



The EQUATOR network and reporting guidelines including for prediction models

Prof. Dr. Willi Sauerbrei

Institute of Medical Biometry and Statistics

Faculty of Medicine and Medical Center – University of
Freiburg, Germany

Overview

- 1 - Introduction
- 2 - Weakness of research
- 3 - EQUATOR
- 4 - REMARK
- 5 - Systematic reviews and meta-analysis
- 6 - Analysis - structured reporting
- 7 - Diagnostic and prognostic models
- 8 - STRATOS
- 9 - PROGRESS

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

1 - Introduction

Poorly conducted trials are a **waste of time, effort, and money**. The **most dangerous risk** associated with poor-quality reporting is an **overestimate of the advantages of a given treatment** ... Whatever the outcome of a study, it is really hard for the average reader to interpret and verify the reliability fo a poorly reported RCT. In turn, this problem could result in changes in clinical practice that are based on false evidence and that may harm patients. The only way to avoid the risk and to be sure that the final message of a RCT can be correctly interpreted is to fulfill the items listed in the CONSORT statement.

Zonta, S. & De Martino, M. (2008) Standard requirements for randomized controlled trials in surgery. *Surgery*, **144**, 838-839

1 - Introduction



2012

WILEY Blackwell

equator
network



CHAPTER 1

Importance of Transparent Reporting of Health Research

Douglas G. Altman¹ and David Moher²

¹Centre for Statistics in Medicine, University of Oxford, Oxford, UK

²Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

1 - Introduction

What do we mean by inadequate reporting of research?

Systematic assessments of published articles highlight frequent, serious shortcomings.

These include but are not limited to

- Omissions of crucial aspects of study methods, such as inclusion and exclusion criteria, precise details of interventions, measurement of outcomes, statistical methods,
- Statistical errors,
- Selective reporting of results for only some of the assessed outcomes,
- Selective reporting of statistical analyses (e.g. subgroup analyses),
- Inadequate reporting of harms,
- Confusing or misleading presentation of data and graphs,
- Incomplete numerical presentation of data precluding inclusion in a later meta-analysis
- Selective presentation of results in abstracts or inconsistency with the main text
- Selective or inappropriate citation of other studies
- Misinterpretation of study findings in the main article and abstract („spin“)

Altman and Moher, 2012

All these issues introduce various types of publication biases

2 - Weakness of research

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

2 - Weakness of research

BMJ

LONDON, SATURDAY 29 JANUARY 1994

The scandal of poor medical research

DOUGLAS G ALTMAN

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

What should we think about a doctor who uses the wrong treatment, either wilfully or through ignorance, or who uses the right treatment wrongly (such as by giving the wrong dose of a drug)? Most people would agree that such behaviour was unprofessional, arguably unethical, and certainly unacceptable.

ethics that excludes scientific issues. Consequently, poor or useless studies pass such review even though they can reasonably be considered to be unethical.⁹

The effects of the pressure to publish may be seen most clearly in the increase in scientific fraud,¹⁰ much of which is relatively minor and is likely to escape detection. There is

As the system encourages poor research it is the system that should be changed. We need less research, better research, and research done for the right reasons.



Why Most Published Research Findings Are False

John P.A. Ioannidis

PLoS Medicine **2005**, 8: 696-701.

2 - Weakness of research

WORLD VIEW

A personal take on events



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 [...].

The main question when reviewing a paper should be whether its conclusions are likely to be correct, not whether it would be important if it were true. Real advances are built with bricks, not straw.

2 - Weakness of research



META-RESEARCH ARTICLE

Where Have All the Rodents Gone? The Effects of Attrition in Experimental Research on Cancer and Stroke

Constance Holman¹, Sophie K. Piper^{2,3}, Ulrike Grittner^{3,4}, Andreas Antonios Diamantaras¹, Jonathan Kimmelman⁵, Bob Siegerink³, Ulrich Dirnagl^{2,3,6,7,8,*}

Given small sample sizes, loss of animals in preclinical experiments can dramatically alter results.

*Where have all the rodents gone?
Ooh ooh, ooh ooh, ooh
To non-random attrition, every one
When will they ever learn?*
—with apologies to Pete Seeger, 1955

We need better animal research, better reported

Fiona Godlee *editor in chief*

The BMJ

BMJ (2018), 360:k124



2 - Weakness of research

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



2 - Weakness of research

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



3 - EQUATOR

Initiatives to improve the situation

Reporting



equator
network

EQUATOR network

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

Enhancing the QUAlity and Transparency Of health Research

Started with: CONSORT statement
Consolidated Standards of Reporting Trials
<http://www.consort-statement.org/>





Your one-stop-shop for writing and publishing high-impact health research

find reporting guidelines | improve your writing | join our courses | run your own training course | enhance your peer review | implement guidelines



Library for health research reporting

The Library contains a comprehensive searchable database of reporting guidelines and also links to other resources relevant to research reporting.



[Search for reporting guidelines](#)



[Not sure which reporting guideline to use?](#)



[Reporting guidelines under development](#)



[Visit the library for more resources](#)



Reporting guidelines for main study types

[Randomised trials](#)

[CONSORT](#)

[Extensions](#)

[Observational studies](#)

[STROBE](#)

[Extensions](#)

[Systematic reviews](#)

[PRISMA](#)

[Extensions](#)

[Study protocols](#)

[SPIRIT](#)

[PRISMA-P](#)

[Diagnostic/prognostic studies](#)

[STARD](#)

[TRIPOD](#)

[Case reports](#)

[CARE](#)

[Extensions](#)

[Clinical practice guidelines](#)

[AGREE](#)

[RIGHT](#)

[Qualitative research](#)

[SRQR](#)

[COREQ](#)

[Animal pre-clinical studies](#)

[ARRIVE](#)

[Quality improvement studies](#)

[SQUIRE](#)

[Economic evaluations](#)

[CHEERS](#)

[See all 418 reporting guidelines](#)

Check the reporting guidelines
that are **UNDER CONSTRUCTION!**

[Visit the page!](#)



3 - EQUATOR

CONSORT extensions

- Crossover trials
- Multi-arm
- Cluster RCT
- Social and psychological interventions
- Within Person RCT
- ..
- Harms
- Patient reported outcome
- ...

STROBE Statement

Strengthening the reporting of observational studies in epidemiology

Extensions of STROBE

- Genetic Association Studies (STREGA)
- Molecular Epidemiology (STROBE-ME)
- STROBE checklist for conference abstracts
- Molecular epidemiology for infectious diseases (STROME-ID)
- Observational Routinely-collected health Data (RECORD)
- Epidemiology for Newborn Infection (STROBE-NI)

... and many more

Issues of (prognostic) biomarker research

- ‘Hot topic’ – many papers.

Nevertheless, only few biomarkers reach clinical application

McShane (2005): „What are we missing?“

Kyzas (2007): „Almost all articles on cancer prognostic markers report statistically significant results“

- **Issues:**

- Lack in well-defined research goal, limited research funding
- Poor study design, e.g. unrepresentative sample, too small study population
- Incorrect methods, but **NOT** restricted to statistical analysis
e.g. inadequate specificity and sensitivity of assays
- Reporting issues

Reporting issues

- **Issues:**
- Non-publication
- Incomplete (poor) reporting
- Selective reporting
- Misinterpretation/mispresentation
- **Effect:**
Bias in any form
- **Way out:**
- Reporting guidelines
- Call for study registry

4 - REMARK

Guidelines

R Eporting 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. McShane², Willi Sauerbrei³, Sheila E. Taube⁴



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)

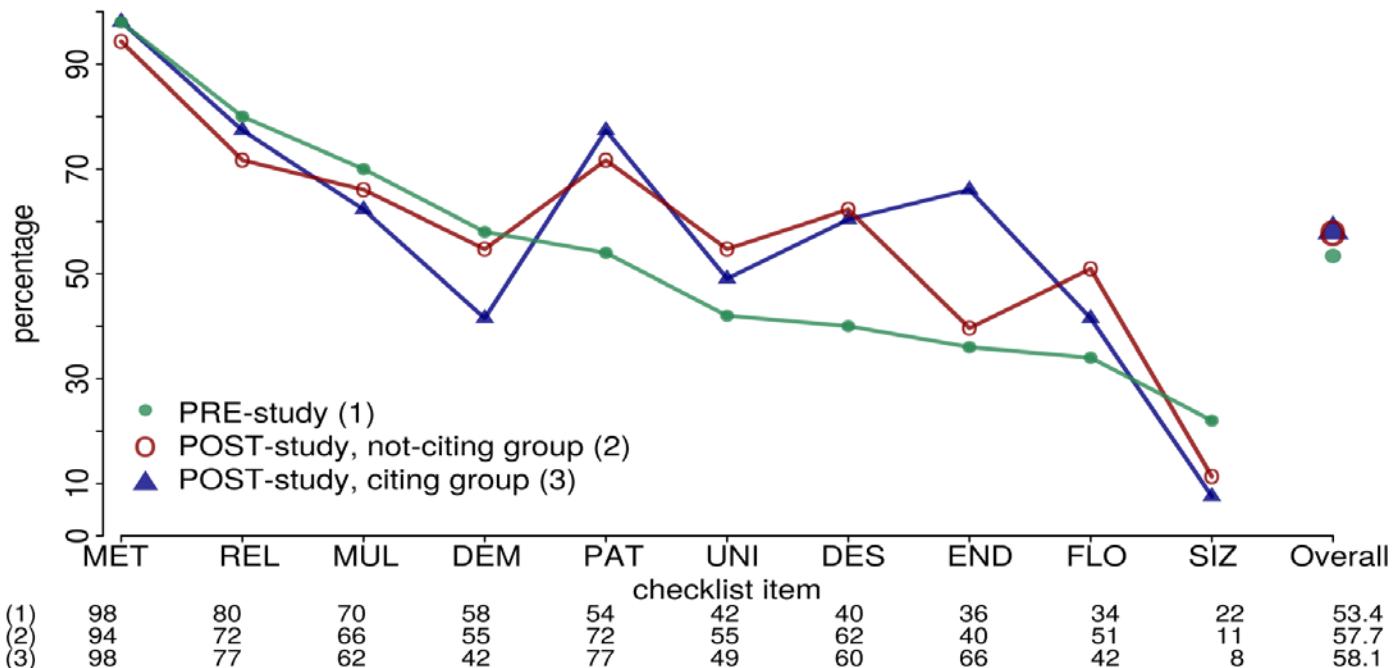
- 53 articles with REMARK citation
- 53 articles w/o citation (matched)

Evaluation: 10 of 20 REMARK checklist items

4 - REMARK

Reporting of tumour marker prognostic studies

- Results (*Sekula et al*):



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

5 - 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, ...

Studies using Individual Patient Data (IPD) need more attention

Reporting of tumour marker prognostic studies

Does poor reporting matter?

→ YES

Meaningful systematic reviews and informative meta-analysis are impossible

5 - Systematic reviews and meta-analysis

Meta-analysis – mission impossible

- **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”.

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?

6 - Analysis - structured reporting

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

Anne-Laure Boulesteix¹, Roman Hornung², Willi Sauerbrei³

In

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

6 - Analysis - structured reporting

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**.*

6 - Analysis - structured reporting

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

6 - Analysis - structured reporting

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

6 - Analysis - structured reporting

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	<i>Disease:</i> Advanced ovarian cancer, stage III and IV <i>Patient source:</i> Surgery 1982 to 1990, University Hospital Freiburg <i>Sample source:</i> Archived specimens available
Excluded	73	General exclusion criteria ^b , non-standard therapy ^b , coefficient of variation >7% ^b
Included	184	Previously untreated. <i>Treatment:</i> all had platinum based chemotherapy after surgery
With outcome events	139	Overall survival: death from any cause

Altman et al. 2012



6 - Analysis - structured reporting

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



6 - Analysis - structured reporting

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

Aim	DFS		OS		Variables considered	Results/remarks
	Patients	Events	Patients	Events		
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 OS: M, v1-v7	Table 4, Table 5 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



6 - Analysis - structured reporting

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 1 st Jan. 1984 to 31 st 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

6 - Analysis - structured reporting

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	FPT 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

7 - Diagnostic and prognostic models

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



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 Reitsma², Douglas G Altman¹ and Karel GM Moons²

Abstract

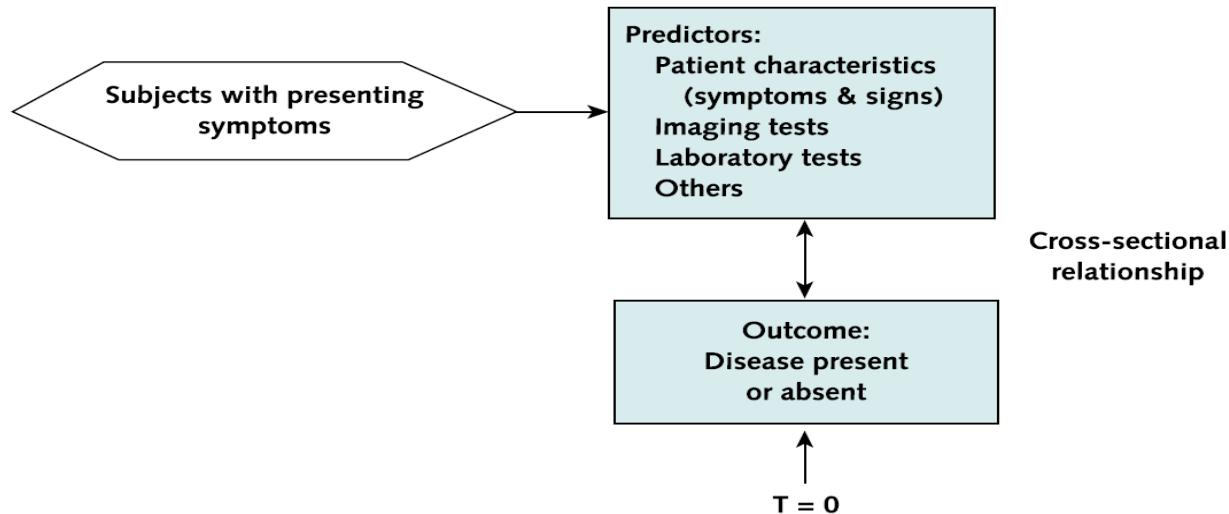
Prediction models are developed to aid health care providers in estimating the probability or risk that a specific disease or condition is present (diagnostic models) or that a specific event will occur in the future (prognostic models), to inform their decision making. However, the overwhelming evidence 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. (...)



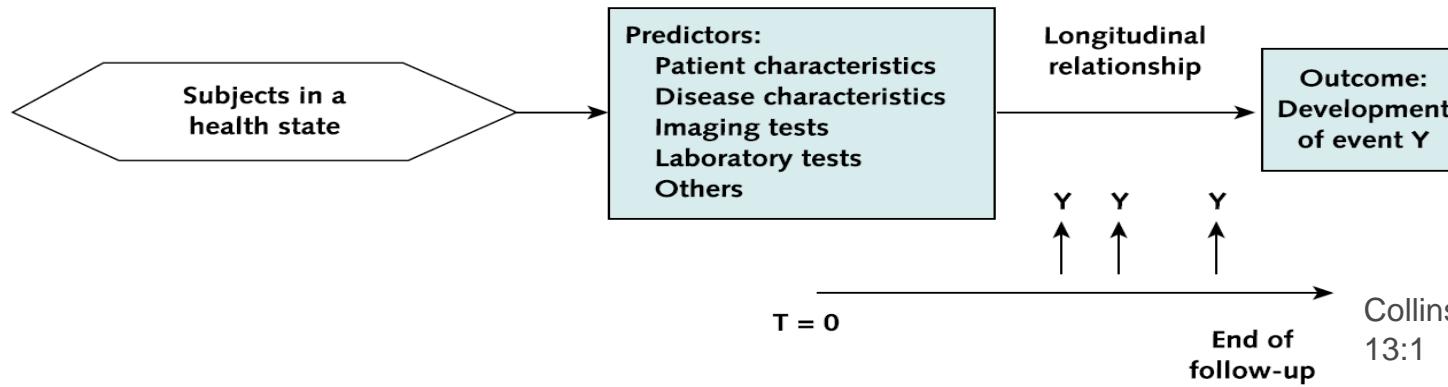
7 - Diagnostic and prognostic models

Figure 1 Schematic representation of diagnostic and prognostic prediction modeling studies.

Diagnostic multivariable modeling study



Prognostic multivariable modeling study

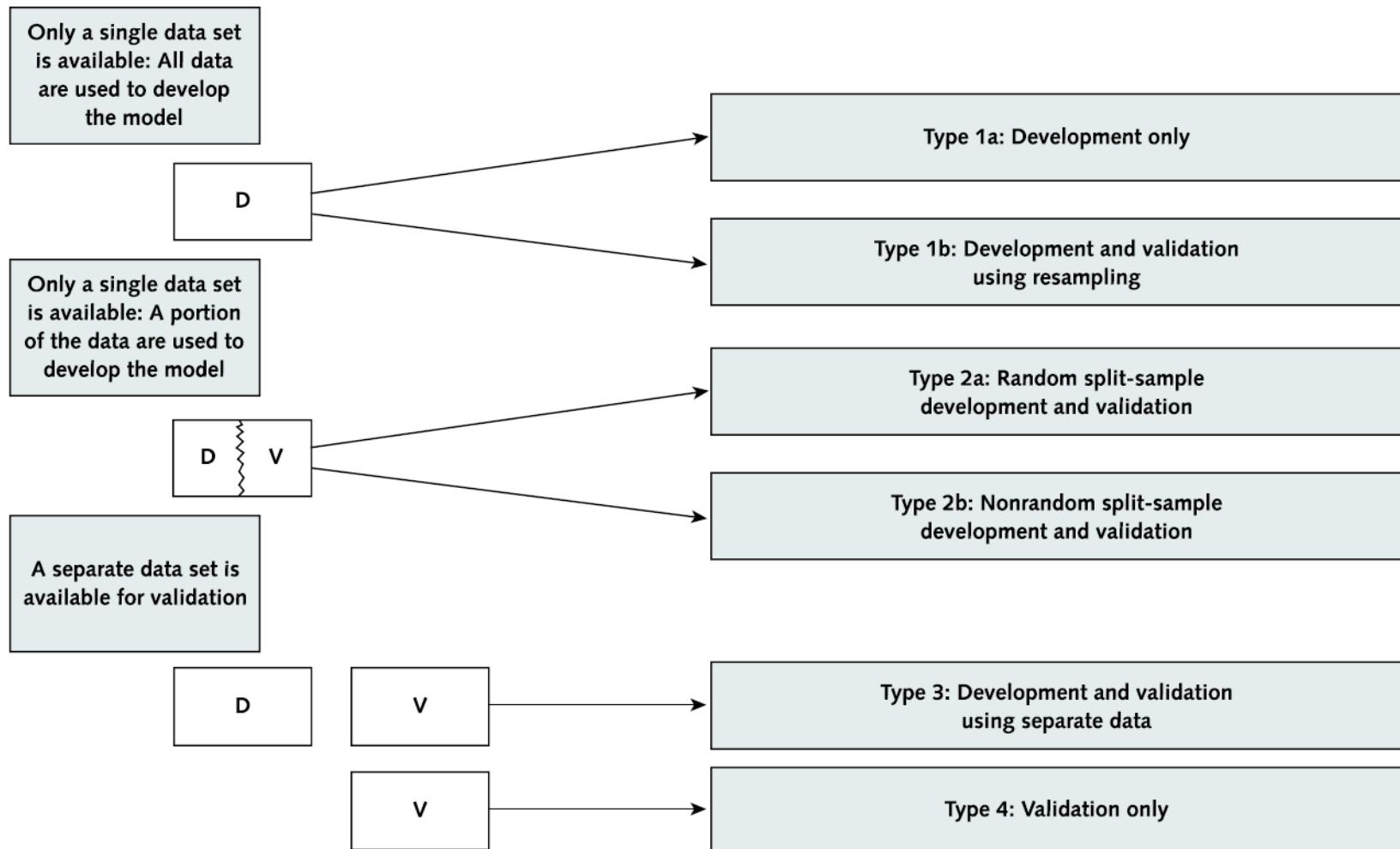


Collins et al. BMC Medicine (2015)
13:1



7 - 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

7 - Diagnostic and prognostic models

Proteomics 2014, 14, 1587–1592

DOI 10.1002/pmic.201300377

1587



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

Belinda Hernández^{1,2}, Andrew Parnell¹ and Stephen R. Pennington²

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



7 - Diagnostic and prognostic models

REVIEW

Open Access



Tufts PACE Clinical Predictive Model Registry: update 1990 through 2015

Benjamin S. Wessler^{1,2}, Jessica Paulus², Christine M. Lundquist², Muhammad Ajlan^{2,3}, Zuhair Natto², William A. Janes², Nitin Jethmalani², Gowri Raman⁴, Jennifer S. Lutz² and David M. Kent^{2*}

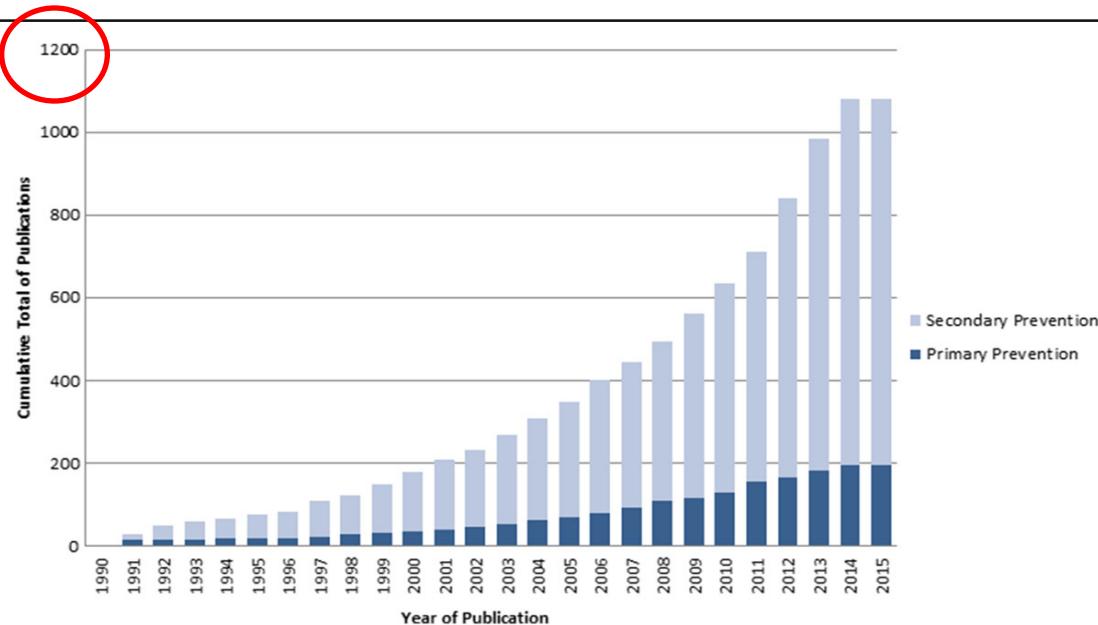


Fig. 2 Cumulative growth in published CPM articles included in the Tufts CPM database over time (January 1990–March 2015). Dark blue represents models derived on CVD-free population samples. Light blue represents models derived on patients with specific cardiovascular conditions at baseline

CPMs for cardiovascular diseases

Wessler et al. *Diagnostic and Prognostic Research* (2017) 1:20,



7 - Diagnostic and prognostic models



Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement

Pauline Heus^{1,2*} , Johanna A. A. G. Damen^{1,2}, Romin Pajouheshnia², Rob J. P. M. Scholten^{1,2}, Johannes B. Reitsma^{1,2}, Gary S. Collins³, Douglas G. Altman³, Karel G. M. Moons^{1,2} and Lotty Hooft^{1,2}

Heus *et al.* *BMC Medicine* (2018) 16:120

170 models:

- 73 (43%) on model development
- 43 (25%) on external validation
- 33 (19%) on incremental value
- 21 (12%) on combined development and external validation of the same model

Overall, publications adhered to a median of 44% of TRIPOD items.

7 - Diagnostic and prognostic models

Completeness of reporting of individual TRIPoD items (n = 170 models)

		Complete reporting (%)
10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation	24
10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models	21
15a	Present the full prediction model to allow predictions for individuals (i.e. all regression coefficients, and model intercept or baseline survival at a given time point)	17
16	Report performance measures (with confidence intervals [CIs]) for the prediction model	14
2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions	2

Heus *et al.* BMC Medicine (2018) 16:120

Some items are very often not reported!

7 - Diagnostic and prognostic models

Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature

Laura E. Cowley* , Daniel M. Farewell, Sabine Maguire and Alison M. Kemp

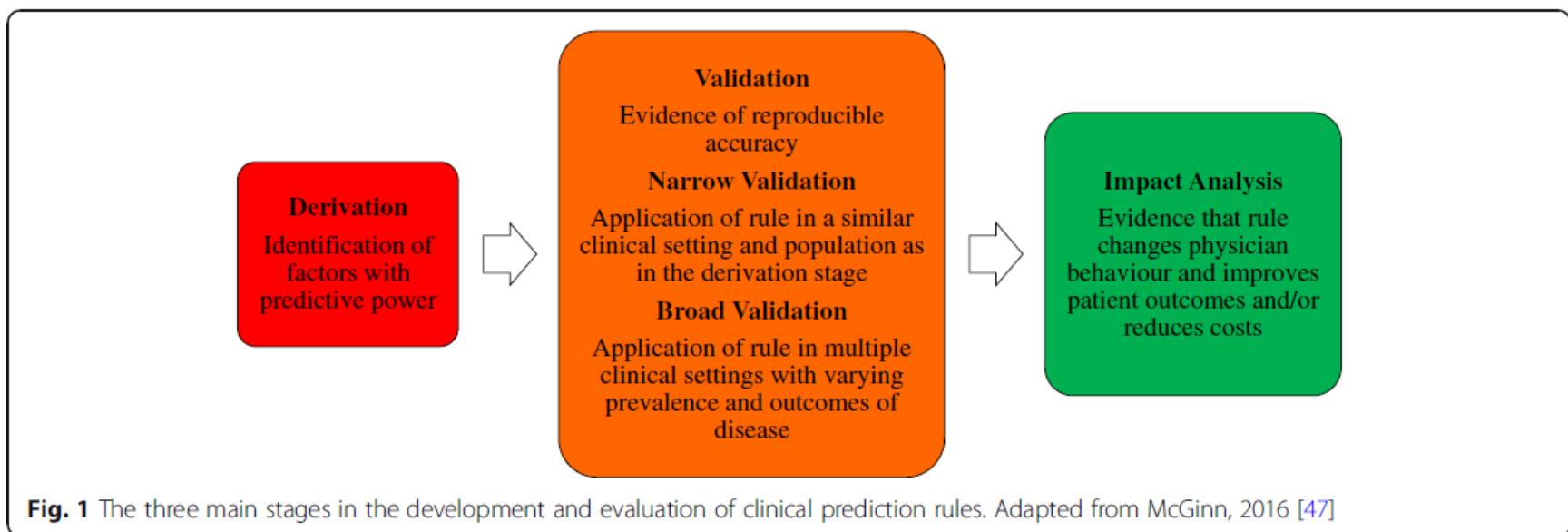


Fig. 1 The three main stages in the development and evaluation of clinical prediction rules. Adapted from McGinn, 2016 [47]

Cowley et al. *Diagnostic and Prognostic Research*

(2019) 3:16

7 - Diagnostic and prognostic models

Table 2 Hierarchies of evidence in the development and evaluation of clinical prediction rules

Level of evidence	Definitions and standards of evaluation	Implications for clinicians
Level 1: Derivation of CPR	Identification of predictors using multivariable model; blinded assessment of outcomes.	Needs validation and further evaluation before it is used clinically in actual patient care.
Level 2: Narrow validation of CPR	Validation of CPR when tested prospectively in one setting; blinded assessment of outcomes.	Needs validation in varied settings; may use CPR cautiously in patients similar to derivation sample.
Level 3: Broad validation of CPR	Validation of CPR in varied settings with wide spectrum of patients and clinicians.	Needs impact analysis; may use CPR predictions with confidence in their accuracy.
Level 4: Narrow impact analysis of CPR used for decision-making	Prospective demonstration in one setting that use of CPR improves clinicians' decisions (quality or cost-effectiveness of patient care).	May use cautiously to inform decisions in settings similar to that studied.
Level 5: Broad impact analysis of CPR used for decision-making	Prospective demonstration in varied settings that use of CPR improves clinicians' decisions for wide spectrum of patients.	May use in varied settings with confidence that its use will benefit patient care quality or effectiveness.

Adapted from Reilly and Evans 2016 [32]. CPR clinical prediction rule

Cowley *et al. Diagnostic and Prognostic Research*

(2019) 3:16



7 - Diagnostic and prognostic models

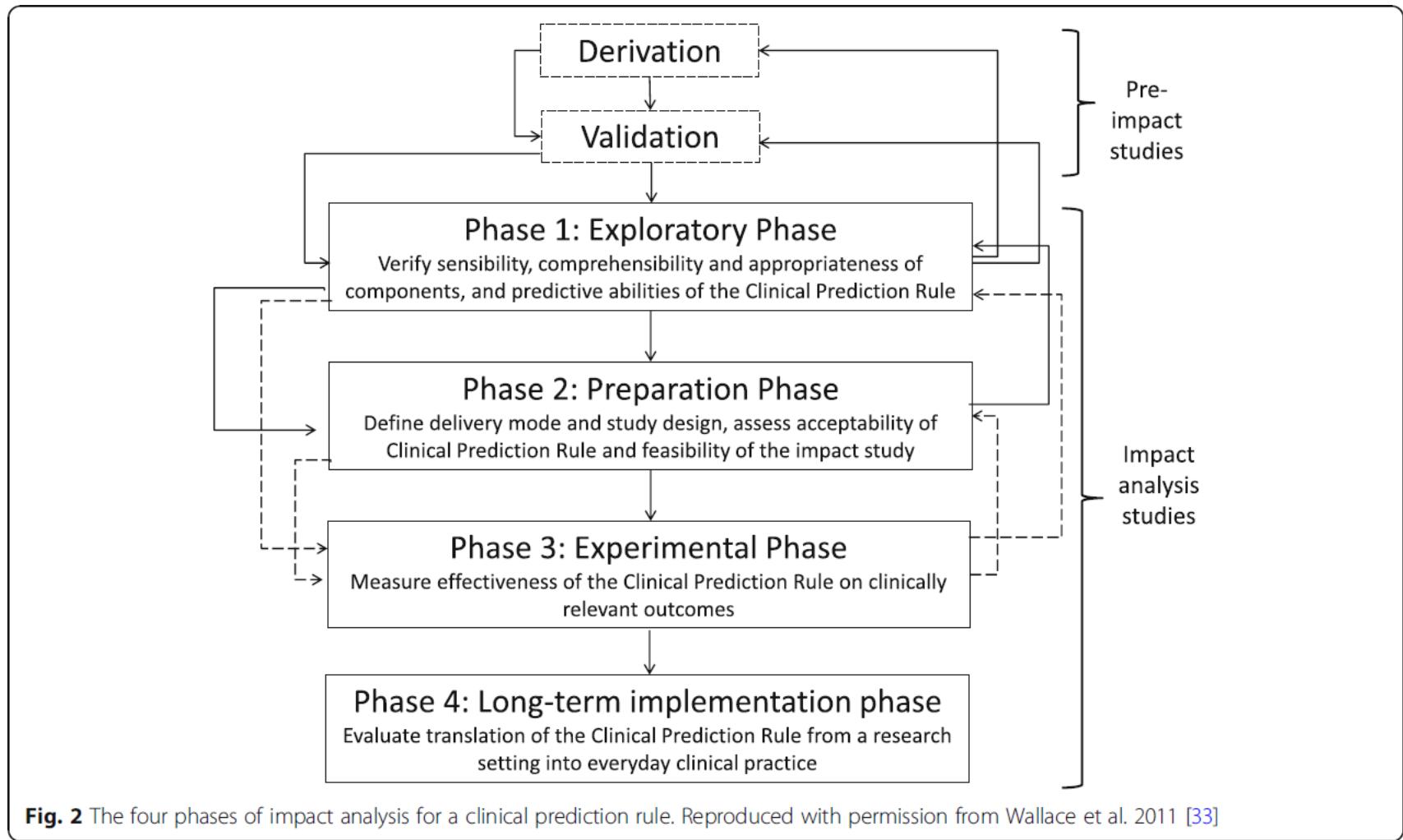


Fig. 2 The four phases of impact analysis for a clinical prediction rule. Reproduced with permission from Wallace et al. 2011 [33]

Cowley et al. *Diagnostic and Prognostic Research*

(2019) 3:16



7 - Diagnostic and prognostic models

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

Emmanuelle Kempf^{1,2}, Jennifer A. de Beyer¹, Jonathan Cook¹, Jane Holmes¹, Seid Mohammed¹, Tri-Long Nguyễn^{1,3}, Iveta Simera⁴, Marialena Trivella¹, Douglas G. Altman¹, Sally Hopewell¹, Karel G. M. Moons^{5,6}, Raphael Porcher⁷, Johannes B. Reitsma^{5,6}, Willi Sauerbrei⁸ 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.**

7 - 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



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

Initiatives to improve the situation

Initiatives to improve the situation

STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative

Willi Sauerbrei,^{a*†} Michal Abrahamowicz,^b
Douglas G. Altman,^c Saskia le Cessie,^d and[‡] James Carpenter^e
on behalf of the STRATOS initiative

Statistics in Medicine 2014, 33: 5413-5432.

Preliminary ideas
Discussions, SG
Initiative launched
Invited Sessions

ISCB 2011, Ottawa
ISCB 2012, Bergen
ISCB 2013, Munich
ISCB 2014, 2015

...
General meetings

...
BIRS 2016, 2019

<http://www.stratos-initiative.org/>

Statistical methodology – Current situation

- Substantial development over last decades
- Computer facilities
- Assess properties of complex models using simulation studies
- Resampling and Bayesian methods now easily available
- Wealth of new statistical software packages
- **Unfortunately, many sensible improvements are ignored in routine analyses**

Reasons:

- 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

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 initiative

- Provide evidence supported guidance for highly relevant issues in the design and analysis of observational studies
- 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
- For the start we will concentrate on state-of-the-art guidance and the necessary evidence
- Help to identify questions requiring much more primary research

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

Topic group 2:

**Selection of variables and their
functional forms in
multivariable analysis**

Many strategies for variable selection available - more new methods needed?

[...] in statistical research and related methodology-oriented fields such as machine learning or bioinformatics, the well-known adage ‘publish or perish’ could be translated into ‘**propose new methods or perish.**’

Such a research paradigm is **not favorable for studies that aim at meaningfully comparing alternative existing methods** or, more generally, studies assessing the behavior and properties of existing methods.

It becomes **more and more difficult to get an overview of existing methods**, not to mention the overview of their respective performances in different settings.”

Boulesteix et al. 2018

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

Relevant issues in deriving evidence-supported state-of-the-art guidance for multivariable model building

- 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?

Guidance for analysis is needed for many stakeholders (analysts with different levels of knowledge, teachers, reviewers, journalists,)

Researchers

First in a Series of Papers for the Biometric Bulletin

STRATOS initiative – Guidance for designing and analyzing observational studies



Willi Sauerbrei¹, Marianne Huebner², Gary S. Collins³, Katherine Lee⁴, Laurence Freedman⁵, Mitchell Gail⁶, Els Goetghebeur⁷, Joerg Rahnenfuehrer⁸ and Michal Abrahamowicz⁹ on behalf of the STRATOS initiative.

Short papers from

TG1 – missing data

→ **TG4** – measurement error and misclassification

TG3 – initial data analysis

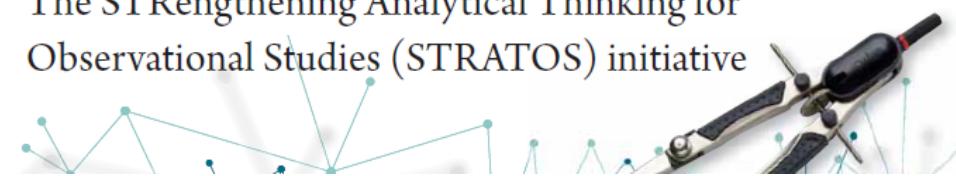
TG2 – Variable and function selection

TG7 – Causal Inference have appeared

Consumers

Guidance for designing and analysing observational studies:

The STREngthening Analytical Thinking for Observational Studies (STRATOS) initiative



Willi Sauerbrei¹, Gary S. Collins²,
Marianne Huebner³, Stephen D. Walter⁴,
Suzanne M. Cadarette⁵, and
Michal Abrahamowicz⁶ on behalf of the
STRATOS initiative

Volume 26 Number 3 | Medical Writing September 2017 | 17

Journal of the European Medical Writers Association
(EMWA)

9 - PROGRESS

Initiatives to improve the situation – PROGRESS partnership

The PROGnosis RESearch Strategy (PROGRESS) Partnership is a UK Medical Research Council (MRC) funded, international, interdisciplinary collaboration developing understanding in research into quality of care outcomes, prognostic factors, risk prediction models, and predictors of differential treatment response.

The objectives of the Partnership are:

- To critically develop concepts, methods and recommendations for improving prognosis research, and systematically apply these across different disease areas, in order to enhance the translational impact of prognosis research;
- Bring together leaders in different clinical disciplines for novel collaborative opportunities;
- To develop guidelines, workshops and prognosis research training courses

<http://progress-partnership.org/>



PROGRESS- framework

Overall Prognosis Research

Prognostic Factor Research

Prognostic Model Research

Stratified Medical Research

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

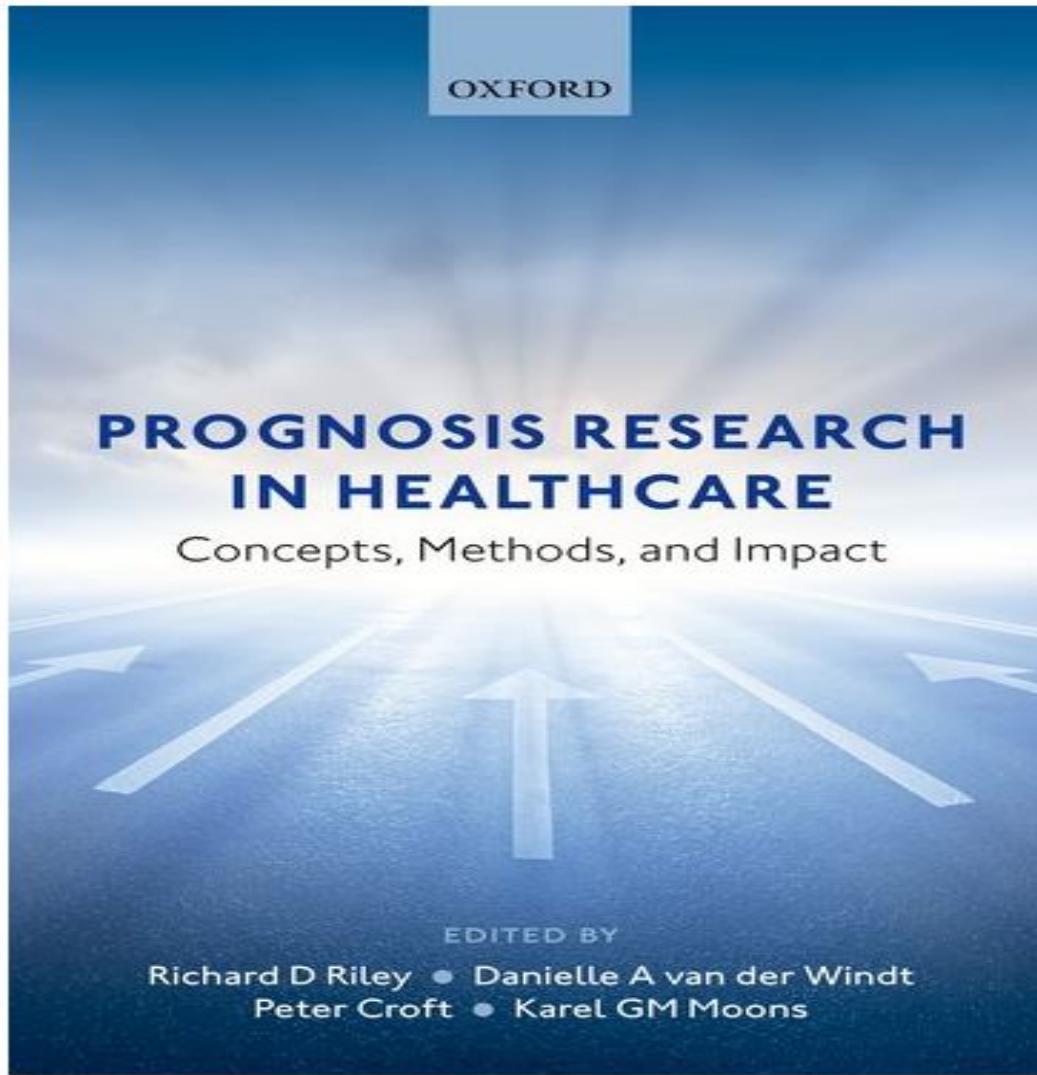
9 - PROGRESS

Recommendations of PROGRESS

Supplementary table 1: Recommendations of PROGRESS (PROgnosis RESearch Strategy)

Recommendation (PROGRESS paper(s) introducing it)	Challenge or opportunity	Recommendation
1 Fundamental shift <u>(1,2,3,4)</u>	Improvements in electronic health records, clinical imaging, and -omic technologies (genotyping and phenotyping) are beginning to challenge current disease taxonomy, clinical pre-occupation with diagnosis (rather than risk) and the focus of health care policy on process (rather than clinical outcomes).	There should be a fundamental shift in clinical practice, translational research and health care policy based on evidence from prognosis research i.e. the prospective relationships between the phenotypic, genomic and environmental assessment of people with a given startpoint and subsequent endpoints.
2 Systems <u>(1,2,3,4)</u>	Over the lifecourse individuals develop multiple diseases (both distinct and related) that often are not reflected in the current organisation of medical research or practice.	There should be an expansion of prognosis research which bridges multiple clinical specialities, health systems, pathological mechanisms, and biological systems and puts the whole patient across their 'journey' as the central unit of concern.
3 Electronic health records <u>(1,2,3,4)</u>	The scope and impact of prognosis research and electronic health records research (in primary and secondary care, and in disease and procedure registries) are intimately related.	There should be new programmes of methodological and empirical prognosis research exploiting electronic health records to define, phenotype and follow up people with different health related conditions.
4 Field <u>(1,2,3,4)</u>	Prognosis research is currently fragmented and not visible as a distinct entity.	Prognosis research should be recognised as a field of enquiry important in translational research, and intrinsic to the practice of clinical medicine and development of health care policy. Efforts should be made to establish prognosis research as a distinct branch of knowledge, with a set of scientific methods aimed at understanding and improving health.
5 Comparing prognosis <u>(1)</u>	The relative impact of having, compared to not having, a health condition on survival or symptom status helps identify priorities for translational research but is uncommonly reported outside the field of cancer.	There should be greater efforts to compare prognosis between those with and without a given condition, and between different conditions.

9 - PROGRESS

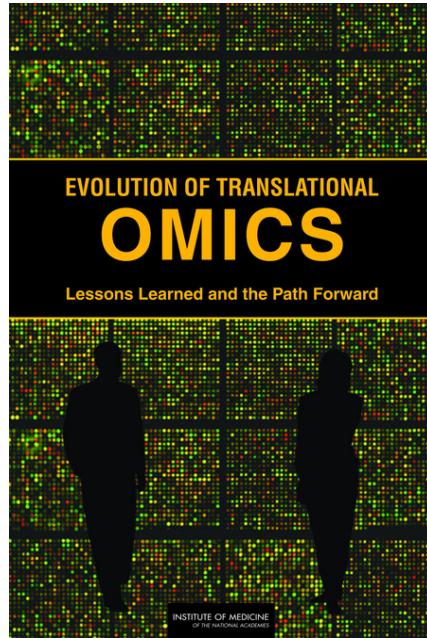


2019



Reproducible Research

Triggered by problems identified in working with
omics data



Further remarks

The Ninth International Congress on Peer Review and Scientific Publication A Call for Research *September 12-14, 2021 in Chicago*

The quantity and quality of scientific research have never been greater, but with unprecedented promise comes unprecedented peril. There are better scientific policies and processes, stronger standards for openness and transparency, and innovative technologies to collaborate and publish. However, the rapidly evolving scientific publication ecosystem that facilitates research dissemination also enables research waste, predation, and piracy. The challenge of distinguishing information from noise, innovation from dystopianlike disruption, and opportunity from threat has created new levels of excitement and angst for those engaged in research and its reporting, publication, and distribution.

*John P. A. Ioannidis, MD, DSc; Michael Berkwits, MD, MSCE; Annette Flanagin, RN, MA; Fiona Godlee, MBBChir, FRCP; Theodora Bloom, PhD
JAMA, September 2019*



Substantial development of statistical methodology

...and application in practice...

“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**.“

Unreliable Research: Trouble at the lab [*The Economist* 2013]

Weaknesses in analyses can have severe consequences for patients

“A mistake in the operating room can threaten the life of one patient; a mistake in statistical analysis or interpretation can lead to hundreds of early deaths. So it is perhaps odd that, while we allow a doctor to conduct surgery only after years of training, we give SPSS to almost anyone.”

Andrew Vickers [*Nat Clin Pract Urol* 2005]

BMJ Lifetime Achievement Award (2015)



Professor Doug Altman, co-founder of the EQUATOR Network, has been awarded the BMJ Lifetime Achievement Award in recognition of his outstanding contribution to the improvement of the scientific and medical research literature. Professor Altman is one of the world's leading experts in health research methodology, statistics and reporting and has spent his career working to improve transparency in the conduct and reporting of health research. Over the years Professor Altman has led or been involved in developing many of the reporting guidelines listed on the EQUATOR website.

The BMJ states "Altman has done more than anybody to raise the standards of medical publication and in the process has transformed the role of statistician from number cruncher to custodian of important but often neglected values".

Fiona Godlee, Editor-in-chief of The BMJ said "he has done more than anyone else to encourage researchers to fully report what they actually did, warts and all, rather than letting the best be the enemy of the good or, worse, pretending that research is perfect".

Some references

- Altman D.G., McShane L., Sauerbrei W., Taube S.E. (2012): Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. PLoS Med; 9(5): E 1001216.
- Altman D.G., Simera I. (2016): A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. J R Soc Med; 109(2):67-77.
- Collins G.S., Reitsma J.B, Altman D.G., Moons K.G. (2015): Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med;162(1):55-63.
- Cowley L. E., Farewell D. M., Maguire S., & Kemp A. M. (2019). Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. *Diagnostic and Prognostic Research*, 3(1), 16.
- Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, Altman DG, Moons KGM, Hooft L (2018): Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. BMC Medicine. 16:120. doi:10.1186/s12916-018-1099-2
- McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM for the Statistics Subcommittee of the NCI-EORTC Working on Cancer Diagnostics (2005): REporting recommendations for tumor MARKer prognostic studies (REMARK). Journal of the National Cancer Institute. 97: 1180-1184
- Moons K.G., Altman D.G., Reitsma J.B., Ioannidis J.P.A., Macaskill P., Steyerberg E.W., Vickers A.J., Ransohoff D.F., Collins G.S. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med. 2015;162:W1–W73.
- Peat G., Riley R.D., Croft P., Morley K.I., Kyza P.A., Moons K.G.M., Perel P., Steyerberg E.W., Schroter S., Altman D.G., Hemingway H., for the PROGRESS Group (2014): Improving the Transparency of Prognosis Research: The Role of Reporting, Data Sharing, Registration, and Protocols. PLoS Medicine; 11(7): e1001671.
- Riley D.R., van der Windt D., Moons K.G.M. (2019): Prognosis Research in Healthcare: Concepts, Methods, and Impact. Oxford University Press; 9780198796619.
- Sauerbrei W., Abrahamowicz M., Altman D.G., le Cessie S. and Carpenter J. on behalf of the STRATOS initiative. (2014) STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. Statistics in Medicine, 33: 5413-5432
- Sekula, P., Mallett, S., Altman, D.G., Sauerbrei, W. (2017): Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. PLoS ONE; 12(6):e0178531.
- Simera I., Moher D., Hirst A., Hoey J., Schulz K.F., Altman D.G. (2010): Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. BMC Med; (8): 24.
- Winzer K.J., Buchholz A., Schumacher M., Sauerbrei W. (2016): Improving the prognostic ability through better use of standard clinical data - the Nottingham Prognostic Index as an example. PLoS ONE; 11(3):e0149977.
- Wolff R. F., Moons K. G., Riley R. D., Whiting P. F., Westwood M., Collins G. S., Reitsma J.B., Kleijnen J. & Mallett S. (2019). PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, 170(1), 51-58.

