Tolerating bad health research: the continuing scandal

Stefania Pirosca
University of Aberdeen
Frances Shiely
University College Cork
Mike Clarke
Queen's University Belfast
Shaun Treweek (streweek@mac.com)
University of Aberdeen https://orcid.org/0000-0002-7239-7241

Research Article

Keywords: Randomised trials, research waste, risk of bias, statisticians, methodologists

Posted Date: December 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1113501/v1

Abstract

Background

At the 2015 REWARD/EQUATOR conference on research waste, the late Doug Altman revealed that his only regret about his 1994 BMJ paper 'The scandal of poor medical research' was that he used the word 'poor' rather than 'bad'. But how much research is bad? And what would improve things?

Main text

We focus on randomised trials and look at scale, participants and cost. We randomly selected up to two quantitative intervention reviews published by all clinical Cochrane Review Groups between May 2020 and April 2021. Data including risk of bias, number of participants, intervention type and country were extracted for all trials included in selected reviews. High risk of bias trials were classed as bad. The cost of high risk of bias trials was estimated using published estimates of trial cost per-participant.

We identified 96 reviews authored by 546 reviewers from 49 clinical Cochrane Review Groups that included 1,659 trials done in 84 countries. Of the 1,640 trials providing risk of bias information, 1,013 (62%) were high risk of bias (bad), 494 (30%) unclear and 133 (8%) low risk of bias. Bad trials were spread across all clinical areas and all countries. Well over 220,000 participants (or 56% of all participants) were in bad trials. The low estimate of the cost of bad trials was £726 million; our high estimate was over £8 billion.

We have five recommendations: trials should be neither funded (1) or given ethical approval (2) unless they have a statistician and methodologist; trialists should use a risk of bias tool at design (3); more statisticians and methodologists should be trained and supported (4); there should be more funding into applied methodology research and infrastructure (5).

Conclusions

Most randomised trials are bad and most trial participants will be in one. The research community has tolerated this for decades. This has to stop: we need to put rigour and methodology where it belongs– at the centre of our science.

Background

At the 2015 REWARD/EQUATOR conference on research waste, the late Doug Altman revealed that his only regret about his 1994 BMJ paper '*The scandal of poor medical research*'[1] was that he used the word 'poor' rather than 'bad'. Towards the end of his life Doug had considered writing a sequel with a title that included not only 'bad' but 'continuing'[2].

That 'continuing' is needed should worry all of us. Ben Van Calster and colleagues have recently highlighted the paradox that science consistently undervalues methodology that would underpin good

research[3]. The Covid-19 pandemic has generated an astonishing amount of research and some of it has transformed the way the virus is managed and treated. But we expect that much Covid-19 research will be bad because much of health research in general is bad[3]. This was true in 1994 and it remains true in 2021 because how research is done allows it to be so. Research waste seems to be baked-in to the system.

In this Commentary, we do not intend to list specific examples of research waste. Rather, we want to talk about scale, participants, money and then finish with five recommendations. All of the latter will look familiar– Doug Altman and others[3-8] have suggested them many times – but we hope our numbers on scale, participants and money will lend the recommendations an urgency they have always deserved but never had.

So, how much research is bad?

That research waste is common is not in doubt[3-8] but we wanted to put a number on something more specific: how much is bad research that is not just wasteful but which we could have done without and lost little or nothing? Rather than trying to tackle all of health research, we have chosen to focus on randomised trials because that's the field we know best and, in addition, they play a central role in decisions regarding the treatments that are offered to patients.

With this in mind, we aimed to estimate the proportion of trials that are bad, how many participants were involved and how much money was spent on them.

Selecting a cohort of trials

We used systematic reviews as our starting point because these bodies of trial evidence often underpin clinical practice through guideline recommendations and policy. We specifically chose Cochrane systematic reviews because they are standardised, high quality systematic reviews. We were only interested in recent reviews because these represent the most up-to-date bodies of evidence.

Moreover, Cochrane reviews record the review authors' judgements about the risk of bias of included trials, in other words they assess the extent to which the trial's findings can be believed[9]. We consider that to be a measure of how good or bad a trial is. Cochrane has three categories of overall risk of bias: high, uncertain and low. We considered a high risk of bias trial to be bad, a low risk of bias trial to be good and an uncertain risk of bias trial to be exactly that, uncertain.

We used the list randomiser at random.org to randomly select two reviews published between May 2020 and April 2021 from each of the 53 clinical Cochrane Review Groups. To be included, a review had to consider intervention effects rather than being a qualitative review or a review of reviews. We then extracted basic information (our full dataset is at https://osf.io/dv6cw/?view_only=0becaacc45884754b09fd1f54db0c495) about every included trial in each review, including the overall risk of bias assessment. Our aim was to make no judgements about risk of bias ourselves but

to take what the review authors had provided. We did not contact review or trial authors for additional information. Extracted data were put into Excel spreadsheets, one for each Cochrane Review Group.

Analysis

To answer our question about the proportion of bad trials and how many participants were in them, we used simple counts across reviews and trials. Counts across spreadsheets were done using R and our code is at https://osf.io/dv6cw/?view_only=0becaacc45884754b09fd1f54db0c495. To estimate how much money might have been spent on the trials, we used three estimates of the cost-per-participant to give a range of possible values for total spend:

1. Estimate 1: An estimate of the cost-per-participant for the UK's National Institute for Health Research Health Technology Assessment (NIHR HTA) Programme trials of 2,987 GBP. This was calculated based on a median cost per NIHR HTA trial of 1,433,978 GBP for 2011-2016[10] and a median final recruitment target for NIHR HTA trials of 480 for 2004-2016[11].
2. Estimate 2: The median cost-per-participant of 41,413 USD found for pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015–2017[12].
3. Estimate 3: The 2012 average cost-per-participant for UK trials of 9,758 EUR found by Europe Economics[13].

These estimates were all converted into GBP using https://www.currency-converter.org.uk to get the exchange rate on 1st January in the latest year of trials covered by the estimate (i.e. 2017 for E2 and 2012 for E3). These were then all converted to 2021 GBP on 11th August 2021 using https://www.inflationtool.com. We acknowledge that these are unlikely to be exact for any given trial in our sample, but they were intended to give ballpark average figures to promote discussion.

Scale, participants and money

1. Scale

We extracted data for 1,659 randomised trials spread across 96 reviews from 49 of the 53 clinical Cochrane Review Groups. The remaining four Review Groups published no eligible reviews in our time period. The 96 included reviews involved 546 review authors. Trials in 84 countries, as well as 193 multinational trials are included. Risk of bias information was not available for 19 trials, meaning our risk of bias sample is 1,640 trials. Almost all reviews (94) exclusively used Cochrane's 'old' risk of bias tool rather than the new Risk of Bias 2 tool[14]. Where the old tool was used, we used review authors' assessment of overall risk of bias. For the two reviews that used Risk of Bias 2 we did not make individual risk of bias judgements for domains, but we did take a view on the overall risk of bias if the review authors did not do this. We did this by looking across the individual domains and making a choice of high, uncertain or low overall risk of bias based on the number of individual domains falling into each category. This was a judgement; we did not use a hard-and-fast rule. We had to do this for 40 trials.

The majority of trials (1,013, or 62%) were high risk of bias (Table 1). These trials were spread across all 49 Cochrane Review Groups and over half of the Groups (28, or 57%) had zero low risk of bias trials included in the reviews we randomly selected. The clinical area covered by the Anaesthesia Review Group had the highest proportion of low risk of bias trials at 60% but this group included 19 trials with no risk of bias information (See Figure 1).

Some of the 84 countries in our sample contributed very few trials but Table 2 shows risk of bias data for the 17 countries that contributed 20 or more trials, as well as for multinational trials. The percentage of a country's trials that were judged as low risk of bias reached double figures for multinational trials (23%) and five individual countries: Australia (10%), France (13%), India (10%), Japan (10%) and the UK (11%). The full country breakdown is given in Supplementary File 1.

2. Participants

The 1,659 included trials involved a total of 398,410 participants. The majority of these (222,850, or 56%) were in high risk of bias trials (Table 1).

Table 1.

	High	Unclear	Low	No RoB
Number of Trials	1013	494	133	19
(total = 1659; % based on 1640 trials with risk of bias information)	(62%)	(30%)	(8%)	
Number of Participants	222,850	127,290	47,138	1132
(total = 398,410)	(55.93%)	(31.95%)	(11.83%)	(0.28%)

Table 2.

Country	Number of Trials	High	Unclear	Low
Multinational	193	100 (52%)	48 (25%)	45 (23%)
Australia	41	27 (66%)	10 (24%)	4 (10%)
Brazil	27	16 (59%)	10 (37%)	1 (4%)
Canada	40	26 (65%)	12 (30%)	2 (5%)
China	87	56 (64%)	23 (26%)	7 (8%)
France	31	11 (35%)	16 (52%)	4 (13%)
Germany	48	40 (83%)	7 (15%)	1 (2%)
India	48	35 (73%)	6 (13%)	5 (10%)
Iran	49	30 (61%)	18 (37%)	0 (0%)
Italy	47	30 (64%)	17 (36%)	0 (0%)
Japan	42	22 (52%)	16 (38%)	4 (10%)
Korea	32	18 (56%)	11 (34%)	1 (3%)
Netherlands	22	17 (77%)	5 (23%)	0 (0%)
Spain	29	25 (86%)	2 (7%)	1 (3%)

Sweden	26	16 (62%)	7 (27%)	2 (8%)
Turkey	32	18 (56%)	9 (28%)	2 (6%)
United Kingdom	114	63 (55%)	38 (33%)	12 (11%)
USA	280	168 (60%)	98 (35%)	12 (4%)

3. Money

Table 3 shows estimates for the amount of money spent on trials in each of the three risk of bias categories.

Using our low estimate for cost-per-participant (Estimate 1 from NIHR HTA trials) we get an estimated spend of £726 million on high risk of bias trials. Our high estimate (Estimate 2 from USA drug approval trials) gives an equivalent figure of over £8 billion. Based on an annual spend of £76 million for the UK's NIHR HTA programme[15], the first figure, our lowest estimate, would be sufficient to fund the programme for almost a decade; while the second figure would fund it for over a century.

While looking at scale, participants and money we made a few other secondary observations. To avoid distracting attention from our main points we present these observations in Supplementary File 2.

Table 3.

Risk of Bias	Number of Trials	Participants	E1 Costs (£)	E2 Costs (£)	E3 Costs (£)
High	1013	222,850	725,599,600	8,004,326,300	2,090,778,700
Uncertain	494	127,290	414,456,240	4,572,002,220	1,194,234,780
Low	133	47,138	153,481,328	1,693,102,684	442,248,716

Discussion

Bad trials– ones where we have little confidence in the results– are not just common, they represent the majority of trials across all clinical areas in all countries. Over half of all trial participants will be in one.

Our estimates suggest that the money spent on these bad trials would fund the UK's largest public funder of trials for anything between a decade and a century. It's a wide range but either way, it's a lot of money. Had our random selection produced a different set of reviews, or we had assessed all those published in the last one, five, ten, or 20 years, we have no reason to believe that the headline result would have been different. Put simply, most randomised trials are bad.

Despite this, we think our measure of bad is actually conservative because we have only considered risk of bias. We have not attempted to judge whether trials asked important research questions, whether they involved the right participants, whether their outcomes were important to decision-makers such as patients and health professionals nor have we attempted to comment on the many other decisions that affect the usefulness of a trial[16,17]. In short, the picture our numbers paint is undoubtedly gloomy, but the reality is probably worse.

Five recommendations for change

Plenty of ideas have been suggested about what must change[1,3-8] but we propose just five here because the scale of the problem is so great that providing focus might avoid being overwhelmed into inaction. We think these five recommendations, if implemented, would reduce the number of bad trials and could do so quite quickly.

Recommendation 1: do not fund a trial unless the trial team contains methodological and statistical expertise

Doing trials is a team sport. These teams need experienced methodologists and statisticians. It's hard to imagine doing, say, bowel surgery without involving people who have been trained in, and know how to do, bowel surgery. Sadly, the same does not seem to be true for trial design and statistical analysis of trial data. Our colleague Darren Dahly, a trial statistician, neatly captured the problem in a series of ironic tweets sent at the end of 2020:

These raise a smile but make a very serious point: we would not tolerate statisticians doing surgery so why do we tolerate the reverse? Clearly this is not about surgeons, it is about not having the expertise needed to do the job properly.

Recommendation 2: do not give ethical approval for a trial unless the trial team contains methodological and statistical expertise

As for Recommendation 1, but for ethical approval. No patient or member of the public should be in a bad trial and ethical committees, like funders, have a duty to stop this happening. Indeed, we think public and patient contributors on ethics committees should routinely ask the question 'Who is the statistician and who is the methodologist?' and if the answer is unsatisfactory, ethical approval is not awarded until a name can be put against these roles.

Recommendation 3: use a risk of bias tool at trial design

This is the simplest of our recommendations. Risk of bias tools were developed to support the interpretation of trial results in systematic reviews. However, as Yordanov wrote in 2015[5], by then the horse has bolted and nothing can be changed. Applying a risk of bias tool at the trial design phase, correctly interpreting the results and making any necessary changes to the trial, would help to avoid some

of the problems we highlight. Funders could ask to see the completed risk of bias tool, as could ethics committees. No trial should be high risk of bias.

Recommendation 4: train and support more methodologists and statisticians

Recommendations 1, 2 and 3 all lead to a need for more methodologists and statisticians. This has a cost but it would probably be much less than the money wasted on bad trials. See Recommendation 5.

Recommendation 5: put more money into applied methodology research and supporting infrastructure

Methodology research currently runs mostly on love not money. This seems odd when over 60% of trials are so methodologically flawed we can't believe their results and we are uncertain whether we should believe the results of another 30%.

In 2015 David Moher and Doug Altman proposed that 0.1% of funders' and publishers' budgets could be set aside for initiatives to reduce waste and improve the quality, and thus value, of research publications[6]. That was for publications but the same could be done for trials, although we'd suggest a figure closer to 10% of funders' budgets. All organisations that fund trials should also be funding applied work to improve trial methodology, including supporting the training of more methodologists and statisticians. There should also be funding mechanisms to ensure methodology knowledge is effectively disseminated and implemented. Dissemination is a particular problem and the UK's only dedicated methodology funder, the Medical Research Council-NIHR 'Better Methods, Better Research' Panel acknowledges this in its Programme Aims[17].

Money will always be tight. Our work, and that of others before us[1,3-8], make clear that a large amount of the money we put into trials globally is being wasted. Some of that money could be repurposed to fund our five recommendations. This might lead to fewer trials overall but it would generate more good trials and mean that a greater proportion of trial data is of the high quality needed to support and improve patient and public health.

Conclusion

That so much research, and so many trials, is and are bad is indeed a scandal. That it continues decades after others highlighted the problem, is a bigger scandal. Even the tiny slice of global research featured in our study describes trials that involved hundreds of thousands of people, and cost hundreds of millions of pounds, but which led to little or no useful information.

The Covid-19 pandemic has been a time for many things, including reflection. As many countries start to look to what can be learnt, all of us connected with trials should put rigour and methodology where it belongs– at the centre of our science. We think our five recommendations are a good place to start.

To quote Doug Altman 'We need less research, better research, and research done for the right reasons'[1]. Quite so.

Abbreviations

BMJ
British Medical Journal
Covid-19
Coronavirus Disease 2019
E1
Estimate 1
E2
Estimate 2
E3
Estimate 3
EUR
Euro
GBP
British Pound Sterling
HTA
Health Technology Assessment
NIHR
National Institute for Health Research
UK
United Kingdom
USA
United States of America
USD
United States Dollar

Declarations

Ethical approval

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All our data are available at https://osf.io/dv6cw/?view_only=0becaacc45884754b09fd1f54db0c495

Competing interests

ST is an Editor-in-Chief of *Trials*. MC, FS and ST are actively involved in initiatives to improve the quality of trials and all seek funding to support these initiatives and therefore have an interest in seeing funding for trial methodology increased. SP has no competing interests.

Funding

This work was funded by Ireland's Health Research Board through the Trial Methodology Research Network (HRB-TMRN) as a summer internship for SP.

Authors' contributions

ST had the original idea for the work. ST, FS and MC designed the study. SP identified reviews and trials, extracted data and did the analysis, in discussion with ST and FS. ST and SP wrote the first draft and all authors contributed to further drafts. All authors approved the final draft.

Acknowledgements

We would like to thank Brendan Palmer for helping SP with R coding and Darren Dahly for confirming that he was happy for us to use his tweets. The Health Services Research Unit, University of Aberdeen, receives core funding from the Chief Scientist Office of the Scottish Government Health Directorates. This work was done as part of the Trial Forge initiative to improve trial efficiency (https://www.trialforge.org).

References

1. Altman DG. The scandal of poor medical research. BMJ 1994;308:283.
2. Matthews R, Chalmers I, Rothwell P. Douglas G Altman: statistician, researcher, and driving force behind global initiatives to improve the reliability of health research. BMJ 2018; 362: k2588.
3. Van Calster B, Wynants L, Riley RD, van Smeden M, Collins GS. Methodology over metrics: Current scientific standards are a disservice to patients and society. J Clin Epi. 2021 May 30, 2021 May 30;S0895-4356(21)00170-0.doi: 10.1016/j.jclinepi.2021.05.018.
4. Glasziou P, Chalmers IC. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. MJ 2018;363:k4645.
5. Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. BMJ 2015 Mar 24;350:h809–9.
6. Moher D, Altman DG (2015) Four Proposals to Help Improve the Medical Research Literature. PLoS Med 12(9): e1001864. doi:10.1371/ journal.pmed.1001864
7. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009 Jul 4;374(9683):86–9.
8. Macleod MR, Michie S, Roberts I, Dirnagl U, Chalmers I, Ioannidis JPA, et al. Biomedical research: increasing value, reducing waste. Lancet 2014; 383(9912):101–4.

9. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. Available from
10. Chinnery, F., Bashevoy, G., Blatch-Jones, A. *et al.* National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme research funding and UK burden of disease. Trials 2018, 19:87. https://doi.org/10.1186/s13063-018-2489-7.
11. Walters SJ, Bonacho dos Anjos Henriques-Cadby I, Bortolami O, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. BMJ Open 2017;7:e015276. doi:10.1136/bmjopen-2016- 015276.
12. Moore TJ, Heyward J, Anderson G, et al. Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015–2017: a cross-sectional study. BMJ Open 2020;**10**:e038863. doi:10.1136/ bmjopen-2020-038863.
13. Hawkes N. UK must improve its recruitment rate in clinical trials, report says. BMJ 2012; 345:e8104 doi: 10.1136/bmj.e8104.
14. Sterne J A C, Savović J, Page M J, Elbers R G, Blencowe N S, Boutron I et al. RoB 2: a revised tool for assessing risk of bias in randomised trials BMJ 2019; 366:l4898 doi:10.1136/bmj.l4898
15. Williams H. The NIHR Health Technology Assessment Programme: Research needed by the NHS. https://www.openaccessgovernment.org/nihr-health-technology-assessment-programme-nhs/85065/ [Accessed 11/10/2021].
16. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet 2014; 383: 156–65.
17. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, Schulz KF, Tibshirani R. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014; 383: 166–75.
18. MRC-NIHR Better Methods, Better Research. Programme Aims. https://mrc.ukri.org/funding/science-areas/better-methods-better-research/overview/#aims [Accessed 30/9/2021].

Figures

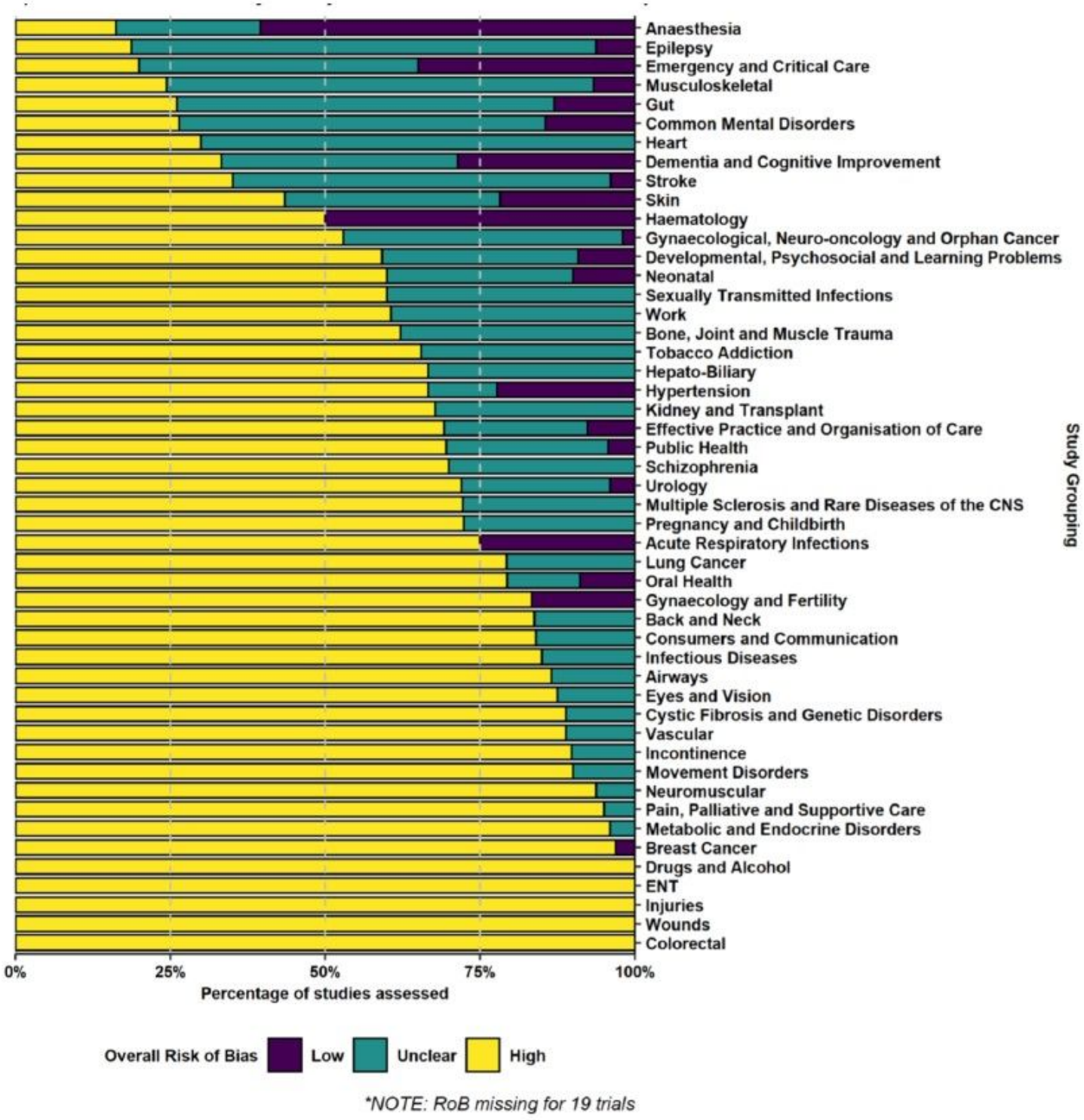

Figure 1

Risk of bias for included trials in randomly selected systematic reviews published between May 2020 and April 2021 by 49 Cochrane Review Groups.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1.xls
- SupplementaryFile2.docx