

All that **glitters** isn't
gold – acknowledging
limitations in scientific
manuscripts

Gerben ter Riet, MD PhD

email: g.ter.riet@hva.nl

twitter: @grietter57

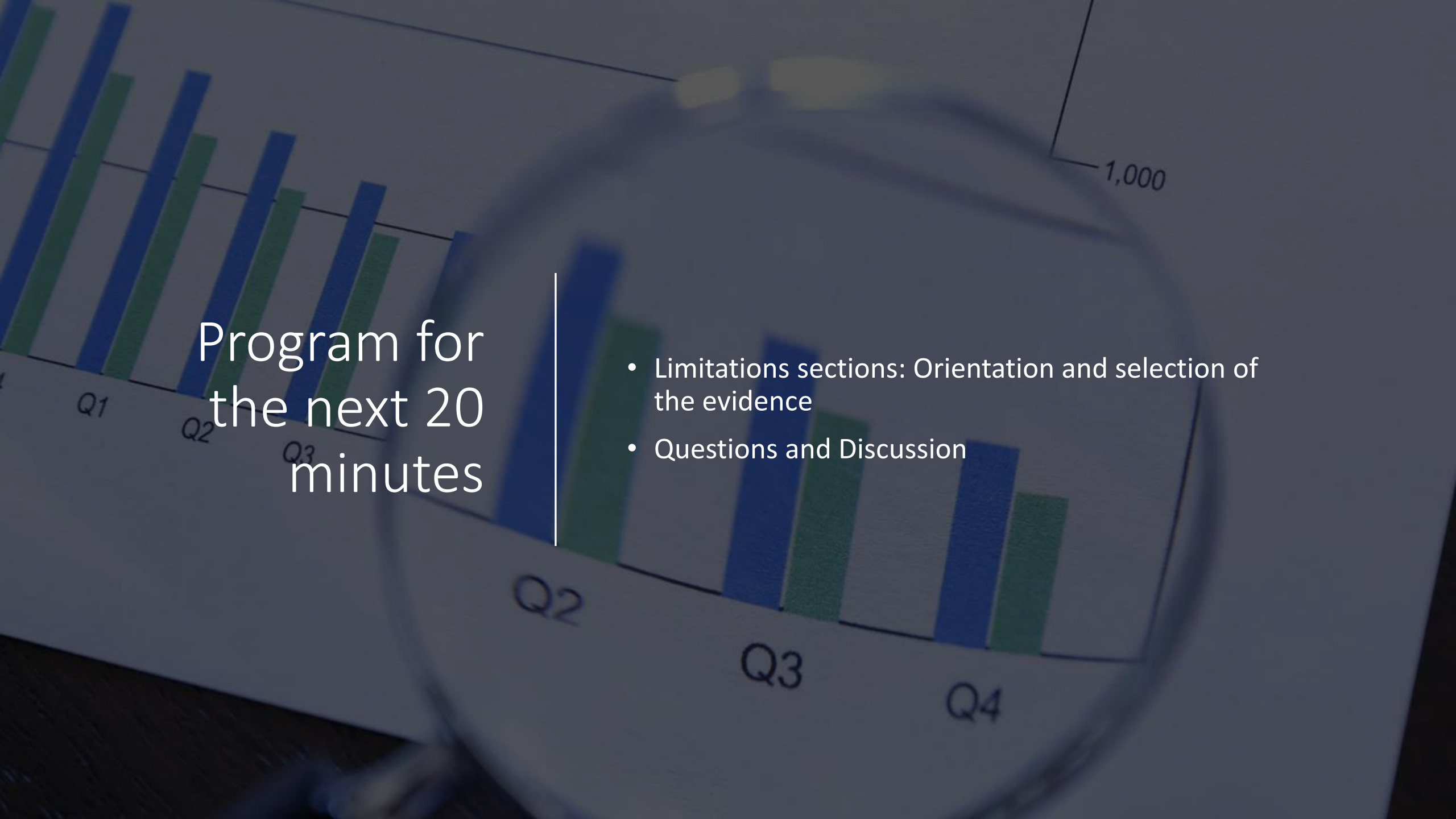
[ORCID](#)



Hogeschool van Amsterdam



Amsterdam UMC
Universitair Medische Centra



Program for the next 20 minutes

- Limitations sections: Orientation and selection of the evidence
- Questions and Discussion

Early interest

Beyond the Discussion Section

Cf. study design paper

Salami slicing?

Problems in the conduct of a randomised clinical trial

A discussion of an investigation of ascorbic acid and ultrasound in the treatment of pressure sores

Problems that arise during the execution of therapeutic trials on pressure sores are not always described in detail in study reports. The results of some trials are therefore difficult to interpret and investigators designing a new trial may make mistakes that could be avoided if previous papers had been more explicit.

G. ter Riet, MD, PhD, Clinical Epidemiologist;
A.G.H. Kessels, MD, MSc, Clinical Epidemiologist;
P. Knipschild, MD, PhD, Professor of Epidemiology;
all at the Department of Epidemiology, Maastricht University, Maastricht, The Netherlands

Randomised controlled trials

group consisting of patients with infected sores. Reports from animal research and research on the effects of ultrasound on

Effervescent tablets containing 10mg or 500mg of ascorbic acid were given daily, in the morning and early evening. This was continued until the sore had healed or 12 weeks had passed, whichever was first.

Pulsed, low-dose ultrasound or sham ultrasound treatment was given five times per week, directly on the wound

JOURNAL OF WOUND CARE MAY, VOL 7, NO 5, 1998

© MA Healthcare Ltd. Downloaded from magonlinelibrary.com by 129.127.145.224 on December 15, 2017.

Limitations: perceived as important for truth and trust

Rank number	Product of frequency and impact on truth (1–25)
1	Insufficiently supervise or mentor junior coworkers (C)
2	Insufficiently report study flaws and limitations (R)
3	Keep inadequate notes of the research process (D)
4	Turn a blind eye to putative breaches of research integrity by others (C)
5	Ignore basic principles of quality assurance (D)
Rank number	Product of frequency and impact on trust (1–25)
1	Use published ideas or phrases of others without referencing (C)
2	Insufficiently report study flaws and limitations (R)
3	Turn a blind eye to putative breaches of research integrity by others (C)
4	Insufficiently supervise or mentor junior coworkers (C)
5	Ignore basic principles of quality assurance (D)

RESEARCH

Open Access



Ranking major and minor research misbehaviors: results from a survey among participants of four World Conferences on Research Integrity

Lex M. Bouter^{1,2*}, Joeri Tjeldink^{2,3}, Nils Axelsen⁴, Brian C. Martinson⁵ and Gerben ter Riet⁶

Numbers are percentages admitting the research practice with a Likert score of 5, 6 or 7 over the last 3 years, where 7 = always.

NSRI 2020 results are out



QRP	Description (In the last three years..)	Disciplinary field				Academic rank			Overall
		Life and medical sciences	Social and behavioural sciences	Natural and engineering sciences	Arts and humanities	PhD candidates and junior researchers	Postdocs and assistant professors	Associate and full professors	
QRP10	Insufficient inclusion of study flaws and limitations in publications	17.8 (16.4,19.4)	17.2 (15.5,19.1)	15.8 (13.9,17.9)	15.2 (12.1,19)	21.2 (19.3,23.3)	16.9 (15.5,18.4)	13.7 (12.2,15.3)	17 (16.1,18)

Source: <https://osf.io/preprints/metaarxiv/vk9yt>

Perceived as important by journals too?

Topic	Percentage* of journals in						Total (n=835, N _w =14,814)
	Arts & Humanities (n=132, N _w =1,052)	Health Sciences (n=153, N _w =3,966)	Life Sciences (n=141, N _w =1,754)	Physical Sciences (n=162, N _w =4,586)	Social Sciences (n=153, N _w =3,350)	Multidisciplinary Sciences (n=94, N _w =106)	
Conflicts of Interest	22	89	67	53	56	61	63
Peer Review Type	61	53	43	41	72	35	52
Plagiarism	20	44	55	52	46	34	46
Errata	5	39	38	37	20	24	31
Data Sharing	3	32	35	33	26	32	29
Ethics Approval	0	74	33	7	13	20	29
COPE	6	31	25	21	26	15	24
ICMJE	0	72	21	3	4	5	24
Replication	0	30	35	24	19	21	24
Preprint	11	26	32	25	14	18	22
ORCID	11	20	22	19	22	16	20
Registration	2	38	22	2	3	10	15
Reporting Guidelines	2	36	19	4	9	6	15
Image Manipulation	3	15	21	11	10	12	12
Limitations	0	22	11	3	5	4	9
Statistics†	1	16	6	0	2	2	6
Null Results	0	5	3	0	3	1	2
Shared Authorship	0	0	8	1	3	3	2
TOP Guidelines	0	1	1	1	4	1	2
Average no. of topics per journal	1	6	5	3	4	3	4



RESEARCH ARTICLE

Journals' instructions to authors: A cross-sectional study across scientific disciplines

Mario Malički ^{1,2*}, IJsbrand Jan Aalbersberg³, Lex Bouter^{4,5}, Gerben ter Riet^{1,2}

Negotiation within teams, team composition

Richard Horton, interviewing authors several months after publication of 10 RCTs in major journals, 2002

“Important weaknesses were often admitted on direct questioning but were not included in the published article.”

AUTHORSHIP AND CONTRIBUTORSHIP

The Hidden Research Paper

Richard Horton, FRCP

Context To determine whether the views expressed in a research paper are accurate representations of contributors' opinions about the research being reported.

Box 1. Questions Asked of Contributors of 10 Selected Research Papers

In your own words, how would you:

1. Summarize the results of your study?
2. Define the strengths of your study?
3. Define the weaknesses of your study?
4. Interpret the results of your study in the context of the totality of available evidence?
5. Assess the implications of your results?
6. Plan further research into the question under investigation?

All That Glitters Isn't Gold: A Survey on Acknowledgment of Limitations in Biomedical Studies

Gerben ter Riet^{1*}, Paula Chesley², Alan G. Gross³, Lara Siebeling¹, Patrick Muggensturm⁴, Nadine Heller⁵, Martin Umbeh^{4,6}, Daniela Vollenweider⁷, Tsung Yu⁸, Elie A. Akl^{9,10,11}, Lizzy Brewster¹², Olaf M. Dekkers¹³, Ingrid Mühlhauser¹⁴, Bernd Richter¹⁵, Sonal Singh¹⁶, Steven Goodman¹⁷, Milo A. Puhan^{8,18}

What about the
true limitations?

- 27% of publications (81/300) no mention of any limitations
- 73% acknowledged 3 (range 1–8) limitations
- 5% mentioned a limitation in the abstract

Impact of editorial handling and peer review

Keserlioglu et al, 2019

- 446 RCT papers from 27 BMC journals and BMJ Open
- Before-after
- Count Self-Acknowledged Limitation (SAL) **sentences**, using software

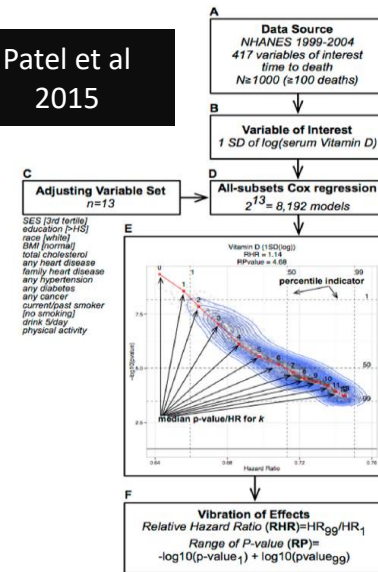
- 2.5 sentences in manuscripts to 3.9 in publications: + 1.4 (1.1 – 1.8)
- 45% (202/446) manuscripts mentioned zero limitations
- After peer review, 31% (63/202) had at least one sentence (SAL)

Quantitative Bias Analysis

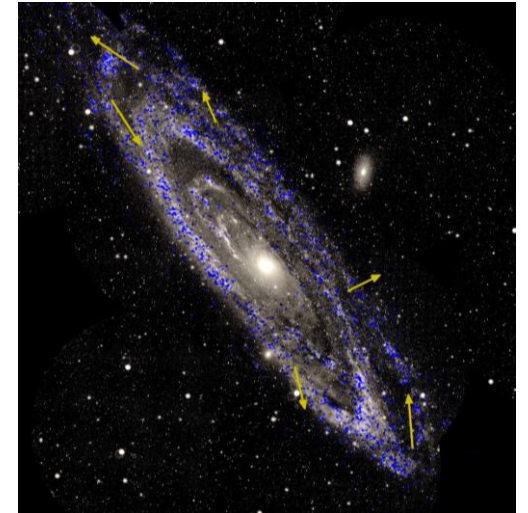
“Move the debate among stakeholders from the realm of qualitative criticism, which is often heavily influenced by politics and polemics, into the realm of quantitative analysis.” Lash 2008

“Reduce the asymmetry of information between analysts and readers, and more transparently show what estimates are possible.” Young & Stewart 2020

Patel et al
2015

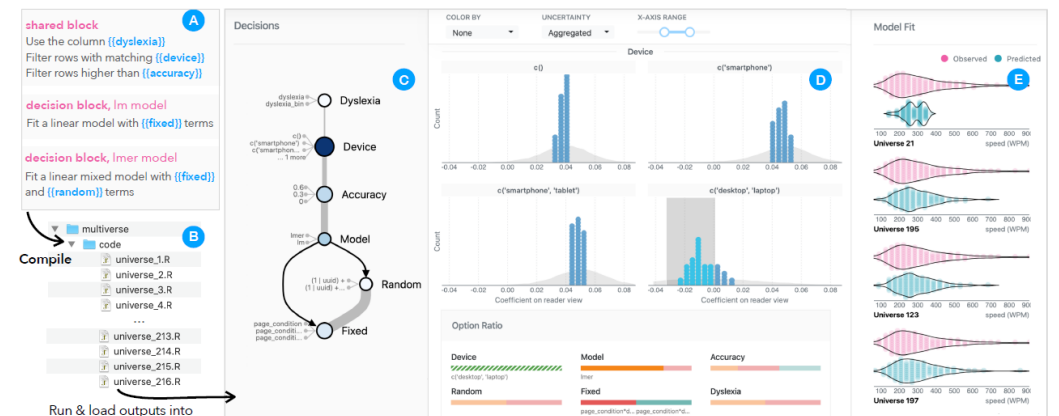


atom & galaxy ~
limitations & research integrity




Boba: Authoring and Visualizing Multiverse Analyses

Yang Liu, Alex Kale, Tim Althoff, and Jeffrey Heer





Questions remaining?

1. Aim of writing: Inform-convince-inspire-activate
 2. Define 'limitation'
 3. Can experts agree on the *true* study limitations?
 4. Do expert readers need SALs?
 5. Restrict to protocol deviations?
 6. What are the forces that hinder or motivate us to list informative limitations
 7. Let others judge (e.g. registered reports)
 8. Multiverse analysis obviates self-acknowledgment of limitations, but what are *its* limitations?
- 

NIRRINGRAZZJAK MOCHCHAKKERAM SALAMAT RAMBI MATRACAT MA MATO KIITOS MOCHCHAKKERAM CHOKRANE
KIITOS CHOKRANE **THANK YOU** OBRIGADO SPASIBO TERMA KASIH MATONDO
ASANTE MOCHCHAKKERAM ARIGATO KIITOS DANKON MULTUMESC **MERCI** MERCI GRAZIE CAM ON BAN OBRIGADO
DANK JE DANKON SPASIBO MOCHCHAKKERAM TERMA KASIH WELALIN KIA ORA SALAMAT VINAKA MATUR NUWUN MATONDO MULTUMESC CHOKRANE MAAKE JUSPAXAR MAAKE ARIGATO SPASIBO KIITOS
OBRIGADO DANKON SPASIBO MOCHCHAKKERAM TERMA KASIH WELALIN KIA ORA SALAMAT VINAKA MATUR NUWUN MATONDO MULTUMESC CHOKRANE MAAKE JUSPAXAR MAAKE ARIGATO SPASIBO KIITOS
MACHCHAKKERAM

What about
true
limitations?
(unpublished)

15 RCTs, 7 (paired) experts: On average the ratio between the number of expert-identified and author-acknowledged limitations was 4.3 (varying between 2.3 and 9.8).

Experts did not agree very often on the nature of the limitations they identified

In only 19/285 (15 publications*19 items) instances did more than one expert judge a limitation as “(very) serious”



Limes, limites

Incomplete history of work on limitations

1974 Richard Feynman (Cargo Cult Science)

1994 Steven Goodman (Impact peer review at Annals Intern Med)

2002 Richard Horton (The hidden research paper)

2004 Annals of Internal Medicine (abstracts SAL required)

2007 John Ioannidis (limitations not properly acknowledged)

2013 Gerben ter Riet & Milo Puhan (N=300 paper, 27% has zero SALs)

2018 Halil Kilicoglu & Gerben ter Riet (Automating detection of SALs)

2019 Kerem Keserlioglu & Gerben ter Riet (Impact peer review & SALs)

2019 Mario Malicki & Gerben ter Riet (Instructions to authors, SALs?)

2019 Isabelle Boutron (spin in publications with limitations as 1 subitem)

20xx RCT on effect of detection software results on peer review

Developing software (nice but imperfect)

Table 2. Automatic limitation recognition results (LR = logistic regression, SVM = support vector machines). The 95% confidence intervals are shown in square brackets. All numbers are percentages. For LR, self-training parameters used were $\alpha_{POS}=0.7$, $\alpha_{NEG}=0.95$, $\beta_{POS}=0.9$. For SVM, they were $\alpha_{POS}=0.7$, $\alpha_{NEG}=0.8$, $\beta_{POS}=0.9$.

Method	Precision	Recall	F ₁ score	Accuracy
Baseline	62.6 [60.2-65.0]	81.2 [79.2-83.2]	70.7 [68.4-73.0]	86.4 [84.7-88.1]
Rules	75.8 [73.6-78.0]	84.8 [83.0-86.6]	80.0 [78.0-82.0]	91.5 [90.1-92.9]
Fully supervised learning with SEED for training				
LR	73.0 [70.8-75.2]	75.9 [73.7-78.1]	74.4 [72.2-76.6]	89.5 [88.0-91.1]
SVM	76.6 [74.5-78.7]	69.3 [67.0-71.6]	72.8 [70.6-75.1]	89.6 [88.1-91.1]
Leveraging UNLABELED for training				
Self-training (LR)	69.4 [67.2-71.8]	84.2 [82.4-86.0]	76.1 [74.0-78.3]	89.4 [87.8-91.0]
Self-training (SVM)	77.1 [75.0-79.2]	71.0 [68.7-73.3]	73.9 [71.7-76.1]	89.9 [88.4-91.4]
Rule-based expansion (LR)	77.4 [75.3-79.5]	81.2 [79.2-83.2]	79.2 [77.2-81.3]	91.4 [90.0-92.8]
Rule-based expansion (SVM)	77.8 [75.7-79.9]	83.5 [81.6-85.4]	80.6 [78.6-82.6]	91.9 [90.5-93.3]

Automatic recognition of self-acknowledged limitations in clinical research literature

Halil Kilicoglu,¹ Graciela Rosemblat,¹ Mario Malicki,^{2,3} and Gerben ter Riet²

Why mention limitations?

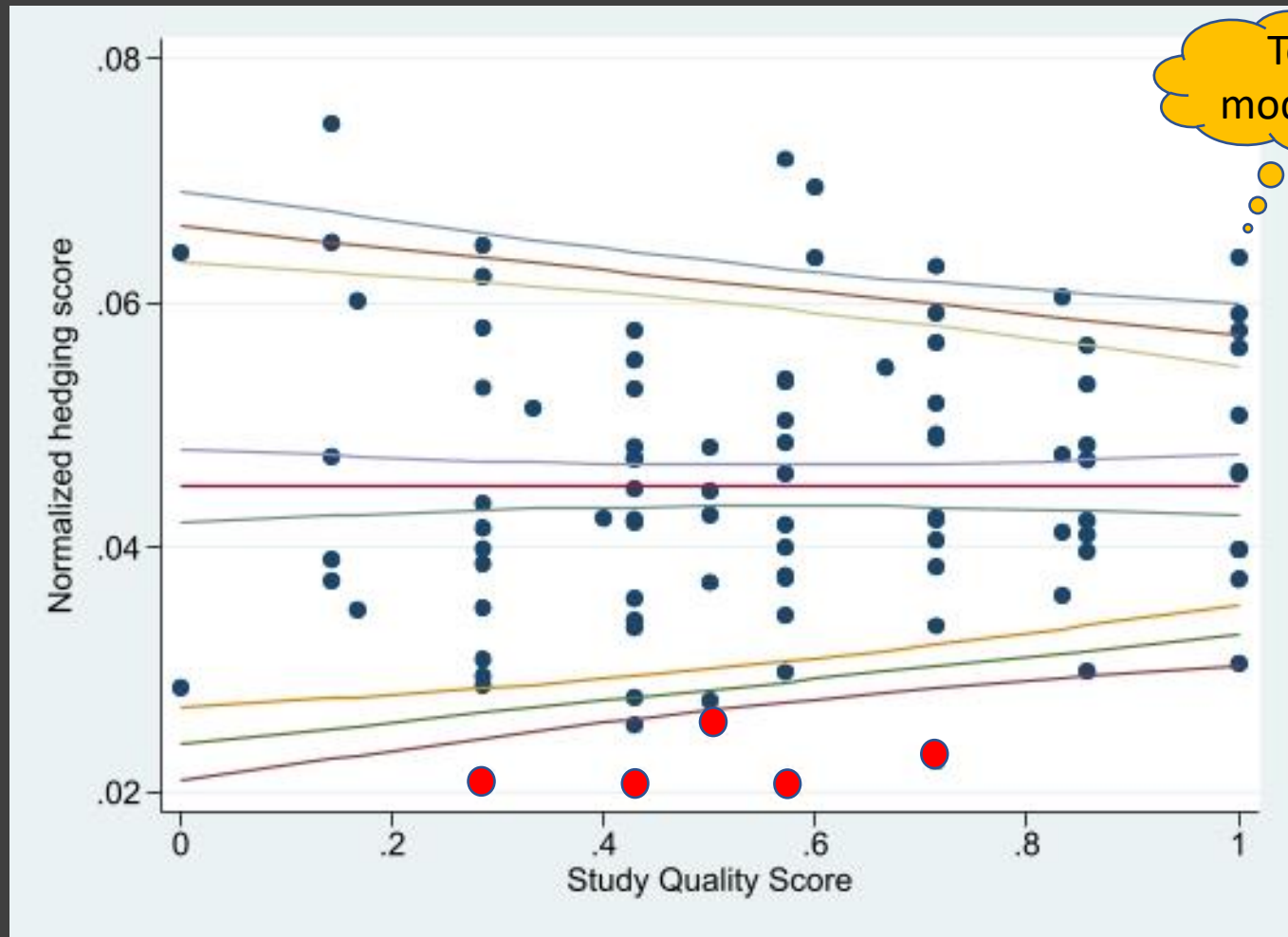
(Ioannidis JP, 2007)

- “[..] do these problems with errors, methods, and validity eventually matter, and, if so, to what extent? ”
- Guide and inform new studies
- Journals’ *instructions to authors tend to emphasize novelty, impact and significance*

“Researchers have an obligation to the academic community to present complete and honest limitations of a presented study.”

“[..] describe the potential limitations, explain the implication of the limitations, provide possible alternative approaches, and describe steps taken to mitigate the limitations.”

“This is not just about being self-critical or particularly humble in presenting our research. Identifying limitations and explaining to the reader what impact these limitations have on the study results, not only demonstrates rigour but also gives the authors a chance to identify clear directions for future research. [...] this [...] paragraph can be one of the most exciting parts to read in a paper, [...]”



Do weaker trials hedge their claims more?: Towards **reference values** for hedging

S. Amini et al

<http://pubmet.unizd.hr/pubmet2017/sessions/weaker-research-make-weaker-claims-towards-automated-detection-linguistic-hedging/>