Improving epidemiological research: avoiding the statistical paradoxes and fallacies that nobody else talks about

Maarten van Smeden, PhD

28th Norwegian Epidemiological Association (NOFE) conference
October 26, 2022

UMC Utrecht
Julius Centrum

Slides available at:

https://www.slideshare.net/MaartenvanSmeden

I have no conflicts of interest to declare

UMC Utrecht

Index

Part I:

- Absence of evidence fallacy
- Table 2 fallacy
- Winner’s curse
- Stein’s paradox

Part II:

- Good statistical practices
- Avoiding statistical fallacies and paradoxes

UMC Utrecht

TRIGGER WARNING

BMJ 2017;357:j2490 doi: 10.1136/bmj.j2490 (Published 2017 June 08)

RESEARCH

Check for updates

Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane rev study

Ag
Da
Pe

BMJ 2018;363:k4645 doi: 10.1136/bmj.k4645 (Published 12 November 2018)

FEATURE

Check for updates

ESSAY

Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers

Progress has been made towards reducing the 85% of wasted effort in medical research—and the huge amounts of money misspent and harm caused to patients—but there's still a long way to go, say **Paul Glasziou** and **Iain Chalmers**

Paul Glasziou *director*[1], Iain Chalmers *coordinator*[2]

Research: increasing value,

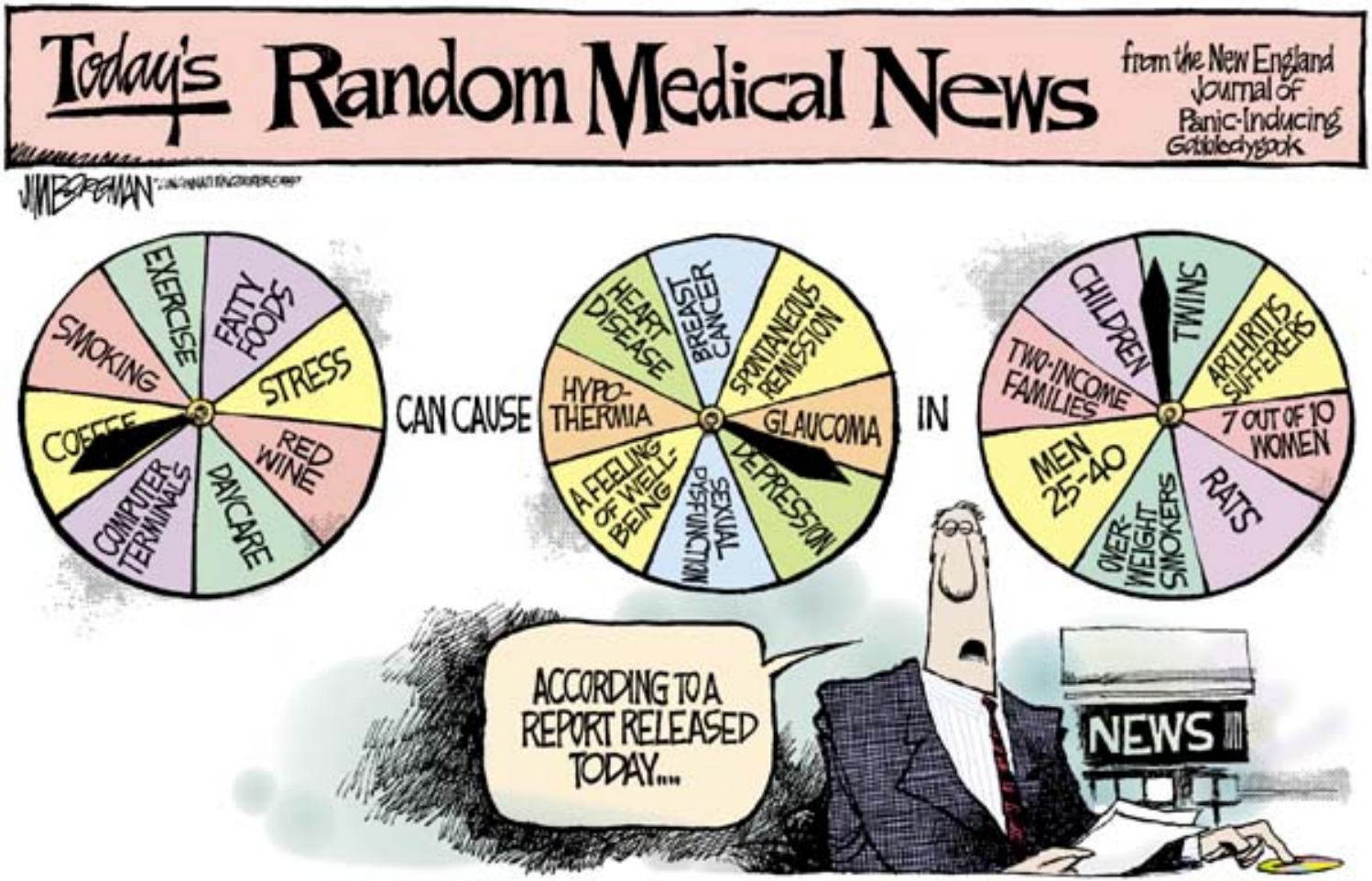
Reducing waste from incor biomedical research

Paul Glasziou, Douglas G Altman, Patrick Bossuyt, Isabelle Bou

Research publication can both communicate and miscommunicate. Unless research is adequately reported, the time

Lancet 2014; 383: 267-76

UMC Utrecht



UMC Utrecht

Cartoon of Jim Borgman, published by the Cincinnati Inquirer and King Features Syndicate April 27 1997

Cookbook review

See corresponding editorial on page 5.

Is everything we eat associated with cancer? A systematic cookbook review[1–3]

Jonathan D Schoenfeld and John PA Ioannidis

ABSTRACT

Background: Nutritional epidemiology is a highly prolific field.

and such discrepancies in the evidence have fueled hot debates (9–12) rife with emotional and sensational rhetoric that can

“We selected 50 common ingredients from random recipes of a cookbook”

Design: We selected 50 common ingredients from random recipes in a cookbook. PubMed queries identified recent studies that evaluated the relation of each ingredient to cancer risk. Information regarding author conclusions and relevant effect estimates were extracted. When >10 articles were found, we focused on the 10 most recent articles.

ploratory, the analyses and protocols are not preregistered, and the findings are selectively reported. It was previously shown in a variety of other fields that “negative” results are either less likely to be published (16–21) or misleadingly interpreted (19, 22). Studies may spuriously highlight results that barely achieve statistical significance (15, 23) or report effect estimates that

UMC Utrecht

Cookbook review

veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, milk, cheese, coffee, bacon, sugar, lobster, potato, beef, lamb, mustard, nuts, wine, peas, corn, cinnamon, cayenne, orange, tea, rum, raisin, bay leaf, cloves, thyme, vanilla, hickory, molasses, almonds, baking soda, ginger, terrapin

UMC Utrecht

Cookbook review

veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, [illegible] coffee, [illegible], sugar, [illegible], [illegible] leaf, [illegible], [illegible], [illegible],

Cookbook review

veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, milk, cheese, coffee, bacon, sugar, lobster, potato, beef, lamb, mustard, nuts, wine, peas, corn, cinnamon, cayenne, orange, tea, rum, raisin, bay leaf, cloves, thyme, vanilla, hickory, molasses, almonds, baking soda, ginger, terrapin

40/50 (80%)

UMC Utrecht

Cookbook review

HOW MANY OF THE INVESTIGATED INGREDIENTS HAVE BEEN REPORTED TO INCREASE OR DECREASE RISK OF CANCER?

Cookbook review

veal, salt, pepper spice, flour, egg, bread, pork, butter, tomato, lemon, duck, onion, celery, carrot, parsley, mace, sherry, olive, mushroom, tripe, milk, cheese, coffee, bacon, sugar, lobster, potato, beef, lamb, mustard, nuts, wine, peas, corn, cinnamon, cayenne, orange, tea, rum, raisin

36/40 (90%)

UMC Utrecht

5. Discussion

Among 14-year olds living in the UK, we found an association between social media use and depressive symptoms and that this was stronger for girls than for boys. The magnitude of these associations reduced when potential explanatory factors were taken into account, sug-

ELSEVIER eclinicalmedicine

hypothesised pathways between social media use and depressive symptoms. Findings are based largely on cross sectional data and thus causality cannot be inferred.

These findings are highly relevant to current policy development on guidelines for the safe use of social media and calls on industry to more tightly regulate hours of social media use for young people [10,

UMC Utrecht

Research waste: 85% (?)

Avoidable waste in the production and reporting of research evidence

Iain Chalmers, Paul Glasziou

Lancet 2009; 374: 86–89

Published Online June 15, 2009 DOI:10.1016/S0140-6736(09)60329-9

James Lind Library, James Lind Initiative, Oxford, UK (Sir I Chalmers DSc); **and Centre for Evidence-Based Medicine, Department of Primary Care, University of Oxford, Oxford, UK** (Prof P Glasziou RACGP)

Correspondence to: Sir Iain Chalmers, James Lind Library, James Lind Initiative, Summertown Pavilion, Middle Way, Oxford OX2 7LG, UK **ichalmers@jameslindlibrary.org**

Without accessible and usable reports, research cannot help patients and their clinicians. In a published Personal View,[1] a medical researcher with myeloma reflected on the way that the results of four randomised trials relevant to his condition had still not been published, years after preliminary findings had been presented in meeting abstracts:

> "Research results should be easily accessible to people who need to make decisions about their own health... Why was I forced to make my decision knowing that information was somewhere but not available? Was the delay because the results were less exciting than expected? Or because in the evolving field of myeloma research there are now new exciting hypotheses (or drugs) to look at? How far can we tolerate the butterfly behaviour of researchers, moving on to the next flower well before the previous one has been fully exploited?"

research involving patients have been powerful disincentives for those who might otherwise have become involved in research in treatment evaluation. In recent years, there has been recognition of the need to address both of these disincentives. In the UK, the Cooksey enquiry concluded that government support for applied research should be increased,[3] and the National Institute for Health Research (NIHR) has responded rapidly to this policy (its funding for clinical trials will soon be £80 million a year).[4] In the USA, a bill currently before Congress calls for federal support for evaluations of treatments independent of industry, and in Italy and Spain, independent research on the effects of drugs is being supported with revenue from a tax on pharmaceutical company drug promotion.[5]

This increased investment in independent treatment evaluation is laudable. Irrespective of who sponsors research, this investment should be protected from the

UMC Utrecht

Statistical illiteracy?

P values
f overall
analyses
partly
atistical
even if
group,
wer, due
al distri-
fference
specific
all death

mate of a 5% decrease in 10-year survival with watchful waiting, 750 men might have died prematurely as a result.

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® (SPSS, Chicago, IL) to almost anyone. Moreover, whilst only a surgeon would comment on surgical technique, it seems that anybody, regardless of statistical training,

day); ar
that on
risk of t
in many
that the
event is

Comp
The aut
no com

UMC Utrecht

Andrew Vickers, Nat Rev Urol, 2005, doi: 10.1038/ncpuro0294

ABSENCE OF EVIDENCE FALLACY

Absence of evidence is not evidence of absence

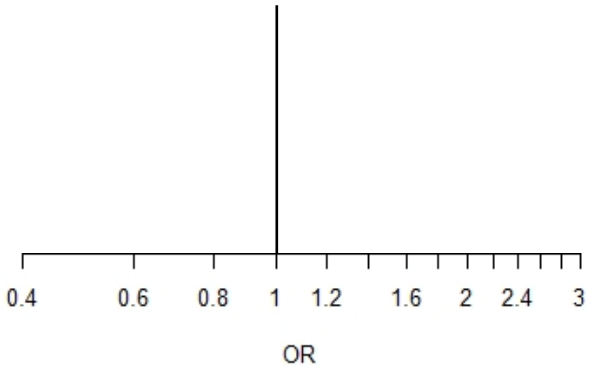
Douglas G Altman, J Martin Bland

By convention a P value greater than 5% (P>0·05) is called "not significant." Randomised controlled clinical trials that do not show a significant difference between the treatments being compared are often called "negative." This term wrongly implies that the study has shown that there is no difference, whereas usually all that has been shown is an absence of evidence of a difference. These are quite different statements.

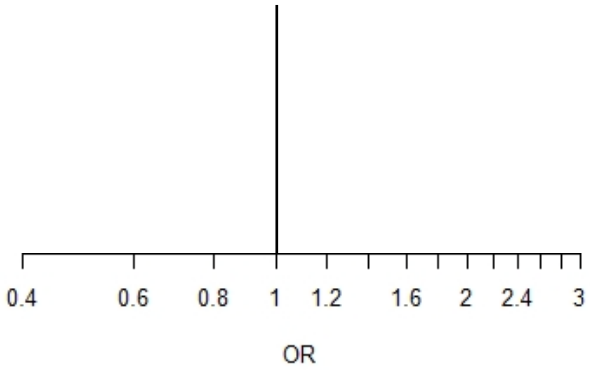
UMC Utrecht

Doi: 10.1136/bmj.311.7003.485

Hypothetical example

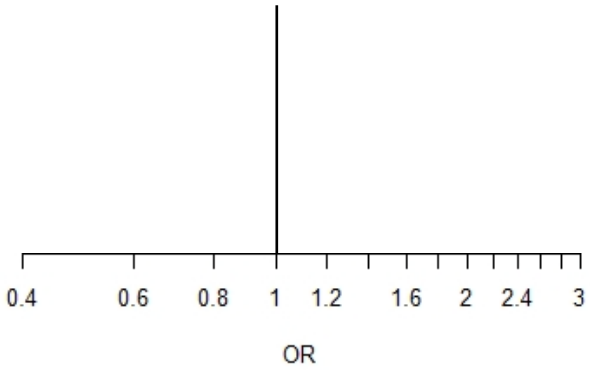


Factor A

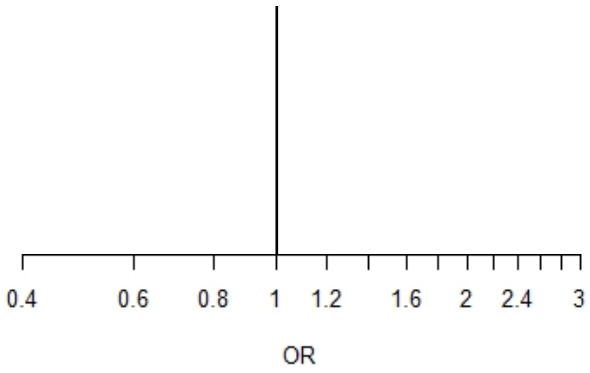


Factor B

Hypothetical example



Coffee



Tea

sky news

24 Oct 17° 12° Watch Live

Home UK World Politics US Climate Science & Tech Business Ents & Arts Travel Offbeat More

Drinking at least two cups of tea a day protects your heart longer, study sugge

The study does not definitively establish th mortality of tea drinkers, because it cannot factors associated with tea consumption.

Tuesday 30 August 2022 09:20, UK

HEALTHY EATING ✓ Evidence-Based

Dangerous Side Effects of Drinking Too Much Tea, According to Science

While tea is one of the healthiest beverages you can enjoy, some drinking habits can have a negative impact on your health.

CNN health Audio Live TV Log In

life but better Food

Coffee lowers risk of heart problems and early death, study says, especially ground and caffeinated

By Sandee LaMotte, C

Published 5:15 AM ED

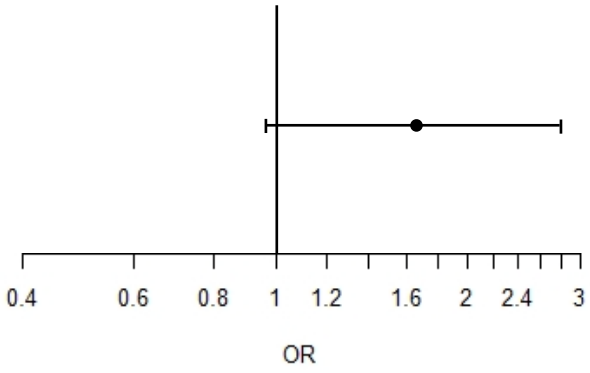
NEWS • HEALTH NEWS

There Are So Many Health Studies on Coffee. Which One Should You Trust?

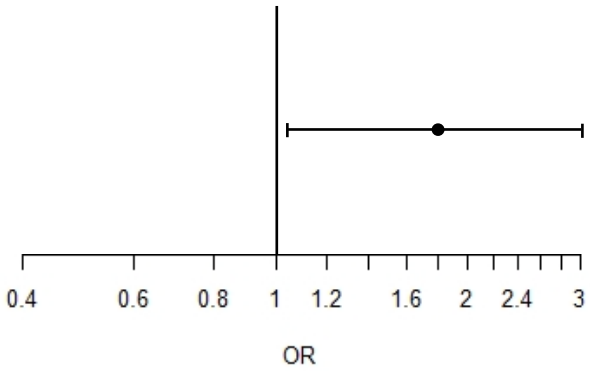
By Stephanie Brown | Published on December 20, 2021

Fact checked by Angela Underwood

Hypothetical example

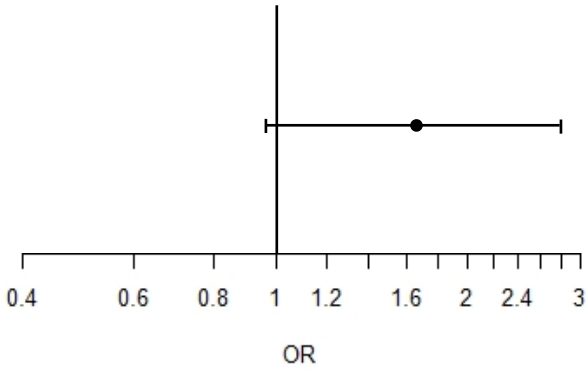


Coffee



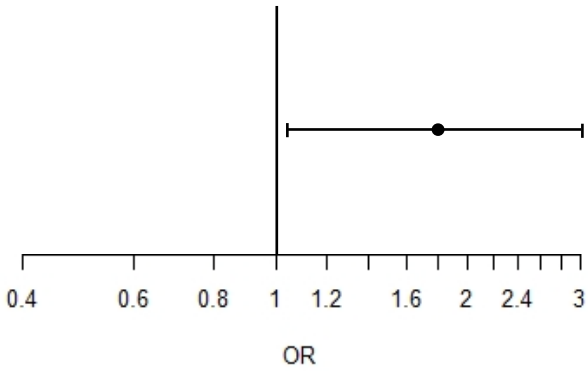
Tea

Hypothetical example



Coffee

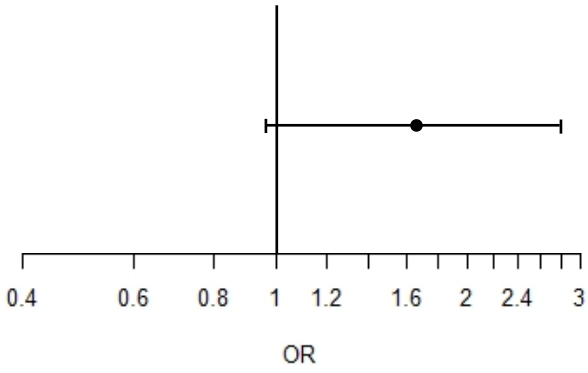
"No health consequences for Coffee"



Tea

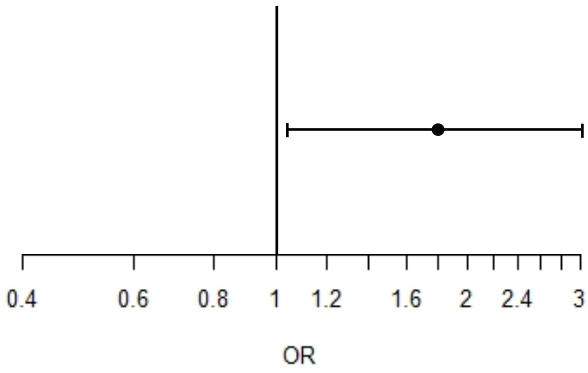
"Negative health consequences for Tea"

Hypothetical example



Coffee

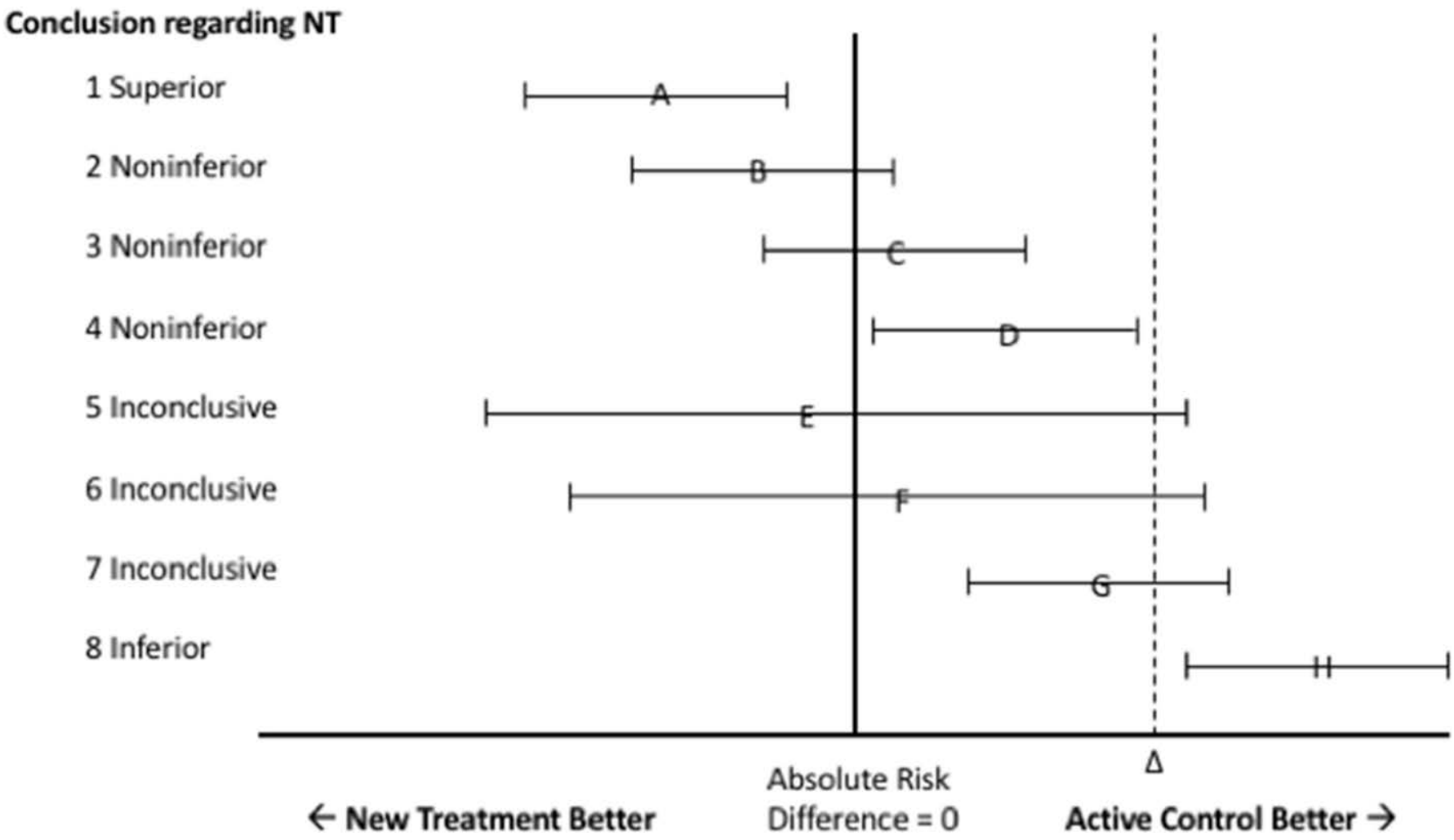
"No health consequences for Coffee"
"No effect of Coffee"
"Coffee is healthy"
"Coffee is better for health than tea"



Tea

"Negative health consequences for Tea"
"Tea is bad for health"
"Tea kills"

Non-inferiority

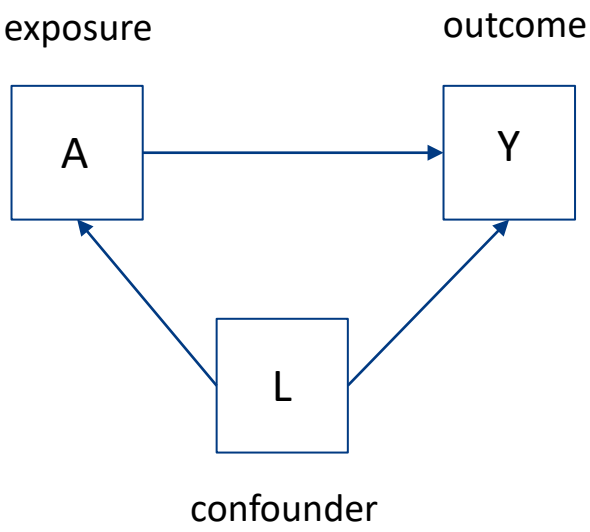


UMC Utrecht

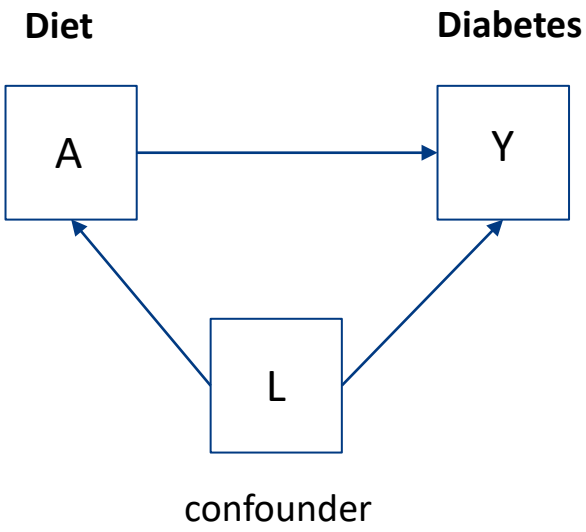
Source fig doi: 10.1007/s11606-018-4813-z

TABLE 2 FALLACY

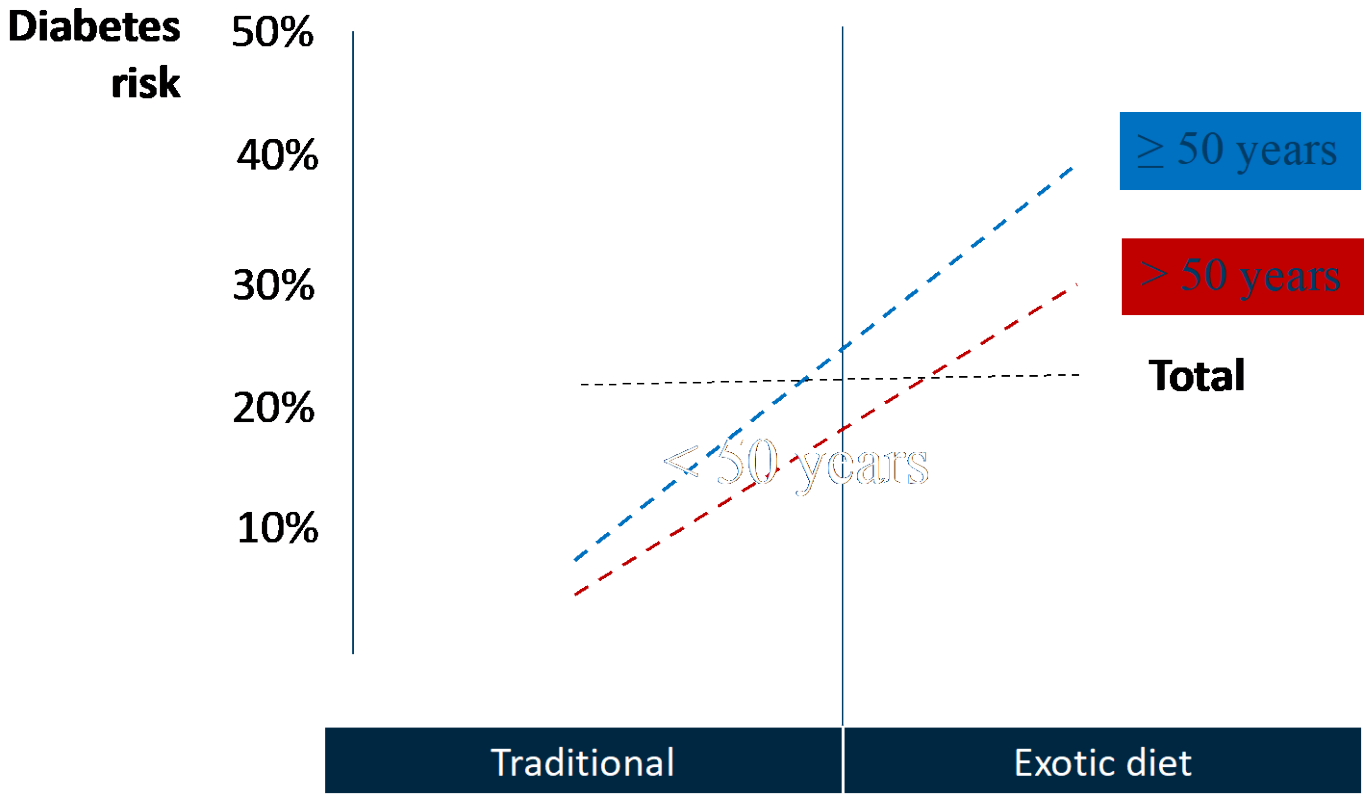
Observational (non-randomized) study



Observational (non-randomized) study



Diet -> diabetes, age a counfounder?

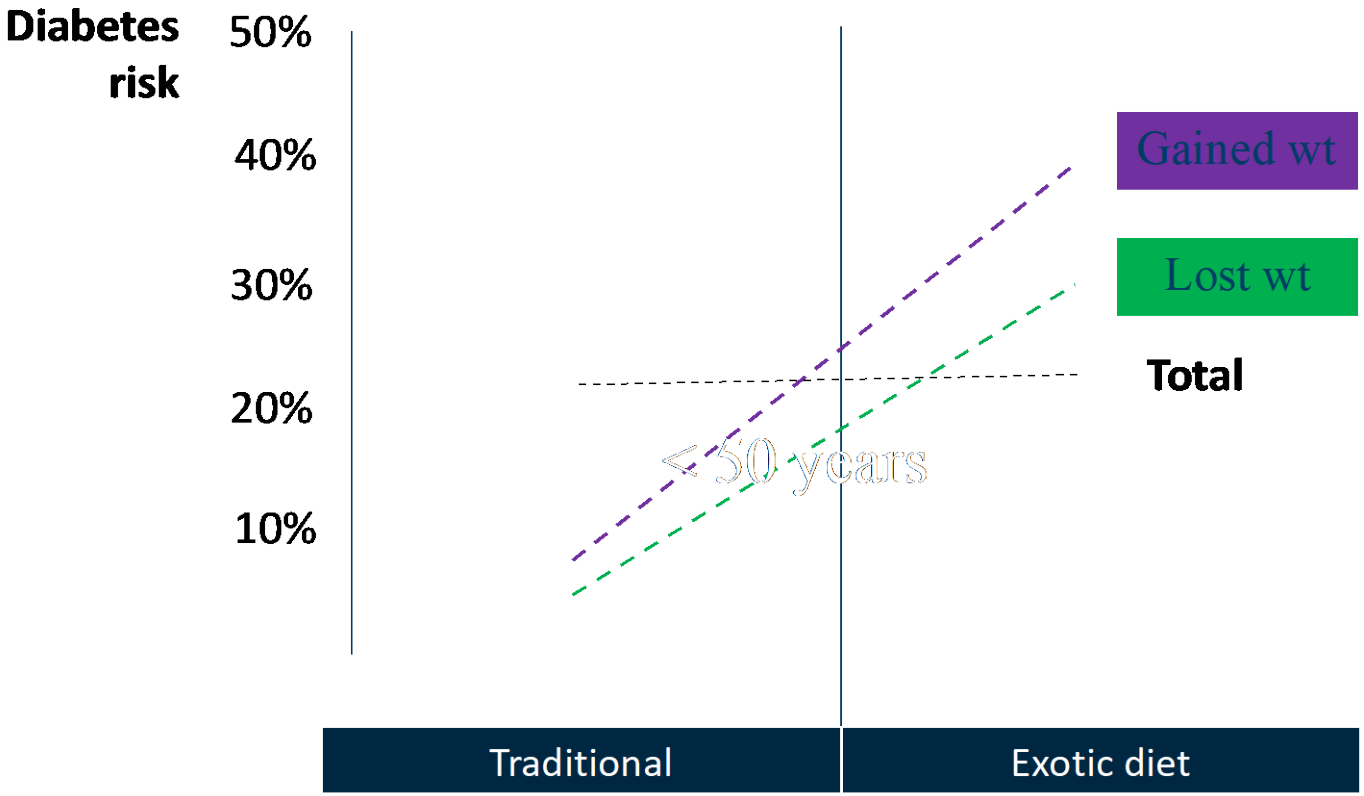


Age	No diabetes	Diabetes	No diabetes	Diabetes	RR
< 50 years	19	1	37	3	1.50
≥ 50 years	28	12	12	8	1.33
Total	**47**	**13**	**49**	**11**	**0.88**

UMC Utrecht

Numerical example adapted from Peter Tennant with permission

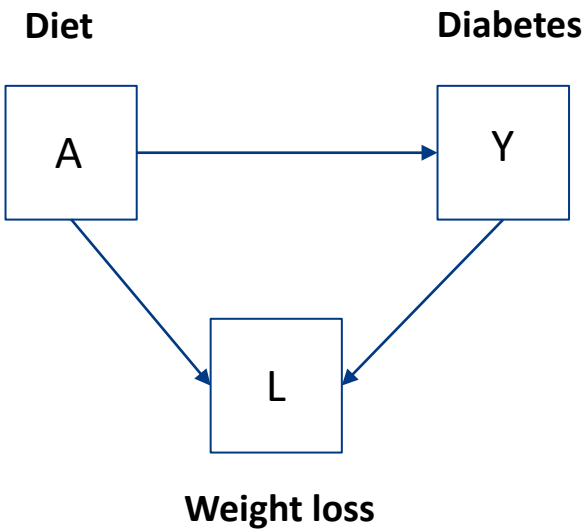
Diet -> diabetes, weight loss a confounder?



Weight	No diabetes	Diabetes	No diabetes	Diabetes	RR
Lost	19	1	37	3	1.50
Gained	28	12	12	8	1.33
Total	**47**	**13**	**49**	**11**	**0.88**

UMC Utrecht

Beware of the colliders



Careful selection of confounders, e.g. through DAGs

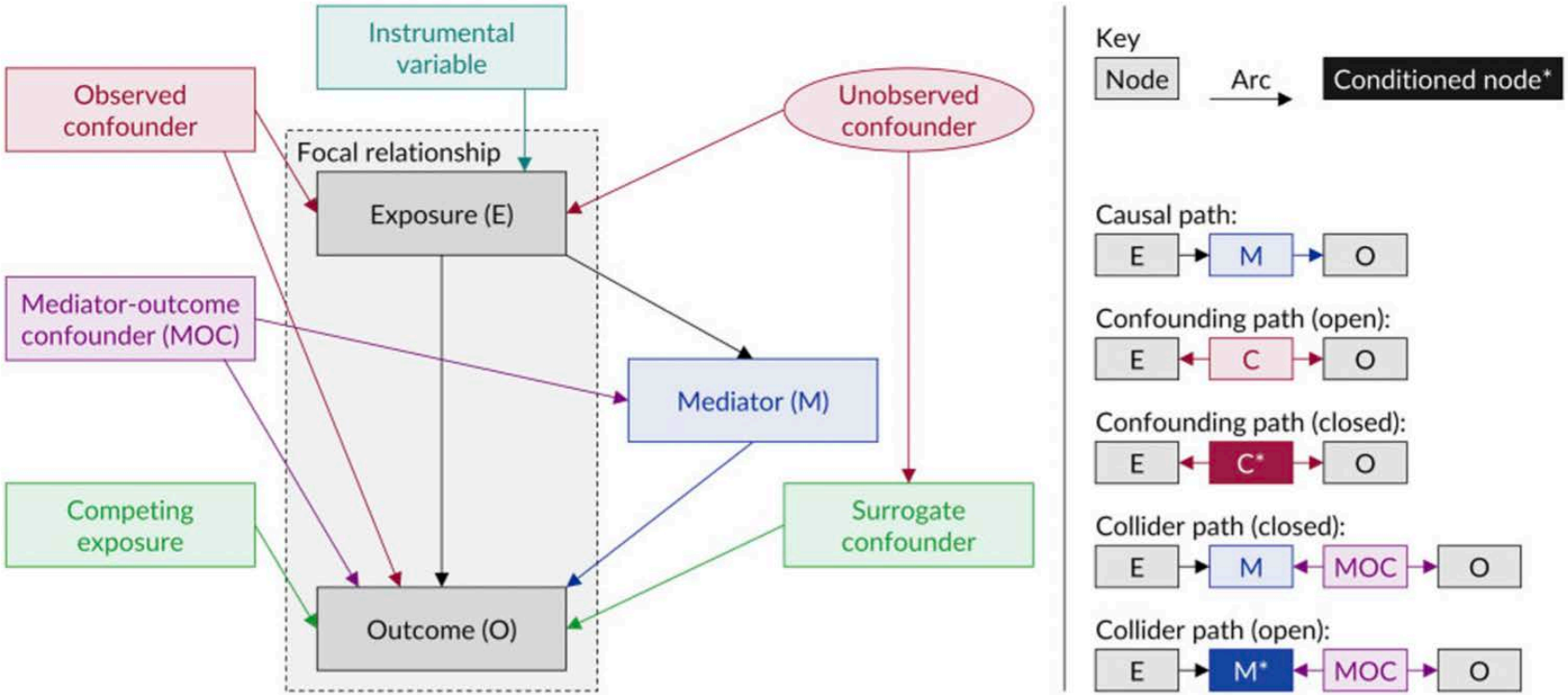


Figure 1 Illustration of the main components of a DAG, the most common types of contextual variables and the most common types of paths. The DAG has been visually arranged so that all constituent arcs flow from top-to-bottom.

UMC Utrecht

From Tennant et al, IJE, 2021, doi: 10.1093/ije/dyaa213

Example of multivariable model table

PLOS O

ar
udy

stitute of
s University
nskog,
Norway

	Odds Ratio	95% Confidence Interval	P–Value
Age (Per Decade)	1.05	1.01–1.08	**0.001**
Sex			
Female	REF	REF	REF
Male	1.19	1.07–1.32	**0.002**
ECI	1.05	1.04–1.06	**0.001**
Clotting Risk Factors			
No Risk Factors	REF	REF	REF
Prior VTE	25.44	19.70–33.29	< **0.001**
Factor V Leiden	24.34	16.96–33.29	< **0.001**
Active Cancer	1.84	1.30–2.60	< **0.001**
Prior MI	1.03	0.71–1.50	0.87
Fracture Type			
Ankle	1.51	1.35–1.69	< **0.001**
Talus	1.07	0.80–1.40	0.63
Calcaneus	1.24	1.00–1.53	**0.048**
Tarsal	0.2	0.69–1.21	0.58
Metatarsal	REF	REF	REF
Multiple F&A Fractures	1.51	1.22–1.85	< **0.001**
Treatment type			
Nonsurgical	REF	REF	REF
Surgical	1.41	1.15–1.72	< **0.001**

Controlled for age, sex, ECI, clotting risk factors, surgical treatment, type of fracture/multiple fractures.

Bolding indicates of $p < 0.05$.

REF = referent variable

VTE = Venous Thromboembolism

ECI = Elixhauser Comorbidity Index

MI = Myocardial Infarction

https://doi.org/10.1371/journal.pone.0276548.t002

Check for updates

UMC Utrecht

Table 2 fallacy

Common approach:

- Fit a multivariable regression model using all “risk factors” of interest
- Presents estimates of regression coefficients as mutually adjusted for each other

American Journal of Epidemiology
Vol. 177, No. 4
DOI: 10.1093/aje/kws412
Advance Access publication: January 30, 2013
© The Author 2013. Published by Oxford University Press on behalf of the Johns Hopkins Bloomberg School of Public Health. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

Commentary

The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients

Daniel Westreich* and Sander Greenland

* Correspondence to Dr. Daniel Westreich, Department of Obstetrics and Gynecology, Duke Global Health Institute, Duke University, DUMC 3967, Durham, NC 27710 (e-mail: daniel.westreich@duke.edu).

Initially submitted January 13, 2012; accepted for publication October 11, 2012.

It is common to present multiple adjusted effect estimates from a single model in a single table. For example, a table might show odds ratios for one or more exposures and also for several confounders from a single logistic regression. This can lead to mistaken interpretations of these estimates. We use causal diagrams to display the sources of the problems. Presentation of exposure and confounder effect estimates from a single model may lead to several interpretative difficulties, inviting confusion of direct-effect estimates with total-effect estimates for covariates in the model. These effect estimates may also be confounded even though the effect estimate for the main exposure is not confounded. Interpretation of these effect estimates is further complicated by heterogeneity (variation, modification) of the exposure effect measure across covariate levels. We offer suggestions to limit potential misunderstandings when multiple effect estimates are presented, including precise distinction between total and direct effect measures from a single model, and use of multiple models tailored to yield total-effect estimates for covariates.

causal diagrams; causal inference; confounding; direct effects; epidemiologic methods; mediation analysis; regression modeling

Abbreviation: HIV, human immunodeficiency virus.

Example of Table 2 fallacy

Article

Factors a

death usi

https://doi.org/10.1038/s41

Received: 15 May 2020

Accepted: 1 July 2020

Published online: 8 July 202

Check for updates

Article

Table 2 | Hazard ratios and 95% confidence intervals for COVID-19-related death

Characteristic	Category	COVID-19 death HR (95% CI)	
		Adjusted for age and sex	Fully adjusted
Age	18–39	0.05 (0.04–0.07)	0.06 (0.04–0.08)
	40–49	0.28 (0.23–0.33)	0.30 (0.25–0.36)
	50–59	1.00 (ref)	1.00 (ref)
	60–69	2.79 (2.52–3.10)	2.40 (2.16–2.66)
	70–79	8.62 (7.84–9.46)	6.07 (5.51–6.69)
	80+	38.29 (35.02–41.87)	20.60 (18.70–22.68)
Sex	Female	1.00 (ref)	1.00 (ref)
	Male	1.78 (1.71–1.85)	1.59 (1.53–1.65)
BMI (kg m^{-2})	Not obese	1.00 (ref)	1.00 (ref)
	30–34.9 (obese class I)	1.23 (1.17–1.30)	1.05 (1.00–1.11)
	35–39.9 (obese class II)	1.81 (1.68–1.95)	1.40 (1.30–1.52)
	≥40 (obese class III)	2.66 (2.39–2.95)	1.92 (1.72–2.13)
Smoking	Never	1.00 (ref)	1.00 (ref)
	Former	1.43 (1.37–1.49)	1.19 (1.14–1.24)
	Current	1.14 (1.05–1.23)	0.89 (0.82–0.97)
Ethnicity[a]	White	1.00 (ref)	1.00 (ref)
	Mixed	1.62 (1.26–2.08)	1.43 (1.11–1.84)
	South Asian	1.69 (1.54–1.84)	1.45 (1.32–1.58)
	Black	1.88 (1.65–2.14)	1.48 (1.29–1.69)
	Other	1.37 (1.13–1.65)	1.33 (1.10–1.61)
IMD quintile	1 (least deprived)	1.00 (ref)	1.00 (ref)
	2	1.16 (1.08–1.23)	1.12 (1.05–1.19)
	3	1.31 (1.23–1.40)	1.22 (1.15–1.30)
	4	1.69 (1.59–1.79)	1.51 (1.42–1.61)
	5 (most deprived)	2.11 (1.98–2.25)	1.79 (1.68–1.91)
Blood pressure	Normal	1.00 (ref)	1.00 (ref)
	High blood pressure or diagnosed hypertension	1.09 (1.05–1.14)	0.89 (0.85–0.93)
Respiratory disease excluding asthma		1.95 (1.86–2.04)	1.63 (1.55–1.71)
Asthma[b] (versus none)	With no recent OCS use	1.13 (1.07–1.20)	0.99 (0.93–1.05)
	With recent OCS use	1.55 (1.39–1.73)	1.13 (1.01–1.26)
Chronic heart disease		1.57 (1.51–1.64)	1.17 (1.12–1.22)
Diabetes[c] (versus none)	With HbA1c < 58 mmol mol^{-1}	1.58 (1.51–1.66)	1.31 (1.24–1.37)
	With HbA1c ≥ 58 mmol mol^{-1}	2.61 (2.46–2.77)	1.95 (1.83–2.08)
	With no recent HbA1c measure	2.27 (2.06–2.50)	1.90 (1.72–2.09)
Cancer (non-haematological, versus none)	Diagnosed <1 year ago	1.81 (1.58–2.07)	1.72 (1.50–1.96)
	Diagnosed 1–4.9 years ago	1.20 (1.10–1.32)	1.15 (1.05–1.27)
	Diagnosed ≥5 years ago	0.99 (0.93–1.06)	0.96 (0.91–1.03)
Haematological malignancy (versus none)	Diagnosed <1 year ago	3.02 (2.24–4.08)	2.80 (2.08–3.78)
	Diagnosed 1–4.9 years ago	2.56 (2.14–3.06)	2.46 (2.06–2.95)
	Diagnosed ≥5 years ago	1.70 (1.46–1.98)	1.61 (1.39–1.87)
Reduced kidney function[d] (versus none)	eGFR 30–60	1.56 (1.49–1.63)	1.33 (1.28–1.40)
	eGFR < 30	3.48 (3.23–3.75)	2.52 (2.33–2.72)
Liver disease		2.39 (2.06–2.77)	1.75 (1.51–2.03)
Stroke or dementia		2.57 (2.46–2.70)	2.16 (2.06–2.27)
Other neurological disease		3.08 (2.85–3.33)	2.58 (2.38–2.79)
Organ transplant		6.00 (4.73–7.61)	3.53 (2.77–4.49)
Asplenia		1.62 (1.19–2.21)	1.34 (0.98–1.83)
Rheumatoid arthritis, lupus or psoriasis		1.30 (1.21–1.38)	1.19 (1.11–1.27)
Other immunosuppressive condition		2.75 (2.10–3.62)	2.21 (1.68–2.90)

Models were adjusted for age using a four-knot cubic spline for age, except for estimation of age-group hazard ratios. Ref, reference group; 95% CI, 95% confidence interval.
[a]Ethnicity hazard ratios were estimated from a model restricted to those with recorded ethnicity.
[b]For OCS use, 'recent' refers to during the year before baseline.
[c]Classification by HbA1c is based on measurements within 15 months of baseline.
[d]eGFR is measured in ml min^{-1} per 1.73 m^2 and taken from the most recent serum creatinine measurement.

lated

eb Bacon[2,6], Chris Bates[3,6],
Peter Inglesby[2],
e Tomlinson[1],
Wong[1], Richard Grieve[1],
lohn Parry[3], Frank Hester[3],
& Ben Goldacre[2,7]✉

UMC Utrecht

Example of Table 2 fallacy

Article

Factors a

death usi

lated

Article

Table 2 | Hazard ratios and 95% confidence intervals for COVID-19-related death

Characteristic	Category	COVID-19 death HR (95% CI)	
		Adjusted for age and sex	Fully adjusted
Age	18–39	0.05 (0.04–0.07)	0.06 (0.04–0.08)
	40–49	0.28 (0.23–0.33)	0.30 (0.25–0.36)
	50–59	1.00 (ref)	1.00 (ref)
	60–69	2.79 (2.52–3.10)	2.40 (2.16–2.66)
	70–79	8.62 (7.84–9.46)	6.07 (5.51–6.69)
	80+	38.29 (35.02–41.87)	20.60 (18.70–22.68)
Sex	Female	1.00 (ref)	1.00 (ref)
	Male	1.78 (1.71–1.85)	1.59 (1.53–1.65)
BMI (kg m^{-2})	Not obese	1.00 (ref)	1.00 (ref)
	30–34.9 (obese class I)	1.23 (1.17–1.30)	1.05 (1.00–1.11)
	35–39.9 (obese class II)	1.81 (1.68–1.95)	1.40 (1.30–1.52)
	≥40 (obese class III)	2.66 (2.39–2.95)	1.92 (1.72–2.13)
Smoking	Never	1.00 (ref)	1.00 (ref)
	Former	1.43 (1.37–1.49)	1.19 (1.14–1.24)
	Current	1.14 (1.05–1.23)	0.89 (0.82–0.97)
Ethnicity[a]	White	1.00 (ref)	1.00 (ref)
	Mixed	1.62 (1.26–2.08)	1.43 (1.11–1.84)
	South Asian	1.69 (1.54–1.84)	1.45 (1.32–1.58)
	Black	1.88 (1.65–2.14)	1.48 (1.29–1.69)
	Other	1.27 (1.12–1.65)	1.22 (1.10–1.61)
Respiratory disease excluding asthma		1.95 (1.86–2.04)	1.63 (1.55–1.71)
Asthma[b] (versus none)	With no recent OCS use	1.13 (1.07–1.20)	0.99 (0.93–1.05)
	With recent OCS use	1.55 (1.39–1.73)	1.13 (1.01–1.26)
Chronic heart disease		1.57 (1.51–1.64)	1.17 (1.12–1.22)
Diabetes[c] (versus none)	With HbA1c < 58 mmol mol^{-1}	1.58 (1.51–1.66)	1.31 (1.24–1.37)
	With HbA1c ≥ 58 mmol mol^{-1}	2.61 (2.46–2.77)	1.95 (1.83–2.08)
	With no recent HbA1c measure	2.27 (2.06–2.50)	1.90 (1.72–2.09)
Cancer (non-haematological, versus none)	Diagnosed <1 year ago	1.81 (1.58–2.07)	1.72 (1.50–1.96)
	Diagnosed 1–4.9 years ago	1.20 (1.10–1.32)	1.15 (1.05–1.27)
	Diagnosed ≥5 years ago	0.99 (0.93–1.06)	0.96 (0.91–1.03)
Haematological malignancy (versus none)	Diagnosed <1 year ago	3.02 (2.24–4.08)	2.80 (2.08–3.78)
	Diagnosed 1–4.9 years ago	2.56 (2.14–3.06)	2.46 (2.06–2.95)
	Diagnosed ≥5 years ago	1.70 (1.46–1.98)	1.61 (1.39–1.87)
Reduced kidney function[d] (versus none)	eGFR 30–60	1.56 (1.49–1.63)	1.33 (1.28–1.40)
	eGFR < 30	3.48 (3.23–3.75)	2.52 (2.33–2.72)
Liver disease		2.39 (2.06–2.77)	1.75 (1.51–2.03)
Stroke or dementia		2.57 (2.46–2.70)	2.16 (2.06–2.27)
Other neurological disease		3.08 (2.85–3.33)	2.58 (2.38–2.79)
Organ transplant		6.00 (4.73–7.61)	3.53 (2.77–4.49)
Asplenia		1.62 (1.19–2.21)	1.34 (0.98–1.83)
Rheumatoid arthritis, lupus or psoriasis		1.30 (1.21–1.38)	1.19 (1.11–1.27)
Other immunosuppressive condition		2.75 (2.10–3.62)	2.21 (1.68–2.90)

Models were adjusted for age using a four-knot cubic spline for age, except for estimation of age-group hazard ratios. Ref, reference group; 95% CI, 95% confidence interval.
[a]Ethnicity hazard ratios were estimated from a model restricted to those with recorded ethnicity.
[b]For OCS use, 'recent' refers to during the year before baseline.
[c]Classification by HbA1c is based on measurements within 15 months of baseline.
[d]eGFR is measured in ml min^{-1} per 1.73 m^2 and taken from the most recent serum creatinine measurement.

Smoking	Never	1.00 (ref)	1.00 (ref)
	Former	1.43 (1.37–1.49)	1.19 (1.14–1.24)
	Current	1.14 (1.05–1.23)	0.89 (0.82–0.97)

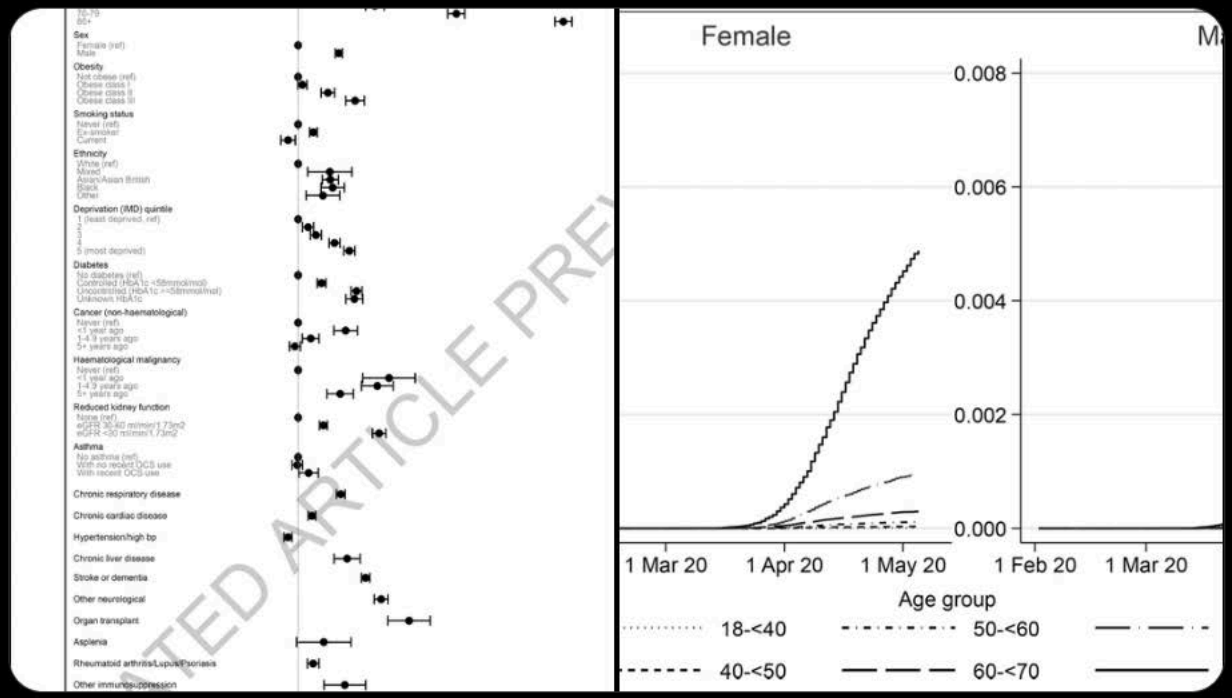
Received: 15 May 2020
Accepted: 1 July 2020
Published online: 8 July 202
Check for updates

e Tomlinson[1],
Wong[1], Richard Grieve[1],
John Parry[3], Frank Hester[3],
& Ben Goldacre[2,7]✉

UMC Utrecht

Eric Topol
@EricTopol

New @nature: the risk factors for dying from #COVID19 from >17 million people and ~11,000 deaths
nature.com/articles/s4158... @bengoldacre and colleagues
importance of age, sex, race, diabetes, obesity, many other conditions
not risk factor: hypertension; current smoker protective



3:14 pm · 8 Jul 2020 · Twitter Web App

443 Retweets **82** Quote Tweets **788** Likes

Received: 14 July 2020 | Revised: 7 December 2020 | Accepted: 13 December 2020
DOI: 10.1111/cdoe.12617

UNSOLICITED SYSTEMATIC REVIEW

Community Dentistry and Oral Epidemiology WILEY

A scoping review of Table 2 fallacy in the oral health literature

Aderonke A. Akinkugbe[1,2,3] | Alyssa M. Simon[2] | Erica R. Brody[4]

[1]Department of Dental Public Health and Policy, School of Dentistry, Virginia Commonwealth University, Richmond, VA, USA
[2]Division of Epidemiology, Department of Family Medicine, School of Medicine

Aderonke A. Akinkugbe, Department of Dental Public Health and Policy, Virginia Commonwealth University. 1101 East Leigh Street, Richmond, VA 23298-0566.
Email: aaakinkugbe@vcu.edu

Funding information
National Institutes of Health/National Institute of Dental and Craniofacial Research, Grant/Award Number: R03DE028403 and L40DE028120

Abstract

Background: Coined by Westreich and Greenland in 2013, Table 2 fallacy refers to the practice of reporting estimates of the primary exposure and adjustment covariates derived from a single model on the same table. This study seeks to describe the

meta-analysis, prediction models or descriptive studies. The remaining 421 articles were eligible for full text reviewed of which, 189 (45%) committed Table 2 fallacy. The prevalence of table 2 fallacy appears high in the oral health literature.

teria. After categorizing the articles, we exported and summarized the results in SAS.

Results: A total of 1358 articles were initially screened of which 937 articles were excluded based on title or abstract for being animal studies, systematic reviews or meta-analysis, prediction models or descriptive studies. The remaining 421 articles were eligible for full text reviewed of which, 189 (45%) committed Table 2 fallacy. The prevalence of table 2 fallacy appears high in the oral health literature.

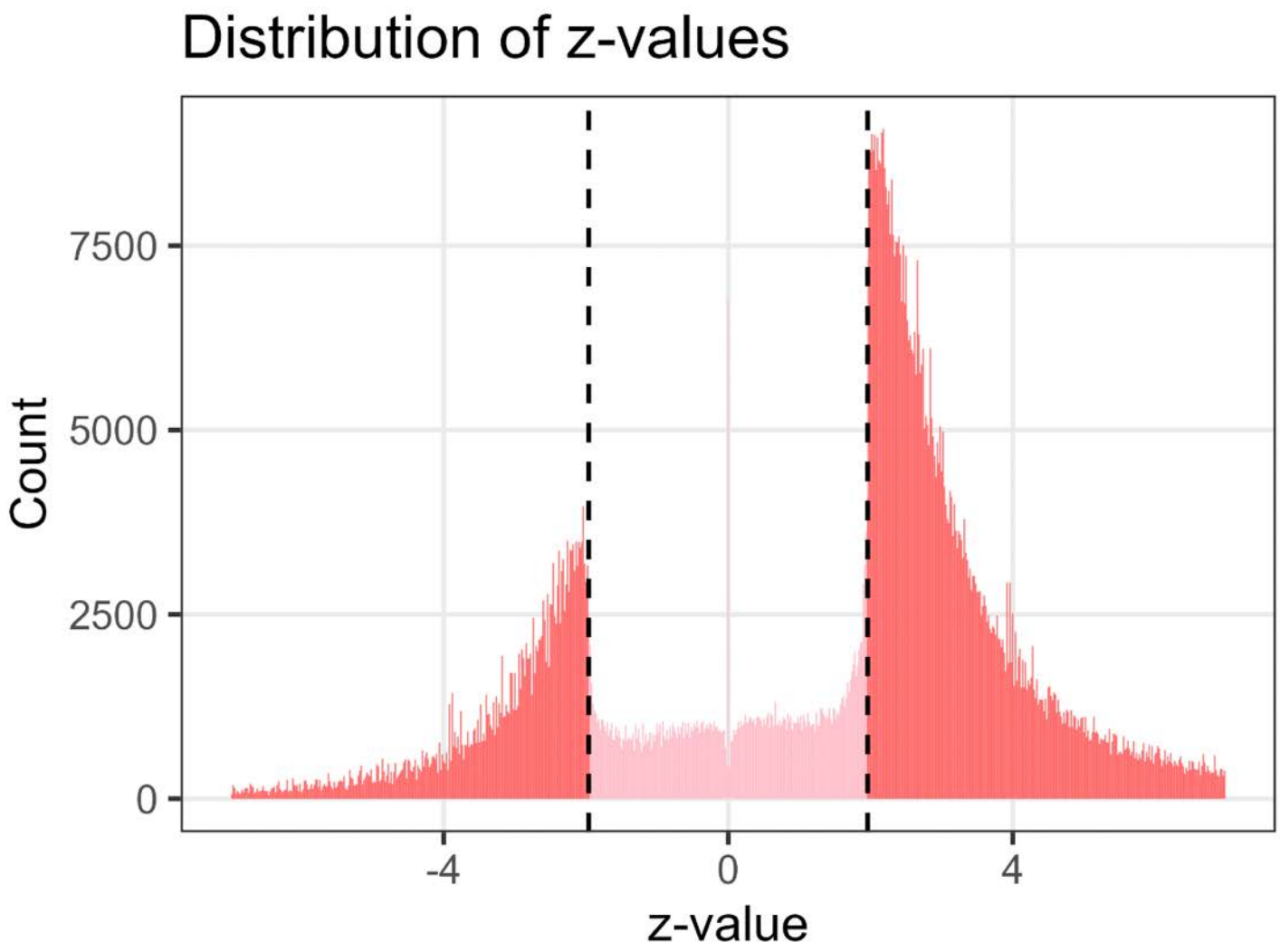
Conclusions: The problem of presenting multiple effect estimates derived from a single model in the same table is that it inadvertently encourages the reader to interpret all estimates the same way, often as total effects. Implications and recommendations are discussed.

KEYWORDS
directed acyclic graph, oral health research, table 2 fallacy

UMC Utrecht

WINNER’S CURSE

Distribution of 1.1m z-values in medical literature



UMC Utrecht

Source: https://agbarnett.github.io/talks/TRI/slides#13

Transplant International

REVIEW

Five myths about variable selection

Georg Heinze & Daniela Dunkler

Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Correspondence
Georg Heinze PhD, Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria.
Tel.: +4314040066880;
fax: +4314040066870;
e-mail: georg.heinze@meduniwien.ac.at

SUMMARY

Multivariable regression models are often used in transplantation research to identify or to confirm baseline variables which have an independent association, causally or only evidenced by statistical correlation, with transplantation outcome. Although sound theory is lacking, variable selection is a popular statistical method which seemingly reduces the complexity of such models. However, in fact, variable selection often complicates analysis as it invalidates common tools of statistical inference such as *P*-values and confidence intervals. This is a particular problem in transplantation research where sample sizes are often only small to moderate. Furthermore, variable selection requires computer-intensive stability investigations and a particularly cautious interpretation of results. We discuss how five common misconceptions often lead to inappropriate application of variable selection. We emphasize that variable selection and all problems related with it can often be avoided by the use of expert knowledge.

Transplant International 2017; 30: 6–10

Key words
association, explanatory models, multivariable modeling, prediction, statistical analysis

Received: 12 September 2016; Revision requested: 14 October 2016; Accepted: 25 November 2016

UMC Utrecht

Transplant International

REVIEW

Myth 1: The number of variables in a model should be reduced until there are 10 events per variables.

Myth 2: Only variables with proven univariable-model significance should be included in a model.

Myth 3: Insignificant effects should be eliminated from a model.

Myth 4: The reported P-value quantifies the type I error of a variable being falsely selected.

Myth 5: Variable selection simplifies analysis.

Transplant International 2017; 30: 6–10

Received: 12 September 2016; Revision requested: 14 October 2016; Accepted: 25 November 2016

UMC Utrecht

Transplant International

REVIEW

Five myths about variable selection

Georg Heinze & Daniela Dunkler

Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

Correspondence
Prof. Georg Heinze, Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria.
Tel.: +431 40400 6680
fax: +431404006870
e-mail: georg.heinze@meduniwien.ac.at

Multivariable regression models are often used in transplantation research to identify or to confirm baseline variables which have an independent association, causally or only evidenced by statistical correlation, with transplantation outcome. Although sound theory is lacking, variable selection is a popular statistical method which seemingly reduces the complexity of such models. However, this fact variable selection often complicates analyses as it invalidates common tools of statistical inference such as P-values and confidence intervals. This is a particular problem in transplantation research where sample sizes are often only small to moderate. Furthermore, variable selection requires computer-intensive stability investigations and a particularly cautious interpretation of results. We discuss how five common misconceptions often lead to inappropriate application of variable selection. We emphasize that variable selection and all problems related with it can often be avoided by the use of expert knowledge.

Transplant International 2017; 30: 6–10

Key words
association, explanatory models, multivariable modeling, prediction, statistical analysis

Received: 12 September 2016; Revision requested: 14 October 2016; Accepted: 25 November 2016

Myth 1: The number of variables in a model should be reduced until there are 10 events per variables.

Myth 2: Only variables with proven univariable-model significance should be included in a model.

Myth 3: Insignificant effects should be eliminated from a model.

Myth 4: The reported P-value quantifies the type I error of a variable being falsely selected.

Myth 5: Variable selection simplifies analysis.

UMC Utrecht

Variable selection can be very instable in low *N*

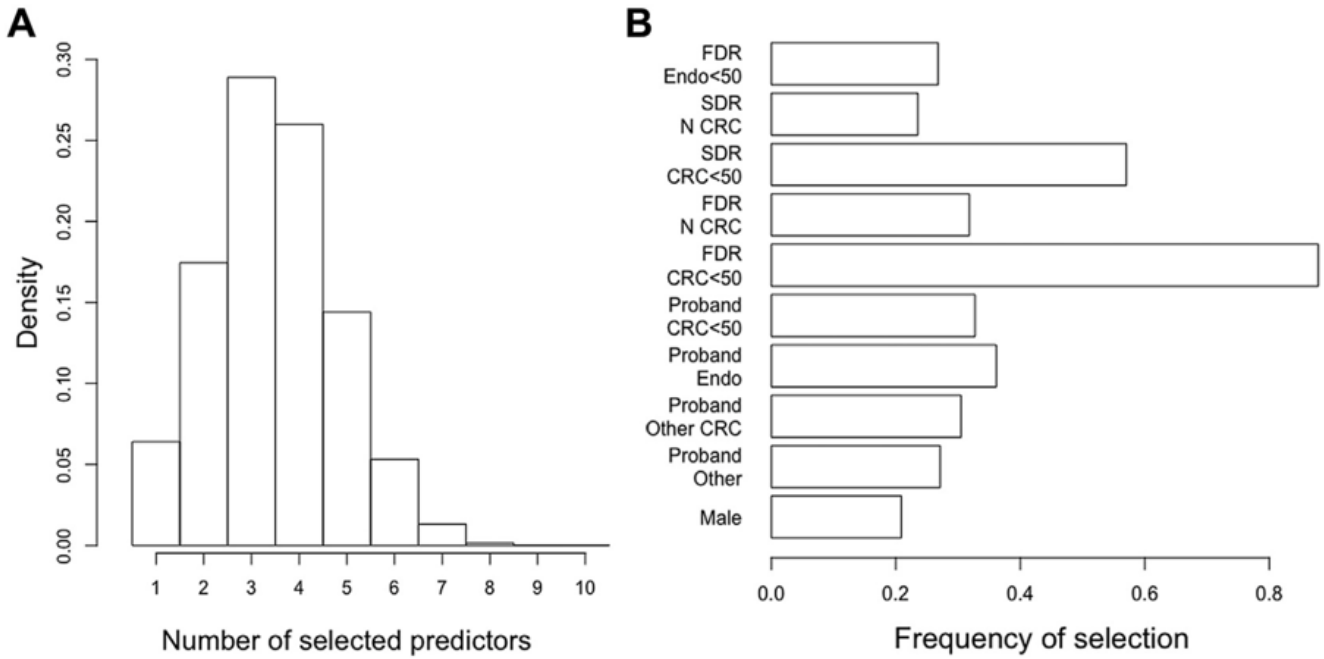
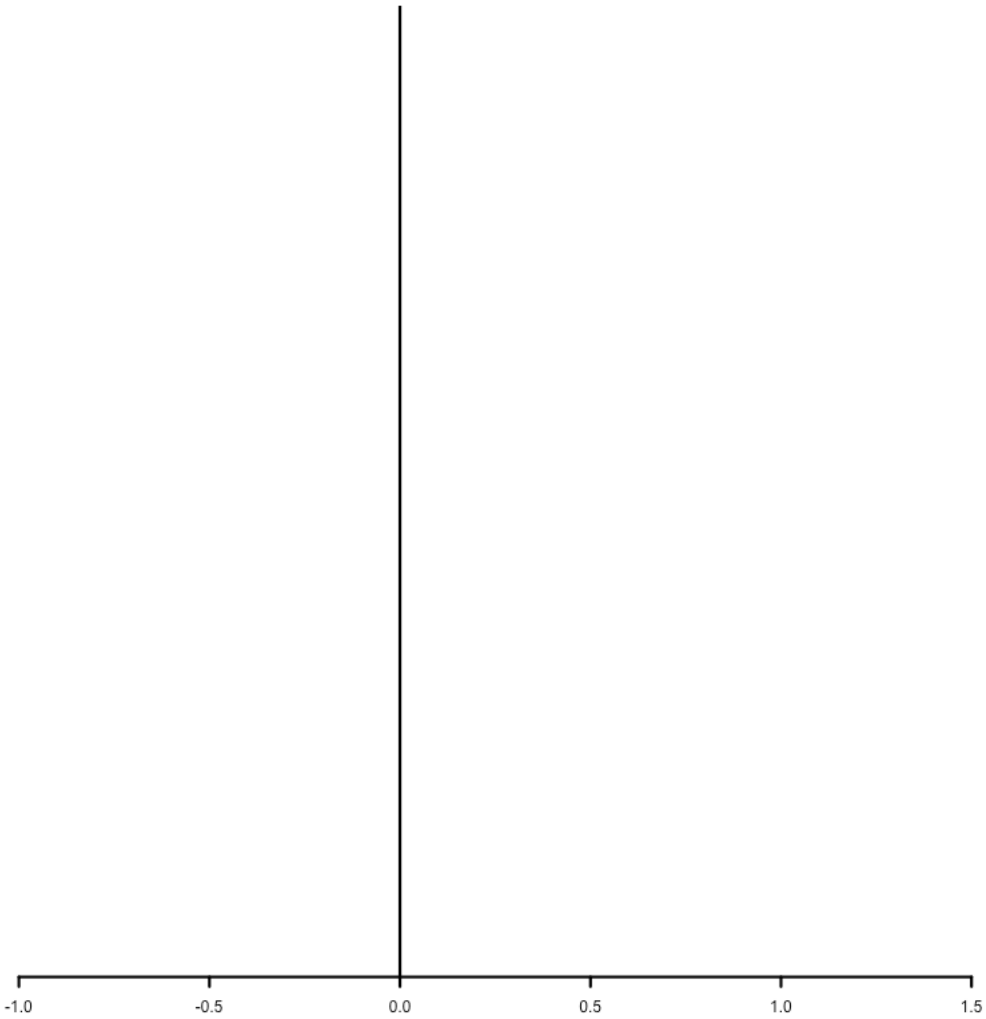


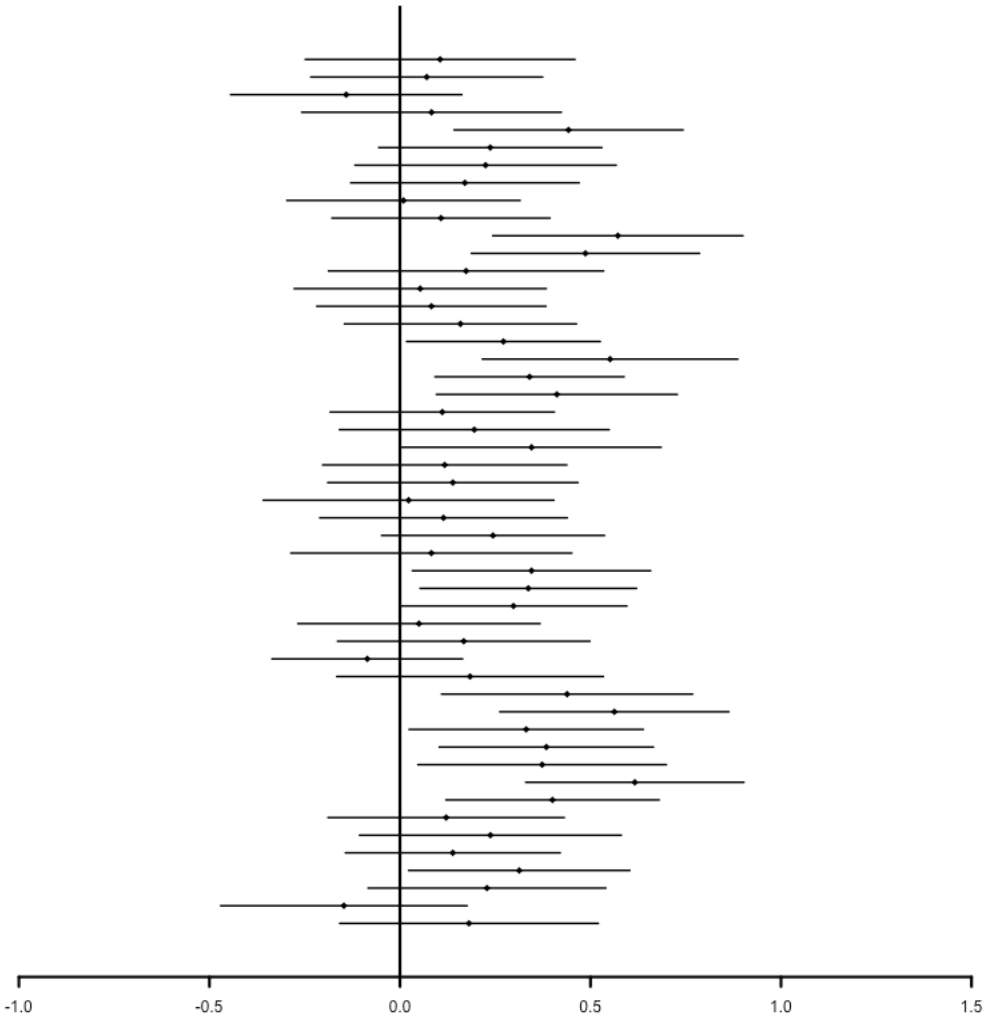
Fig. 2. Number of predictors (panel A) and top 10 predictors (panel B) selected in models among 5,000 samples of 870 probands with 38 mutation carriers. FDR, first degree relatives; SDR, second degree relatives; CRC, colorectal cancer; Endo, Endometrial cancer.

UMC Utrecht

Illustrative example

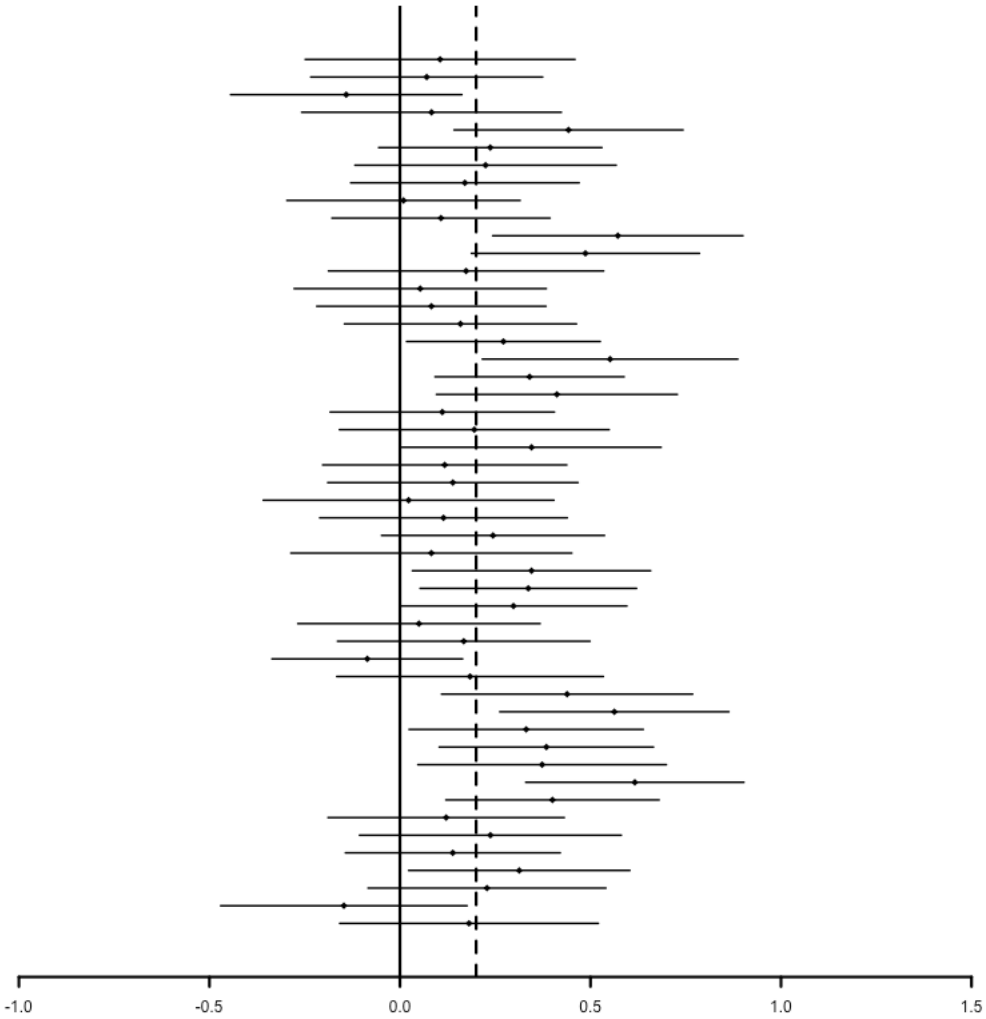


Illustrative example



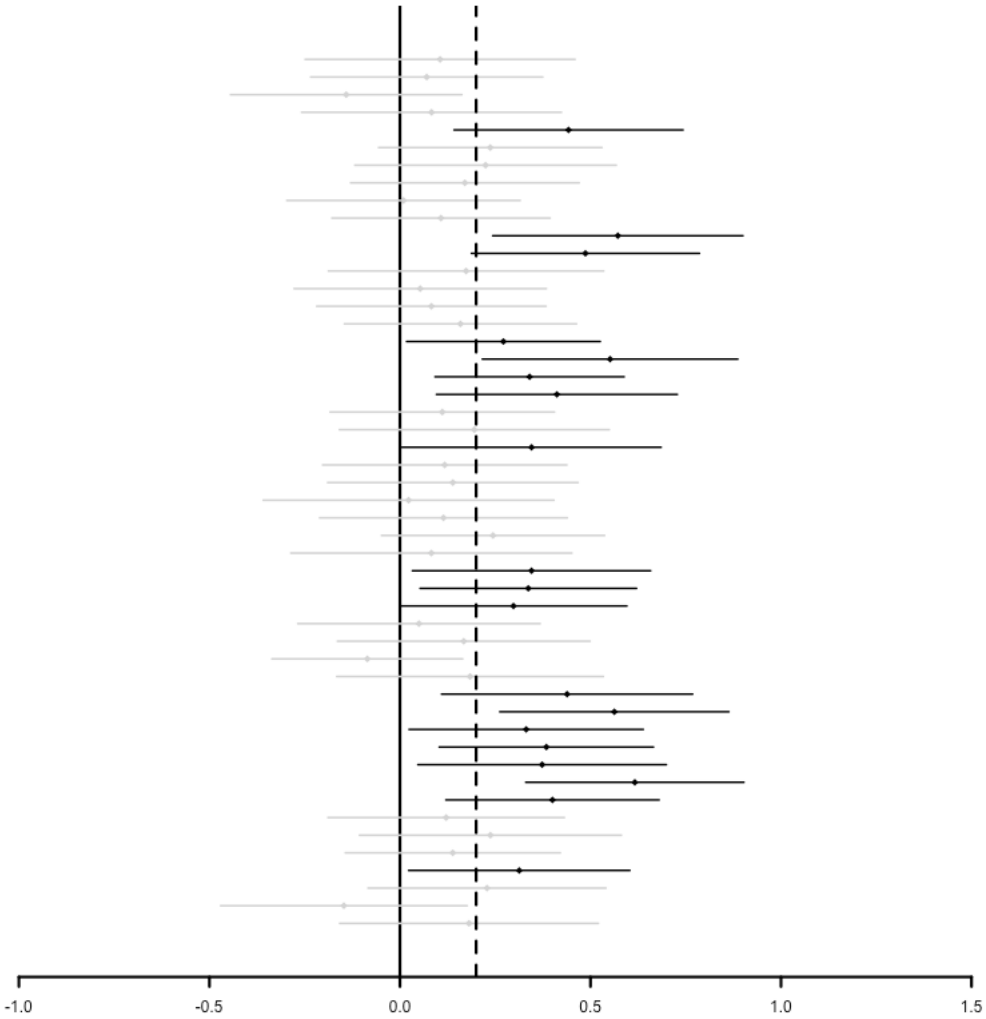
UMC Utrecht

Illustrative example



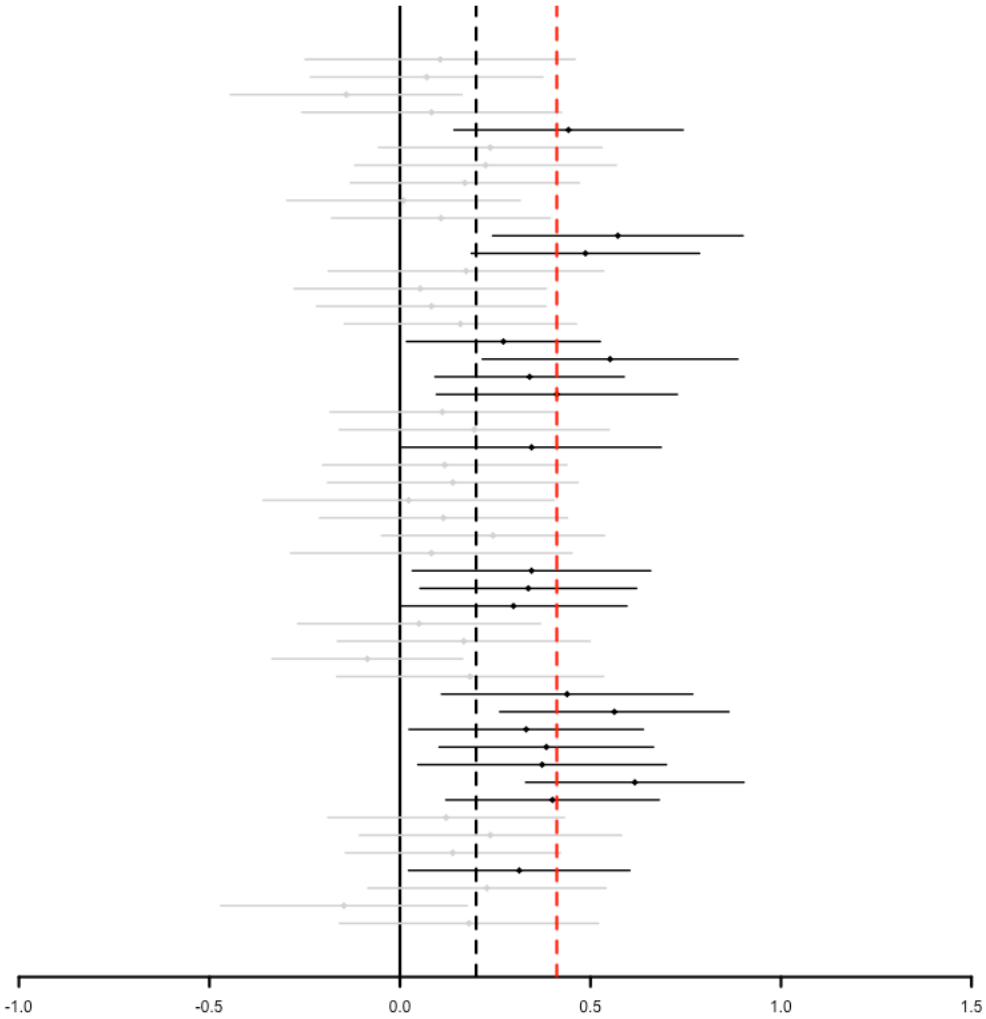
UMC Utrecht

Illustrative example



UMC Utrecht

Winner’s curse



UMC Utrecht

Variable selection often makes things worse

Received: 5 August 2021 | Revised: 10 December 2021 | Accepted: 5 February 2022

DOI: 10.1002/bimj.202100237

Biometrical Journal

RESEARCH ARTICLE

A comparison of full model specification and backward elimination of potential confounders when estimating marginal and conditional causal effects on binary outcomes from observational data

Kim Luijken[1] | Rolf H.H. Groenwold[1,2] | Maarten van Smeden[1,3] | Susanne Strohmaier[4,5] | Georg Heinze[4]

[1]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

[2]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

[3]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, University of Utrecht, Utrecht, The Netherlands

[4]Section for Clinical Biometrics, Center for Medical Statistics, Informatics and Intelligent Systems, Medical University of Vienna, Vienna, Austria

[5]Department of Epidemiology, Center for Public Health, Medical University of Vienna, Vienna, Austria

Investigated **3960 scenario's**, backward elimination made exposure effect estimation **worse 97% of the time**
(in remaining 3% improvements were neglegible)

UMC Utrecht

Variable selection…

- Is often **instable** (e.g. small N or high collinearity)
- Can create **testimation bias**
- Can **invalidate inferential statistics**: default p-values and confidence intervals not valid (post-selection inference literature)
- Can be a source of model **overfitting**

Received: 18 April 2017 | Revised: 13 November 2017 | Accepted: 17 November 2017

DOI: 10.1002/bimj.201700067

REVIEW ARTICLE

Biometrical Journal

Variable selection – A review and recommendations for the practicing statistician

Georg Heinze | Christine Wallisch | Daniela Dunkler

UMC Utrecht

STEIN’S PARADOX

1955: Stein's paradox

INADMISSIBILITY OF THE USUAL ESTIMATOR FOR THE MEAN OF A MULTIVARIATE NORMAL DISTRIBUTION

CHARLES STEIN
STANFORD UNIVERSITY

PROCEEDINGS *of the* THIRD BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

UMC Utrecht

Stein's paradox in words (rather simplified)

When one has **three or more units** (say, individuals), and for each unit one can calculate an **average score** (say, average blood pressure), then **the best guess of future** observations for each unit (say, blood pressure tomorrow) is **NOT the average score**.

UMC Utrecht

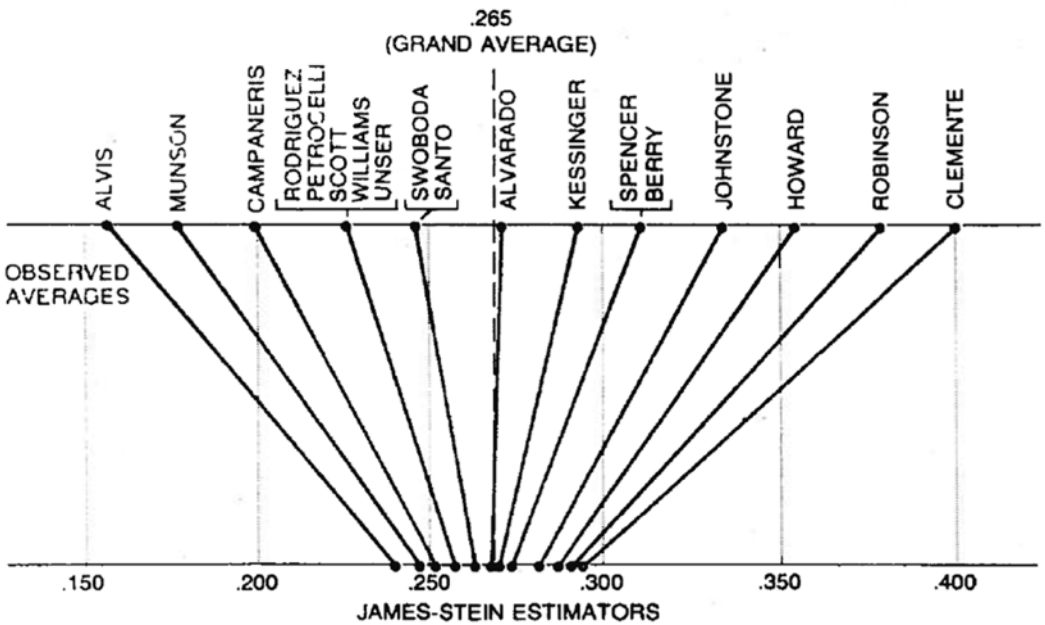
1961: James-Stein estimator: the next Berkley Symposium

ESTIMATION WITH QUADRATIC LOSS

W. JAMES
FRESNO STATE COLLEGE
AND
CHARLES STEIN
STANFORD UNIVERSITY

- James and Stein. Estimation with quadratic loss. Proceedings of the fourth Berkeley symposium on mathematical statistics and probability. Vol. 1. 1961.

1977: Baseball example



JAMES-STEIN ESTIMATORS for the 18 baseball players were calculated by "shrinking" the individual batting averages toward the overall "average of the averages." In this case the grand average is .265 and each of the averages is shrunk about 80 percent of the distance to this value. Thus the theorem on which Stein's method is based asserts that the true batting abilities are more tightly clustered than the preliminary batting averages would seem to suggest they are.

Squared error reduced from .077 to .022

Stein's paradox

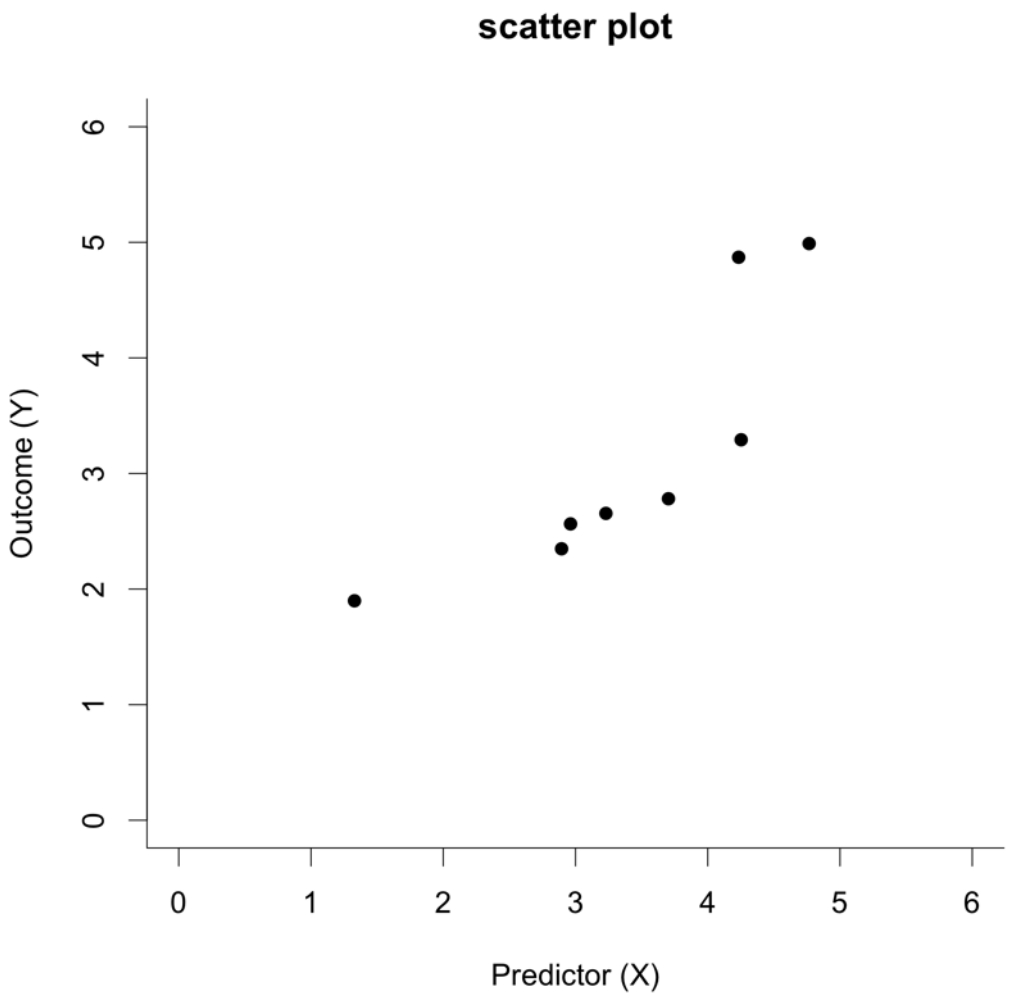
- Probably among the most **surprising** (and initially doubted) phenomena in statistics

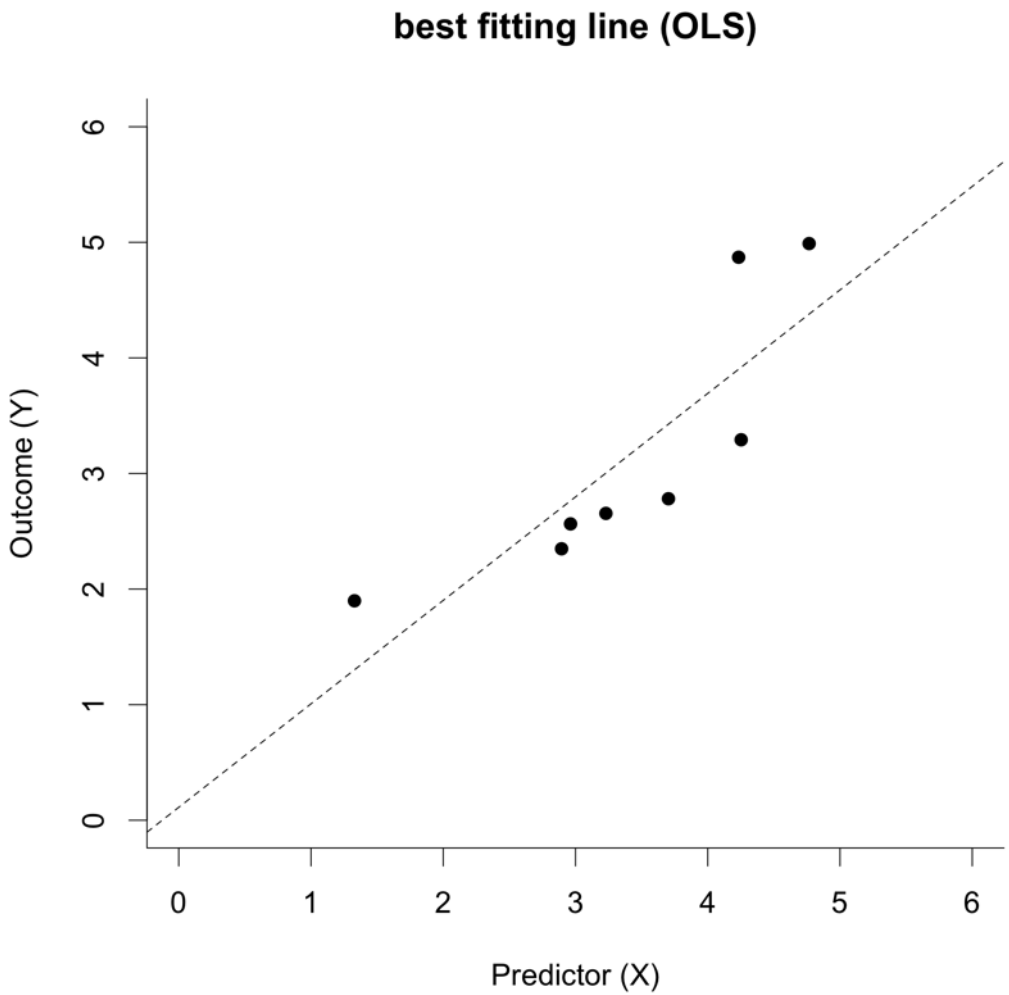
- Now a large "family": **shrinkage estimators** reduce prediction variance to an extent that typically outweighs the bias that is introduced

- **Bias/variance trade-off** principle has motivated many statistical and machine learning developments

Expected prediction error = irreducible error + bias^2 + variance

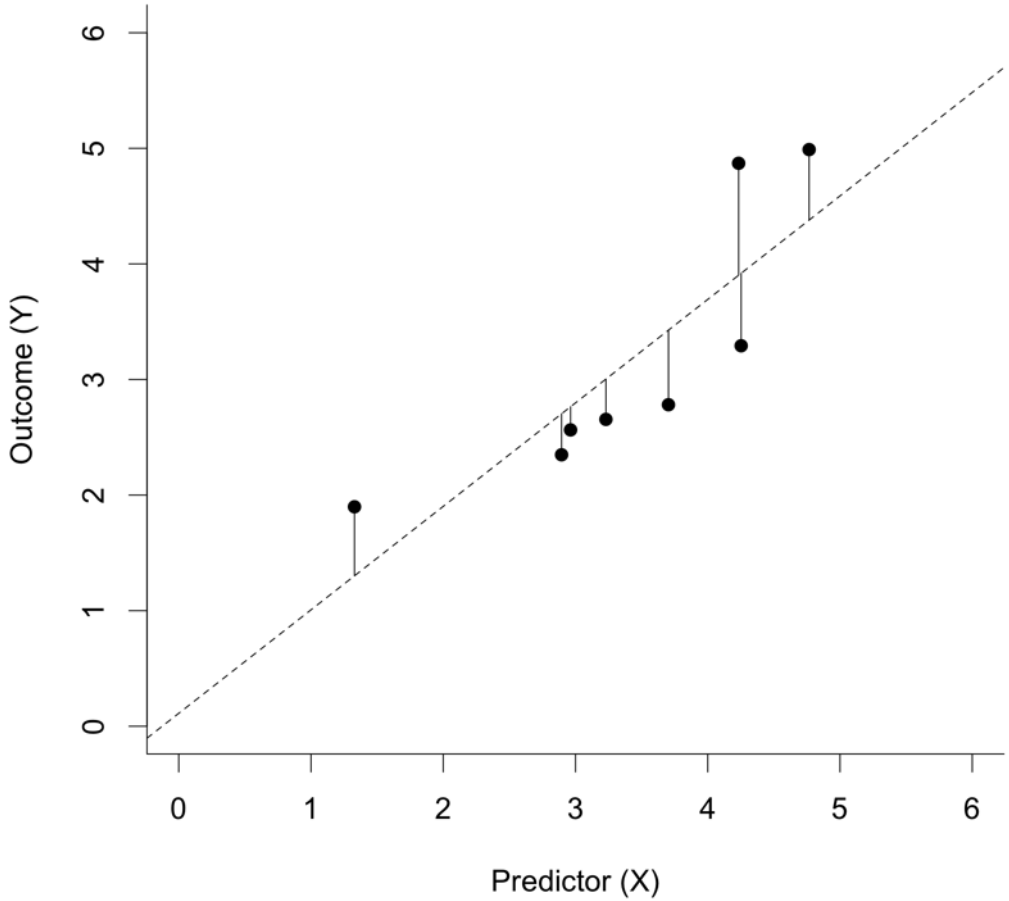
UMC Utrecht

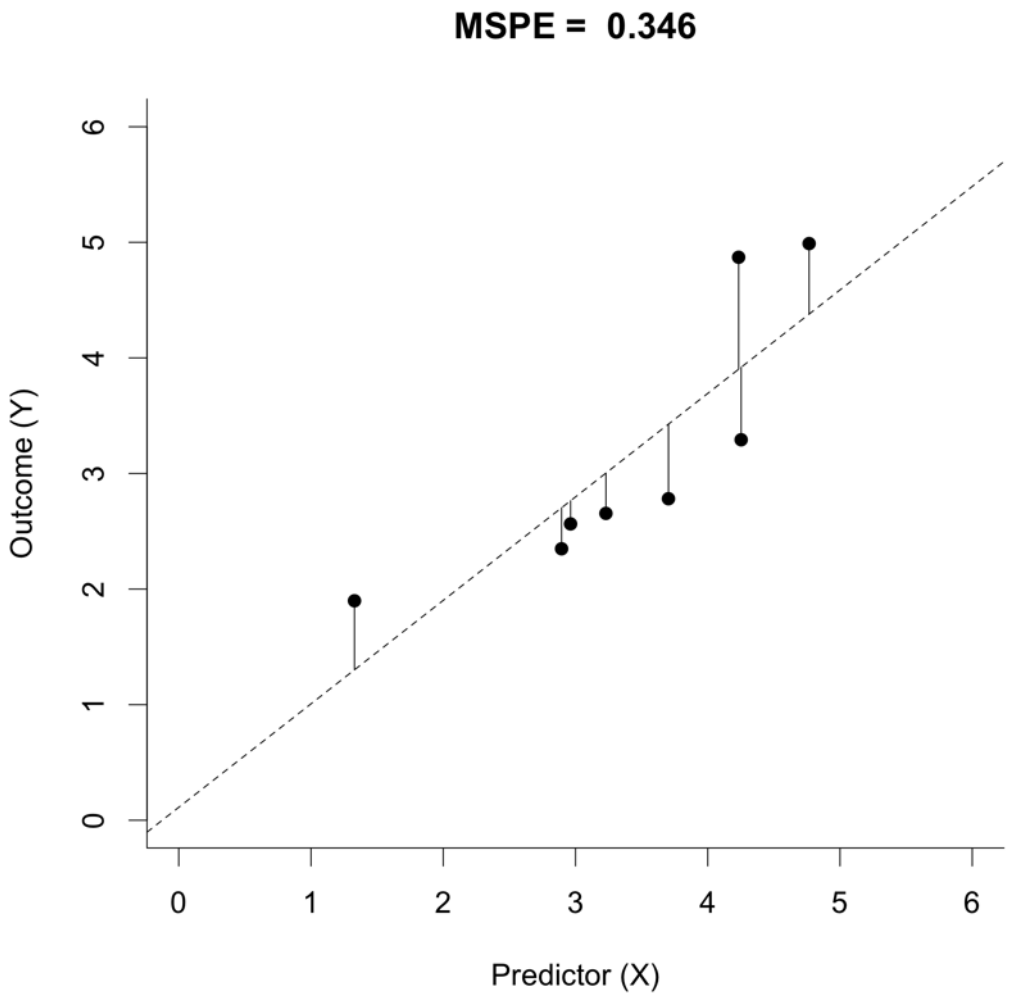

scatter plot
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6

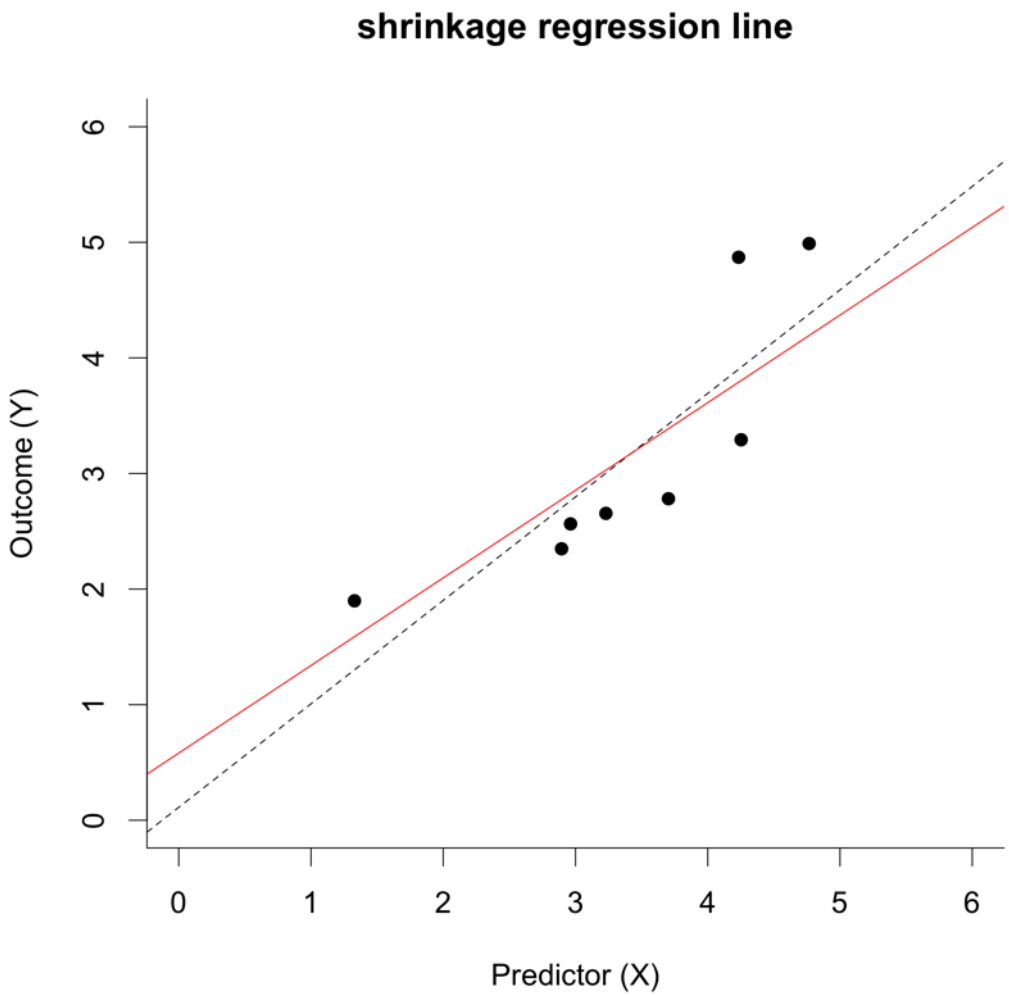


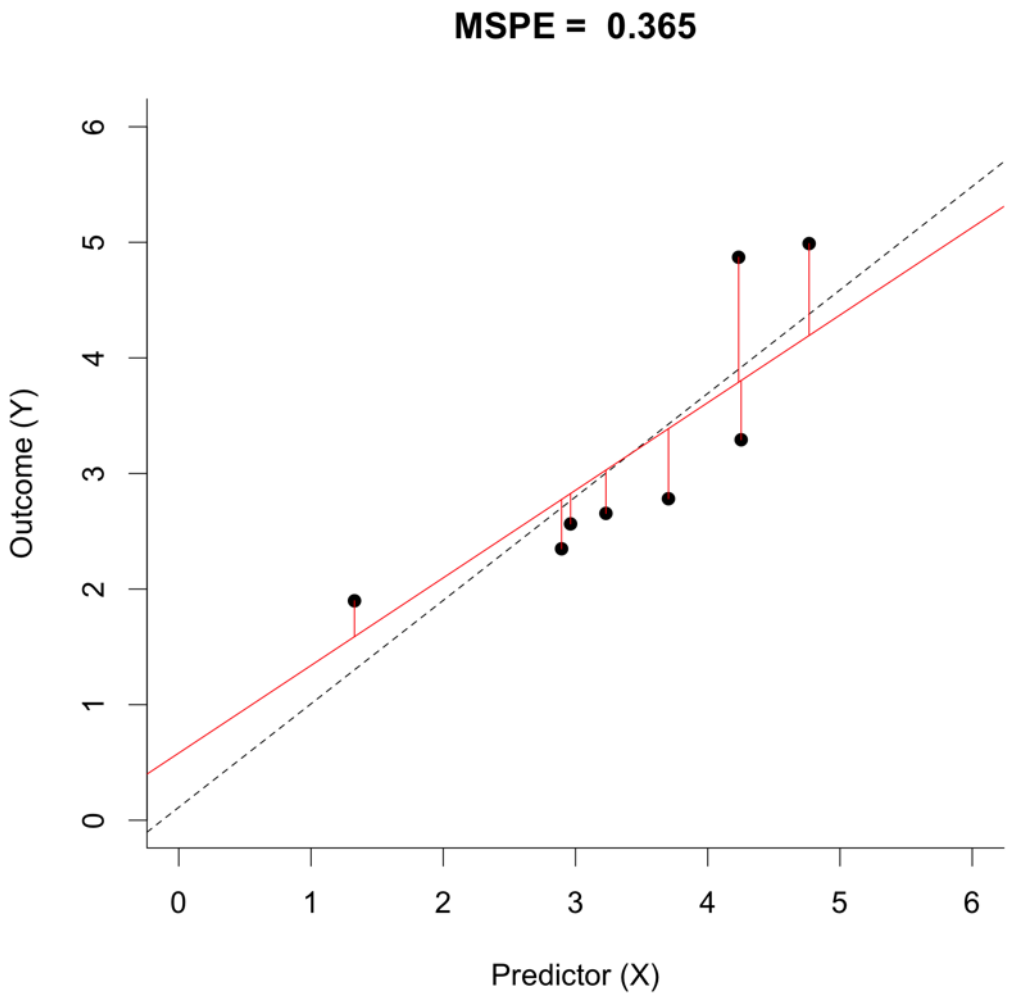
UMC Utrecht

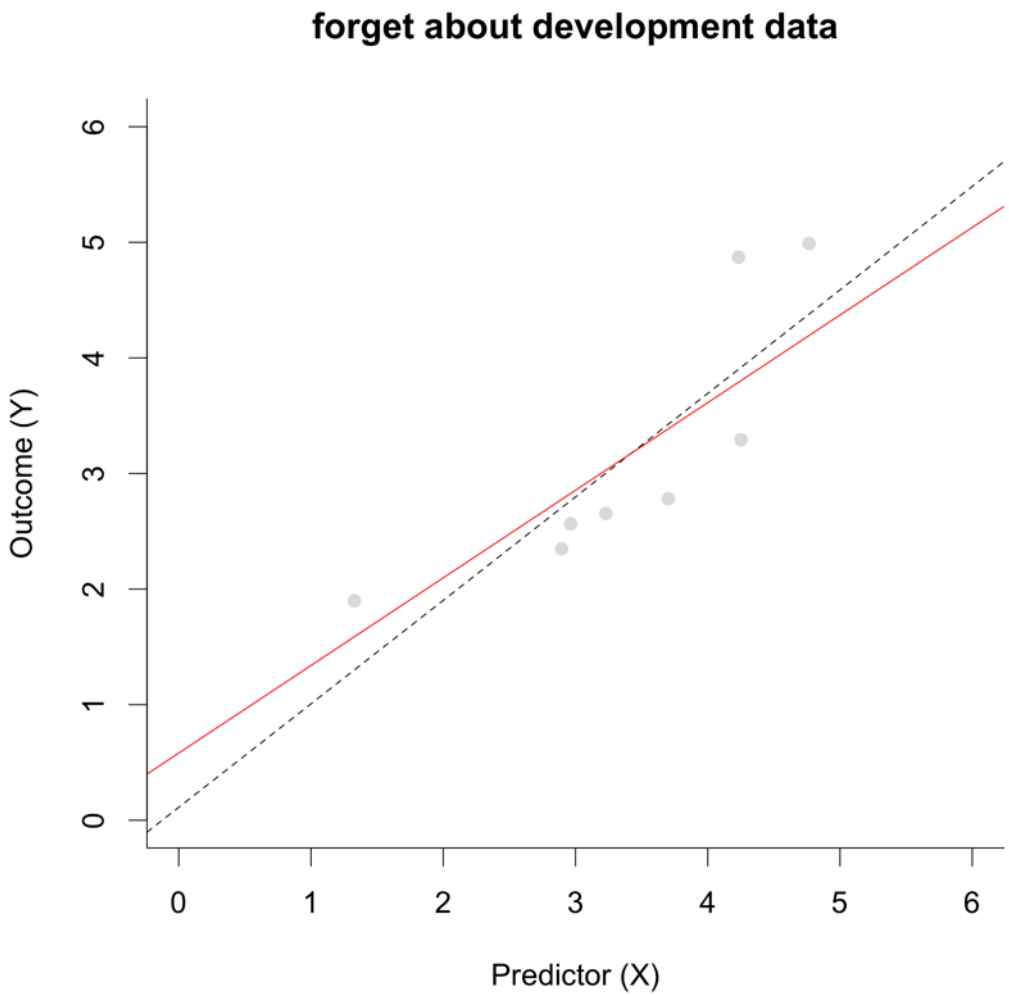
minimizing squared error (MSPE)

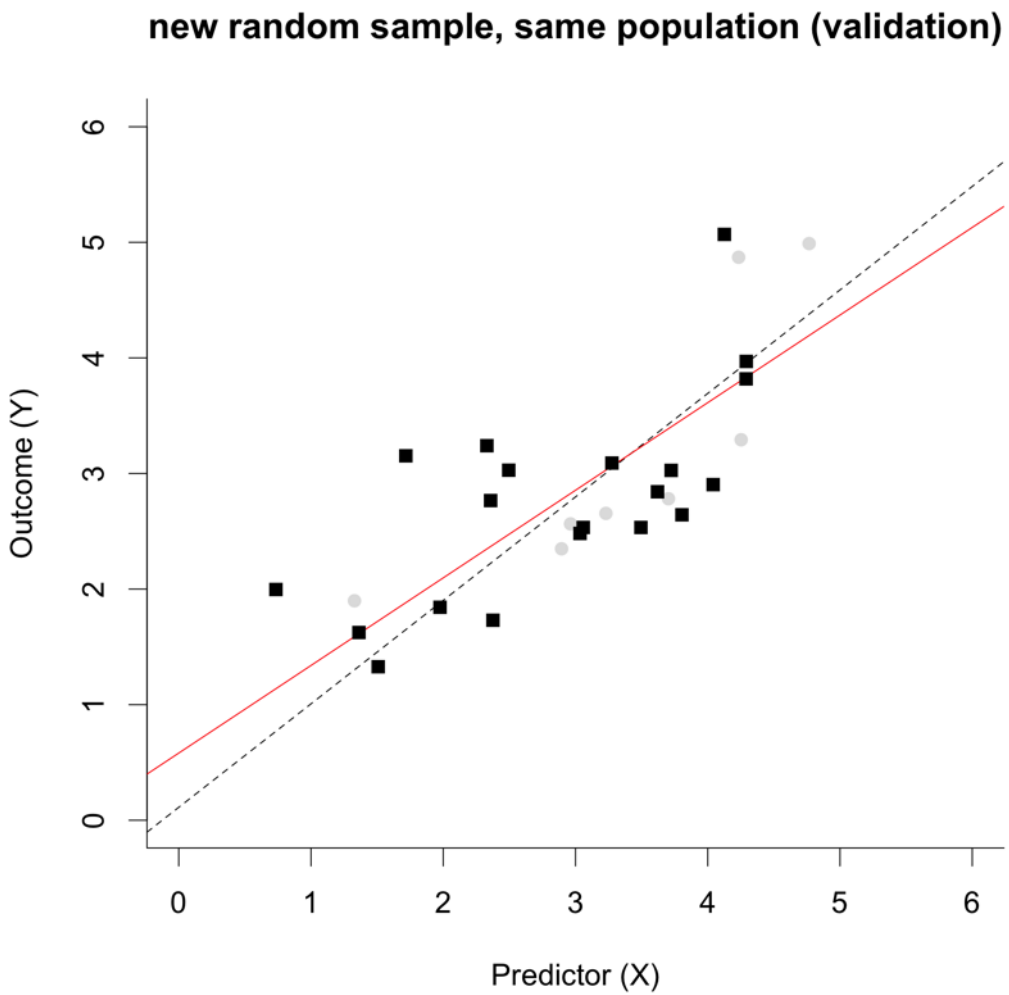


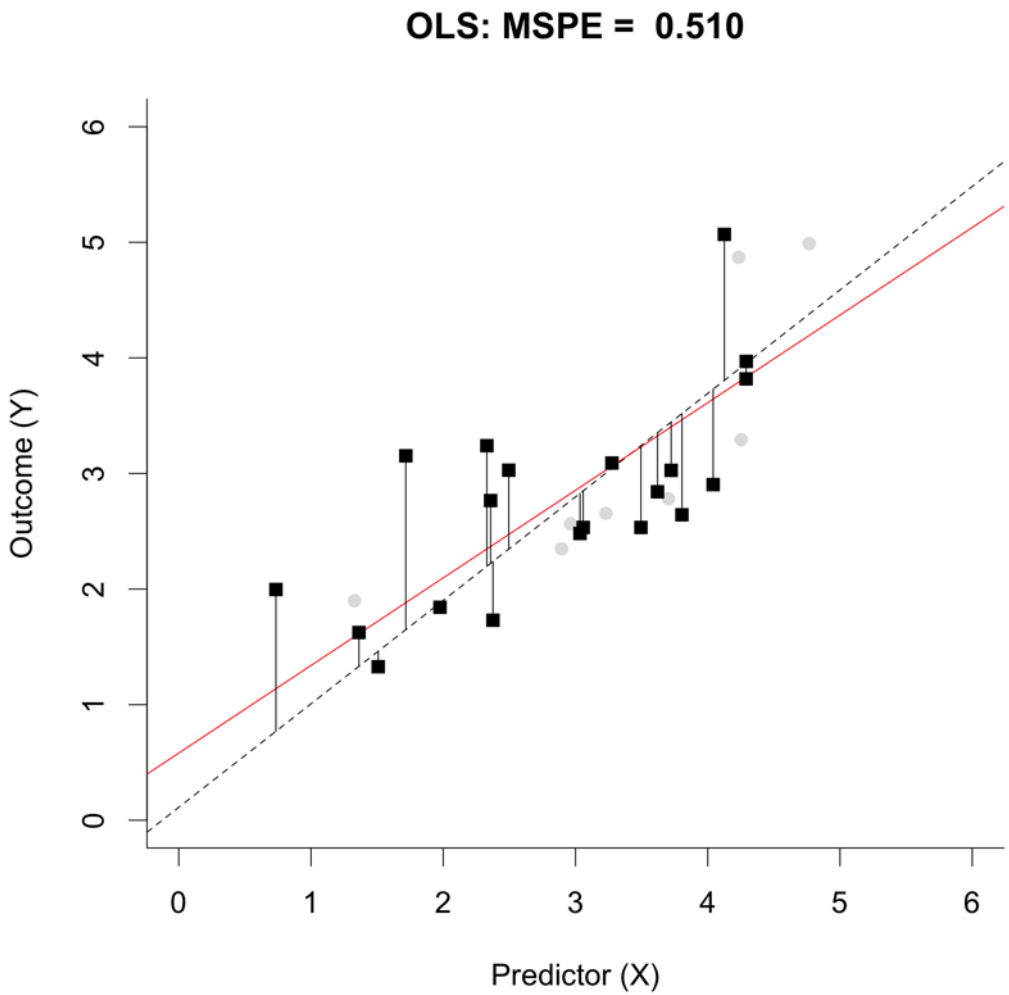

MSPE = 0.346
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6


shrinkage regression line
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6

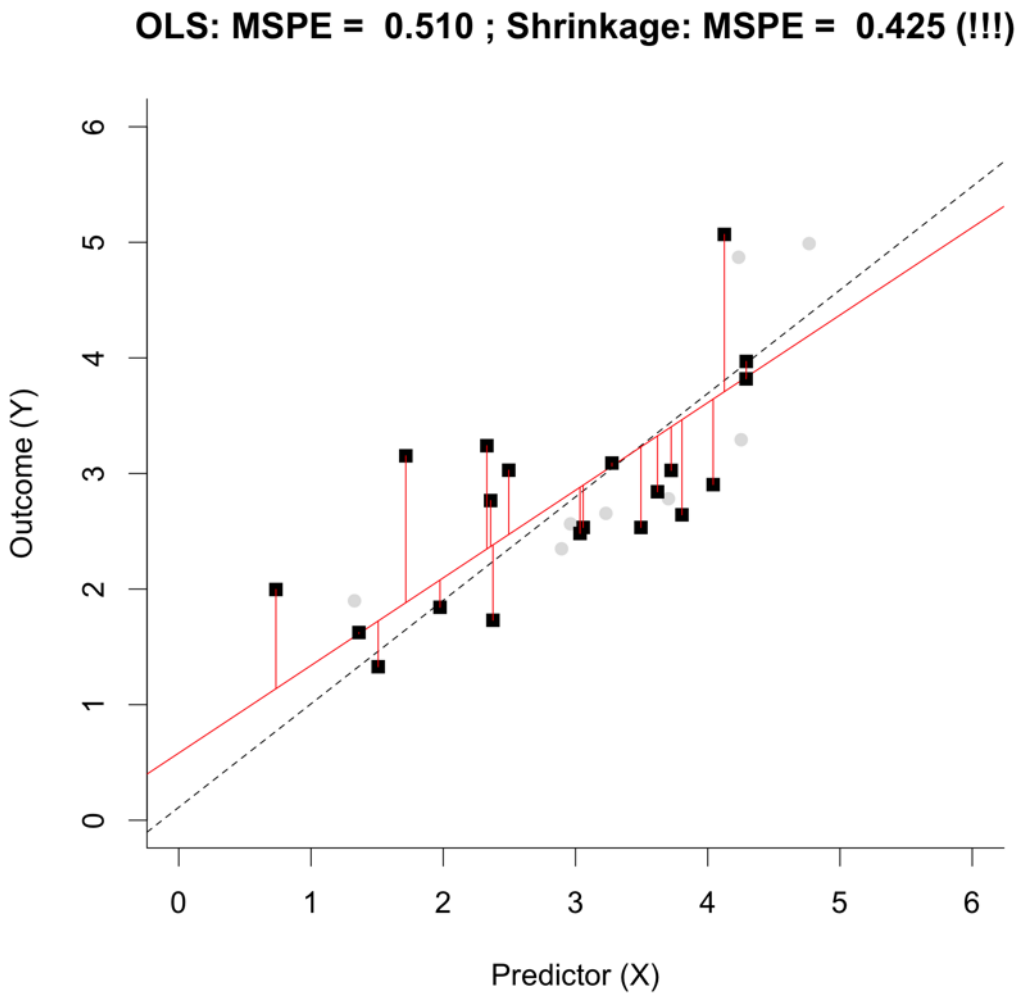

MSPE = 0.365
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6


forget about development data
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6


new random sample, same population (validation)
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6

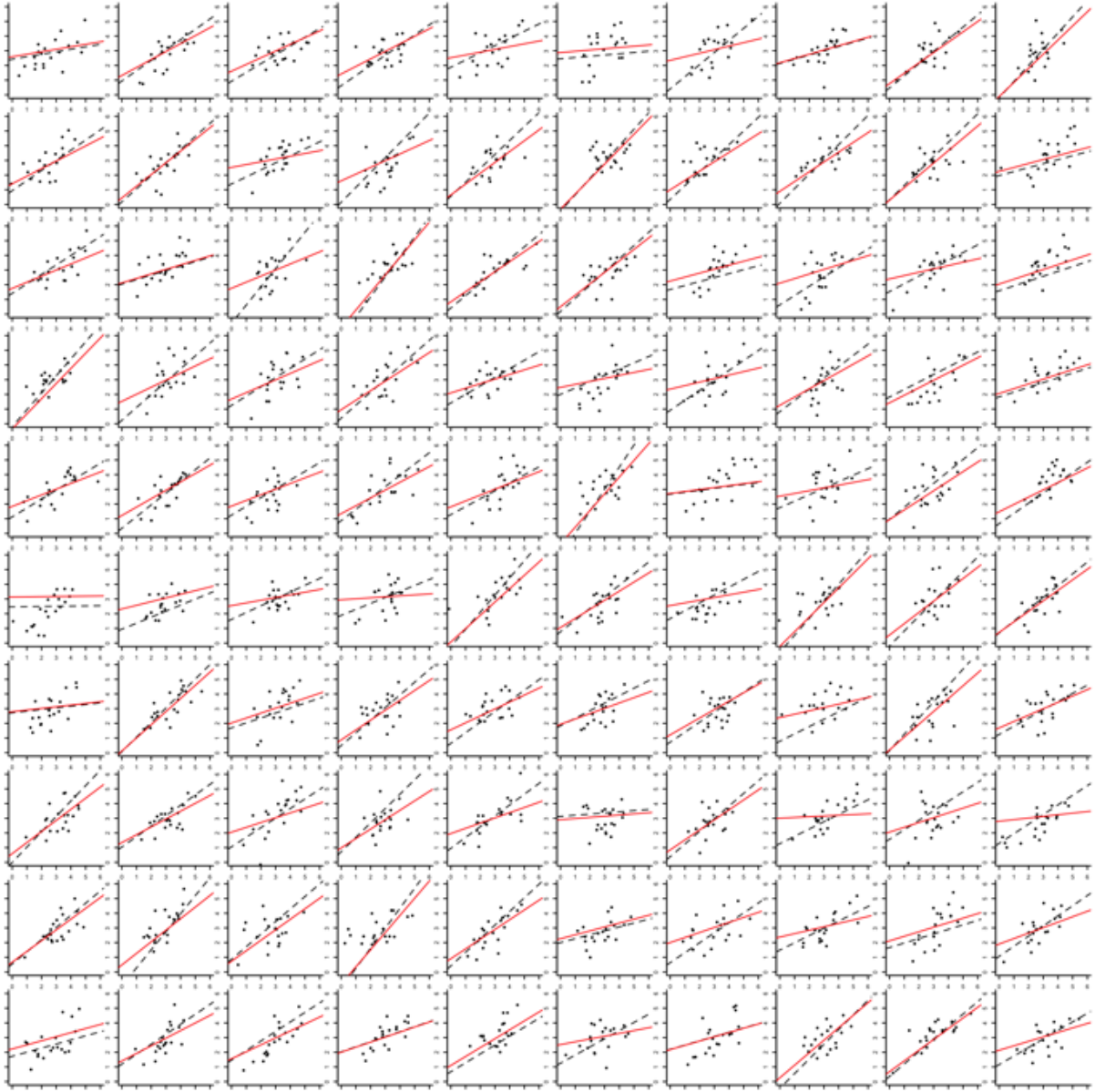


UMC Utrecht


OLS: MSPE = 0.510 ; Shrinkage: MSPE = 0.425 (!!!)
Outcome (Y)
Predictor (X)
0
1
2
3
4
5
6

UMC Utrecht

Simulation: 100 times



UMC Utrecht

Not just lucky

- 5% reduction in MSPE just by shrinkage estimator
- Van Houwelingen and le Cessie's heuristic shrinkage factor

STATISTICS IN MEDICINE, VOL. 9, 1303–1325 (1990)

PREDICTIVE VALUE OF STATISTICAL MODELS

J. C. VAN HOUWELINGEN AND S. LE CESSIE

Department of Medical Statistics, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands

SUMMARY

A review is given of different ways of estimating the error rate of a prediction rule based on a statistical model. A distinction is drawn between apparent, optimum and actual error rates. Moreover it is shown how cross-validation can be used to obtain an adjusted predictor with smaller error rate. A detailed discussion is given for ordinary least squares, logistic regression and Cox regression in survival analysis. Finally, the split-sample approach is discussed and demonstrated on two data sets.

Shrinkage

Post-estimation shrinkage factor estimation

- Van Houwelingen & Le Cessie, 1990: uniform shrinkage factor
- Sauerbrei 1999: parameterwise shrinkage factors

Regularized regression (shrinkage during estimation)

- Ridge regression: L2-penalty on regression coefficient
- Lasso: L1 penalty
- Elastic net: L1 and L2 penalty

UMC Utrecht

Shrinkage

Post-estimation shrinkage factor estimation

$$\Pr(Y = 1) = \text{expit}[\beta_0^* + S_{VH}(\beta_1 X_1 + \ldots + \beta_P X_P)]$$

Regularized regression (shrinkage during estimation)

$$\text{Ln}\mathcal{L}_p = \text{Ln}\mathcal{L}_{ml} - \lambda \left[(1-\alpha) \sum_{p=1}^{P} \beta_p^2 + a \sum_{p=1}^{P} |\beta_p| \right]$$

Ridge regression: a = 0, Lasso: a = 1, Elastic net $0 < a < 1$

Consequences of shrinkage

- Can **improve** the accuracy of **predictions** on average[1]
- Can **reduce** (part of) the detrimental effects of **overfitting**
- In specific situations (e.g. Lasso) it can be used for **automated variable selection** at **reduced risk of winner's curse**

- No free lunch principle: shrinkage often introduces (by design) a **negative bias** in regression coefficients

- Exception: Firth's correction, e.g. see:

van Smeden *et al. BMC Medical Research Methodology* (2016) 16:163
DOI 10.1186/s12874-016-0267-3

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

No rationale for 1 variable per 10 events criterion for binary logistic regression analysis

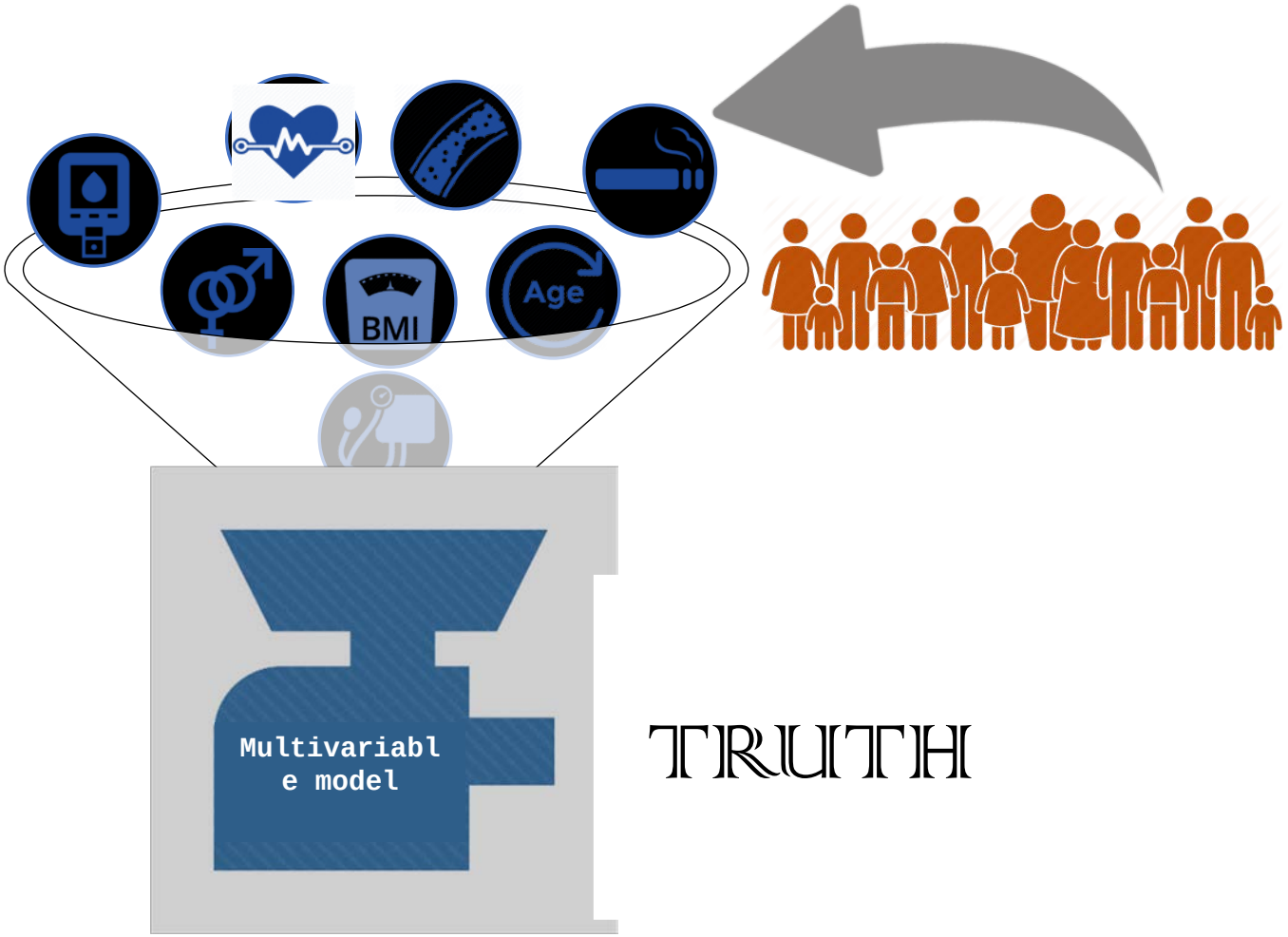
Maarten van Smeden[1*], Joris A. H. de Groot[1], Karel G. M. Moons[1], Gary S. Collins[2], Douglas G. Altman[2], Marinus J. C. Eijkemans[1] and Johannes B. Reitsma[1]

UMC Utrecht

[1]On average is important here, see:

PART II:
GOOD STATISTICAL PRACTICE
AVOIDING FALACIES/PARADOXES

Utopia



Utopia

"SOMETHING USEFUL"

Statistical Science
2010, Vol. 25, No. 3, 289–310
DOI: 10.1214/10-STS330
© Institute of Mathematical Statistics, 2010

To Explain or to Predict?

Galit Shmueli

Abstract. Statistical modeling is a powerful tool for developing and testing theories by way of causal explanation, prediction, and description. In many disciplines there is near-exclusive use of statistical modeling for causal explanation and the assumption that models with high explanatory power are inherently of high predictive power. Conflation between explanation and prediction is common, yet the distinction must be understood for progressing scientific knowledge. While this distinction has been recognized in the philosophy of science, the statistical literature lacks a thorough discussion of the many differences that arise in the process of modeling for an explanatory versus a predictive goal. The purpose of this article is to clarify the distinction between explanatory and predictive modeling, to discuss its sources, and to reveal the practical implications of the distinction to each step in the modeling process.

Key words and phrases: Explanatory modeling, causality, predictive modeling, predictive power, statistical strategy, data mining, scientific research.

UMC Utrecht

To explain or to predict?

Explanatory models

- Theory: interest in regression coefficients
- Testing and comparing existing causal theories
  - e.g. aetiology of illness, effect of treatment

Prediction models

- Interest in (risk) predictions of future observations
- Causality not a primary concern
- Concerns about overfitting and optimism
  - e.g. prognostic or diagnostic prediction model

Descriptive models

- Capture the data structure

UMC Utrecht

To explain or to predict?

Explanatory models

- Theory: interest in regression coefficients
- Testing and comparing existing causal theories
  - e.g. aetiology of illness, effect of treatment

Prediction models

- Interest in (risk) predictions of future observations
- Causality not a primary concern
- Concerns about overfitting and optimism
  - e.g. prognostic or diagnostic prediction model

Descriptive models

- Capture the data structure

UMC Utrecht

To explain or to predict?

Explanatory models

- *Absence of absence fallacy*: e.g**. non-significant effect of exposure** interpreted as ”not working” (tx) or “not bad for health”

- *Table 2 fallacy:* e.g. regression **coefficients of confounding variables** interpreted as themselves “adjusted” for confounding

- *Winner’s curse:* e.g. selected factors on average **too extreme values** for the regression coefficients (i.e. **biased**)

- *Stein’s paradox:* shrinkage may **lead to a bias** that may not be beneficial for inference
  (but not always, see[1])

UMC Utrecht

[1]https://bit.ly/3VZQFJV

To explain or to predict?

Prediction models

- *Absence of absence fallacy*: e.g. non-significant result on **measure for miscalibration** misinterpreted as good calibration

- *Table 2 fallacy:* e.g. predictors misinterpreted as **causal effects**

- *Winner's curse:* e.g. final model with selected predictors results in **overfitting**

- *Stein's paradox:* shrinkage **may improve predictions** (but not always, see[1])

UMC Utrecht

[1]https://bit.ly/3spjPo5

Explanatory vs prediction models

International Journal of Epidemiology, 2020, 338–347
doi: 10.1093/ije/dyz251
Advance Access Publication Date: 10 December 2019
Education Corner

Education Corner

Reflection on modern methods: five myths about measurement error in epidemiological research

Maarten van Smeden,[1*] Timothy L La

[1]Department of Clinical Epidemiology, Leiden Univ
[2]Department of Epidemiology, Rollins School of Publ
[3]Department of Biomedical Data Sciences, Leiden U

*Corresponding author. Albinusdreef 2, Leiden 2333 ZA, The N

Editorial decision 29 October 2019; Accepted 16 November 2019

Abstract

Epidemiologists are often confronted with d
ment error due to, for instance, mistaken
measurement instrument or procedural error
judged, the data analyses are hampered and t
affected. In this paper, we describe five myt
measurement error, regarding expected stru
problems resulting from mismeasurements.
error misconceptions. We show that the influ
logical data analysis can play out in ways t
heuristics about whether or not to expect at
we encourage epidemiologists to deliberate
measurement error in their analyses, we a
making claims about the magnitude or even
not accompanied by statistical measurement
sis. Suggestions for alleviating the problems
tude of measurement error are given.

Key words: Measurement error, misclassification, bias, bia

Leeuwenberg *et al. Diagnostic and Prognostic Res*
https://doi.org/10.1186/s41512-021-00115-5

RESEARCH

Performance of b high-correlation l comparison of m

Artuur M. Leeuwenberg[1*], Maarten va
Murielle E. Mauer[3], Karel G. M. Moons[1],

Abstract

Background: Clinical prediction model
models are highly collinear, unexpectec
reducing face-validity of the prediction
but when there is no a priori motivatio
approach is arbitrary and possibly inapp

Methods: We compare different metho
constrained optimization. The effectiver

Results: In the conducted simulations,
Intercept, Slope) across methods. Howe
was found, affecting all compared meth
Lasso). Methods for which the included set of predictors remained most stable under increased collinearity were Ridge, PCLR, LAELR, and Dropout.

Conclusions: Based on the results, we would recommend refraining from data-driven predictor selection approaches in the presence of high collinearity, because of the increased instability of predictor selection, even in relatively high events-per-variable settings. The selection of certain predictors over others may disproportionally give the impression that included predictors have a stronger association with the outcome than excluded predictors.

Keywords: Multi-collinearity, Prediction models, Normal-tissue complication probability models

BMJ 2020;368:m441 doi: 10.1136/bmj.m441 (Published 18 March 2020) Page 1 of 12

RESEARCH METHODS & REPORTING

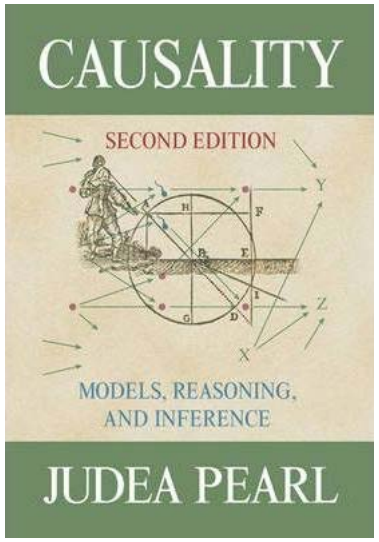
Calculating the sample size required for developing a clinical prediction model

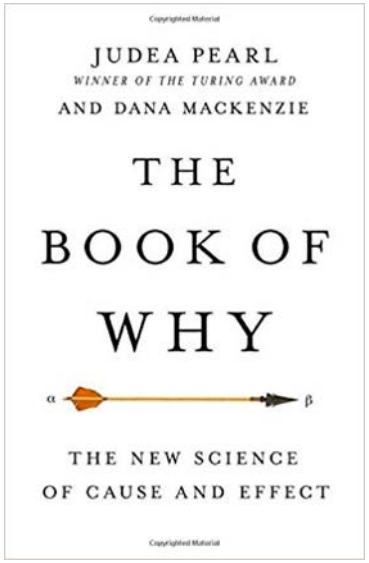
Clinical prediction models aim to predict outcomes in individuals, to inform diagnosis or prognosis in healthcare. Hundreds of prediction models are published in the medical literature each year, yet many are developed using a dataset that is too small for the total number of participants or outcome events. This leads to inaccurate predictions and consequently incorrect healthcare decisions for some individuals. In this article, the authors provide guidance on how to calculate the sample size required to develop a clinical prediction model.

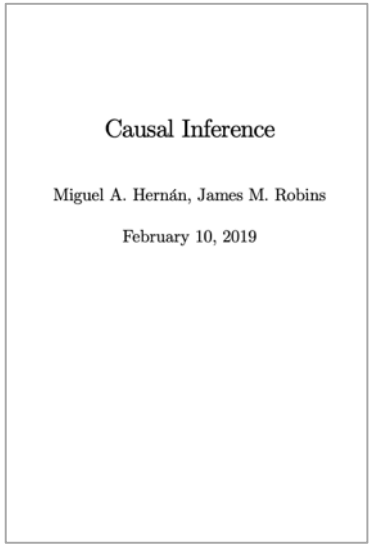
Richard D Riley *professor of biostatistics*[1], Joie Ensor *lecturer in biostatistics*[1], Kym I E Snell *lecturer in biostatistics*[1], Frank E Harrell Jr *professor of biostatistics*[2], Glen P Martin *lecturer in health data sciences*[3], Johannes B Reitsma *associate professor*[4], Karel G M Moons *professor of clinical epidemiology*[4], Gary Collins *professor of medical statistics*[5], Maarten van Smeden *assistant professor*[4 5 6]

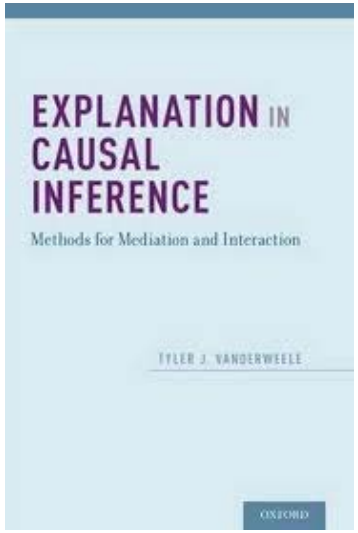
UMC Utrecht

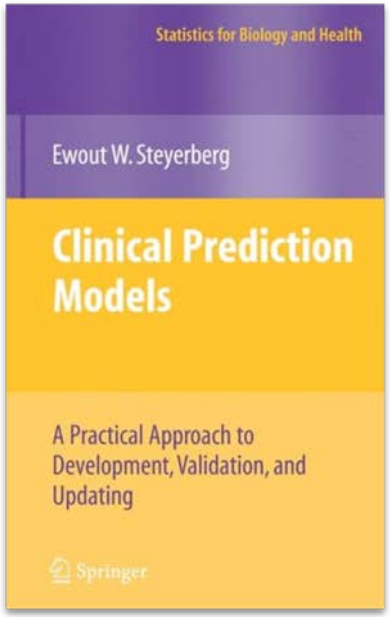
Specific guidance on conduct

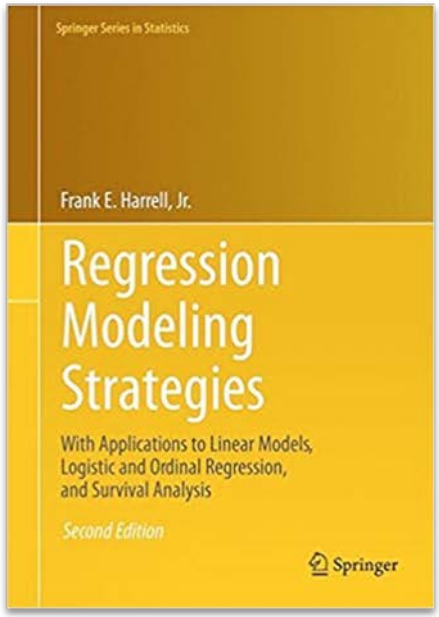


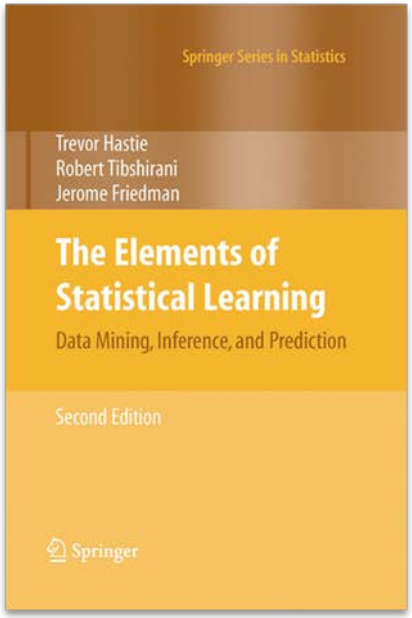


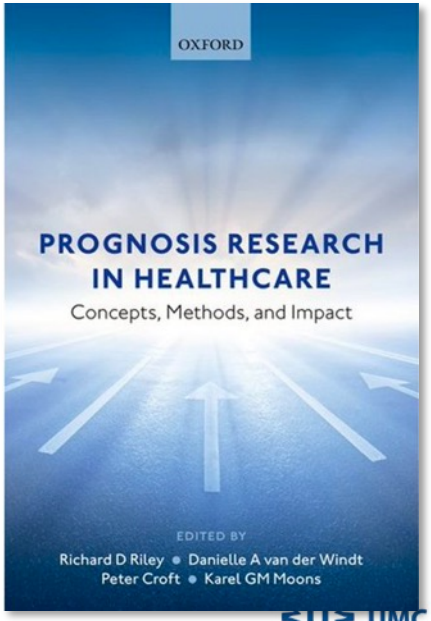












UMC Utrecht

Specific guidance on reporting and risk of bias

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration

Karel G.M. Moons, Ph
Petra Macaskill, PhD;

Annals of Internal Medicine RESEARCH AND REPORTING METHODS

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

equator network

Enhancing the QUAlity and Transparency Of health Research

Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies

Erik von Elm, MD; Douglas
for the STROBE Initiative

RESEARCH METHODS AND REPORTING

OPEN ACCESS

ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions

CrossMark

Jonathan AC Sterne,[1] Miguel A Hernán,[2] Barnaby C Reeves,[3] Jelena Savović,[1,4] Nancy D Berkman,[5] Meera Viswanathan,[6] David Henry,[7] Douglas G Altman,[8] Mohammed T Ansari,[9] Isabelle Boutron,[10] James R Carpenter,[11] An-Wen Chan,[12] Rachel Churchill,[13] Jonathan J Deeks,[14] Asbjørn Hróbjartsson,[15] Jamie Kirkham,[16] Peter Jüni,[17] Yoon K Loke,[18] Theresa D Pigott,[19] Craig R Ramsay,[20] Deborah Regidor,[21] Hannah R Rothstein,[22] Lakhbir Sandhu,[23] Pasqualina L Santaguida,[24] Holger J Schünemann,[25] Beverly Shea,[26] Ian Shrier,[27] Peter Tugwell,[28] Lucy Turner,[29] Jeffrey C Valentine,[30] Hugh Waddington,[31] Elizabeth Waters,[32] George A Wells,[33] Penny F Whiting,[34] Julian PT Higgins[35]

UMC Utrecht

Many other fallacies and paradoxes to consider

- Ecological fallacy
- Lord’s paradox
- Simpson’s paradox
- Berkson’s paradox
- Prosecutors fallacy
- Gambler’s fallacy
- Lindley’s paradox
- Low birthweight paradox
- Noisy data fallacy
- Will Rogers phenomenon
- …

UMC Utrecht

Improving epidemiological research: avoiding the statistical paradoxes and fallacies that nobody else talks about

Maarten van Smeden, PhD

28th Norwegian Epidemiological Association (NOFE) conference
October 26, 2022

UMC Utrecht
Julius Centrum

Improving epidemiological research: avoiding the statistical paradoxes and fallacies that ~~nobody else talks about~~ everyone should talk about and are well described in literature

Maarten van Smeden, PhD

28th Norwegian Epidemiological Association (NOFE) conference
October 26, 2022

UMC Utrecht
Julius Centrum

A gentle (1000 words) introduction

PRiMER
peer-reviewed reports in medical education research
special article

A Very Short List of Common Pitfalls in Research Design, Data Analysis, and Reporting

Maarten van Smeden, PhD

PRiMER. 2022;6:26.

Published: 8/10/2022 | DOI: 10.22454/PRiMER.2022.511416

Introduction

Performing scientific research without falling victim to one of the many research design, analysis, and reporting pitfalls can be challenging. As a medical statistician with research experience in a variety of medical disciplines, I regularly come across (and sometimes have been the cause of) avoidable errors and inaccuracies. Without such errors, research would, at the very least, be more informative to the readership of the research manuscript. In this article I present a short, nonexhaustive list of issues to consider.

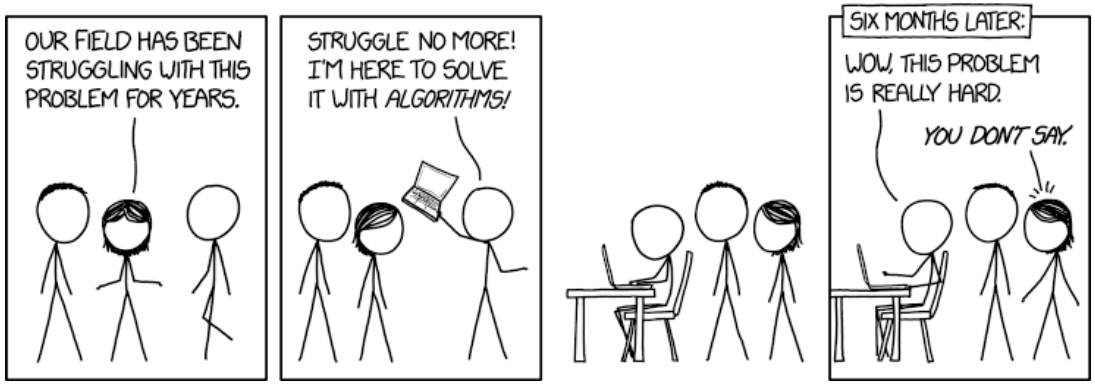
Research Questions and Aims

As the starting point of all scientific endeavors, it is incontrovertibly important to clearly define the research questions and aims. The subsequent planning of the collection of useful data and formulating adequate statistical analysis often becomes easier once it is clarified whether the ultimate aim is to *predict*, *explain*, or *describe*.[1] If the ultimate aim is to *explain*, the ideal design is often an *experiment* (eg, a randomized controlled trial). Conversely, for many health-related research questions, nonexperimental data are the only viable source of information. This type of data is subject to factors that hamper our ability to distinguish between true causes of outcomes and mere correlations. For instance, for a nonexperimental before-after study, a change in the health for some individuals over time is easily mistaken as evidence for the effectiveness of a particular curative treatment, which may just be caused by regression to the mean.[2] To avoid such errors, studies with an explanatory aim may benefit from applying *causal inference methodology*.[3]

Collecting Enough Data

A too-small-for-purpose sample size may result in *overfitting*,[4] *imprecision*, and lack of *power*, which can ruin a study of any kind. It is worthwhile to calculate the minimal sample size required to avoid disappointment.[5] It is usually wise to be skeptical about *rules of thumb* for sample size.[6]

UMC Utrecht

Email: M.vanSmeden@umcutrecht.nl
Twitter: @MaartenvSmeden

Slides available at:
https://www.slideshare.net/MaartenvanSmeden