closer looks
     - E.g. ‚rationale for sample size' – positively assessed in all studies by Papadakis et al, vs. 22%, 11% and 8% in Sekula et al.
   - Several shortcomings in reporting of the primary literature found - examples:
     - Rationale for sample size:
       - *‚All patients with histopathological confirmation of breast cancer, diagnosed […] between 1995 and 2001, were included [only 70 patients included].'*
     - Multivariable analysis:
       - No effect estimates, only p-values in several studies or indication of non-significance

Sauerbei & Haeussler (2018), BJC

Comment on Papadakis et al (2017)

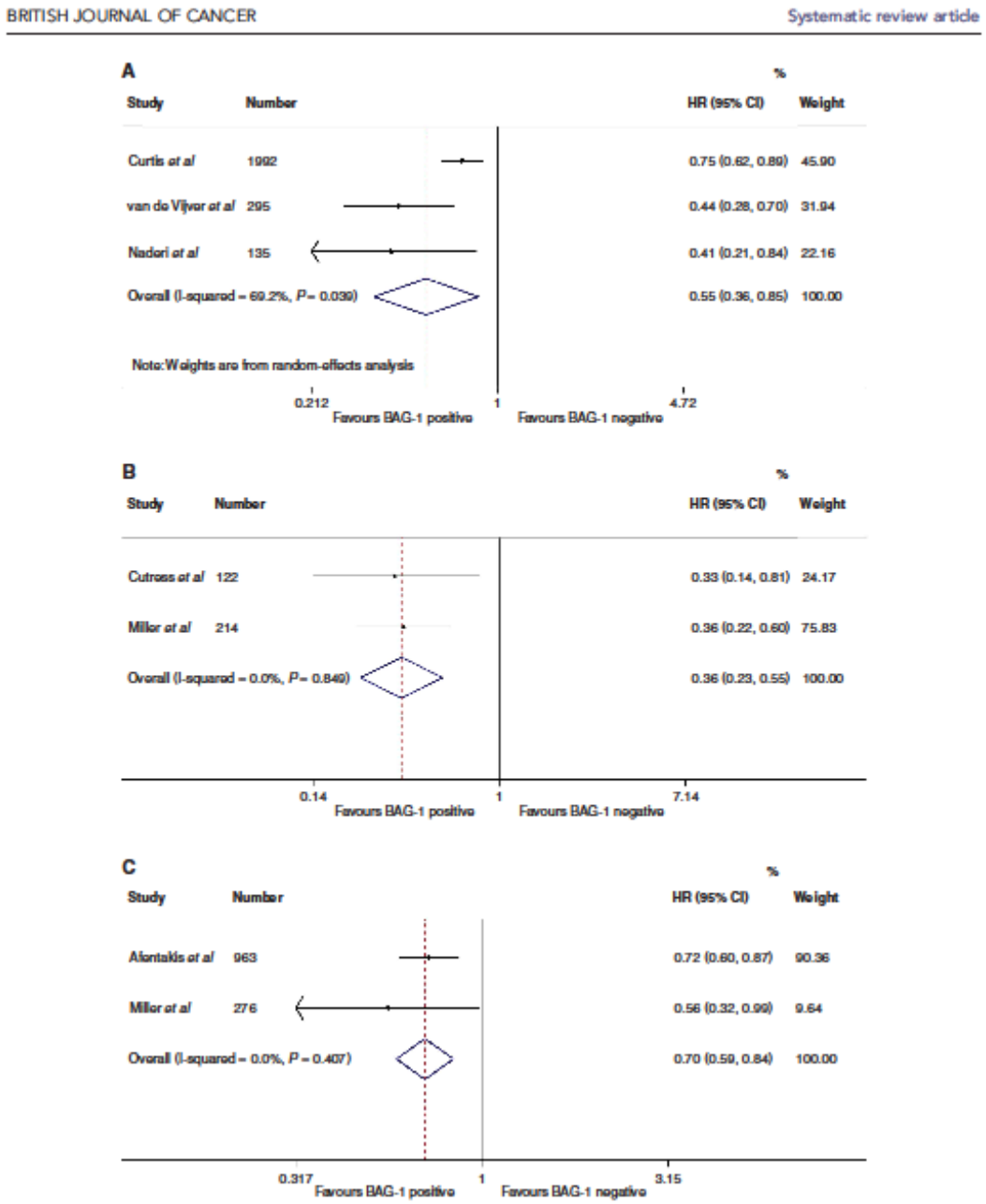
2. Meta-analysis

'In general, data were too heterogenous, and outcome measures were too varied to perform meta-analyses for the majority of studies. Meta-analyses of mRNA expression from the two data sets analysed in Millar et al (2009) and the data set analysed in Papadakis et al (2016) including a total of 2422 patients produced a HR of 0.55 (95% CI 0.36–0.85) favouring improved BCSS with high expression of BAG-1 (Figure 2a).' (Papadakis et al.)

Nevertheless…

Three “meta-analyses” published

- Several issues
  - 14 out of 18 papers ignored
  - Combination of multivariable and univariate analyses
  - Variable definitions of BAG-1 positivity



Comment on Papadakis et al (2017)

3. Meaningful meta-analyses of biomarkers – individual participant data (IPD) required
   - Primary study – multivariable model required (effect adjusted for potential confounders)
- Meta-analysis – combine ‚adjusted effects‘

Collaboration between study groups and IPD required

4. Publication bias and the need for a comprehensive biomarker study registry

Meta-analyses based on published data

Primary studies:

Use different cutpoints for continuous variables

Adjust for different confounders

Reporting is insufficient. Estimates from multivariable models are needed but are often not provided

Different measurement techniques are used – which studies can be combined?

IPD meta-analyses – are they feasible?

IPD projects are difficult but many good projects have been started.

Abo-Zaid et al found 48 published IPD meta-analyses of prognostic factors (published 1991 – March 2009, several inclusion criteria).

However, it is obvious that reporting and analysis of IPD projects need improvement.

Individual participant data meta-analysis of prognostic factor studies: *state of the art?*

Abo-Zaid *et al. BMC Medical Research Methodology* 2012, **12**:56

Cooperative IPD projects are possible (1)

In traumatic brain injury, researchers initiated IMPACT (International Mission for Prognosis and Analysis of Clinical Trials) and meta-analysed IPD from 11 studies including 9,205 patients [Marmarou et al, 2007].

http://www.tbi-impact.org/?p=publications

62 publications listed.

Probably more, most recent listed is from 2013.

Cooperative IPD projects are possible (2)

The **Emerging Risk Factors Collaboration (ERFC)** is a CEU-led consortium of >130 prospective studies from >30 countries

IPD collated and harmonized from ~2.5M participants

Cardiovascular diseases risk factors and cause-specific mortality studied in greater detail by IPD meta-analysis.

Risk factors studied included: circulating lipid markers, inflammatory markers, glycaemia markers, adiposity markers, diabetes, and cardio-metabolic multi-morbidity.

Analyses concern etiological hypothesis or risk prediction assessment in subsets of studies/participants with relevant data, **with methodological developments** occurring in parallel as necessary.

http://www.phpc.cam.ac.uk/ceu/erf

ERFC
Emerging Risk Factors Collaboration

OPEN ACCESS Freely available online

PLOS | MEDICINE

Guidelines and Guidance

Improving the Transparency of Prognosis Research: The Role of Reporting, Data Sharing, Registration, and Protocols

George Peat[1*¶], Richard D. Riley[2¶], Peter Croft[3], Katherine I. Morley[4,5], Panayiotis A. Kyzas[6], Karel G. M. Moons[7], Pablo Perel[8], Ewout W. Steyerberg[9], Sara Schroter[10], Douglas G. Altman[11], Harry Hemingway[12], for the PROGRESS Group[‡]

1 Arthritis Research UK Primary Care Research Centre, Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, United Kingdom, **2** School of Health and Population Sciences, University of Birmingham, United Kingdom, **3** Arthritis Research UK Primary Care Research Centre, Research Institute for Primary Care & Health Sciences, Keele University, Keele, Staffordshire, United Kingdom, **4** Department of Epidemiology and Public Health, University College London, London, United Kingdom, **5** Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Victoria, Australia, **6** Department of Oral and Maxillofacial Surgery, North Manchester General Hospital, Pennine Acute NHS Trust, Manchester, United Kingdom, **7** Julius Center for Health Sciences and Primary Care, UMC Utrecht, Utrecht, Netherlands, **8** London School of Hygiene & Tropical Medicine, London, United Kingdom, **9** Department of Public Health, Erasmus MC, Rotterdam, Netherlands, **10** BMJ, London, United Kingdom, **11** Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Oxford, United Kingdom, **12** Department of Epidemiology and Public Health and Director of the Farr Institute of Health Informatics Research at UCL Partners, London, United Kingdom

Peat et al. (2014) PLoS Med 11(7): e1001671

Improving the Transparency of Prognosis Research: The Role of Reporting, Data Sharing, Registration, and Protocols

Summary Points

- Prognosis research is concerned with predicting outcomes to make health care more effective. It has a crucial role to play in clinical and policy decision-making.
- The quality of much prognosis research is poor, evidenced by incomplete reporting, poor data sharing, incomplete registrations, and absent study protocols.
- Initiatives to improve transparency in trials include reporting guidelines, data pooling, registers, and journal requirements for protocols. Prognosis research could be transformed by similar initiatives.
- Routine registration of all prognostic studies, linked to an accessible study protocol using agreed reporting guidelines, would improve transparency and promote data sharing.
- Concern about applying transparency methods to observational research could be resolved by flexibility to update date-stamped protocols during prognosis studies.

Peat et al. (2014) PLoS Med 11(7): e1001671

Potential benefits of study registration, protocol publication, better study reporting, and data sharing of prognosis research studies

Potential Benefit	Registration	Protocols	Reporting	Sharing
Ethical				
Respect the investigator-participant covenant to generate new, publicly accessible biomedical knowledge of potential value to future patients	X	X	X	X
Facilitate monitoring and accountability in relation to global standards for ethical research, including informed consent	X	X	X	
Cost-effective use of public money	X	X	X	X
Scientific				
Improve the quality and reliability of evidence from prognosis research, (and thereby enhance impact on health and health care)	X	X	X	X
Help accelerate knowledge creation through easier identification of and access to full study details, including data, in order to increase opportunities for collaboration including systematic reviews and meta-analysis	X	X	X	X
Answer research questions only possible through collaboration				X
Reduce unnecessary duplication of invested research resources through awareness of existing studies	X	X		
Establish intellectual property		X		
Provide a denominator against which publication bias can be assessed	X	X		
Provide means for identification and prevention of biased under-reporting or over-reporting of research		X	X	
Involve patients in studies, including enrolment	X	X		
Peer review of protocols to improve study quality and refine methods		X		
Methodological issues sufficiently detailed to, in principle, allow study replication (details not always allowable in published reports)		X		

doi:10.1371/journal.pmed.1001671.t002

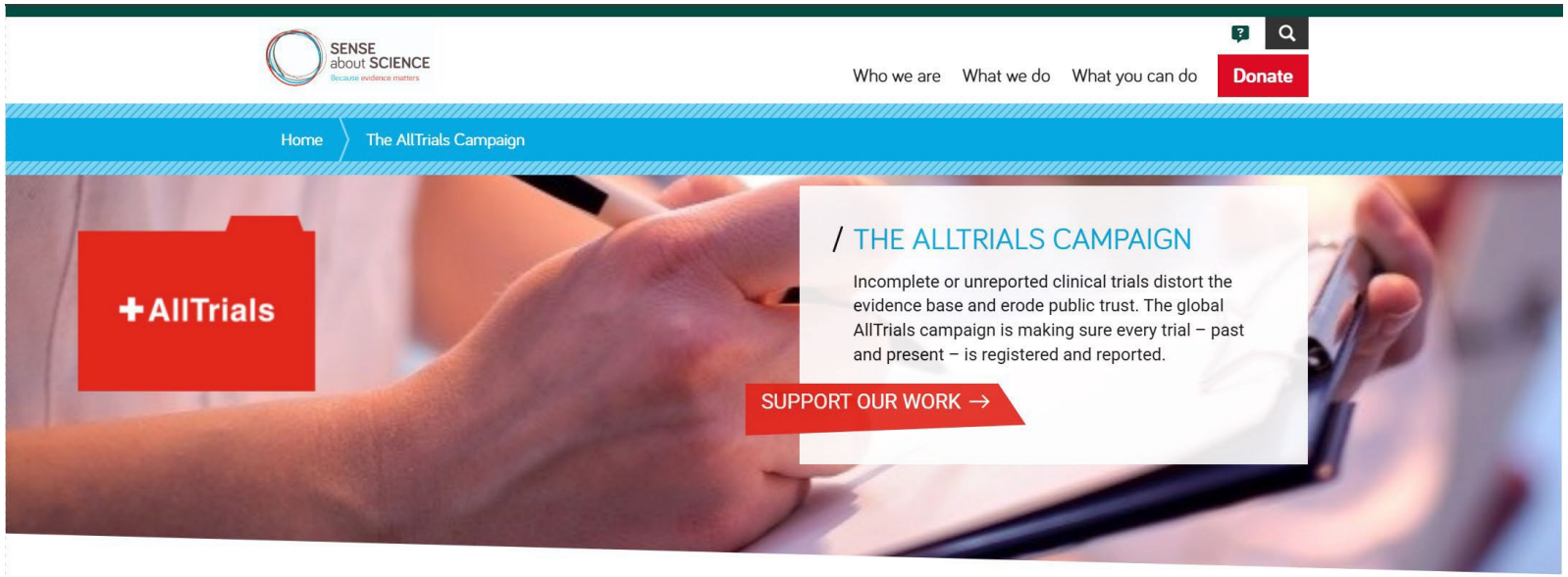
Peat et al. (2014) PLoS Med 11(7): e1001671

PROGRESS recommendations

1. Full study reporting through use of guidelines
2. Facilitate and expect data sharing
3. Routine registration of all prognosis studies using existing registers
4. Protocols for all prognosis studies made public
5. Promote systematic development and evaluation of methods and value of transparency

Peat et al. (2014) PLoS Med 11(7): e1001671

All Trials Campaign



In 2013, we estimated over half of all clinical trial results were never reported. Our response was the global AllTrials campaign — an international initiative of Ben Goldacre, the BMJ, the Centre for Evidence-based Medicine, the Cochrane Collaboration, the James Lind Initiative, PLOS and Sense about Science. Our goal: for every clinical trial to be registered and their full methods and results reported – even if they are negative.

/ WHY REGISTRATION AND REPORTING OF CLINICAL TRIALS MATTERS

When the results of clinical trials are not reported, the public trust in science is damaged — a loss we cannot afford in an era where mis- and disinformation run rampant.

Over 95,000 people and 747 organizations have signed the AllTrials petition.
www.alltrials.net

Guidelines for Code and Data Submission

Benjamin Hofner, Fabian Scheipl (RR Editors, Biometrical Journal)
E-mail: fabian.scheipl@stat.uni-muenchen.de

Document Version: 1.7 (2016/10/28)

Reporting

The Lancet Research:
Increasing Value, Reducing Waste Series

How should medical science change?

In 2009, we published a Viewpoint by Iain Chalmers and Paul Glasziou called "Avoidable waste in the production and reporting of research evidence", which made the extraordinary claim that as much as 85% of research investment was wasted.

Our belief is that research funders, scientific societies, school and university teachers, professional medical associations, and scientific publishers (and their editors) can use this Series as an opportunity to examine more forensically why they are doing what they do—the purpose of science and science communication—and whether they are getting the most value for the time and money invested in science.

Kleinert and Horton, 2014

The Lancet Research:
Increasing Value, Reducing Waste
Series

Biomedical research: increasing value, reducing waste

Of 1575 reports about cancer prognostic markers published in 2005, 1509 (96%) detailed at least one significant prognostic variable. However, few identified biomarkers have been confirmed by subsequent research and few have entered routine clinical practice.

….

Global biomedical and public health research involves billions of dollars and millions of people. In 2010, expenditure on life sciences (mostly biomedical) research was US$240 billion. The USA is the largest funder, with about $70 billion in commercial and $40 billion in governmental and non-profit funding annually, representing slightly more than 5% of US health-care expenditure. Although this vast enterprise has led to substantial health improvements, many more gains are possible if the waste and inefficiency in the ways that biomedical research is chosen, designed, done, analysed, regulated, managed, disseminated, and reported can be addressed.

Macleod et al., 2014

Initiatives to improve the situation

Reporting

EQUATOR network

http://www.equator-network.org/

Enhancing the **QU**Ality and **T**ransparency **O**f health **R**esearch

Reporting: Important improvements

Started with RCTs

Begg et al JAMA (1996) Improving the Quality of Reporting of Randomized Controlled Trials – The CONSORT Statement

Moher et al JAMA (2001), Revised Recommendations
Schulz et al Ann Int Med (2010) Updated…

Observational studies

Mc Shane et al, JNCI (2005) REMARK for tumor markers

von Elm et al Lancet (2007), STROBE for epidemiological studies

REMARK

Guidelines

REporting recommendations for tumour MARKer prognostic studies (REMARK)

Lisa M. McShane *, Douglas G. Altman, Willi Sauerbrei, Sheila E. Taube, Massimo Gion, Gary M. Clark, for the Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics

OPEN ACCESS Freely available online

PLoS MEDICINE

Guidelines and Guidance

Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): Explanation and Elaboration

May 2012 | Volume 9 | Issue 5 | e1001216

Douglas G. Altman[1]*, Lisa M. McShane[2], Willi Sauerbrei[3], Sheila E. Taube[4]

Explanation and Elaboration papers

- Good examples
- Basic background of analysis issues

  For example, REMARK

  - BOX 1 – SUBGROUPS AND INTERACTIONS: THE ANALYSIS OF JOINT EFFECTS
  - BOX 2 – CLINICAL OUTCOMES
  - BOX 3 – MISSING DATA
  - BOX 4 – CONTINUOUS VARIABLES
  - BOX 5 – SELECTIVE REPORTING

Explanation and Elaboration papers

- More about analysis issues
  - Item 10 - All Statistical Methods
    - Preliminary Data Preparation
    - Association of Marker Values With Other Variables
    - Methods to Evaluate a Marker's Univariable Association With Clinical Outcome
    - Multivariable Analyses
    - Missing Data
    - Variable Selection
    - Checking Model Assumptions
    - Model Validation

Reporting of tumour marker prognostic studies

- **Mallett *et al* (2010):** *pre-REMARK area*

Conclusion: ‚Current reporting … is poor.‘

- **Sekula *et al* (2017):** *post-REMARK area*

Aim: to assess whether reporting quality improved

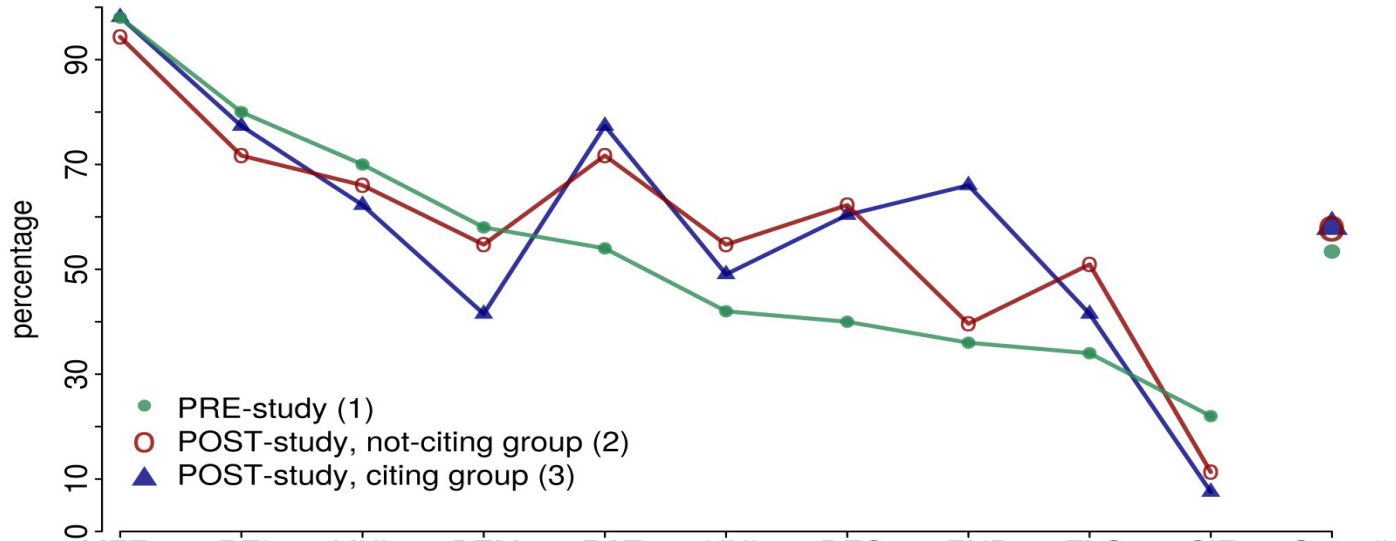
Design: Evaluation of 106 published studies (2007-2012)

- 53 articles with REMARK citation
- 53 articles w/o citation (matched)

Evaluation: 10 of 20 REMARK checklist items

Reporting of tumour marker prognostic studies

- **Results (Sekula *et al*):**



	MET	REL	MUL	DEM	PAT	UNI	DES	END	FLO	SIZ	Overall
(1)	98	80	70	58	54	42	40	36	34	22	53.4
(2)	94	72	66	55	72	55	62	40	51	11	57.7
(3)	98	77	62	42	77	49	60	66	42	8	58.1

➡ **Conclusion:** (1) studies still poorly reported
(2) call for combined effort

More structured reporting is required

Participant flow diagram is well accepted

but what about reporting of statistical analyses?

Often, many analyses (e.g. subgroups, additional outcomes) are hidden in the text.

What about checks of important assumptions (e.g. proportional hazards in the Cox model)? Done?

Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK)

Lisa M. McShane, Douglas G. Altman, Willi Sauerbrei, Sheila E. Taube, Massimo Gion, Gary M. Clark for the Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics

Journal of the National Cancer Institute, 2005

Item 12.

*Describe the **flow of patients** through the study, including the number of patients included in **each stage of the analysis** (a diagram may be helpful) and reasons for dropout. Specifically, both overall and **for each subgroup** extensively examined report the **number of patients and the number of events**.*

Reporting of Item 12 is still bad

Percentage of adequate reporting

Period 2006/07		34
Period 2011/12	Not citing REMARK	51
Period 2011/12	Citing REMARK	42

Sekula et al. PLOS one 2017

REMARK profile as an instrument to improve reporting of flow of patients and of all analyses conducted

A two part study profile

a) Patients, treatment, and variables

b) Statistical analyses

REMARK profile – part a

Table 2. Example of the REMARK profile illustrated using data from a study of ploidy in patients with advanced ovarian cancer [157] (from [20]).

a) Patients, treatment and variables

Study and marker	**Remarks**	
Marker (If non-binary: how was marker analyzed? continuous or categorical. If categorical, how were cutpoints determined?)	M = ploidy (diploid, aneuploid)	
Further variables (variables collected, variables available for analysis, baseline variables, patient and tumor variables)	v1 = age, v2 = histologic type, v3 = grade, v4 = residual tumor, v5 = stage, v6 = ascites[a], v7 = estrogen[a], v8 = progesterone[a], v9 = CA-125[a]	
Patients	**n**	**Remarks**
Assessed for eligibility	257	*Disease*: Advanced ovarian cancer, stage III and IV *Patient source*: Surgery 1982 to 1990, University Hospital Freiburg *Sample source*: Archived specimens available
Excluded	73	General exclusion criteria[b], non-standard therapy[b], coefficient of variation >7%[b]
Included	184	Previously untreated. *Treatment*: all had platinum based chemotherapy after surgery
With outcome events	139	Overall survival: death from any cause

Altman et al. 2012

REMARK profile – part b

Relatively simple example

b) Statistical analyses of survival outcomes				
Analysis	**Patients**	**Events**	**Variables considered**	**Results/remarks**
A1: Univariable	184	139	M, v1 to v5	Table 2, Figure 1
A2: Multivariable	174	133	M, v1, v3 to v5	Table 3 [v2 omitted because many missing data; Backward selection, see text]
A3: Effect for ploidy adjusted for v4	184	139	M, v4	Figure 2 [Based on result of A2]
A4: Interaction: ploidy and stage	175	133	M, v1, v2, v4, v5	See text
A5: Ploidy in stage subgroups				
v5 = III	128	88	M	Figure 3
v5 = IV	56	51	M	Figure 4

Altman et al. 2012

REMARK profile – another simple example

Two outcomes - structure needs to be adapted

Table 3. Example of the REMARK profile illustrated using data from a study of expression of epithelial membrane protein-2 in patients with endometrial adenocarcinoma [158].

a) Patients, treatment and variables						
136 Patients with endometrial adenocarcinoma assessed for eligibility, 37 excluded (33 no informative immune histochemistry, 4 without clinical information)						
99 Patients included, stages IA to IVB						
Formalin fixed, paraffin embedded endometrial tissue samples, Department of Pathology, UCLA Los Angeles, USA						
Marker (and how was the marker handled in analysis?)					M = epithelial membrane protein-2 Immunoreactive score obtained by multiplying subscores for intensity (0 to 3+) and distribution of immunoreactivity (0 to 4+) grouped as negative (score 0), weak (1 to 3) or moderate-to-strong (4 to 12)	
Outcomes:					DFS (97 patients, 42 events), OS (99 patients, 32 events)	
Further variables:					v1 = age, v2 = ER, v3 = PR, v4 = vascular invasion, v5 = stage, v6 = histology, v7 = grade	
b) Statistical analyses of survival outcomes						
	DFS		OS			
Aim	Patients	Events	Patients	Events	Variables considered	Results/remarks
A1: Univariable	97	42	99	32	M, v1-v7	Figure 3, Figure 4, Table 2, Table 3
						DFS: except v1 all significant
						OS: all significant
A2: Multivariable	97	42	99	32	DFS: M, v2-v7	Table 4, Table 5
					OS: M, v1-v7	In multivariable analysis: all significant in A1, then stepwise selection
						Variables in final models: DFS: M, v5, v6; OS: v4, v6, v7 (M is not included)

Altman et al. 2012

REMARK profile

An extension to improve completeness and transparency of reporting all steps of the analysis

a) Patients, treatments and variables		
Study and marker	**Remarks**	
Markerhandled	M = NPI *Continuous and categorical. Cutpoints as predefined in the literature. For details see Blamey et al [27].*	
Further variables	v1 = Tumor Size, v2 = No. of pos. Lymph Nodes, v3 = Tumor Grade, v4 = Age, v5 = Histology, v6 = Hormone Receptor Status, v7 = Menopausal Status, v8 = Vessel Invasion, v9 = Lymphatic Vessel Invasion	
Patients	**n**	**Remarks**
Assessed for eligibility	2062	**Disease:** Primary Breast Cancer **Patient source:** Database Surgical clinic Charité, Berlin. All patients with surgery from 1st Jan. 1984 to 31st Dec. 1998.
Excluded	502	63 metastasis, 73 previous carcinoma other than breast cancer, 86 primary breast cancer prior to the study, 134 breast cancer in situ, 8 pt0, 123 older than 80 years, 20 neo-adjuvant chemotherapy, 71 death within first months of surgery, three or more standard prognostic factors missing. For some patients, more than one exclusion criterion applied.
Included	1560	Previously untreated. ***Treatment:*** *Local therapy: BCT or mastectomy with or without radiotherapy, adjuvant therapy: chemo (y/n), hormone (y/n). For details see Add file 1 and Table 2 in Winzer et al [28]*
With outcome events	221	Overall survival: death from any cause

Winzer et al. 2016

REMARK profile – prospectively it helps to write SAP

b) Statistical analyses. All analyses using a Cox model are stratified for strata according to therapy. There are 8 strata defined by the combination of surgery, radiotherapy (y/n) and systemic therapy (y/n (no chemotherapy and no hormone therapy))

Analysis	Patients	Events	Variables considered	Results/ remarks
IDA 1[1]: Imputation for missing values	1560	NR [2]	v1(94), v2 (68), v3(217), v6 (490), v7(54)	Variables (number of patients) with imputed values
A1[3]: NPI (3)	1560	221	NPI	Prognostic value of NPI in 3 categories (Table 2, Fig 1, Table 3)
A2: NPI (6)	1560	221	NPI	6 categories (Fig 1, Table 3)
C1[4]: Check of PH[5] in NPI (3) and in NPI (6)	1560	221	NPI	Fig 2, S2 Fig and non-significant result of FPT (see last paragraph 4.2).
A3: NPIcont.	1560	221	NPI	More information from continuous data? (Table 3)
C2: NPIcont. has a linear effect	1560	221	NPI	FP2 function not significantly better, see 4.3.1
C3: Check of PH[5] in NPIcont.	1560	221	NPI	Non-significant result of FPT (see last paragraph 4.3.1)
A4: MFP[6] of the three NPI variables (univ. and multivariable)	1560	221	v1, v2, v3	Table 4
A5: Functional form for nodes	1560	221	v2	Fig 3
A6: Prognostic value and additional value of further variables (univ. and multiv.)	1560	221	NPI, v4, v5, v6, v7, v8, v9	Table 5, Fig 4
A7: MFP using all available information	1560	221	v1, v2, v3, v4, v5, v6, v7, v8, v9	Final MFP model in Table 6, see 4.5
A8: Measures of separation	1560	221	NPI, v1, v2, v3, v4, v5, v6, v7, v8, v9	Table 7, see 4.6
C4: Check of PH[5] in MFP model	1560	221	v1, v2, v3, v6	Non-significant result of FPT (see end of 4.5)

Winzer et al. 2016

Diagnostic and prognostic models

Collins *et al. BMC Medicine* (2015)13:1
DOI 10.1186/s12916-014-0241-z

BMC Medicine

GUIDELINE Open Access

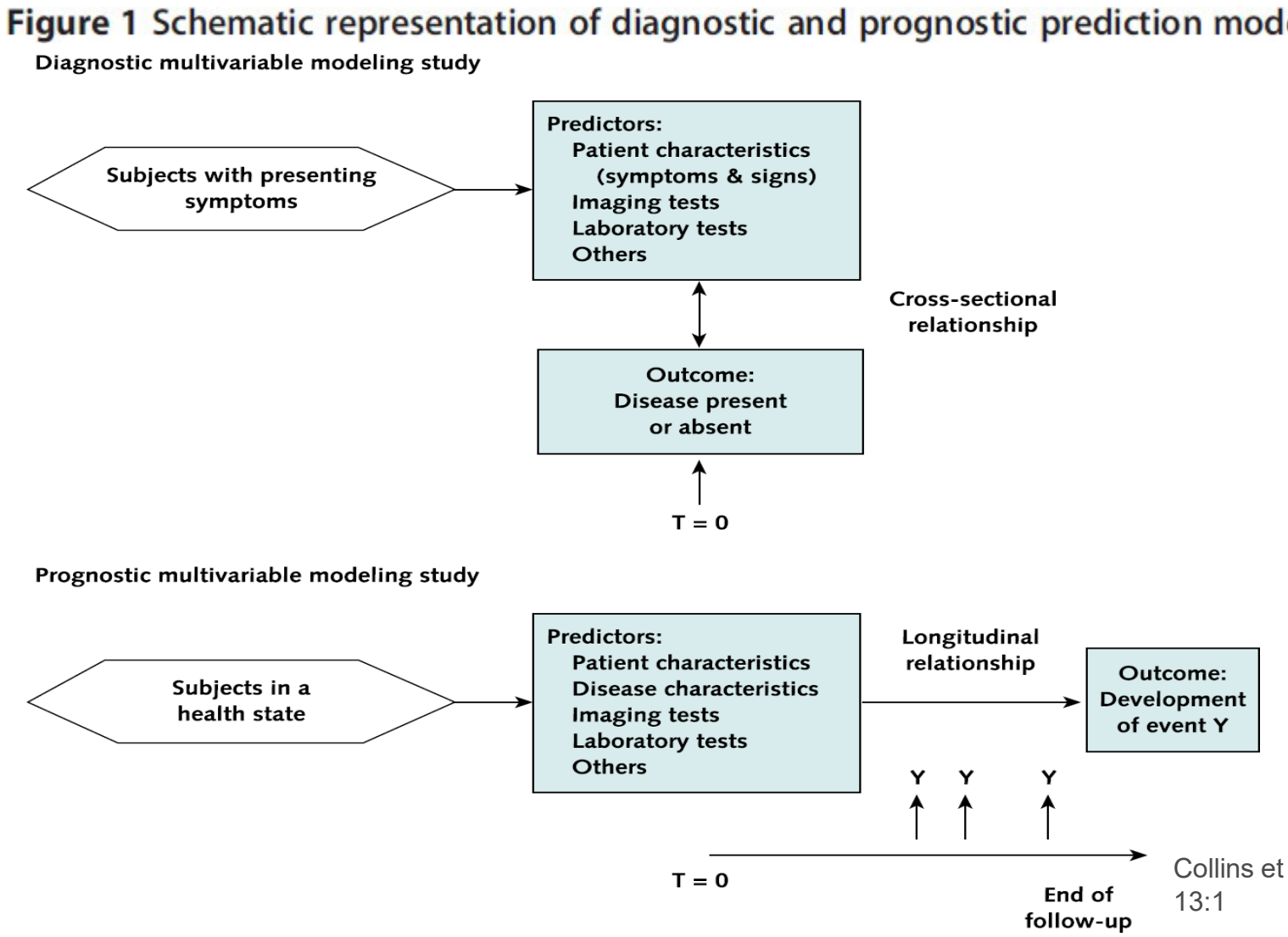
Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement

Gary S Collins[1*], Johannes B Reitsma[2], Douglas G Altman[1] and Karel GM Moons[2]

Abstract

Prediction models are developed to aid health care providers in estimating the probability or risk that a specific disease or condition is present (diagnostic models) or that a specific event will occur in the future (prognostic models), to inform their decision making. However, the overwhelming evidence shows that the quality of reporting of prediction model studies is poor. Only with full and clear reporting of information on all aspects of a prediction model can risk of bias and potential usefulness of prediction models be adequately assessed.(…)

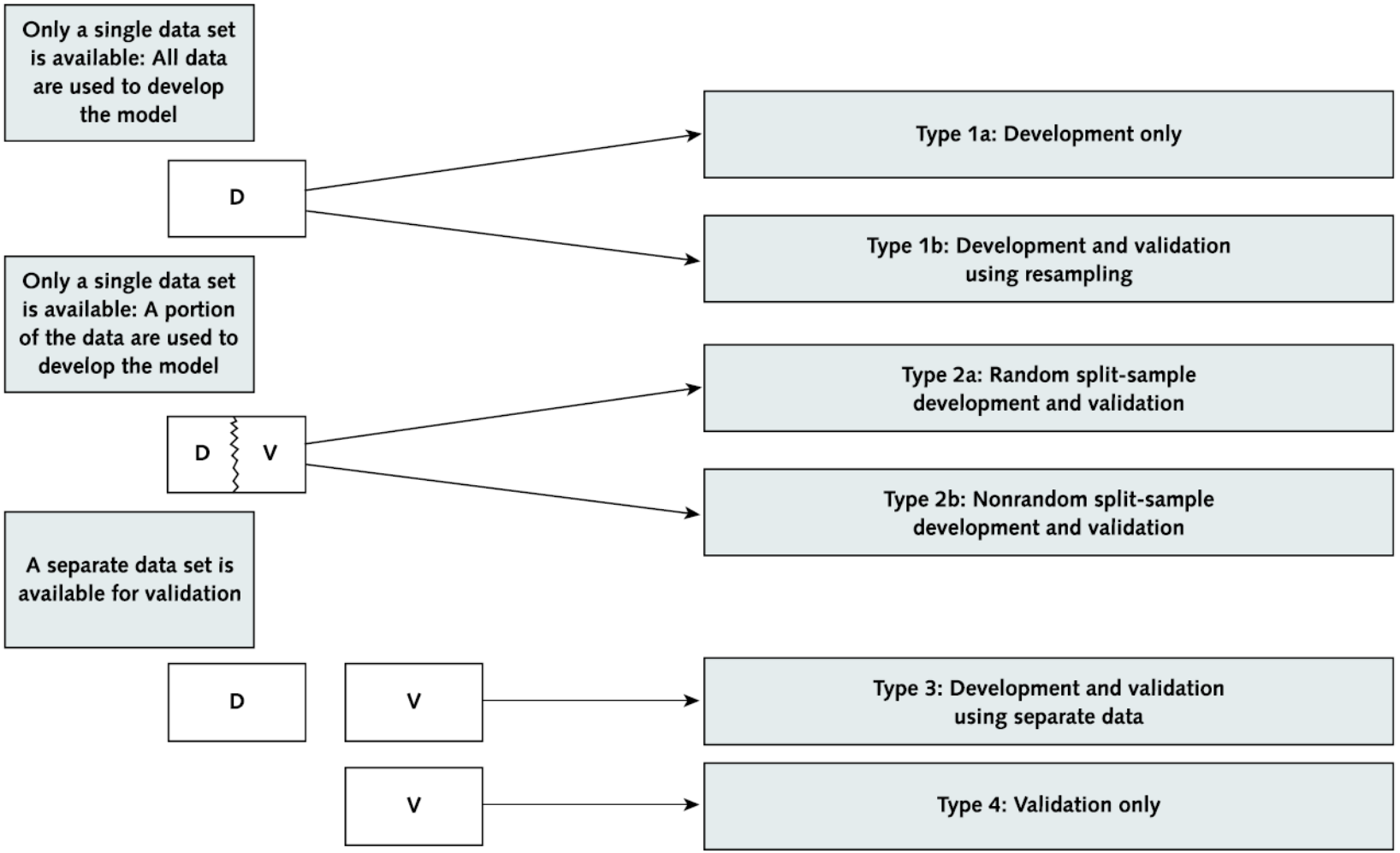
Diagnostic and prognostic models



Collins et al. BMC Medicine (2015) 13:1

Diagnostic and prognostic models

Figure 3 Types of prediction model studies covered by the TRIPOD statement. D = development data; V = validation data.



Collins et al. BMC Medicine (2015) 13:1

Diagnostic and prognostic models

Proteomics 2014, *14*, 1587–1592 DOI 10.1002/pmic.201300377 1587

VIEWPOINT

Why have so few proteomic biomarkers "survived" validation? (Sample size and independent validation considerations)

Belinda Hernández[1,2], *Andrew Parnell*[1] *and Stephen R. Pennington*[2]

[1] Complex and Adaptive Systems Laboratory, School of Mathematical Sciences (Statistics), University College Dublin, Dublin, Ireland
[2] School of Medicine and Medical Science, UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland

ajog.org Obstetrics Clinical Opinion

Prognostic models in obstetrics: available, but far from applicable

C. Emily Kleinrouweler, MD, PhD; Fiona M. Cheong-See, MRCOG; Gary S. Collins, PhD; Anneke Kwee, MD, PhD; Shakila Thangaratinam, PhD; Khalid S. Khan, MSc, MRCOG; Ben Willem J. Mol, MD, PhD; Eva Pajkrt, MD, PhD; Karel G. M. Moons, PhD; Ewoud Schuit, PhD

Am J Obstet Gynecol. 2016

Diagnostic and prognostic models

Overinterpretation and misreporting of prognostic factor studies in oncology: a systematic review

Emmanuelle Kempf[1,2], Jennifer A. de Beyer[1], Jonathan Cook[1], Jane Holmes[1], Seid Mohammed1, Tri-Long Nguyên[1,3], Iveta Simera[4], Marialena Trivella[1], Douglas G. Altman[1], Sally Hopewell[1], Karel G. M. Moons[5,6], Raphael Porcher[7], Johannes B. Reitsma[5,6], Willi Sauerbrei[8] and Gary S. Collins[1,9]

British Journal of Cancer 119, pp.1288–1296 (2018)

Cancer prognostic biomarkers have shown disappointing clinical applicability. The objective of this study was to classify and estimate how study results are overinterpreted and misreported in prognostic factor studies in oncology.

[…] 17 oncology journals with an impact factor above 7.

[…] 98 studies included […] the prognostic factors' effects were selectively and incompletely reported in 35/98 and 24/98 full texts, respectively.

One in five articles had discussion and/or abstract conclusions that were inconsistent with the study findings. Sixteen reports had discrepancies between their full-text and abstract conclusions.

CONCLUSIONS: Our study provides evidence of frequent overinterpretation of findings of prognostic factor assessment in high-impact medical oncology journals.

Diagnostic and prognostic models

PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies

Robert F. Wolff, MD*; Karel G.M. Moons, PhD*; Richard D. Riley, PhD; Penny F. Whiting, PhD; Marie Westwood, PhD; Gary S. Collins, PhD; Johannes B. Reitsma, MD, PhD; Jos Kleijnen, MD, PhD; and Sue Mallett, DPhil; for the PROBAST Group†

Ann Intern Med. 2019;170:51-58

PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration

Karel G.M. Moons, PhD*; Robert F. Wolff, MD*; Richard D. Riley, PhD; Penny F. Whiting, PhD; Marie Westwood, PhD; Gary S. Collins, PhD; Johannes B. Reitsma, MD, PhD; Jos Kleijnen, MD, PhD; and Sue Mallett, DPhil

Ann Intern Med. 2019;170:W1-W33

Some recommendations have been extended to include AI:

CONSORT-AI

Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension

Xiaoxuan Liu, MBChB [a,c,d,e,g] · Samantha Cruz Rivera, PhD [b,c] · David Moher, PhD [h,k] · Prof Melanie J Calvert, PhD [b,c,g,i,j] · Prof Alastair K Denniston, PhD [a,b,c,d,f,g] and the SPIRIT-AI and CONSORT-AI Working Group [†]

Liu *et al.*, 2020

TRIPOD+AI

TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods

BMJ 2024 ; 385 doi: https://doi.org/10.1136/bmj-2023-078378 (Published 16 April 2024)
Cite this as: *BMJ* 2024;385:e078378

Collins *et al.*, 2024

…and others.

Reporting guidelines - summary

- Research in health sciences needs to improve
- Though many parts are difficult (timewise, costly) good reporting is easy:

  Follow reporting guidelines !
- The importance of complete and transparent reporting of all statistical analyses (otherwise “fishing for significance”) is still underrated
- REMARK type profile is a suitable instrument for improvement

However, good reporting does not help if a study is badly designed or analyzed

Guidance for analysis

STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative

http://stratos-initiative.org/

STRATOS INITIATIVE

2019, Marseille, France

Statistical methodology – problems are well known

The severeness of problems is even discussed in the public press:

The Economist ‘Unreliable research: Trouble at the lab.’ (October 2013):

“Scientists’ grasp of statistics has not kept pace with the development of complex mathematical techniques for crunching data. Some scientists use inappropriate techniques because those are the ones they feel comfortable with; others latch on to new ones without understanding their subtleties. Some just rely on the methods built into their software, even if they don’t understand them.”

The STRATOS initiative

Current situation in statistical methodology

- Statistical methodology has seen substantial development

Resampling and Bayesian methods allow investigations that were impossible two decades ago

- Computer facilities can be viewed as the cornerstone
- Possible to compare complex model building strategies using simulation studies
- Wealth of new statistical software packages allows a rapid implementation and verification of new statistical ideas

Current situation in practical analyses

Unfortunately, many sensible improvements are ignored, such as the use of spline functions in regression analyses

Reasons why improved strategies are ignored

Overwhelming concern with **theoretical aspects**

•Very **limited guidance** on key issues that are **vital in practice**, discourages analysts from utilizing more sophisticated and possibly more appropriate methods in their analyses

How should medical science change?

Many more gains are possible if the waste and inefficiency in the ways that biomedical research is chosen, **designed**, done, **analysed**, regulated, managed, **disseminated**, and **reported** can be addressed.” *The Lancet Research, Macleod et al., 2014*

Better use of statistical methods

At least two tasks are essential:

1. **Experts** in specific methodological areas have to work towards **developing guidance**
2. An ever-increasing need for **continuing education** at all stages of the career

For busy applied researchers it is often difficult to follow methodological progress even in their principal application area

- Reasons are diverse
- Consequence is that analyses are often deficient

Knowledge gained through research on statistical methodology needs to be **transferred** to the broader community

Many **analysts** would be **grateful for** an overview on the current **state of the art** and for **practical guidance**

Aims of the STRATOS Initiative

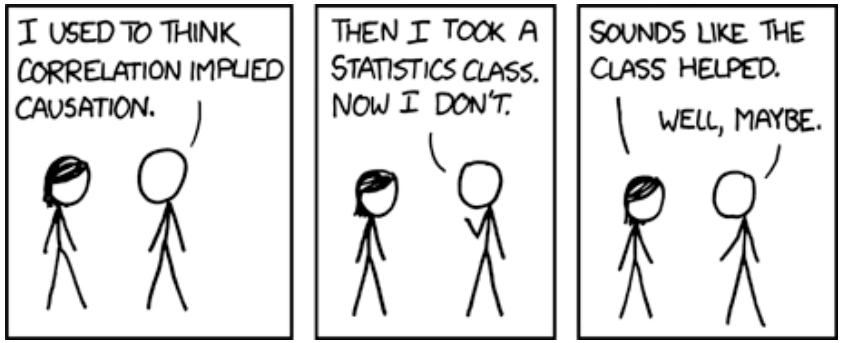
- **Provide evidence supported guidance** for highly relevant issues in the design and analysis of observational studies
- For the **start** we will concentrate on **state-of-the-art** guidance and the necessary evidence

As the **statistical knowledge** of the analyst **varies** substantially, guidance has to keep this background in mind. **Guidance** has to be provided **at several levels** (Low statistical knowledge, Experienced statistician, Expert in a specific area)

The overarching long-term aim is to improve key parts of design and statistical analyses of observational studies in practice

Relevance of guidance for statistical analyses of observational studies

Identifying causal effects is the aim of many studies, but how?



Comic from xkcd.com

In general, complex model building is required. Which confounders are required?

What about the functional form of continuous variables?

Is there a „state of the art“?

STRATOS initiative – TG2

Sauerbrei et al. Diagnostic and Prognostic Research (2020) 4:3
https://doi.org/10.1186/s41512-020-00074-3

Diagnostic and Prognostic Research

COMMENTARY Open Access

State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues

Willi Sauerbrei[1*], Aris Perperoglou[2], Matthias Schmid[3], Michal Abrahamowicz[4], Heiko Becher[5], Harald Binder[1], Daniela Dunkler[6], Frank E. Harrell Jr[7], Patrick Royston[8], Georg Heinze[6] and for TG2 of the STRATOS initiative

Towards state of the art – research required!

Table 1 Relevant issues in deriving evidence-supported state of the art guidance for multivariable modelling

No.	Item
1	Investigation and comparison of the properties of variable selection strategies
2	Comparison of spline procedures in both univariable and multivariable contexts
3	How to model one or more variables with a 'spike-at-zero'?
4	Comparison of multivariable procedures for model and function selection
5	Role of shrinkage to correct for bias introduced by data-dependent modelling
6	Evaluation of new approaches for post-selection inference
7	Adaption of procedures for very large sample sizes needed?

Prognostic research

Based on **observational** studies.

Usually **retrospective** studies, which **increases problems** related to design, sample size, data quality, statistical knowledge of analyst, reporting, publication bias, …

Even before the omics time started, hundreds of prognostic markers and many prognostic models were proposed

Only a small number of markers and models is **validated** and used in practice.

Omics data offer promising opportunities but with severe **challenges and problems**.

Obviously, **evidence-based** investigations concerning the value of markers and models are needed. Consequently, **systematic reviews and meta-analyses** are needed.

An important step forward: Pre-registration

Open Science Foundation (OSF)

There's a better way to manage your research

OSF is a free, open platform to support your research and enable collaboration

How OSF Supports your research

Search And Discover
Find papers, data, and materials to inspire your next research project. Search public projects to build on the work of others and find new collaborators

Design Your Study
Start a project and add collaborators, giving them access to protocols and other research materials. Built-in version control tracks the evolution of your study

Collect and Analyze Data
Store data, code, and other materials in OSF Storage, or connect your Dropbox or other third-party account. Every file gets a unique, persistent URL for citing and sharing

Publish Your Reports
Share papers in OSF Preprints or a community-based preprint provider, so others can find and cite your work. Track impact with metrics like downloads and view counts

Final remarks

At least for evidence based assessments closer collaboration among disciples and among study groups is required.

Data sharing is required.

Funders of prognosis research should require data sharing with appropriate governance (Peat et al 2014).

To improve analyses, methodologists need to work and agree on guidance for many relevant relevant issues.

Partly it may help to borrow ideas and suitable instruments from clinical research.

The lowest hanging fruit: GOOD REPORTING! http://www.equator-network.org/