

# BMJ Open Trends and predictors of biomedical research quality, 1990–2015: a meta-research study

Maryaline Catillon 

**To cite:** Catillon M. Trends and predictors of biomedical research quality, 1990–2015: a meta-research study. *BMJ Open* 2019;**9**:e030342. doi:10.1136/bmjopen-2019-030342

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2019-030342>).

Received 09 March 2019

Revised 13 May 2019

Accepted 01 August 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

Ph.D. Program in Health Policy, Harvard University, Cambridge, Massachusetts, USA

## Correspondence to

Maryaline Catillon;  
mcatillon@g.harvard.edu

## ABSTRACT

**Objective** To measure the frequency of adequate methods, inadequate methods and poor reporting in published randomised controlled trials (RCTs) and test potential factors associated with adequacy of methods and reporting.

**Design** Retrospective analysis of RCTs included in Cochrane reviews. Time series describes the proportion of RCTs using adequate methods, inadequate methods and poor reporting. A multinomial logit model tests potential factors associated with methods and reporting, including funding source, first author affiliation, clinical trial registration status, study novelty, team characteristics, technology and geography.

**Data** Risk of bias assessments for random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data and selective reporting, for each RCT, were mapped to bibliometric and funding data.

**Outcomes** Risk of bias on six methodological dimensions and RCT-level overall assessment of adequate methods, inadequate methods or poor reporting.

**Results** This study analysed 20 571 RCTs. 5.7% of RCTs used adequate methods (N=1173). 59.3% used inadequate methods (N=12 190) and 35.0% were poorly reported (N=7208). The proportion of poorly reported RCTs decreased from 42.5% in 1990 to 30.2% in 2015. The proportion of RCTs using adequate methods increased from 2.6% in 1990 to 10.3% in 2015. The proportion of RCTs using inadequate methods increased from 54.9% in 1990 to 59.5% in 2015. Industry funding, top pharmaceutical company affiliation, trial registration, larger authorship teams, international teams and drug trials were associated with a greater likelihood of using adequate methods. National Institutes of Health funding and university prestige were not.

**Conclusion** Even though reporting has improved since 1990, the proportion of RCTs using inadequate methods is high (59.3%) and increasing, potentially slowing progress and contributing to the reproducibility crisis. Stronger incentives for the use of adequate methods are needed.

## INTRODUCTION

The quality and reliability of biomedical research are of paramount importance to treatment decisions and patient outcomes. Flawed research conclusions can lead to poor treatment and harm patients. As much as 85% of the annual US\$265 billion spent on

## Strengths and limitations of this study

- This work combines the strengths of expert human assessments with data science techniques to build a comprehensive database on biomedical research quality, including the full-text and systematic assessment of randomised controlled trial (RCT) methods with bibliometric and funding information in a sample of 20 571 RCTs.
- The study analyses trends in methods and reporting over 25 years and identifies factors associated with biomedical research quality including funding source, first author affiliation, clinical trial registration status, study novelty, team characteristics, technology and geography.
- PubMed identifier, full-text and/or funding information were not available for all RCTs. 30.5% of RCTs (unpublished or published in journals not indexed in PubMed) did not have a PubMed identifier. 43.2% of RCTs with PubMed identifier did not have a full text available from the Harvard Library. 23.6% of included RCTs were reported in articles disclosing National Institutes of Health or industry funding. Classification of sectors relies on primary reported affiliation.
- Cochrane reviewers may have been able to obtain more information on more recent RCTs (from authors, registries or protocols rather than the primary report), suggesting some of the apparent improvement in reporting may reflect an improvement in access to study details.
- This study does not identify causal mechanisms explaining biomedical research quality.

biomedical research may be wasted due to inadequate methods.<sup>1–8</sup>

Previous scientific work aiming to evaluate the reliability of biomedical research has been limited by data and methodological issues. Data challenges included the time and resources necessary to assess methods and reporting, resulting in the use of small selected samples and/or limited information available for each scientific article evaluated in larger samples.<sup>9–28</sup> As a result, it remains unknown what is the overall magnitude of waste due to inadequate methods and reporting in biomedical research and what

factors are associated with the use of adequate vs inadequate research methods.

To address these questions, this study combines the full text of randomised controlled trials (RCTs) and systematic assessment of study methods with bibliometric and funding information in a large sample of RCTs included in 'gold-standard' systematic reviews. The study describes the evolution of adequate research methods and reporting over time. A multinomial logit model tests potential factors associated with methods and reporting, including funding source, first author affiliation, clinical trial registration status, study novelty, team characteristics, technology and geography.

## METHODS

This work combines the strengths of human expert assessments with data science techniques to build a comprehensive database on biomedical research quality, including full text, systematic assessment of study methods, bibliometric and funding information in a sample of 20 571 RCTs. Python V.3.6 and Stata V.15 were used to assemble the database and conduct the analysis.

### Data

Cochrane reviews constitute a valuable data source to assess biomedical research quality as they follow strict methods and precise reporting guidelines defined in the Cochrane Handbook.<sup>29 30</sup> This study does not involve new assessment of the methods and reporting of included RCTs, but relies entirely on the assessments available in the Cochrane reviews, which are systematically performed by two expert reviewers who compare their assessments and reach consensus on the final assessment.<sup>29</sup> The research method dimensions evaluated in Cochrane reviews include random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data and selective reporting (detailed in online supplementary table A1).<sup>31</sup>

The database assembly had seven steps: (1) All included references were extracted from each review, including PubMed identifiers, (2) all risk of bias assessments on the six dimensions of the 2011 update of the Cochrane Risk of Bias Assessment Tool (see online supplementary table A1) were extracted from each review. Each assessment included three variables: bias type (eg, random sequence generation), judgment (eg, low risk) and support for judgment (eg, computer random number generator), (3) each RCT was matched with its main published reference as identified by Cochrane reviewers, (4) PubMed records corresponding to these publications, including bibliometric information and first author affiliation, were retrieved using the E-utilities public application programming interface (API), (5) affiliation information for other authors (not available from PubMed over the study period) was retrieved from SCOPUS, (6) full text for references with PubMed identifier were retrieved from

the Harvard Library, and (7) industry funding information was extracted from the full-text.

Sector affiliation with university, government, hospital, non-profit, top pharmaceutical company or other firm, as well as geographical variables were derived from the first author affiliation address. Top 25 universities were identified using the 2007 Academic Ranking of World Universities in Clinical Medicine and Pharmacy (see online supplementary material appendix A). Firms were classified as top pharmaceutical companies or other firms using the listing of pharmaceutical companies with a revenue greater than US\$10 billion in any year since 2011 (see online supplementary material appendix B). Technologies were retrieved from the keywords and abstracts of the Cochrane Reviews. Private funding information was retrieved from the full-text of the main reference.

### Sample

Figure 1 summarises the data flow. All RCTs assessed for risk of bias after 2011 (update of the Risk of Bias Assessment Tool) and through October 2017 were considered for inclusion (63 748 RCTs included in 4195 reviews). This list of Cochrane reviews is reported in online supplementary appendix C.

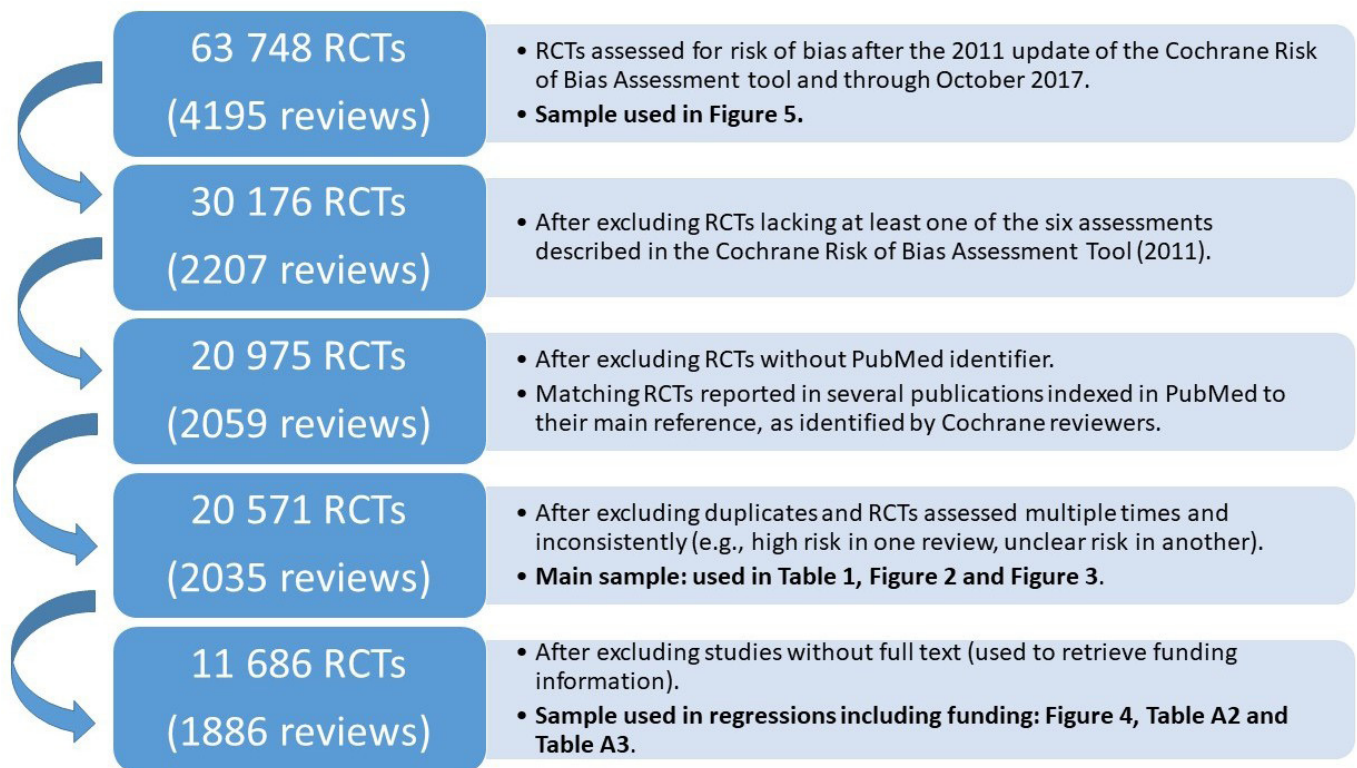
Criteria for study inclusion were: (1) the review included all six assessments (to allow comparison of the overall use of adequate methods, inadequate methods and poor reporting across reviews) (1988 reviews dropped), (2) the article reporting the study was referenced in PubMed (to allow bibliometric data to enter the analysis) (9201 RCTs dropped), (3) duplicates were removed and (4) RCTs assessed multiple times with different outcomes (eg, high risk in one review, unclear risk in another) were dropped (404 RCTs dropped).

Applying these criteria, the analysis sample for the descriptive statistics and the time series of methods included 20 571 RCTs. A full-text PDF was available from the Harvard Library for 11 686 RCTs. This subsample was needed to retrieve private funding information from the full text of the paper and constitutes the analysis sample for those regressions including funding information.

### Analysis

The outcomes were risk of bias on the six assessed methodological dimensions and RCT-level assessment of adequate methods, inadequate methods or poor reporting. The six methodological dimensions assessed included (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data and (6) selective reporting (detailed in online supplementary table A1). The category 'other bias' was not used in this study, as it includes concerns not necessarily about methods or reporting, such as conflicts of interest.

Following guidelines for assessing the quality of evidence<sup>31</sup> and previous empirical work,<sup>7</sup> the RCT-level



**Figure 1** Data flow. RCTs, randomised controlled trials.

assessment was ‘adequate methods’ if the study was at low risk of bias on all dimensions assessed. It was ‘inadequate methods’ if the study was at high risk of bias for one or more reasons. It was ‘poorly reported’ if the reviewers did not have enough information to assess whether the methods used were adequate or inadequate (if the study was at ‘unclear’ risk of bias).

Several reasons support the use of at least one high risk of bias assessment as the definition for inadequate methods. Some risk of bias domains might translate into more statistical bias than others, but empirical evidence on the relative importance of the risk of bias domains is limited, and the effect of several versus one high risk assessment on research outcomes is unknown.<sup>32 33</sup> The empirical relationship between risk of bias assessments and research outcomes (including actual statistical bias) requires further research.

There is also a theoretical reason to use at least one high risk of bias assessment as the definition of method inadequacy. Cochrane risk of bias domains can be mapped to important conditions to make RCTs valuable. If not truly randomised or if differences between the treatment and control group are introduced post randomisation, an RCT does not produce an unbiased estimate of the treatment effect.<sup>34</sup> These two conditions imply that one inadequacy in the randomisation process (non-random sequence generation or inadequate allocation concealment), or one difference introduced post randomisation between the treatment and control groups (through inadequate blinding of participants,

personnel or outcome assessors) or after the trial (due to incomplete outcome data or selective reporting) should be the default threshold for assessing methods adequacy.

Two analyses were performed. The first reports the time series of the proportion of RCTs using adequate methods, inadequate methods and poor reporting, for each dimension and in aggregate. The second tests whether adequate methods, inadequate methods and poor reporting are associated with funding source (National Institutes of Health (NIH) grant or industry funding), sector affiliation of first author (top university, other university, government, hospital, non-profit, top pharmaceutical company and other Firm), other industry affiliation, clinical trial registration status, study novelty (first or subsequent study on a particular research question), team characteristics (number of authors and international collaboration), technology (drug, device, surgery, behavioural intervention or other intervention) and geography of first author (Canada, Europe, UK, USA or other country). A multinomial logit model using these variables predicts overall adequate methods, inadequate methods and poor reporting, as well as risk of bias along each dimension assessed.

#### Patient involvement

Patients were not involved in any aspect of the study design, conduct or in the development of the research question or outcome measures. As a meta-research study

based on existing published research, there was no patient recruitment for data collection.

## RESULTS

### Prevalence of adequate methods, inadequate methods and poor reporting

**Table 1** presents descriptive statistics. Only 5.7% of RCTs used adequate methods on all six dimensions (n=1173). 59.3% used inadequate methods on at least one dimension (n=12 190) and 35.0% were poorly reported (n=7208).

**Figure 2** shows the proportion of RCTs at low, high or unclear risk of bias for random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data and selective reporting, for all RCTs assessed on all six dimensions (n=20 571). Thirty-eight per cent of trials used inadequate methods for blinding of participants and personnel. A total of 15%–20% of trials used inadequate methods for blinding of outcome assessment (20%), incomplete outcome data (19%) and selective reporting (15%). The proportion of trials using inadequate methods for random sequence generation and allocation concealment was lowest (respectively, 5% and 7%), but these two dimensions were frequently poorly reported (respectively, 47% and 58% of trials).

### Methods and reporting over time

**Figure 3** shows the overall proportion of RCTs using adequate methods, inadequate methods and poorly reported methods by year of publication. The proportion of poorly reported RCTs decreased, five percentage points per decade, from 42.5% in 1990 to 30.2% in 2015. The proportion of RCTs using adequate methods increased linearly, three percentage points per decade, from 2.6% in 1990 to 10.3% in 2015. The proportion of RCTs using inadequate methods increased from 54.9% in 1990 to 59.5% in 2015.

Reporting improved on all dimensions. The proportion of RCTs using adequate methods for random sequence generation, allocation concealment, blinding of outcome assessment, incomplete outcome data and selective reporting increased. However, the proportion of trials using inadequate methods for blinding of participants and personnel increased.

**Figure 4** provides graphs similar to **figure 3** for all RCTs assessed on at least one dimension (n=63 748). Similar patterns suggest that the evolution over time observed for the RCTs assessed on all dimensions (n=20 571) reflects the evolution over time in all RCTs assessed on at least one dimension.

### Factors associated with methods and reporting

**Figure 5** reports regression results from a multinomial logit model predicting overall quality. Online supplementary tables A2 and A3 report all regression results.

Public funding was not associated with the overall use of adequate methods. However, NIH-funded RCTs were less likely to use inadequate methods for random sequence

generation (RR=0.29,  $p<0.001$ ) and allocation concealment (RR=0.51,  $p<0.001$ ). Industry-funded RCTs were slightly more likely to use adequate methods (RR=0.84,  $p<0.05$ ), because of better blinding of participants and personnel (RR=0.87,  $p<0.05$ ).

First author affiliation with a top pharmaceutical company was associated with increased use of adequate methods (RR=0.43,  $p<0.01$ ). First author affiliation with a top university was not.

Registered trials (RR=0.42,  $p<0.001$ ), larger authorship teams (RR=0.95,  $p<0.001$ ), international teams (RR=0.51,  $p<0.01$ ) and RCTs on drugs (RR=0.50,  $p<0.001$ ) were less likely to use inadequate methods. RCTs on medical devices were more likely to use inadequate methods (RR=1.71,  $p<0.01$ ).

## DISCUSSION

In 1951, the first review assessing the quality of clinical trials found that only 27 of 100 were well controlled.<sup>35 36</sup>

Since, a steady stream of scholarly work periodically voiced concerns about the quality of medical research.<sup>1–8 37 38</sup>

Recent medical reversals<sup>39</sup> and the reproducibility crisis<sup>40</sup> have sharpened focus on medical research quality. Newly available large scale data and data science techniques provide powerful tools to measure the overall magnitude of the problem, investigate its determinants and provide an evidence base to inform the design and evaluation of future interventions. This study assessed whether methods and reporting improved over time and identified the characteristics of better and worse RCTs.

This study has six main results. First, in a large sample of RCTs assessed in systematic reviews, only 5.7% used adequate methods, 59.3% used inadequate methods, and 35.0% were poorly reported. Since the 1990s, reporting has improved. But in parallel with this improvement in reporting, the proportion of trials using both adequate and inadequate methods has increased.

The overall proportion of poorly reported trials decreased by about five percentage points per decade. This is good news but much remains to be done. At the current rate of improvement, it would take 50 years for 95% of RCTs to be adequately reported. These results are consistent with previous research finding improvements in reporting in several clinical areas such as physiotherapy<sup>10</sup> and dentistry.<sup>26</sup> The trends for each dimension assessed separately are also very similar to those found in another large sample of RCTs.<sup>25</sup>

This improvement in reporting happened over a period of time when the Consolidated Standards of Reporting Trials (CONSORT) statement, a minimum set of evidence-based reporting recommendations, and other initiatives, such as the Enhancing the QUALity and Transparency Of health Research (EQUATOR) Network, developed to improve reporting practices.<sup>41–45</sup> Since the 1990s, the CONSORT statement has been endorsed by over 50% of the core clinical journals indexed in PubMed and may improve reporting of RCTs they publish.<sup>46</sup>

**Table 1** Descriptive statistics

	All	Adequate	Inadequate methods	Poor reporting
Sample 1, (%)	20571 (100)	1173 (5.7)	12 190 (59.3)	7208 (35.0)
Sample 2 (with full text)	11 686 (56.8)	833 (7.1)	6783 (58.0)	4070 (34.8)
<b>Funder type, (%)</b>				
NIH grant	2147 (10.4)	146 (6.8)	1282 (59.7)	719 (33.5)
Industry funding	2725 (13.2)	283 (10.2)	1464 (52.6)	978 (35.1)
<b>First author affiliation, (%)</b>				
Top university	1063 (5.2)	51 (4.8)	601 (56.5)	411 (38.7)
Other university	11 120 (54.1)	677 (6.1)	6589 (59.3)	3854 (34.7)
Hospital	4450 (21.6)	185 (4.2)	2608 (58.6)	1657 (37.2)
Government	1744 (8.5)	108 (6.2)	1071 (61.4)	565 (32.4)
Non-profit	751 (3.7)	48 (6.4)	454 (60.5)	249 (33.2)
Top pharma	239 (1.2)	26 (10.9)	115 (48.1)	98 (41.0)
Other firm	195 (1.0)	13 (6.7)	115 (59.0)	67 (34.3)
Other research institution	200 (1.0)	18 (9.0)	120 (60.0)	62 (31.0)
Other industry affiliation	570 (2.8)	44 (7.7)	287 (50.4)	239 (41.9)
Registered RCTs (NCT), (%)	1888 (9.2)	298 (15.8)	1011 (53.6)	579 (30.7)
<b>Novelty, (%)</b>				
First study	2284 (11.1)	126 (5.5)	1390 (60.9)	768 (33.6)
Second study	2124 (10.3)	127 (6.0)	1262 (59.4)	735 (34.6)
<b>Team characteristics</b>				
No of authors—avg (Std)	6.15 (3.9)	8.04 (5.5)	5.99 (3.8)	6.13 (6.8)
International, (%)	748 (3.6)	60 (8.0)	379 (50.7)	309 (41.3)
<b>Technology*, (%)</b>				
Drug	13 485 (65.6)	914 (6.8)	7306 (54.2)	5265 (39.0)
Device	5347 (26.0)	235 (4.4)	3366 (63.0)	1746 (32.7)
Procedure	8710 (42.3)	460 (5.3)	4925 (56.5)	3325 (38.2)
Behavioural	4543 (22.1)	122 (2.7)	3239 (71.3)	1182 (26.0)
Other	1199 (5.8)	78 (6.5)	819 (68.3)	302 (25.2)
<b>Geography†, (%)</b>				
Canada	680 (3.3)	61 (9.0)	362 (53.2)	257 (37.8)
Europe	4467 (21.7)	254 (5.7)	2693 (60.3)	1520 (34.0)
UK	2306 (11.2)	154 (6.7)	1399 (60.7)	753 (32.7)
USA	4465 (21.7)	284 (6.4)	2592 (58.1)	1589 (35.6)
Other	4165 (20.3)	253 (6.1)	2444 (58.7)	1468 (35.3)
Publication year—avg (Std)	2001 (10.2)	2005 (8.1)	2001 (10.4)	2001 (9.9)
Study age at review—avg (Std)	13.44 (10.1)	9.81 (8.0)	13.39 (10.3)	14.14 (9.9)

Unless otherwise specified, column 1 reports the number of RCTs and their proportion as of the total number of RCTs ( $n=20571$ ). An RCT uses adequate methods if it is at 'low risk of bias' on all six dimensions assessed (see online supplementary table A1). Methods are inadequate if an RCT is at 'high risk of bias' for at least one reason. Methods are poorly reported if there is no evidence of methods inadequacy, but at least one assessment is 'unclear risk of bias'. Columns 2–4 report the number of RCTs in each category and their proportion as of the number of RCTs in column 1. For number of authors, publication year and study age at time of review, table 1 reports the average and standard deviation.

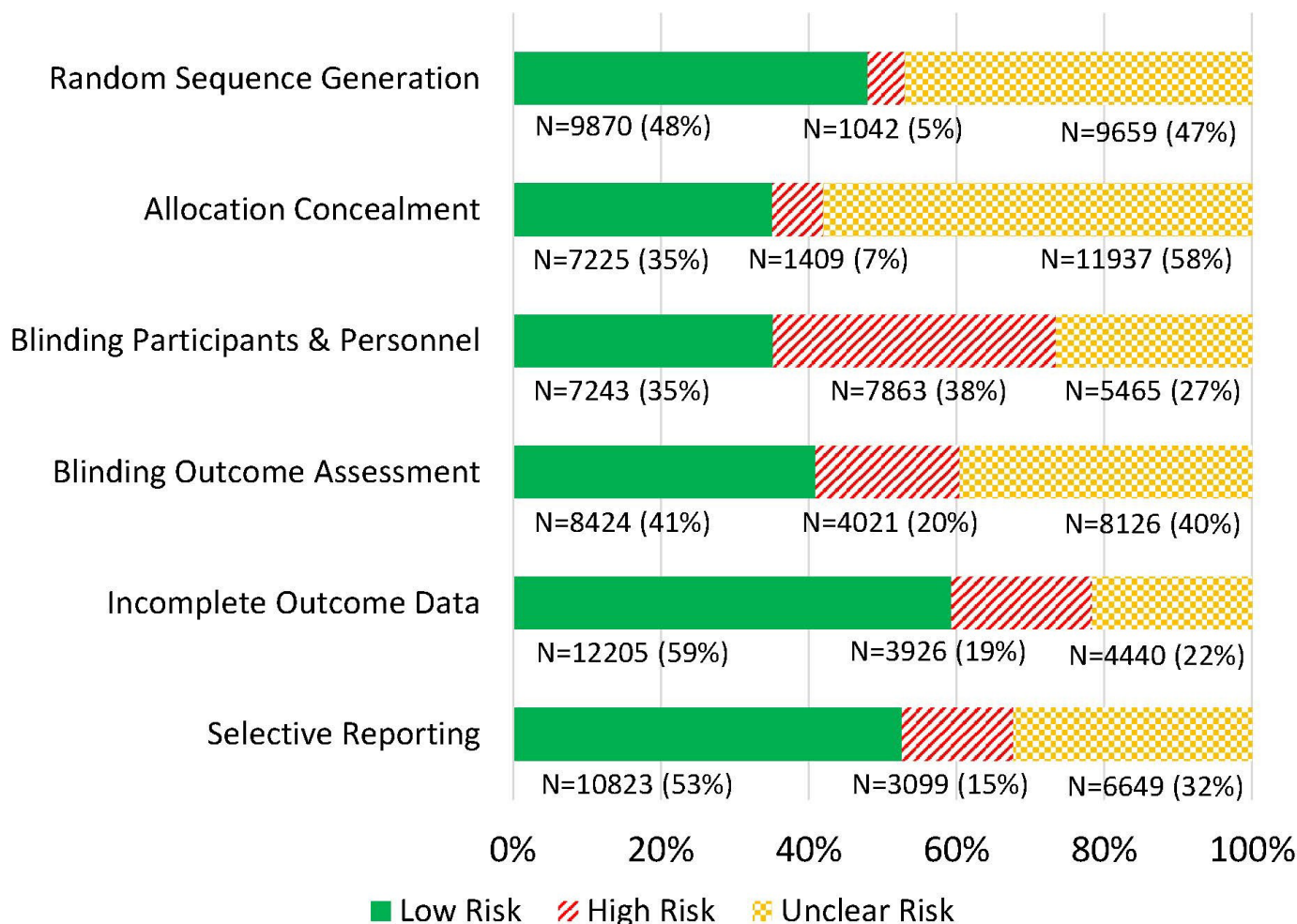
\*One RCT can belong to several technology categories.

†For some RCTs, affiliation address is not provided.

NCT, National clinical Trial number in ClinicalTrials.gov; NIH, National Institutes of Health; RCT, randomised controlled trial.

Spurred by the CONSORT statement, the EQUATOR Network was launched in 2008 in the UK to improve the reliability of medical publications by promoting

transparent and accurate reporting of health research.<sup>47</sup> Since, it has developed into a global initiative aiming to improve research reporting worldwide.<sup>36</sup>



**Figure 2** Number and proportion of RCTs at low risk of bias, high risk of bias and unclear risk of bias (N=20 571). RCTs, randomised controlled trials.

In parallel with this improvement in reporting, the proportion of trials using both adequate and inadequate methods has increased. The linear increase in the proportion of RCTs using adequate methods is heartening. However, improvement in the use of adequate methods is even slower than improvement in reporting. At the current rate of improvement (three percentage points per decade), it would take more than a century for half the RCTs to use adequate methods. This finding is consistent with previous empirical results in small samples,<sup>23</sup> but contrasts with research in larger samples analysing each methodological dimension separately to conclude that methods improved over time.<sup>25</sup>

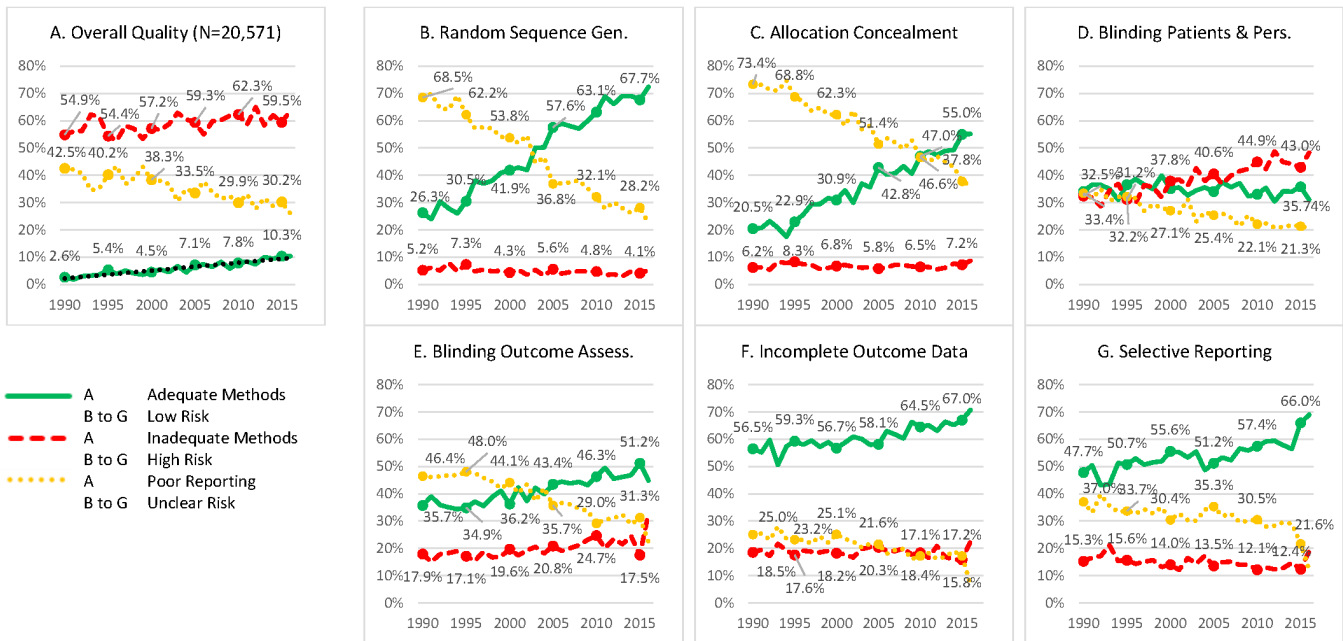
Second, NIH-funded RCTs were not more likely to use adequate methods. This is surprising given the rigorous grant application process, shown to select better scientific proposals,<sup>48</sup> and the public stakes in the reliability of publicly funded research.<sup>49</sup> Notably, the efforts of the NIH to address the reproducibility crisis began just at the end of the study period.<sup>50</sup>

Third, top pharmaceutical company affiliation was significantly associated with better methods. Affiliation with other companies was not. Heterogeneity across firms may explain inconsistency of previous research on

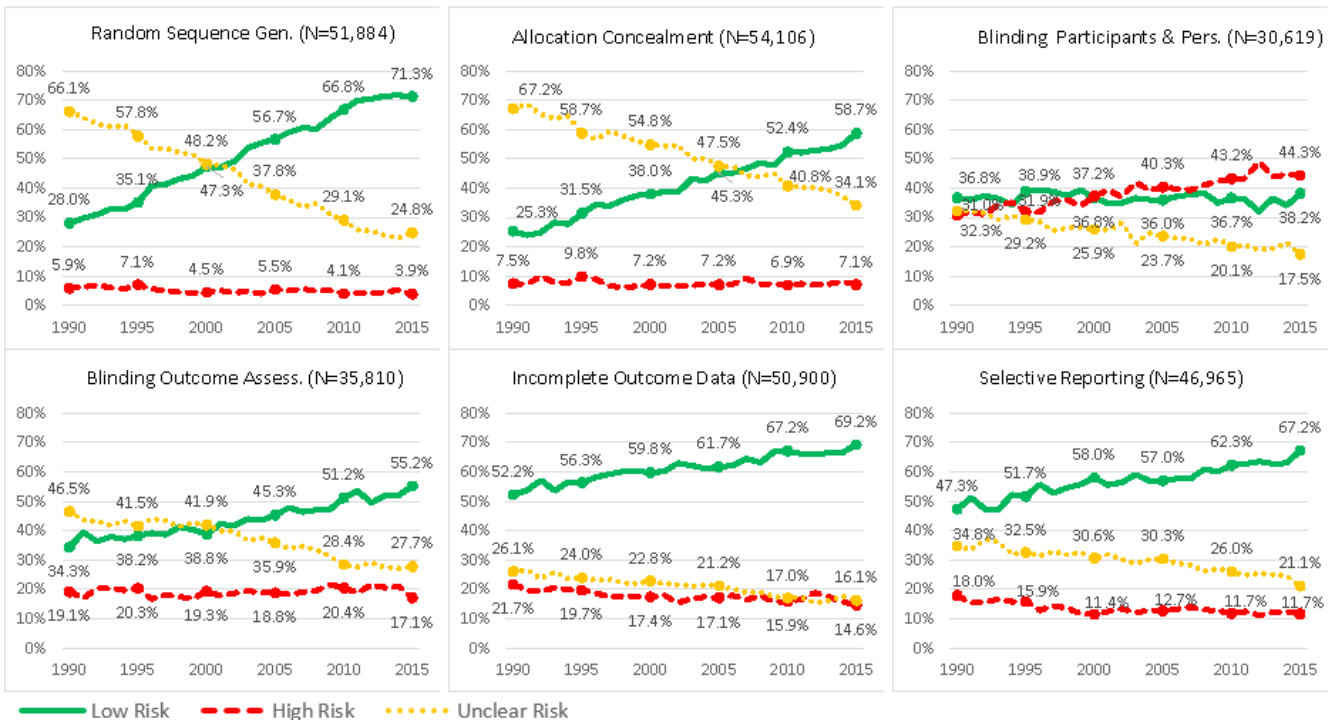
the effect of industry funding or affiliation on research methods and outcomes.<sup>28 51</sup>

Fourth, university prestige was not associated with greater use of adequate methods. The current scientific reward system focuses on numbers of publications and citations rather than the assessment of research methods.<sup>52</sup> The resulting incentives affect both scientists and institutions, as through the allocation of grant funding.<sup>53 54</sup> Thus, in a climate of hypercompetition,<sup>55</sup> the use of adequate methods and reporting might yield little reward while exposing scientists to better informed scrutiny.

Fifth, team size and international collaboration were associated with greater use of adequate methods. Increasing the number of authors by one was associated with a small, but highly significant improvement in methods and reporting. Many RCTs are published by large teams so it is not surprising that the effect of one additional author was small. But this effect was also highly significant, consistent with previous research finding that larger teams and international teams produce more frequently cited research.<sup>56 57</sup> Team characteristics associated with performance in other settings open avenues for future research.<sup>58 59</sup>



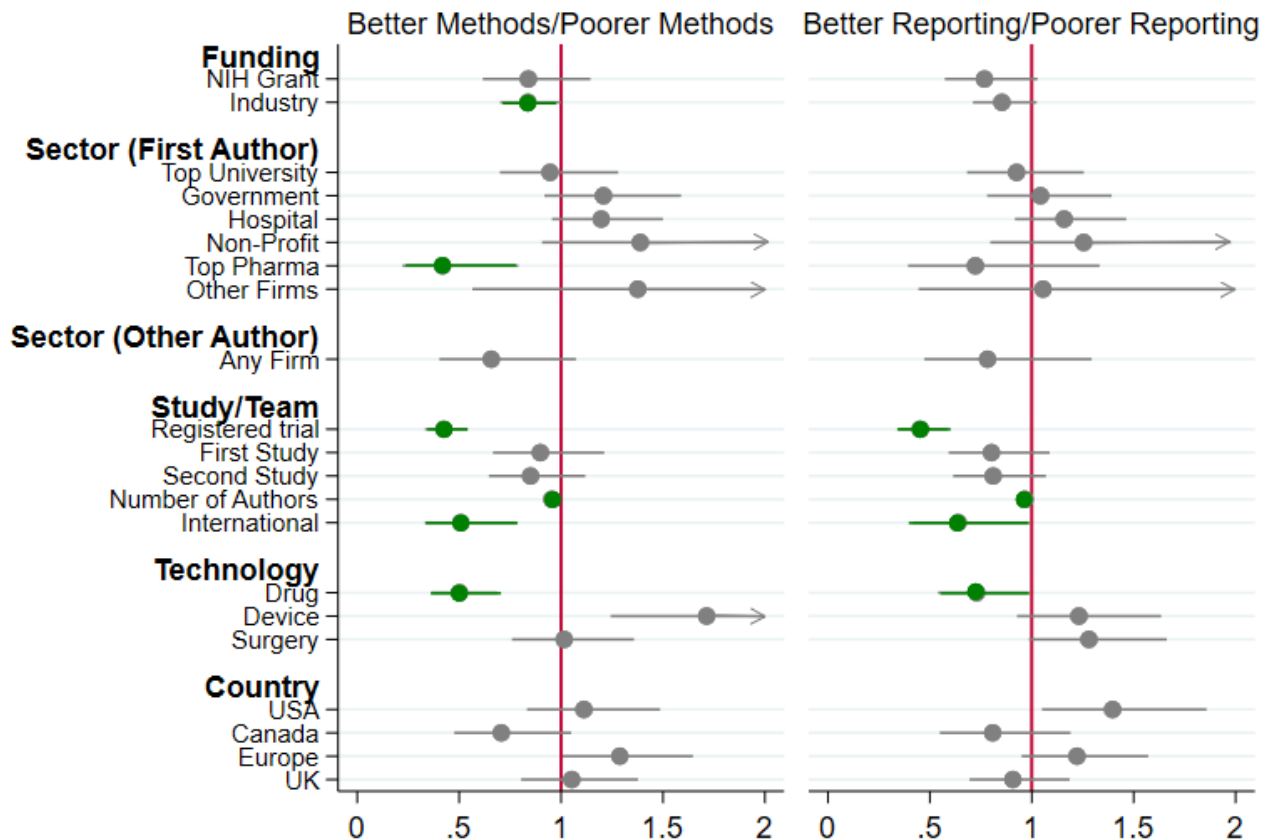
**Figure 3** Evolution of methods and reporting over time. (A) Proportion of RCTs using adequate methods, inadequate methods and poorly reported. (B–G) Proportion of RCTs at low risk of bias, high risk of bias and unclear risk of bias for each dimension assessed. See online supplementary table A1 for the definition of each dimension. N=20 571 RCTs. An observation is an RCT assessed on all six dimensions. See figure 1 for more detailed information about the sample. RCTs, randomised controlled trials.



**Figure 4** Same figure as figure 3, but including all RCTs assessed on at least one dimension (as opposed to RCTs assessed on all six dimensions). Evolution of methods and reporting over time. (A) Proportion of RCTs using adequate methods, inadequate methods and poorly reported. (B–G) Proportion of RCTs at low risk of bias, high risk of bias and unclear risk of bias for each dimension assessed. See online supplementary table A1 for the definition of each dimension. RCTs, randomised controlled trials.

## Factors Associated with Better Methods and Reporting

Multinomial Logit Model Reference Category: Adequate Methods, Relative Risk Ratios (RRR)



**Figure 5** Main regression results (relative risk ratios and 95% CIs) from the multinomial logit model predicting overall RCT quality. The arrow heads on the CIs indicate that the upper bound of the 95% CI is greater than 2. The dependent variable is a categorical variable and can take three values: adequate methods, inadequate methods and poor reporting. Adequate methods is the reference outcome category. The regression sample includes 11 686 RCTs with accessible full text. See figure 1 for more detailed information about the sample. The relative risk ratios represent the likelihood of an RCT with specific funding, sector, study/team, technology and country characteristics using inadequate methods (or being poorly reported), as compared with the likelihood of an RCT in a reference group without these characteristics using inadequate methods (or being poorly reported). In the regression, sector, technology and country are categorical variables. The omitted category for sector is other university. The omitted category for technology is other interventions. The omitted category for country is other countries. The regression includes topic and year fixed effects. The regression coefficients are reported in online supplementary table A2. Regression results predicting relative risk ratios for high or unclear risk of bias on each dimension assessed (as opposed to overall quality) are reported in online supplementary table A3. NIH, National Institutes of Health; RCT, randomised controlled trial.

Finally, RCTs on drugs were more likely to use adequate methods than RCTs on other interventions, while RCTs on devices were more likely to use inadequate methods. In many countries, trials on drugs are more tightly regulated than trials on devices. In the USA, under the Federal Food, Drug and Cosmetic Act (1938), drugs and devices face different premarket review and postmarket compliance requirements. The finding is also consistent with specific barriers to the conduct of RCTs on medical devices, in particular for randomisation and blinding, and with the lack of scientific advice and regulations for medical device trials.<sup>60</sup> RCTs on drugs were using better methods and reporting than RCTs on other interventions,

but much remains to be done. This finding is consistent with previous work showing that even RCTs used in the drug approval process frequently use inadequate methods and reporting.<sup>61</sup>

Future research should carefully evaluate the effect of method adequacy on research outcomes, and identify successful strategies and incentives to accelerate the diffusion of good reporting practices and the adoption of adequate methods. Given the size of the medical research industry and its effect on human lives, successful evidence based policies could have tremendous impact.



## Limitations

PubMed identifier, full-text and/or funding information were not available for all RCTs. 30.5% of RCTs (unpublished or published in journals not indexed in PubMed) did not have a PubMed identifier. 43.2% of RCTs with PubMed identifier did not have a full text available from the Harvard Library. 23.6% of included RCTs were reported in articles disclosing NIH or industry funding. Classification of sectors relies on primary reported affiliation. This paper does not identify causal mechanisms explaining biomedical research quality.

Cochrane reviewers may have been able to obtain more information on more recent RCTs (from authors, registries or protocols rather than the primary report), suggesting that some of the apparent improvement in reporting may in fact be an improvement in access to study details.

## CONCLUSION

Even though reporting has improved since 1990, the proportion of RCTs using inadequate methods is high (59.3%) and increasing, potentially slowing progress and contributing to the reproducibility crisis. Stronger incentives for the use of adequate methods are needed.

**Acknowledgements** The author thanks David Cutler, Richard Freeman, Mack Lipkin, Tim Simcoe, Ariel Stern, Griffin Weber, Richard Zeckhauser and the participants at the NBER-IFS International Network on the Value of Medical Research meetings for helpful conversations and feedback, and Harvard Business School Research Computing Services for technical advice and support. The author also thanks the editors and the two referees, Paul Glasziou and Simon Gandevia, for their most helpful and generous feedback and clear guidance.

**Contributors** MC designed the study, performed the analysis, interpreted the results, wrote the manuscript and approved the final version to be published. MC accepts full responsibility for the work and the conduct of the study, had access to the data and controlled the decision to publish.

**Funding** MC gratefully acknowledges support by the National Institute on Aging of the National Institutes of Health under Award Number R24AG048059 to the National Bureau of Economic Research (NBER).

**Disclaimer** The content of this article is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health or the NBER.

**Competing interests** MC has completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declares: the author reports grants from the National Institute on Aging of the National Institutes of Health during the conduct of the study.

**Patient consent for publication** Not required.

**Ethics approval** This is a meta-research study.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** All data sources necessary to reproduce the analysis are described in the main text or the online supplementary material. No additional data available.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *The Lancet* 2009;374:86–9.
- Macleod MR, Michie S, Roberts I, *et al*. Biomedical research: increasing value, reducing waste. *The Lancet* 2014;383:101–4.
- Chalmers I, Bracken MB, Djulbegovic B, *et al*. How to increase value and reduce waste when research priorities are set. *The Lancet* 2014;383:156–65.
- Glasziou P, Altman DG, Bossuyt P, *et al*. Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet* 2014;383:267–76.
- Ioannidis JPA, Greenland S, Hlatky MA, *et al*. Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet* 2014;383:166–75.
- Moses H, Matheson DHM, Cairns-Smith S, *et al*. The anatomy of medical research: US and international comparisons. *JAMA* 2015;313:174–89.
- Yordanov Y, Dechartres A, Porcher R, *et al*. Avoidable waste of research related to inadequate methods in clinical trials. *BMJ* 2015;350:h809.
- Moher D, Glasziou P, Chalmers I, *et al*. Increasing value and reducing waste in biomedical research: who's listening? *The Lancet* 2016;387:1573–86.
- Lee KP, Schotland M, Bacchetti P. Association of Journal quality indicators with methodological quality of clinical research articles. *JAMA* 2002;287:2805–8.
- Moseley AM, Herbert RD, Maher CG, *et al*. Reported quality of randomized controlled trials of physiotherapy interventions has improved over time. *J Clin Epidemiol* 2011;64:594–601.
- Bala MM, Akl EA, Sun X, *et al*. Randomized trials published in higher vs. lower impact journals differ in design, conduct, and analysis. *J Clin Epidemiol* 2013;66:286–95.
- Lee S-Y, Teoh PJ, Camm CF, *et al*. Compliance of randomized controlled trials in trauma surgery with the CONSORT statement. *J Trauma Acute Care Surg* 2013;75:562–72.
- Agha RA, Camm CF, Doganay E, *et al*. Randomised controlled trials in plastic surgery: a systematic review of reporting quality. *Eur J Plast Surg* 2014;37:55–62.
- Chen B, Liu J, Zhang C, *et al*. A retrospective survey of quality of reporting on randomized controlled trials of metformin for polycystic ovary syndrome. *Trials* 2014;15:128.
- Chen X, Zhai X, Wang X, *et al*. Methodological reporting quality of randomized controlled trials in three spine journals from 2010 to 2012. *Eur Spine J* 2014;23:1606–11.
- Kim KH, Kang JW, Lee MS, *et al*. Assessment of the quality of reporting for treatment components in Cochrane reviews of acupuncture. *BMJ Open* 2014;4:e004136.
- Lempesi E, Koletsi D, Fleming PS, *et al*. The reporting quality of randomized controlled trials in orthodontics. *J Evid Based Dent Pract* 2014;14:46–52.
- Yao AC, Khajuria A, Camm CF, *et al*. The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. *Eye* 2014;28:1341–9.
- Zhuang L, He J, Zhuang X, *et al*. Quality of reporting on randomized controlled trials of acupuncture for stroke rehabilitation. *BMC Complement Altern Med* 2014;14:151.
- Chen Z, Chen Y, Zeng J, *et al*. Quality of randomized controlled trials reporting in the treatment of melasma conducted in China. *Trials* 2015;16:156.
- Glujovsky D, Boggino C, Riestra B, *et al*. Quality of reporting in infertility journals. *Fertil Steril* 2015;103:236–41.
- Kloukos D, Papageorgiou SN, Doulis I, *et al*. Reporting quality of randomised controlled trials published in prosthodontic and implantology journals. *J Oral Rehabil* 2015;42:914–25.
- Revez L, Chapman E, Asial S, *et al*. Risk of bias of randomized trials over time. *J Clin Epidemiol* 2015;68:1036–45.
- Zhai X, Wang Y, Mu Q, *et al*. Methodological reporting quality of randomized controlled trials in 3 leading diabetes journals from 2011 to 2013 following consort statement: a system review. *Medicine* 2015;94:e1083.
- Dechartres A, Trinquart L, Atal I, *et al*. Evolution of poor reporting and inadequate methods over time in 20920 randomised controlled trials included in Cochrane reviews: research on research study. *BMJ* 2017;357:j2490.
- Saltaji H, Armijo-Olivo S, Cummings GG, *et al*. Randomized clinical trials in dentistry: risks of bias, risks of random errors, reporting quality, and methodologic quality over the years 1955–2013. *PLoS One* 2017;12:e0190089.
- Karlsen APH, Dahl JB, Mathiesen O. Evolution of bias and sample size in postoperative pain management trials after hip and knee arthroplasty. *Acta Anaesthesiol Scand* 2018;62:666–76.

28. Salandra R. Knowledge dissemination in clinical trials: exploring influences of institutional support and type of innovation on selective reporting. *Res Policy* 2018;47:1215–28.
29. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions [updated March 2011]. The Cochrane Collaboration, 2011. Available: <http://handbook.cochrane.org/>
30. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med* 2016;13:e1002028.
31. Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
32. Savović J, Turner RM, Mawdsley D, et al. Association between risk-of-bias assessments and results of randomized trials in Cochrane reviews: the ROBES meta-epidemiologic study. *Am J Epidemiol* 2018;187:1113–22.
33. Rhodes KM, Turner RM, Savović J, et al. Between-trial heterogeneity in meta-analyses may be partially explained by reported design characteristics. *J Clin Epidemiol* 2018;95:45–54.
34. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Soc Sci Med* 2018;210:2–21.
35. Ross OB. Use of controls in medical research. *J Am Med Assoc* 1951;145:72–5.
36. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR network. *J R Soc Med* 2016;109:67–77.
37. DerSimonian R, Charette LJ, McPeck B, et al. Reporting on methods in clinical trials. *N Engl J Med* 1982;306:1332–7.
38. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. *N Engl J Med* 1987;317:426–32.
39. Prasad V, Vandross A, Toomey C, et al. A decade of reversal: an analysis of 146 contradicted medical practices. In: *In Mayo clinic proceedings*. Elsevier, 2013: 88. 790–8.
40. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature* 2016;533:452–4.
41. Begg C, Cho M, Eastwood S. Improving the quality of reporting of randomized controlled trials. *JAMA* 1996;276:637–9.
42. Moher D, Schulz KF, Altman DG, et al. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *BMC Med Res Methodol* 2001;1:2.
43. Schulz KF, Altman DG, Moher D. Consort 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010;8.
44. Altman DG, Schulz KF, Moher D, et al. The revised consort statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663–94.
45. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:e1–37.
46. Turner L, Shamseer L, Altman DG, et al. Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. *Syst Rev* 2012;1:60.
47. Altman DG, Simera I, Hoey J, et al. EQUATOR: reporting guidelines for health research. *The Lancet* 2008;371:1149–50.
48. Li D, Agha L. Research funding. big names or big ideas: do peer-review panels select the best science proposals? *Science* 2015;348:434–8.
49. Sampat BN. Mission-oriented biomedical research at the NIH. *Res Policy* 2012;41:1729–41.
50. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature* 2014;505:612–3.
51. Lundh A, Sismondo S, Lexchin J, et al. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2012;12.
52. Stephan PE. *How economics shapes science*. Cambridge, MA: Harvard University Press, 2012: 228–41.
53. Ali MM, Bhattacharyya P, Olejniczak AJ. The effects of scholarly productivity and institutional characteristics on the distribution of federal research grants. *J Higher Educ* 2010;81:164–78.
54. Alberts B, Kirschner MW, Tilghman S, et al. Rescuing us biomedical research from its systemic flaws. *Proc Natl Acad Sci U S A* 2014;111:5773–7.
55. Edwards MA, Roy S. Academic research in the 21st century: maintaining scientific integrity in a climate of perverse incentives and Hypercompetition. *Environ Eng Sci* 2017;34:51–61.
56. Wuchty S, Jones BF, Uzzi B. The increasing dominance of teams in production of knowledge. *Science* 2007;316:1036–9.
57. National Research Council. *Enhancing the effectiveness of team science*. National Academies Press, 2015.
58. Huckman RS, Staats BR, Upton DM. Team Familiarity, role experience, and performance: evidence from Indian software services. *Manage Sci* 2009;55:85–100.
59. Huckman RS, Staats BR. Fluid tasks and fluid teams: the impact of diversity in experience and team familiarity on team performance. *Manuf Serv Oper Manag* 2011;13:310–28.
60. Neugebauer EAM, Rath A, Antoine S-L, et al. Specific barriers to the conduct of randomised clinical trials on medical devices. *Trials* 2017;18:427.
61. Downing NS, Aminawung JA, Shah ND, et al. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005–2012. *JAMA* 2014;311:368–77.