Incentives for Bad Science:

How Inadequate Methods Affect Experimental Results and Publication Outcomes of Randomized Controlled Trials

Please click here for the most recent version

Maryaline Catillon*

January 20, 2020

Abstract

Randomized controlled trials (RCTs) inform medical practice, health care delivery, follow-on research, regulation, and health policy. Yet, many RCTs are inadequately randomized, blinded, and reported. To analyze scientists' and firms' incentives to meet clinical trial standards, I assemble a detailed database on research methods, experimental results, bibliometric information, and funding for 23,321 RCTs published between 1990 and 2018. I estimate the impact of meeting scientific standards on three outcomes: (1) the direction and significance of experimental results; (2) the impact factor of the publishing journal; and (3) the number of citations the publication receives. I find that increasing numbers of inadequacies increase the probability of finding support for product adoption by 7% per inadequacy, but decrease journal prestige and citations. Publication bias and strategic non-disclosure do not appear to drive the results. I conclude that individual scientists benefit marginally from higher quality research, but that pharmaceutical companies lack strong incentives to drive improvement in trial quality.

Keywords: Innovation, Science Policy, Research and Development, Biomedical Research, Pharmaceutical Industry

*Harvard Business School, Boston, Massachusetts, 02163 (mcatillon@hbs.edu). I am grateful to David Cutler, Richard Freeman, Ariel Stern, Jay Bhattacharya, Tim Simcoe, Richard Zeckhauser, and to the participants in the Harvard Health Policy Seminar, the Harvard Business School Science Based Business Initiative Seminar, the Columbia University Health Policy and Management Seminar, the NYU School of Medicine Seminar, the Annual Conference of the American Society of Health Economists, the INFORMS Healthcare Conference, and the INFORMS Annual Meeting for helpful feedback and suggestions.

1. Introduction

Valid and rigorous randomized controlled trials (RCTs) are critical to medical practice, and to managers and policy makers that fund research, make health care delivery and investment decisions, and design regulation. Results from RCTs inform medical treatment decisions, medical education, insurance design and coverage, health policy, and follow-on research. Invalid or unreliable science generates major societal costs through uninformative or even wrong conclusions, irreproducible results, suboptimal or harmful decisions, and slowed innovation (Freedman et al., 2015; Baker, 2016; Ioannidis, 2005, 2016; Zarin et al., 2019). Yet less than 25% of drug trials meet scientific standards intended to minimize the risk of bias[1] (Chalmers and Glasziou, 2009; Catillon, 2019; Zarin et al., 2019).

The disconnect between the high societal value of reliable science and the high prevalence of inadequate methods is not only due to lack of knowledge, infeasibility or the cost of meeting standards in drug trials. Meeting standards would have been feasible in 96% of inadequately conducted trials, and would have been "easy adjustments with no or minor cost" in 50% of the cases (Yordanov et al., 2015). Top university researchers are not more likely to meet standards than other researchers (Catillon, 2019). One theory is that scientists' and firms' incentives contribute to the high prevalence of inadequate methods in drug trials (Lacetera and Zirulia, 2011; Di Tillio et al., 2017; Gall et al., 2017; Kiri et al., 2018; Lindner et al., 2018). In this paper, I empirically investigate scientists', and firms' incentives to adequately randomize, blind and report drug trials.

To evaluate academic and industry scientists' payoffs for compliance with scientific standards, I build on prior literature in economics, management, and sociology, suggesting that scientists value the intrinsic reward of doing science, academic prestige, and monetary rewards (Stern, 2004; Stephan, 2012; Bikard et al., 2015). Scholars in sociology and economics argue that academic scientists value the intrisic reward of problem solving and seek to increase the rate of production and diffusion of scientific knowledge (Merton, 1973; Dasgupta and David, 1994; David, 1998). Research in the management of science

[1]Risk of bias, defined as the risk of "a systematic error or deviation from the truth, in results or inferences," is interchangeable with internal validity, defined as "the extent to which the design and conduct of a study are likely to have prevented bias" (Viswanathan et al., 2012).

suggests that industry scientists primarily serve the private purpose of their firm though they also value their participation in science[2] (Stern, 2004; Roach and Sauermann, 2010; Sauermann and Stephan, 2013).

Research in economics and management has also established that scientists value recognition and academic prestige, often measured using journal impact factor and citations. In academia, prestige translates into promotion, tenure and funding (Dasgupta and David, 1994; Judge et al., 2007). In industry, academic prestige and scientific publication in top journals are valued as they allow researchers to maintain valuable connections with the academic community and facilitate knowledge flows between academia and industry (Cockburn and Henderson, 1998; Stern, 2004).

Moreover, scholars have documented that scientists respond to monetary incentives. In academia, prestige is just one driver of income differences across scientists. Some scientists receive rewards contingent on research results, such as royalties from patents, equity in start-ups, consulting fees, or research sponsorship (Stephan, 2012). In industry, publication in academic journals helps scientists elicit favorable assessments from outside parties that position products in the marketplace (Azoulay, 2002; Polidoro Jr and Theeke, 2012; Simeth and Cincera, 2015; Salandra, 2018).

This study estimates academic and industry scientists' payoffs for meeting methodological standards in clinical trials along two dimensions: science (experimental results) and academic prestige (journal impact factor and citations), and considers monetary implications for scientists and firms. I assemble a novel dataset of 23,321 RCTs of drugs published between 1990 and 2018, and included in Cochrane Reviews[3], which are widely used "gold standard" systematic reviews in clinical medicine. Each Cochrane Review addresses a specific research question and exhaustively identifies, assesses and summarizes the findings of all available RCTs on the topic. Expert human reviewers systematically assess the methodological quality each of these RCTs using pre-specified clearly defined methodological standards.

[2]Industry scientists "pay" to be scientists, Stern (2004) shows that there is a negative relationship between wages and science.

[3]The background section provides institutional information about these systematic reviews and explains their potential for research in economics and management. The data section provides further details on the program I created in Python 3.6. to automate the data collection.

My empirical strategy relies on a comparison of trials that use adequate versus inadequate methods within narrowly-defined groups. Trials on the same topic (e.g., beta-blockers for hypertension), compare the effect of the same drug(s) (e.g., a beta-blocker, such as atenolol) to the same comparator(s) (e.g., a placebo), on the same outcome(s) (e.g., all-cause mortality, cardiovascular events, and stroke). I estimate the effect of research quality on (1) scientific results, (2) journal placement, and (3) citations.

The study has two key findings. First, compliance with scientific standards shapes scientific results: inadequate methods do not merely generate statistical noise, they generate bias and increase the probability of a positive result[4], such as finding that a drug significantly reduces death rates compared to a placebo. The relationship between methods and results does not appear to be driven by publication bias or strategic non-disclosure: it did not change with the introduction of pre-registration and reporting requirements for RCTs in the early 21st century and does not depend on government versus industry sponsorship.

Second, good science yields higher academic prestige in publication outcomes, considering both journal placement and citations. Each inadequacy in a RCT is associated with an approximately 10% penalty in impact factor[5] in the study's publication outlet compared to a more adequate RCT on the same topic. Journals are more lenient with inadequacies from prestigious authors, but not with inadequacies in more novel trials. The relationship between inadequacies and scientific impact in publication outcomes goes beyond the editorial and peer-review process selecting RCTs using better methods into better journals. Even conditional on journal impact factor, RCTs with more inadequacies receive fewer citations.[6] However, inadequacies have no effect on net citations to positive results, a citation weighted measure of scientific evidence previously found to affect drug sales (Azoulay, 2002), and generate slightly fewer citations to negative results.

[4]I define a "positive result" in the following way: the outcome in the treatment group is significantly superior to the outcome in the control group (e.g., lower death rate in the treatment group receiving a drug compared to the control group receiving a placebo).

[5]Journal impact factor (IF) is a measure of the frequency with which the average article in a journal has been cited in a particular year. It is used to measure the importance or rank of a journal by calculating how many citations its articles receive.

[6]RCTs with one to three inadequacies receive 10% fewer citations, and RCTs with four or more inadequacies receive 35% fewer citations than adequately conducted RCTs on the same topic in the same journal

Taken together, the results indicate that inadequate trials generate a negative externality on society, but scientists' incentives to meet clinical research standards are likely weak, possibly explaining some of the prevalence of inadequate trials. For an academic scientist, I find that better research yields higher publication prestige, but low estimated monetary value[7]. For an industry scientist, incentives to meet scientific standards depend on how scientific disclosure creates value for the firm. For a firm, the lower probability to reach a positive result may disincentivize compliance with clinical trial standards if the purpose of the study is to back up marketing efforts. I discuss the implications of these findings for public and private policy, suggesting that journal editors and regulators should strengthen incentives to meet drug trial standards, and, in the meantime, healthcare decision makers and patients should be wary of inadequate drug trials supporting adoption.

The next section summarizes the scientific importance of the topic and builds on prior literature on scientists' motivations to propose a framework to describe scientists' incentives to meet scientific standards in drug trials. I then present the data, methods, and results of empirical analysis. The last section discusses implications.

[7]To estimate the monetary value of meeting standards in academia, the discussion builds on published estimates of the effect of journal impact factor and citations on academic scientists' salary.

2. Background

2.1. Scientific importance

Biomedical science and new pharmaceutical products explain a large part of health improvement and life expectancy gains over the last century (Cutler et al., 1998, 2006, 2019; Heidenreich and McClellan, 2001; Murphy and Topel, 2003; Luce et al., 2006; Lichtenberg, 2016, 2018, 2019). But this does not mean that each biomedical research project produces valuable knowledge.

Invalid or unreliable science generates major societal costs, through wrong conclusions, irreproducible results, suboptimal or harmful treatment decisions, and slowed innovation (Ioannidis, 2005, 2016). Most attempts to reproduce published studies fail (Freedman et al., 2015; Baker, 2016); for example, in cancer, 47 of 53 landmark papers could not be reproduced (Begley and Ellis, 2012). The conclusions of many published RCTs are either uninformative or wrong, damaging patients and misleading subsequent research (Altman, 1994, 2002; Zarin et al., 2019). For example, atenolol, a beta-blocker approved by the Food and Drug Administration (FDA) in 1981, sold for hundreds of millions of dollars and became standard of care, before it was shown to be no better than a placebo for hypertension more than twenty-five years later (Prasad and Cifu, 2011; Carlberg et al., 2004). The trial that first questioned atenolol's efficacy met higher scientific standards than almost all previous trials (Wiysonge et al., 2017).

One possible reason why many results are irreproducible or wrong, is that a high proportion of randomized controlled trials use inadequate methods for randomization, blinding and reporting. A recent study concluded that: "the proportion of RCTs using inadequate methods is high (59.3%) and increasing, potentially slowing progress and contributing to the reproducibility crisis." (Catillon, 2019)

Lack of knowledge or high cost of meeting standards do not fully explain such a high prevalence of inadequate methods. There is no difference in compliance with scientific standards between prestigious universities and other institutions, suggesting that investigator training is not a main driver of inadequate methods (Catillon, 2019). Indeed, meeting standards would have been feasible at no or minor cost[8] in most inadequate

[8]Minor cost was defined as less than 1% of the cost of the trial.

trials (Yordanov et al., 2015). One theory is that scientists and firms' incentives may contribute to the high prevalence of inadequate methods (Lacetera and Zirulia, 2011; Kiri et al., 2018; Gall et al., 2017; Lindner et al., 2018).

2.2. Literature on scientists' motivations

The literature identifies three key drivers of scientists' motivation: the intrinsic reward of doing science, academic prestige, and monetary rewards, highlighting that these motivations may not align with the production of high quality science.

A first motivation of scientists is the intrinsic reward of doing science, which has been described as a "taste" for science, a satisfaction derived from "puzzle solving", or a pleasure from "finding things out" (Merton, 1973; Dasgupta and David, 1994; Roach and Sauermann, 2010; Stephan, 2012). Science can be modeled as a process of search for the "true state of the world", in which high (low) effort yields a perfect (imperfect) signal about the true state of the world, and scientists' intrinsic preference for science can be expressed as a low cost of effort (Kiri et al., 2018).

Academic prestige, the second motivation of scientists, is often measured using journal impact factor and citations[9] (Stephan, 1996; Gibson et al., 2017; Hamermesh, 2018). Academic prestige matters to scientists both directly, as a form of recognition, and indirectly as it translates into career and monetary rewards, such as promotion, tenure, salary and ability to change jobs (Macri and Sinha, 2006; Hamermesh and Pfann, 2012; Stephan, 2012). In industry, academic prestige helps increase knowledge flows, persuade regulators and expert adopters, and back up marketing efforts (Cockburn and Henderson, 1998; Stern, 2004; Simeth and Cincera, 2015).

The pursuit of academic prestige does not necessarily translate into high quality research. First, prestige may reflect publication quantity and bibliometric measures of impact more than methodological quality (Nosek et al., 2012; Edwards and Roy, 2017; Ravenscroft et al., 2017). Second, prestige may be contingent on results. Papers are more likely to be published (Song et al., 2000; Dwan et al., 2013; Hopewell et al., 2007;

[9]Journal impact factor reflects experts' assessment during the peer review process. Citations provide a collective assessment and measure impact over time (Sgroi and Oswald, 2013). Overall, the literature suggests that journal prestige matters more to scientists' careers than citations, but the relative effect of journal impact factor and citations on salary varies by school rank. (Gibson et al., 2017)

Nosek et al., 2012; Scherer et al., 2018), to be cited, (Kjaergard and Gluud, 2002; Etter and Stapleton, 2009; Leimu and Koricheva, 2005) and to be accepted in better journals (Murtaugh, 2002; Easterbrook et al., 1991; Hopewell et al., 2009; Andrews and Kasy, 2019) if they report "positive" results supporting the experimental hypothesis.

Money also motivates scientists. In academia, prestige is only one driver of income differences across scientists; some scientists receive royalties from patents, equity in start-ups, consulting fees or research sponsoring, and may have conflicts of interest or receive monetary rewards contingent on results (Bekelman et al., 2003; Stephan, 2012; Dunn et al., 2016). Industry scientists are directed towards high-payoff activities in the firm's interest (Aghion et al., 2008; Sauermann and Stephan, 2013). A key question about monetary rewards contingent on results is whether the choice of methods affects how much scientific publication can provide support for product approval and adoption, and back up marketing efforts[10] (Hicks, 1995; Azoulay, 2002; Venkataraman and Stremersch, 2007; Polidoro Jr and Theeke, 2012; Salandra, 2018).

Academic and industry scientists, as well as different scientists within academia (e.g., scientists at different career stages) and within industry (e.g., in large pharmaceutical companies versus start-ups), may weigh science, prestige, and monetary rewards differently. Prior literature on the differences between academia and industry is mixed. Industry scientists have a "taste for science", but they show a greater concern for monetary rewards than academic scientists (Stern, 2004; Roach and Sauermann, 2010). Moreover, the boundaries between academia and industry are blurry, especially in biomedical research, and the theoretical differences between academia and industry may be overstated. Scientists' work in both environments share many overlapping features and there is great heterogeneity within industry and within academia (Sauermann and Stephan, 2013). To account for these similarities and differences, the scientists' utility function will be modeled as a function depending on science, academic prestige and monetary rewards, but allowing scientists to put different weights on each of these dimensions.

[10] Azoulay (2002), shows that, controlling for marketing spending, scientific evidence, as measured by citation weighted positive results affects drug sales. Azoulay (2002) defines "market expanding citations in the following way: "I score each RCT using a three-step Likert scale (+1 0 −1) to assess the negative, neutral, or positive impact of the article: +1 (respectively −1) is assigned if the treatment effect is significant and favors (respectively does not favor) the drug studied. A score of 0 is assigned if the treatment effect fails to reach statistical significance. [...] I weight the treatment effect score by the cumulative number of forward citations to the original article."

2.3. Institutional background: Cochrane systematic reviews

This study leverages unique characteristics of Cochrane Reviews. Founded in 1993, the Cochrane Collaboration is an independent nonprofit multinational organization that produces systematic reviews of healthcare interventions. Cochrane reviews are considered the "gold standard" in systematic reviews by the National Library of Medicine in the US, and the Center for Review Dissemination in the UK. Both classify systematic reviews depending on whether they are Cochrane reviews or other reviews. There is also empirical evidence of the higher quality of the methods and reporting of Cochrane Reviews compared to other systematic reviews (Page et al., 2016). The commercial funding of review groups is not allowed.

Because of the process used to produce them, Cochrane Reviews constitute a rich data source for research in the economics and management of science. Cochrane reviews aim to answer a specific scientific question by using an explicit, pre-planned protocol to identify, assess, and summarize the findings of similar but separate studies (Higgins and Green, 2011). They systematically locate all available studies on a specific research question. Then, at least two reviewers appraise the methods used in each included study following pre-specified clearly defined scientific standards (see Table A1), and extract the data on its experimental results for analysis. Thus, each Cochrane Review contains the methodological assessments, the experimental results, and the references of publications, for the universe of studies available on the topic of the review.

One limitation of Cochrane Reviews explains why they have not been used more often in economic and management studies, except for studies relying on small samples of reviews. The data from the reviews is not available in a standardized aggregated database. The format of each review is partially standardized, allowing automation of the extraction of contents. However, most of the information is entered as free text, requiring intensive classification work to make comparisons across reviews possible. For instance, the methods assessed, or the outcomes of the RCTs, even when they are similar, can be worded slightly differently.[11] The data section provides further information on the intensive process used to assemble and clean the database.

[11] e.g., blinding participants and personnel vs blinding (patients & physicians)

3. Conceptual framework

Consider a scientist with preferences for science, academic prestige, and monetary payoffs, deciding whether to meet scientific standards in a randomized controlled trial. In the context of RCTs, scientists' intrinsic rewards for doing science can be modeled as a low cost of effort for compliance with standards[12] (Kiri et al., 2018). Prestige payoffs depend on both methods and results, and can be measured using journal impact factor and citations. Contingent monetary payoffs[13] depend on both experimental results and academic prestige of the article where the results are reported.

Under these assumptions, the researcher chooses methods m^* to maximize the objective function f[14]:

$$\theta w_s 1[m] + \theta w_p P\big(m, R + b(m)\big) + \theta w_g G\Big(R + b(m), P\big(m, R + b(m)\big)\Big) - c(m)$$

θ	is the exogenous value of the research question, the importance of the topic.
w_s, w_p, w_g	represent the scientist's preference for science, academic prestige and money.
$1[m]$	represents the intrinsic reward for doing science. If a represents adequate methods and i inadequate method $1[a] = 1$ and $1[i] = 0$.
R	is the real value of the drug, which is a random variable characterized by a distribution $g(R)$.
$b(.)$	is a function mapping inadequacies into expected bias. Meeting all standards yields $b(0) = 0$.
$P(.)$	represents how much methods and results are rewarded by academic prestige. The function $P(.)$ reflects both the preference of journal editors for articles meeting standards, the results reported in the paper $(R + b(m))$ and their bayesian inference about the real value of R.

[12] In a general model of production of high quality science, Kiri et al. (2018) model the preference for science as a low cost of effort for the production of a perfect signal about the true state of the world. In that model low effort yields an imperfect signal about the true state of the world.

[13] The payoffs contingent on results correspond to the monetary payoffs of academic scientists with conflicts of interest or industry scientists pursuing the firms' interest in the drug under study.

[14] I assume f a continuous non-decreasing utility function.

$G(.)$ represents how much the article yields monetary payoffs, which depends on methods, results and prestige. The function $G(.)$ represents the adopters Bayesian inference about the real value of R.

$c(.)$ represents the cost (or effort involved) in using method m.

If methods do not impact experimental results (if $b(m) = 0$ for all m), then incentives depend only on how much the scientist, journal editors and adopters value scientific standards, and on the difference in cost between adequate and inadequate trials.

If methods affect experimental results (if $b(m) > 0$ for some m), then the scientist's incentives are more complex. The preference for science involves a disutility from the expected bias generated by inadequate methods. But inadequacies also inflate the importance of the result, potentially increasing expected prestige and monetary payoffs if journal editors, readers and/or adopters are misinformed about the effect of methods on experimental results (unable to correct for the researcher's optimal choice). In such situation, the increase in expected prestige and/or monetary gain associated with inflated results may desincentivize adequate methods.

Consider adequate method a and inadequate method i.

Then, scientist decides to meet standards if:

$$w_p\Big(P(a,R) - P\big(i, R+b(i)\big)\Big) + w_g\Big(G\big(R, P(a,R)\big) - G\big(R+b(i), P(i, R+b(i))\big)\Big) > \Big(c(a) - c(i) - \theta w_s\Big)/\theta$$

In this framework, besides the exogenous importance of the research question (which is reflected in the topic of the reviews), three parameters are critical to academic and industry scientists' decisions about how many and which standards to meet, depending on their preferences for science, academic impact, and results: (1) the effect of meeting standards on experimental results $b(i)$, (2) the effect of meeting standards on academic prestige $P(a,R) - P(i, R+b(i))$, which may be valued both directly by the scientist, but also indirectly, as academic prestige affects monetary rewards, and (3) the effect of meeting standards on monetary rewards $\mathrm{G}\big(\mathrm{R},\mathrm{P}(\mathrm{a},\mathrm{R})\big)$-$\mathrm{G}\Big(\mathrm{R}+\mathrm{b}(\mathrm{i}),\mathrm{P}\big(\mathrm{i},\mathrm{R}+\mathrm{b}(\mathrm{i})\big)\Big)$, depending on the joint effect of methods, experimental results and academic prestige on expected gains. The next section presents the data assembled to estimate these parameters.

4. Data

I assembled detailed data on research methods, experimental results and scientific impact in a large sample of drug trials to compare adequate and inadequate RCTs within narrowly defined subgroups. The database combines three sources: (1) assessments of research methods and experimental results extracted from Cochrane Reviews[15], (2) bibliometric information from PubMed[16], and (3) journal impact factor and citations from Web of Science[17]. The definition of each variable is reported in Appendix Table A2.

Systematic reviews focus on a specific question, for example "Beta-blockers for hypertension" (Wiysonge et al., 2017). Included trials compare a treatment group receiving a drug (e.g., atenolol) to a control group receiving a comparator (e.g., placebo) on one or several outcomes (e.g., all-cause mortality, cardiovascular events and stroke). The list and distribution of outcome categories is provided in Table 1.

The analysis sample includes 23,321 RCTs on drugs, published between 1990 and 2018, from 3,199 Cochrane reviews. Figure 1 summarizes the data flow. All trials included in Cochrane Reviews available in October 2018, and including a statistical file were considered for inclusion ($N = 75,526$ RCTs from 5,788 reviews). Criteria for inclusion were: (1) outcomes were dichotomous or continuous and classifiable as positive or negative without context[18], so the treatment effect could be summarized as a standardized z-score ($N = 46,126$ RCTs from 4,900 reviews) (2) the intervention was a drug ($N = 36,754$ RCTs from 3,878 reviews) (3) the results were published between 1990 and 2018 ($N = 29,682$ RCTs from 3,638 reviews) and (4) the trial was assessed on at least one of the six dimensions of the Cochrane Risk of Bias Assessment Tool. Duplicates were removed ($N = 23,321$ RCTs from 3,199 reviews).

Table 2 presents descriptive statistics at the RCT level for the main variables. At least two human expert reviewers assess all included trials on pre-specified scientific stan-

[15]The scraper, built by the author, uses Python 3.6.5. The code will be made publicly available on GitHub when the papers based on these data are published.

[16]Information from PubMed was retrieved using the Entrez Programming Utilities public Application Programming Interface (API).

[17]Cochrane reviews include an up-to-date Web of Science citation link for each reference. Journal titles are extracted from PubMed using the PubMed identifier listed in the reference link for each included reference. Journal titles are matched to the InCites Journal Citation Reports available through the Harvard Library.

[18]For example, death or heart attack are negative without context, but weight requires context.

dards, including random sequence generation (allocation to the treatment and control group is truly random), allocation concealment (intervention allocation could not have been foreseen before or during enrollment), blinding of participants and personnel (participants and personnel don't know which intervention the participant receives), blinding of outcome assessment (outcome assessors don't know which intervention the participant received), complete outcome data (attrition and exclusion in each group, and corresponding reasons are reported), and selective reporting (all results on all prespecified outcomes are reported). Further details about each dimension are available in Table A1. On each of these standards, a trial can be "adequate", "inadequate", "unclear", or the assessment can be missing[19].

Overall, 17% of RCTs used adequate methods, 45% used inadequate methods and 37% were poorly reported. Inadequacies, by decreasing prevalence, came from blinding of participants and personnel (23%), incomplete outcome data (16%), selective reporting (11%), blinding of outcome assessment (10%), allocation concealment (5%) and random sequence generation (3%) (see Table 2). The distribution of inadequacies is presented in Figure 2, for the sample of RCTs assessed on all six dimensions ($N = 11,600$). In this sample, 45% of RCTs have zero inadequacies ($N = 5,266$), 31% have one inadequacy ($N = 3,561$), 17% have two inadequacies ($N = 1,960$), 5% have three inadequacies ($N = 609$), 1% have four inadequacies ($N = 161$), and less than 1% have five inadequacies ($N = 34$) and six inadequacies ($N = 9$). Because of their small numbers, the analyses group together the RCTs with four and more inadequacies.

Some inadequacies are more prevalent than others. Blinding of participants and personnel is the most common issue. It concerns 43% of RCTs with one inadequacy, 79% of RCTs with two inadequacies, and 95% of RCTs with three or more inadequacies. Among RCTs with one inadequacy, the two next most frequent issue are incomplete outcome data and selective reporting, present in respectively 28% and 22% of RCTs. Among RCTs with two inadequacies, inadequate blinding of participants and personnel, present in 79% of RCTs, is followed by inadequate blinding of outcome assessment (53%),

[19]Some Cochrane reviews do not include assessments on all the dimensions of the Cochrane risk of bias assessment tool published in 2011.

incomplete outcome data (32%), and selective reporting (25%). Among RCTs with three inadequacies, the relative prevalence of inadequacies remains similar.

The inadequacies are not independent from each other. Researchers who use inadequate methods on one dimension may be more likely to use inadequate methods on another dimension. Each cell in Table A3 reports the pseudo R-squared from the logit regression of the indicator for one type of inadequacy (e.g. random sequence generation) on another type of inadequacy (e.g., allocation concealment). Two pairs of inadequacies exhibit relatively large correlations: (1) random sequence generation with allocation concealment (pseudo R-squared = 0.20) and (2) blinding of participants and personnel with blinding of outcome assessment (pseudo R-squared = 0.12). To account for these correlations, the analyses use two alternative independent variables for inadequacies: the number of inadequacies and indicators for each inadequacy.

RCT results are classified and coded into normed z-scores so that a higher value of z indicates a better outcome for the treatment group compared to the control group (see Appendix table A2). The distribution of results for all outcomes is presented in Figure 3a. Overall 15% of all results are statistically significant and positive ($z > 1.96$), 12% are statistically significant and negative ($z < -1.96$) and 74% are null ($-1.96 \leq z \leq 1.96$). The mean of z across all outcomes is 0.16, with a standard deviation of 3.84 (see also Table 2). The distribution of RCT results varies by outcome. Figure A1 shows the distribution of results for six categories of outcomes: death, adverse event, adverse effect, physical health, mental health and other efficacy outcomes.

Journal impact factor[20] and citations through 2018 are retrieved from Web of Science for the main article reporting the trial as identified by Cochrane reviewers. The distribution of journal impact factor is presented in Figure 3b. The journal impact factor distribution is highly skewed, with a mean of 5.34 and a standard deviation of 8.95. The top 0.1% corresponds to RCTs published in the *New England Journal of Medicine*. The top 1% adds RCTs published in *JAMA* and *the Lancet*. The top decile adds RCTs published in other high impact journals in general and internal medicine, such as *the Annals of Internal Medicine* and *the Archives of Internal Medicine*, as well as top specialty jour-

[20]Journal impact factor is a time varying variable from 1997 to 2018. For years before 1997, the 1997 value is imputed. For journals not listed in Web of Science, a value of zero is imputed.

nals such as the *Journal of Clinical Oncology*, *Gastroenterology*, and *Circulation*. 80% of the RCTs are published in journals with an impact factor of less than 5.3, and 20% of RCTs are published in journals with an impact factor less than 1.

The distribution of annual citations is highly skewed, and similar to the distribution of journal impact factor (Figure 3c). The annual citation mean is 4.78, with a standard deviation of 18.43. On average, the top 1% RCTs receive 133 annual citations, RCTs in the 96-99 percentiles receive 27 annual citations, and RCTs in the 90-95 percentiles receive 12 annual citations. RCTs in the top 80-90 decile receive just above 5 annual citations. 80% of RCTs receive less than 5 annual citations. RCTs in the bottom 20% receive zero annual citations.

Cochrane reviews offer two ways to measure novelty within and across topics. All reviews of prescription drugs include trials on drugs that are approved at the time of the review. Considering early verus late science along the R&D spectrum, there is no "early stage" research in the sample. However, within topics, the first RCT is more "early science" than the twentieth RCT. The "order of entry" in a review measures this dimension of "within topic" novelty. Across reviews, some topics are also more novel than others. For instance, RCTs included in a review on immunotherapy address a more novel topic than RCTs included in a review on statins.

The next sections explains how my empirical analysis leverages these data to compare trials within narrowly defined groups and to estimate the effect of meeting standards on experimental results, journal placement, and citations.

5. Empirical Analysis

My empirical analysis relies on the comparison of RCTs using adequate versus inadequate methods within narrowly defined groups. RCTs on the same topic (e.g., Active drug versus a placebo for hypertension) compare the effect of the same drug(s) (e.g., a beta-blocker, such as atenolol) to the same comparator(s) (e.g., a placebo), on the same outcome(s) (e.g., all-cause mortality, cardiovascular events, and stroke). Within these groups, the study compares experimental results and scientific impact of adequate versus inadequate RCTs.

Using the equation below, I estimate the effect of meeting six scientific standards[21] on three main outcomes: (1) the sign and significance of experimental results; (2) the impact factor of the publishing journal; and (3) the number of citations the paper receives:

$$Y_{\text{irot}} = \alpha + \beta M_i + \delta X_i + \rho_r + \omega_o + \tau_t + \epsilon_{\text{irot}} \tag{1}$$

Y_{irot} Outcome for RCT i published in year t, included in review r, on outcome o (e.g., death or stroke).

M_i Indicators for each number (or type) of inadequacies.

X_i RCT level controls (e.g., sample size)

ρ_r Review fixed effect (topic, treatments, comparators, outcomes, subgroups)

ω_o Outcome type fixed effect (death, adverse event, physical health...)

τ_t Publication year fixed effect

ϵ_{irot} Error term

The first regression[22] predicts whether, within reviews, RCTs with n inadequacies (or with each type of inadequacy) have a higher probability to yield a positive result (i.e., a statistically significant result supporting the use of the subject drug) than a more adequate RCT on the same topic. The regressions control for sample size, missing assessments, and include outcome and year fixed effects.

[21] The six standards include: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, complete outcome data, and absence of selective reporting, and are defined in Table A1.

[22] I used a logistic regression to predict the binary outcome of a statistically significant positive result.

To investigate differences in publication bias and strategic non-disclosure by sponsor type (NIH versus industry), equation (1) is modified to include indicators for sponsor type and interactions between the number of inadequacies and sponsor type. To study whether the relationship between inadequacies and results changed with pre-registration and reporting requirements, equation (1) is modified to include an indicator identifying the period after each new requirement (post 2005, 2007 and 2012) and an interaction between the number of inadequacies and the time variables. [23]

Using an equation similar to (1), regression models[24] estimate the difference in journal placement and citations between adequate and inadequate RCTs. The models include the same controls as (1), and add controls for sponsor (NIH, industry), novelty (as measured by indicators for order of entry in the review) and prestige[25] (based on first author affiliation).

A first set of models estimates the effect of inadequacies on journal prestige conditional on results. To get the net global effect accounting for both the effect of methods on results and the effect of results on journal prestige, a second set of models estimates the effect of inadequacies on journal prestige unconditional on results.

Regressions including indicators for sponsor type (NIH, industry) and interactions between sponsor and inadequacies explore whether the effect varies by sponsor (*i.e.* whether journal editors are more or less tolerant of inadequacies for publicly sponsored trials or privaltely sponsored trials). Regressions with interactions for prestige (based on first author affiliation) investigate whether journal editors are more lenient towards prestigious authors' inadequacies. Regressions with interactions for top pharma company first author and other type of firm involvement explore whether the relationship between inadequacies and journal impact factor depends on the saliency of firm involvement. Re-

[23] In 2005, the International Committee of Medical Journal Editors (ICMJE) required trial registration before patient enrollment for trials to be eligible for publication in one of its member journals (De Angelis et al., 2004). In 2007, the U.S. Food and Drug Administration Amendments Act of 2007 (FDAAA 801), required registration for most studies of FDA-regulated drug, biological or device products. In 2012, detailed reporting of outcome measures became mandatory for study records submitted to ClinicalTrials.gov (Tse et al., 2018).

[24] Negative binomial models are reported in the result section. OLS models are reported in appendix.

[25] Observable characteristics of the author or the research, other than the intrinsic scientific quality of the article, affect its journal placement and citations. For instance, prestigious authors receive proportionally more credit and reward for their work than less prestigious authors (the "Matthew effect") (Merton, 1968; Azoulay et al., 2013)

gressions with interactions for novelty (as measured by the first or first to third RCT(s) on a topic) investigate whether journal editors are more or less tolerant of methodological inadequacies in more novel trials.

Journal prestige is likely to explain a part of the difference in citations between RCTs using adequate versus inadequate methods. To get the net effect of inadequacies on citations, a first set of models predicts the effect unconditional on impact factor. To study whether, comparing articles published in similar journals, RCTs using better methods receive more citations, a second set of models includes impact factor in the regressions.[26]

Regression models test whether citations to positive and negative results, and net positive citations, are different for RCTs with n (or each type of) inadequacies compared to adequate RCTs on the same topic. "Positive citations" include all citations to articles with at least one positive and statistically significant result supporting the use of the drug under study (and no statistically significant negative finding). "Positive citations including citations to non-significant results" add non-significant results, as these may be used in marketing campaigns. "Negative citations" include all citations to articles with at least one significant result against the use of the drug. "Net positive citations" account for both citations to positive results (supporting the use of the drug) and citations to negative results (against the use of the drug).[27]

[26] The second set of models aims to explore whether RCTs published in a specific journal (e.g., the *New England Journal of Medicine*) and meeting scientific standards receive more citations than similar articles on the same topic published in that journal but using inadequate methods.

[27] "Net positive citations" are defined by $+1\times$ citations if the result is positive (i.e. finding that the drug is superior to the comparator), $-1\times$ citations if the result is negative (i.e. finding that the drug is inferior to the comparator), and zero otherwise. This is the same approach used to define "market expanding citations" or "weighted evidence driving drug sales" in Azoulay (2002).

6. Results

6.1. The impact of inadequacies on experimental results

Figure 4 shows the impact of inadequacies on the probability of a positive result. The formal regression results are presented in Table 3. On average, each inadequacy increases the probability of a positive result by 7%. The probability of a positive result in RCTs with no inadequacies is 15%. RCTs with one inadequacy have a 17% change of finding a postive result (+13.3%). RCTs with three or more inadequacies have a 19% chance of finding a positive result (+26.6%). These effects are large and statistically significant.

Table 3, column 3 presents the regression results, adding interactions between each number of inadequacies and sponsorship by NIH or industry. Positive results are more likely in trials sponsored by the NIH or by industry than in all other trials. All interactions between number of inadequacies and sponsor are non significant, and the coefficients on each number of inadequacies remain very similar to the coefficients in the main model without interactions. There is no difference in the relationship between inadequacies and results by sponsor type.

Table 3, columns 4-6 report the regressions with time indicators as well as interactions between each number of inadequacies and time indicators (post 2005, post 2007 and post 2012). The variables indicating the second time period are all negative and statistically significant. In all three models, however, the interactions between inadequacies and time are not significant, except for three inadequacies, suggesting that the relationship between inadequacies and positive results has not changed with new reporting policies.

The effect of each inadequacy, ordered by prevalence[28], on the probability of a positive result is presented in Figure 4b and Table A4. Inadequate blinding of participants and personnel, which is the most common inadequacy, also has the largest positive effect on the probability of a positive result (+4 percentage points), followed by allocation concealment (+3 percentage points). Inadequate blinding of outcome assessment has a negative effect on the probability of a positive result (−3 percentage points).

[28](1) blinding of participants and personnel, (2) incomplete outcome data, (3) selective reporting, (4) blinding of outcome assessment, (5) allocation concealment and (6) random sequence generation.

Figure A2 reports estimates of the effect of the number of inadequacies on the probability of a positive result for six outcome categories[29]. When the outcome is death, adverse effect and adverse event, one inadequacy is enough to significantly increase the probability of a positive result. This is not the case for other physical and mental health outcomes and other (more subjective) measures of efficacy. Suprisingly, this suggests inadequacies matter even more for objective outcomes than subjective outcomes.

Inadequate methods increase the probability of a positive result, and inadequacies matter even more for objective outcomes than subjective outcomes. Inadequate blinding of participants and personnel, which is the most common inadequacy, also has the largest positive effect on the probability of a positive result. The relationship between inadequacies and positive results did not change with the introduction of pre-registration and reporting requirements and does not vary by sponsor type, suggesting that publication bias and strategic non-reporting do not drive the results.

6.2. The impact of inadequacies on journal placement

Table 4[30] and Figure 5a report the regressions of journal impact factor on number of inadequacies. Conditional on results, RCTs with one, two, three, and four or more inadequacies are published in journals with an impact factor respectively 87%, 81%, 71% and 64% of the expected journal impact factor for an adequate RCT included in the same review.[31] The effect of each additional inadequacy on journal impact factor is large and highly significant.

Table 4, column 3, presents the regressions of journal impact factor on number of inadequacies, with interactions between number of inadequacies and NIH or industry sponsor. NIH sponsored trials are published in much better journals than other trials, by a large and highly significant coefficient ($IRR = 1.44$). Industry sponsored trials are published in better journals than other trials, but the effect is not as large and only marginally significant ($IRR = 1.10$). Journal editors judge inadequacies more harshly

[29]Death, adverse event, adverse effect, physical health, mental health and other measures of efficacy

[30]Table 4 reports negative binomial models. Table A5 report the same regressions using OLS.

[31]Unconditional on results, the journal impact factor penalty on each number of inadequate methods is very similar (see Figure 5).

in NIH sponsored trials compared to other trials. Surprisingly, this is not the case for industry sponsored trials.[32]

Table 4, column 4, reports the regressions predicting journal impact factor with interaction between the number of inadequacies and academic prestige (as measured by first author affiliated with a top university or another university). Articles by prestigious authors are published in much better journals, with a 30% boost for authors affiliated with top universities and a 8% boost for authors affiliated with other universities. Interaction coefficients between one or two inadequacies and first author affiliated with a top university are large and significant[33].

Table 4, column 5, reports the regression of journal prestige on number of inadequacies, with interaction for first author affiliation with top pharma company and other type of firm involvement. None of the interaction coefficients are significant.[34]

Table 4, columns 6 and 7, report regressions of journal prestige on the number of inadequacies, with interactions for novelty. Novelty is associated with a large and significant reward in journal prestige[35]. However, the relationship between inadequacies and journal prestige does not depend on novelty. None of the interaction coefficients is significant.

Regressions of journal impact factor on each inadequacy are presented in Figure 5b and Table A4. Conditional on results, inadequate blinding of participants and personnel yields 88% of the expected journal impact factor for an adequate RCT included in the same review, incomplete outcome data 92%, selective reporting 90%, blinding of outcome assessment 83%, allocation concealment 83%, and random sequence generation 78%.[36]

RCTs meeting scientific standards are published in better journals. Less frequent inadequacies are more penalized. Journal editors judge inadequacies more harshly in NIH sponsored trials compared to other trials. Surprisingly, this is not the case for industry sponsored trials. Furthermore, the saliency of industry involvement does not

[32]Three out of four interaction coefficients on NIH with number of inadequacies are less than 1. Three out of four interaction coefficient on industry with number of inadequacies are more than 1.

[33]$IRR = 1.06$ for one inadequacy and $IRR = 1.19$ for two.

[34]Given the relatively small sample of RCTs with a top pharma first author ($N = 226$), this result should be interpreted with caution.

[35]First: $IRR = 1.37$, First to third: $IRR = 1.34$.

[36]The effect of each inadequacy on journal impact factor unconditional on results is very similar (see Figure 5).

lead journal editors to judge inadequacies more harshly. Journal editors are more lenient with prestigious authors, especially when the number of inadequacies is limited to one or two. Novel trials are published in much better journals, but inadequate trials, novel or not, receive the same journal impact factor penalty.

6.3. The impact of inadequacies on citations

Table 5 and Figure 6show the impact of inadequacies on total annual citations. The relationship between inadequacies and total annual citations is non-linear, in contrast to the relationship between inadequacies and journal prestige. The first inadequacy yields a citation penalty of about 10% compared to an adequate trial. The second and third inadequacies do not make much difference compared to one inadequacy. Four or more inadequacies yields a 20% additional penalty. As shown in Table 5, controlling for impact factor in the regressions predicting citations eliminates the difference between trials with one, two and three inadequacies.

Table A6 and Figure 6b show the effect of each inadequacy on total annual citations. RCTs with inadequate blinding receive 91% of the expected citations for an adequate RCT included in the same review, incomplete outcome data 98%, selective reporting 94%, inadequate blinding of outcome assessment 93%, inadequate allocation concealment 87%, inadequate random sequence generation 78%.[37]

Table 6 and Figure 7 show the effect of the number of inadequacies on citations to positive results, citations to negative results and net citations. Only one number of inadequacies (two inadequacies) is associated with a decrease in citations to positive results, and this effect is only marginally significant. No number of inadequacies has any effect on citations to positive results including non-significant results (which are sometimes used in marketing campaigns, especially in the absence of significant positive results). One or two inadequacies decrease citations to negative results.

Compliance with scientific standards affects academic prestige beyond the editorial and peer-review process. RCTs with one or more inadequacies receive significantly fewer citations than adequate RCTs on the same topic, and a main reason why better trials get more citations is that they get published in better journals. Different inadequacies have

[37]Conditional and unconditional on results, the estimates of the effect of each inadequacy on citations are very similar.

different effects on total citations, and the effect of each inadequacy on total citations is an attenuated version of the effect of each inadequacy on journal impact factor. However, inadequacies have no effect on net positive citations, and the number of citations to positive results does not change with the number of inadequacies. One or two inadequacies decrease citations to negative results.

7. Discussion and conclusion

In this paper, using a novel dataset of 23,321 clinical trials of drugs assessed in Cochrane Reviews, I investigate academic and industry scientists' incentives to meet clinical trial standards. Within narrowly defined subgroups, I compare experimental results and scientific impact of RCTs using adequate versus inadequate methods for randomization, blinding and reporting.

The first contribution of this paper is to show that methodological inadequacies in drug trials generate measurable bias (not merely noise), through an empirical estimate of the effect of inadequacies on experimental results. The effect is large (+7% per inadequacy) and highly significant, suggesting that physicians, patients, firms, governments or other decision makers using the results of the research as input in decision making need interpret results of inadequate trials with caution. Notably, the relationship between inadequacies and experimental results did not change with the introduction of pre-registration and reporting requirements for clinical trials, and does not depend on government or industry sponsorship. This suggests that publication bias or strategic non-disclosure do not drive the results.

A second contribution of the paper is to estimate whether and how much valid and reliable trials get published in better journals and receive more citations. I find that the effect of inadequacies on journal impact factor is large and journal impact factor decreases linearly in the number of inadequacies (-10% per inadequacy). However, the effect of inadequacies on citations is more limited and non-linear, and more valid and reliable research is not associated with more citations to positive results. Journal editors are sensitive to the number of inadequacies. Peer scientists do differentiate adequate and inadequate trials in their citations, but are not sensitive to the number of inadequacies between one and three. Only four or more inadequacies received a much larger citation penalty than one inadequacy. These results are consisten with prior research finding, in other domains, that better methods are associated with publication in better journals and more citations (Antonakis and Lalive, 2008; Antonakis et al., 2014; Bergh et al., 2006).

Scientists who directly derive utility from a publication in a better journal or more expected citations have incentive to produce better research, if they do not have a strong

preference for results supporting the drug. However, published estimates of the monetary value of a publication in a better journal or a few additional citations are low (Sandnes, 2018; Gibson et al., 2014). Across studies, a 10% increase in total citations increases salary between 0.1% and 1.4% (Hamermesh, 2018). Consider a scientist producing one paper a year every year over a 40 year career[38]. Based on previously described published estimates, assuming her annual salary is $100K and there is no quantity/quality tradeoff involved in compliance with scientific standards in drug trials[39], the monetary value of a better RCT—receiving a 10% citation boost compared to an RCT with inadequacies—is between $2.5 and $35 annually or a maximum gain between $100 and $1,400 over the scientist's career[40]. For an academic scientist motivated by expected monetary gain, even if the cost of meeting standards is very small, the monetary incentives deriving from academic prestige may not be strong enough to encourage the production of high quality drug trials.

For scientists with monetary incentives contingent on results, the decrease in the probability of positive results associated with the use of better methods may disincentivize compliance with high standards. If scientists are interested in the success of a company's product—because they work for the company or have conflicts of interest—their incentives are aligned with the firm's payoff for meeting standards. If publication aims to support marketing efforts, industry incentives to meet standards may be low. Each inadequacy increases the probability to find support for the drug by 7%, and inadequate studies lead to slightly fewer citations to negative results.

These findings have implications for journal editors and regulators. Journal editors may need to strengthen incentives for compliance with scientific standards in drug trials. Inadequacies that are less frequent are also more penalized. The least frequent inadequacies also correspond to standards that are easiest to follow (such as random sequence generation, which can be easily done with a computer at no cost). If compliance with a scientific standard is the norm, the lower frequency of violation of this standard might

[38]Less than 1% of all scientists publish every year, but these scientists account for about 40% of all publications (Ioannidis et al., 2014)

[39]I assume here that the scientist can produce one paper a year, whether the paper meets scientific standards or not, which will overestimate the gain from meeting standards if there is a quality/quantity tradeoff.

[40]Between (1/40)*0.1%*100K=$2.5 and (1/40)*1.4%*100K=$35 annually.

lead to its higher saliency to the editor and to a higher penalty. Scientists may also be more likely to comply with scientific standards that lead to higher penalties.

Journals were at the forefront of the improvement in registration and reporting. The 2004 policy of the International Committee of Medical Journals' (ICMJE) on mandatory registration of clinical trials as a precondition for manuscript submission led a dramatic increase in the number of registered trials (Zarin and Keselman, 2007; Huić et al., 2011). The endorsement of the Consolidated Standards of Reporting Trials (CONSORT) statement by medical journals[41] has beneficially influenced the reporting of trials (Turner et al., 2012). As several scientific standards (including those in this study) are supposed to be reported according to the CONSORT checklist, a next step could be for journals to publish the assessment of the paper[42] on each of these dimensions along with the paper. Such assessments could be beneficial for three reasons: as an incentive in the form of public feedback, it could increase compliance with standards; to increase awareness about risk of bias and facilitate the interpretation of results by readers using the article as an input in decisions about treatment choice or future research; to create a broad data base for future meta-research on the impact of compliance with scientific standards.

Regulators may also want to take measures to strengthen incentives to meet standards in drug trials because inadequate research generates a large externality for society (Zarin et al., 2019). Regulators could consider the mandatory reporting of methods used to meet scientific standards in drug trials, for instance in the ClinicalTrials.gov database. This would make public feedback possible, even for unpublished trials, and facilitate the interpretation of trial results. Limits to this strategy include that already mandated items are not fully and completely reported in the database so far (Zarin et al., 2016), and that changes to registration elements in registries are not reflected in published articles (Pranić and Marušić, 2016).

Future research should study the relationship between inadequacies and medical reversals to inform priority setting in replication projects, and assess welfare effects of possible changes to mandatory requirements and drug approval policies.

[41] The CONSORT statement is now endorsed by over 50% of the medical journals listed in the Abridged Index Medicus on PubMed. See: `http://www.consort-statement.org/about-consort/endorsers1`

[42] This assessment could be as simple as "Adequate", "Inadequate", or "Unclear", with a quote from the paper to support the judgment, as currently in Cochrane reviews.

References

Aghion, Philippe, Mathias Dewatripont, and Jeremy C Stein, "Academic freedom, private-sector focus, and the process of innovation," *The RAND Journal of Economics*, 2008, *39* (3), 617–635.

Altman, Douglas G, "The scandal of poor medical research," 1994.

_ , "Poor-quality medical research: what can journals do?," *Jama*, 2002, *287* (21), 2765–2767.

Andrews, Isaiah and Maximilian Kasy, "Identification of and correction for publication bias," *American Economic Review*, 2019, *109* (8), 2766–94.

Angelis, Catherine De, Jeffrey M Drazen, Frank A Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, AJ Overbeke et al., "International committee of medical journal editors. Clinical trial registration: a statement from the international committee of medical journal editors. Clinical trial registration: a statement from the International Committee of Medical Journal Editors [Editorial]," *JAMA*, 2004, *292*, 13631364.

Antonakis, John and Rafael Lalive, "Quantifying scholarly impact: IQp versus the Hirsch h," *Journal of the American Society for information Science and Technology*, 2008, *59* (6), 956–969.

_ **, Nicolas Bastardoz, Yonghong Liu, and Chester A Schriesheim**, "What makes articles highly cited?," *The Leadership Quarterly*, 2014, *25* (1), 152–179.

Azoulay, Pierre, "Do pharmaceutical sales respond to scientific evidence?," *Journal of Economics & Management Strategy*, 2002, *11* (4), 551–594.

_ **, Toby Stuart, and Yanbo Wang**, "Matthew: Effect or fable?," *Management Science*, 2013, *60* (1), 92–109.

Baker, Monya, "Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the'crisis rocking science and what they think will help," *Nature*, 2016, *533* (7604), 452–455.

Begley, C Glenn and Lee M Ellis, "Drug development: Raise standards for preclinical cancer research," *Nature*, 2012, *483* (7391), 531.

Bekelman, Justin E, Yan Li, and Cary P Gross, "Scope and impact of financial conflicts of interest in biomedical research: a systematic review," *Jama*, 2003, *289* (4), 454–465.

Bergh, Donald D, John Perry, and Ralph Hanke, "Some predictors of SMJ article impact," *Strategic Management Journal*, 2006, *27* (1), 81–100.

Bikard, Michaël, Fiona Murray, and Joshua S Gans, "Exploring trade-offs in the organization of scientific work: Collaboration and scientific reward," *Management science*, 2015, *61* (7), 1473–1495.

Carlberg, Bo, Ola Samuelsson, and Lars Hjalmar Lindholm, "Atenolol in hypertension: is it a wise choice?," *The Lancet*, 2004, *364* (9446), 1684–1689.

Catillon, Maryaline, "Trends and predictors of biomedical research quality, 1990–2015: a meta-research study," *BMJ open*, 2019, *9* (9), e030342.

Chalmers, Iain and Paul Glasziou, "Avoidable waste in the production and reporting of research evidence," *The Lancet*, 2009, *374* (9683), 86–89.

Cockburn, Iain M and Rebecca M Henderson, "Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery," *The Journal of Industrial Economics*, 1998, *46* (2), 157–182.

Cutler, David M, Allison B Rosen, and Sandeep Vijan, "The value of medical spending in the United States, 1960–2000," *New England journal of medicine*, 2006, *355* (9), 920–927.

_ , Kaushik Ghosh, Kassandra L Messer, Trivellore E Raghunathan, Susan T Stewart, and Allison B Rosen, "Explaining The Slowdown In Medical Spending Growth Among The Elderly, 1999–2012," *Health Affairs*, 2019, *38* (2), 222–229.

_ , **Mark McClellan, Joseph P Newhouse, and Dahlia Remler**, "Are medical prices declining? Evidence from heart attack treatments," *The Quarterly Journal of Economics*, 1998, *113* (4), 991–1024.

Dasgupta, P and PA David, "Toward a New Economics of Science'," *Research Policy*, 1994, *23*, 487–521.

David, Paul A, "Common agency contracting and the emergence of" open science" institutions," *The American Economic Review*, 1998, *88* (2), 15–21.

Dunn, Adam G, Enrico Coiera, Kenneth D Mandl, and Florence T Bourgeois, "Conflict of interest disclosure in biomedical research: a review of current practices, biases, and the role of public registries in improving transparency," *Research integrity and peer review*, 2016, *1* (1), 1.

Dwan, Kerry, Carrol Gamble, Paula R Williamson, and Jamie J Kirkham, "Systematic review of the empirical evidence of study publication bias and outcome reporting biasan updated review," *PloS one*, 2013, *8* (7), e66844.

Easterbrook, Phillipa J, Ramana Gopalan, JA Berlin, and David R Matthews, "Publication bias in clinical research," *The Lancet*, 1991, *337* (8746), 867–872.

Edwards, Marc A and Siddhartha Roy, "Academic research in the 21st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition," *Environmental Engineering Science*, 2017, *34* (1), 51–61.

Etter, Jean-François and John Stapleton, "Citations to trials of nicotine replacement therapy were biased toward positive results and high-impact-factor journals," *Journal of Clinical Epidemiology*, 2009, *62* (8), 831–837.

Freedman, Leonard P, Iain M Cockburn, and Timothy S Simcoe, "The economics of reproducibility in preclinical research," *PLoS biology*, 2015, *13* (6), e1002165.

Gall, Thomas, John PA Ioannidis, and Zacharias Maniadis, "The credibility crisis in research: Can economics tools help?," *PLoS biology*, 2017, *15* (4), e2001846.

Gibson, John, David L Anderson, and John Tressler, "Which journal rankings best explain academic salaries? Evidence from the University of California," *Economic Inquiry*, 2014, *52* (4), 1322–1340.

_ , _ , and _ , "Citations or journal quality: Which is rewarded more in the academic labor market?," *Economic Inquiry*, 2017, *55* (4), 1945–1965.

Hamermesh, Daniel S, "Citations in economics: Measurement, uses, and impacts," *Journal of Economic Literature*, 2018, *56* (1), 115–56.

_ and Gerard A Pfann, "Reputation and earnings: the roles of quality and quantity in academe," *Economic Inquiry*, 2012, *50* (1), 1–16.

Heidenreich, Paul A and Mark McClellan, "Trends in treatment and outcomes for acute myocardial infarction: 1975–1995," *The American journal of medicine*, 2001, *110* (3), 165–174.

Hicks, Diana, "Published papers, tacit competencies and corporate management of the public/private character of knowledge," *Industrial and corporate change*, 1995, *4* (2), 401–424.

Higgins, Julian PT and Sally Green, *Cochrane handbook for systematic reviews of interventions*, Vol. 4, John Wiley & Sons, 2011.

Hopewell, Sally, Kirsty Loudon, Mike J Clarke, Andrew D Oxman, and Kay Dickersin, "Publication bias in clinical trials due to statistical significance or direction of trial results," *Cochrane Database of Systematic Reviews*, 2009, (1).

_ , Mike J Clarke, Lesley Stewart, and Jayne Tierney, "Time to publication for results of clinical trials," *Cochrane database of systematic reviews*, 2007, (2).

Huić, Mirjana, Matko Marušić, and Ana Marušić, "Completeness and changes in registered data and reporting bias of randomized controlled trials in ICMJE journals after trial registration policy," *PloS one*, 2011, *6* (9), e25258.

Ioannidis, John PA, "Why most published research findings are false," *PLoS medicine*, 2005, *2* (8), e124.

_ , "Why most clinical research is not useful," *PLoS medicine*, 2016, *13* (6), e1002049.

_ , **Kevin W Boyack, and Richard Klavans**, "Estimates of the continuously publishing core in the scientific workforce," *PloS one*, 2014, *9* (7), e101698.

Jr, Francisco Polidoro and Matt Theeke, "Getting competition down to a science: The effects of technological competition on firms' scientific publications," *Organization Science*, 2012, *23* (4), 1135–1153.

Judge, Timothy A, Daniel M Cable, Amy E Colbert, and Sara L Rynes, "What causes a management article to be citedarticle, author, or journal?," *Academy of management journal*, 2007, *50* (3), 491–506.

Kiri, Bralind, Nicola Lacetera, and Lorenzo Zirulia, "Above a swamp: A theory of high-quality scientific production," *Research Policy*, 2018, *47* (5), 827–839.

Kjaergard, Lise L and Christian Gluud, "Citation bias of hepato-biliary randomized clinical trials," *Journal of clinical epidemiology*, 2002, *55* (4), 407–410.

Lacetera, Nicola and Lorenzo Zirulia, "The economics of scientific misconduct," *The Journal of Law, Economics, & Organization*, 2011, *27* (3), 568–603.

Leimu, Roosa and Julia Koricheva, "What determines the citation frequency of ecological papers?," *Trends in ecology & evolution*, 2005, *20* (1), 28–32.

Lichtenberg, Frank R, "The impact of pharmaceutical innovation on premature cancer mortality in Switzerland, 1995–2012," *The European Journal of Health Economics*, 2016, *17* (7), 833–854.

_ , "The Impact of Biomedical Research on US Cancer Mortality," *Measuring and Modeling Health Care Costs*, 2018, *76*, 475.

_ , "The long-run impact of new medical ideas on cancer survival and mortality," *Economics of Innovation and New Technology*, 2019, *28* (7), 722–740.

Lindner, Mark D, Karina D Torralba, and Nasim A Khan, "Scientific productivity: An exploratory study of metrics and incentives," *PloS one*, 2018, *13* (4), e0195321.

Luce, Bryan R, Josephine Mauskopf, Frank A Sloan, Jan Ostermann, and L Clark Paramore, "The return on investment in health care: from 1980 to 2000," *Value in Health*, 2006, *9* (3), 146–156.

Macri, Joseph and Dipendra Sinha, "Rankings methodology for international comparisons of institutions and individuals: An application to economics in Australia and New Zealand," *Journal of Economic Surveys*, 2006, *20* (1), 111–156.

Merton, Robert K, "The Matthew effect in science: The reward and communication systems of science are considered," *Science*, 1968, *159* (3810), 56–63.

_ , *The sociology of science: Theoretical and empirical investigations*, University of Chicago press, 1973.

Murphy, Kevin M and Robert H Topel, "The economic value of medical research," *Measuring the gains from medical research: An economic approach*, 2003, *15* (30), 125–146.

Murtaugh, Paul A, "Journal quality, effect size, and publication bias in meta-analysis," *Ecology*, 2002, *83* (4), 1162–1166.

Nosek, Brian A, Jeffrey R Spies, and Matt Motyl, "Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability," *Perspectives on Psychological Science*, 2012, *7* (6), 615–631.

Page, Matthew J, Larissa Shamseer, Douglas G Altman, Jennifer Tetzlaff, Margaret Sampson, Andrea C Tricco, Ferran Catala-Lopez, Lun Li, Emma K Reid, Rafael Sarkis-Onofre et al., "Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study," *PLoS medicine*, 2016, *13* (5), e1002028.

Pranić, Shelly and Ana Marušić, "Changes to registration elements and results in a cohort of Clinicaltrials. gov trials were not reflected in published articles," *Journal of clinical epidemiology*, 2016, *70*, 26–37.

Prasad, Vinay and Adam Cifu, "Medical reversal: why we must raise the bar before adopting new technologies," *The Yale journal of biology and medicine*, 2011, *84* (4), 471.

Ravenscroft, James, Maria Liakata, Amanda Clare, and Daniel Duma, "Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements," *PloS one*, 2017, *12* (3), e0173152.

Roach, Michael and Henry Sauermann, "A taste for science? PhD scientists academic orientation and self-selection into research careers in industry," *Research policy*, 2010, *39* (3), 422–434.

Salandra, Rossella, "Knowledge dissemination in clinical trials: Exploring influences of institutional support and type of innovation on selective reporting," *Research Policy*, 2018, *47* (7), 1215–1228.

Sandnes, Frode Eika, "Do Norwegian academics who publish more earn higher salaries?," *Scientometrics*, 2018, *115* (1), 263–281.

Sauermann, Henry and Paula Stephan, "Conflicting logics? A multidimensional view of industrial and academic science," *Organization science*, 2013, *24* (3), 889–909.

Scherer, Roberta W, Joerg J Meerpohl, Nadine Pfeifer, Christine Schmucker, Guido Schwarzer, and Erik von Elm, "Full publication of results initially presented in abstracts," *Cochrane Database of Systematic Reviews*, 2018, (11).

Sgroi, Daniel and Andrew J Oswald, "How should peer-review panels behave?," *The Economic Journal*, 2013, *123* (570), F255–F278.

Simeth, Markus and Michele Cincera, "Corporate science, innovation, and firm value," *Management science*, 2015, *62* (7), 1970–1981.

Song, Fujian, A Eastwood, Simon Gilbody, Lelia Duley, and A Sutton, "Publication and related biases: a review," *Health technology assessment*, 2000, *4* (10).

Stephan, Paula E, "The economics of science," *Journal of Economic literature*, 1996, *34* (3), 1199–1235.

_ , *How economics shapes science*, Vol. 1, Harvard University Press Cambridge, MA, 2012.

Stern, Scott, "Do scientists pay to be scientists?," *Management science*, 2004, *50* (6), 835–853.

Tillio, Alfredo Di, Marco Ottaviani, and Peter Norman Sørensen, "Persuasion bias in science: can economics help?," 2017.

Tse, Tony, Kevin M Fain, and Deborah A Zarin, "How to avoid common problems when using Clinicaltrials. gov in research: 10 issues to consider," *bmj*, 2018, *361*, k1452.

Turner, Lucy, Larissa Shamseer, Douglas G Altman, Kenneth F Schulz, and David Moher, "Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review a," *Systematic reviews*, 2012, *1* (1), 60.

Venkataraman, Sriram and Stefan Stremersch, "The debate on influencing doctors' decisions: are drug characteristics the missing link?," *Management Science*, 2007, *53* (11), 1688–1701.

Viswanathan, Meera, Mohammed T Ansari, Nancy D Berkman, Stephanie Chang, Lisa Hartling, Melissa McPheeters, P Lina Santaguida, Tatyana Shamliyan, Kavita Singh, Alexander Tsertsvadze et al., "Assessing the risk of bias of individual studies in systematic reviews of health care interventions," in "Methods guide for effectiveness and comparative effectiveness reviews [Internet]," Agency for Healthcare Research and Quality (US), 2012.

Wiysonge, Charles S, Hazel A Bradley, Jimmy Volmink, Bongani M Mayosi, and Lionel H Opie, "Beta-blockers for hypertension," *Cochrane database of systematic reviews*, 2017, (1).

Yordanov, Youri, Agnes Dechartres, Raphaël Porcher, Isabelle Boutron, Douglas G Altman, and Philippe Ravaud, "Avoidable waste of research related to inadequate methods in clinical trials," *bmj*, 2015, *350*, h809.

Zarin, Deborah A and Alla Keselman, "Registering a clinical trial in ClinicalTrials. gov," *Chest*, 2007, *131* (3), 909–912.

_ **, Steven N Goodman, and Jonathan Kimmelman**, "Harms From Uninformative Clinical Trials," *Jama*, 2019, *322* (9), 813–814.

_ **, Tony Tse, Rebecca J Williams, and Sarah Carr**, "Trial reporting in ClinicalTrials. govthe final rule," *New England Journal of Medicine*, 2016, *375* (20), 1998–2004.

Figures

Figure 1: Data flow

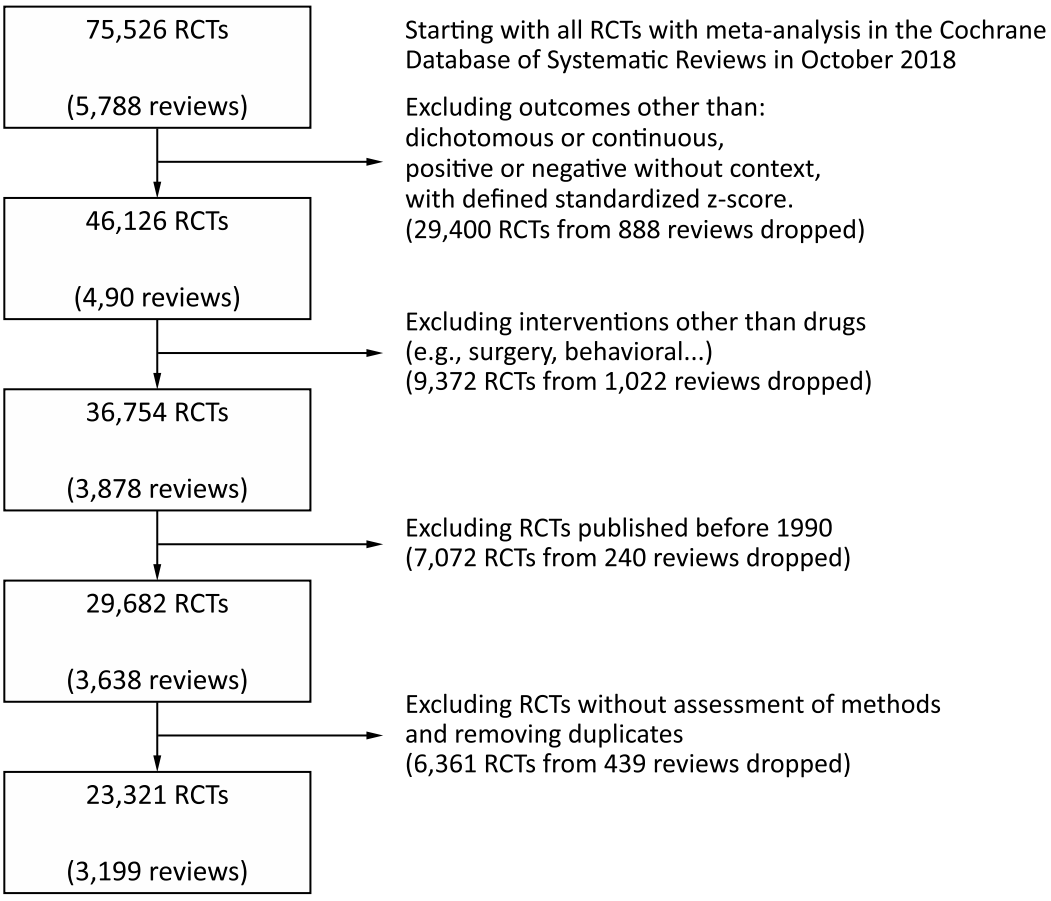


Figure 2: Distribution of inadequacies, by number of inadequacies

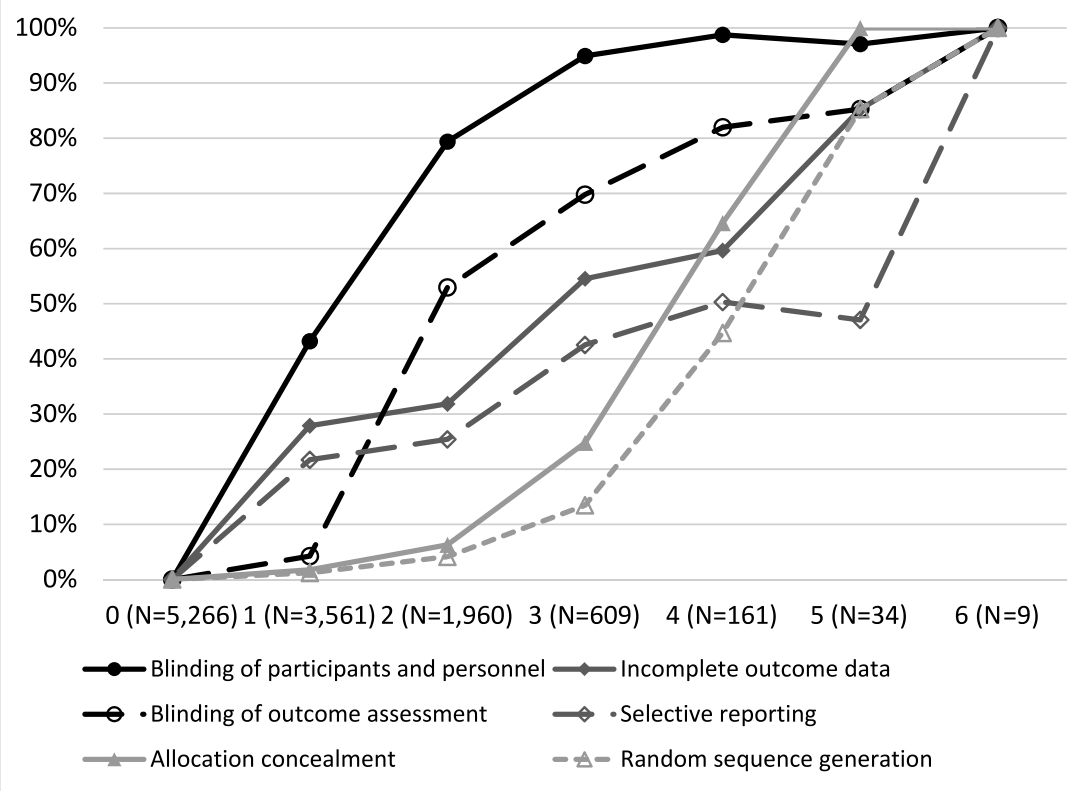
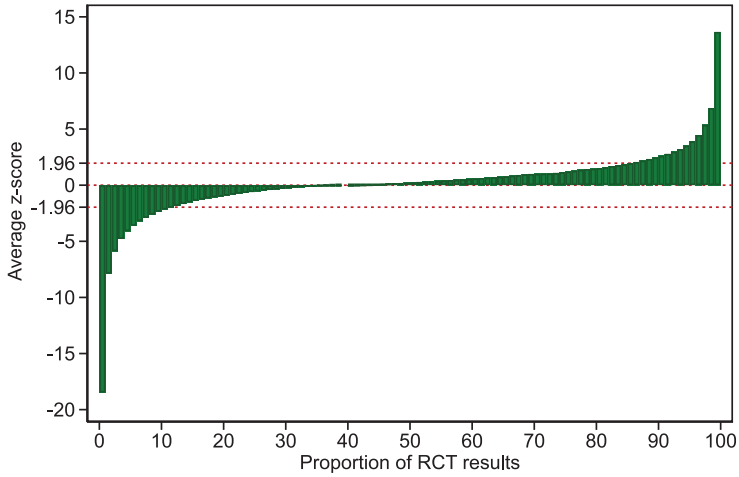


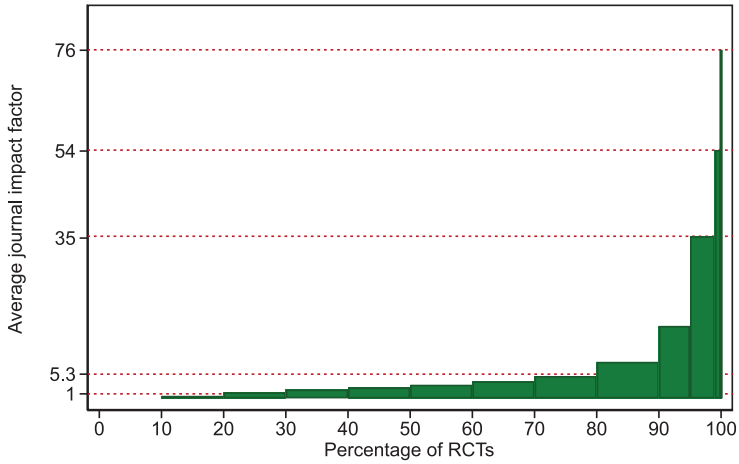
Figure 2 shows the distribution of inadequacies, by number of inadequacies, for all RCTs assessed on six dimensions ($N = 11,600$).

Figure 3: Distribution of experimental results, journal impact factor and annual citations

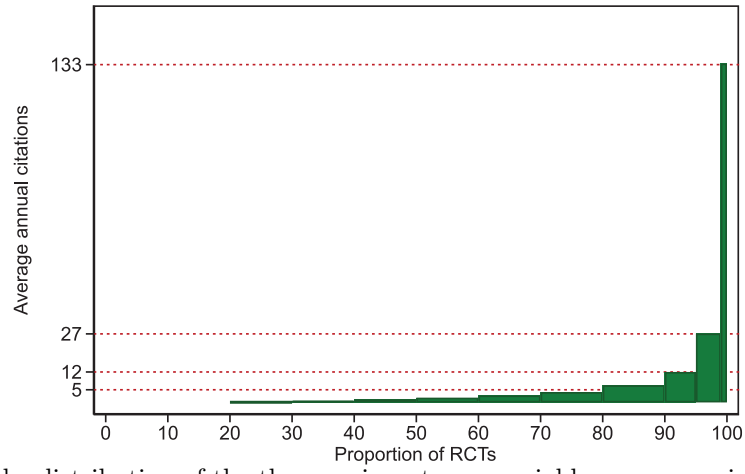
(a) Distribution of RCT results (z-scores)



(b) Distribution of journal impact factor



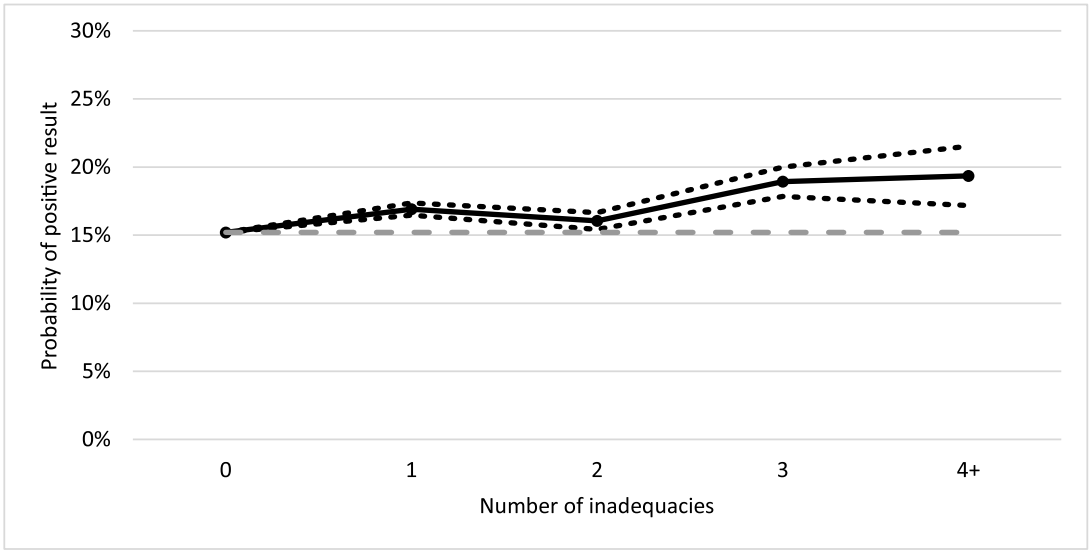
(c) Distribution of annual citations



This figure shows the distribution of the three main outcome variables: z-scores, journal impact factor and total annual citations. Z-scores are classified and coded so that a higher value corresponds to better results. $N = 177,333$ RCT results corresponding to 23,321 RCTs included in 3,199 reviews.

Figure 4: Effect of inadequacies on the probability of positive result

(a) Effect of number of inadequacies on the probability of positive results



(b) Effect of each inadequacy on the probability of positive results (regression with all inadequacies)

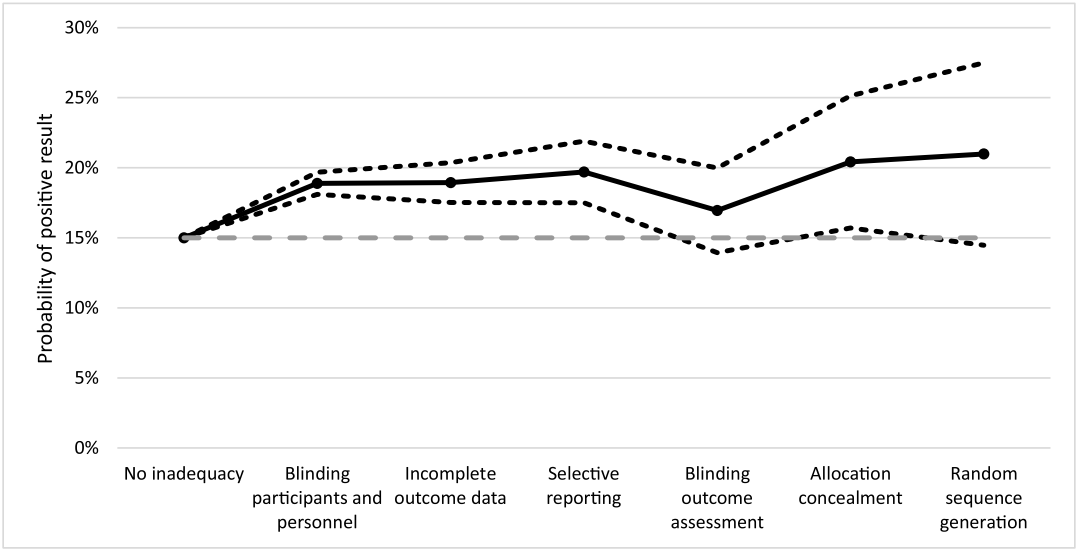
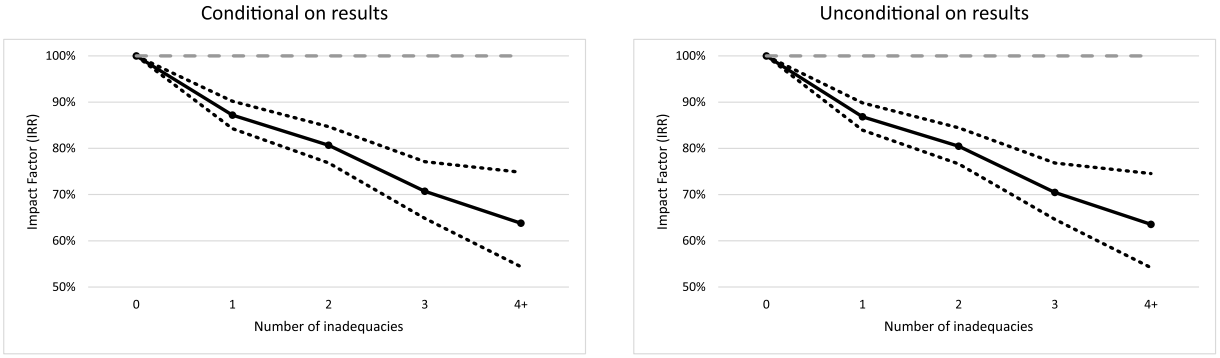


Figure 5: Effect of inadequacies on journal impact factor

(a) Effect of number of inadequacies on journal impact factor



(b) Effect of each inadequacy on journal impact factor (regression with all inadequacies)

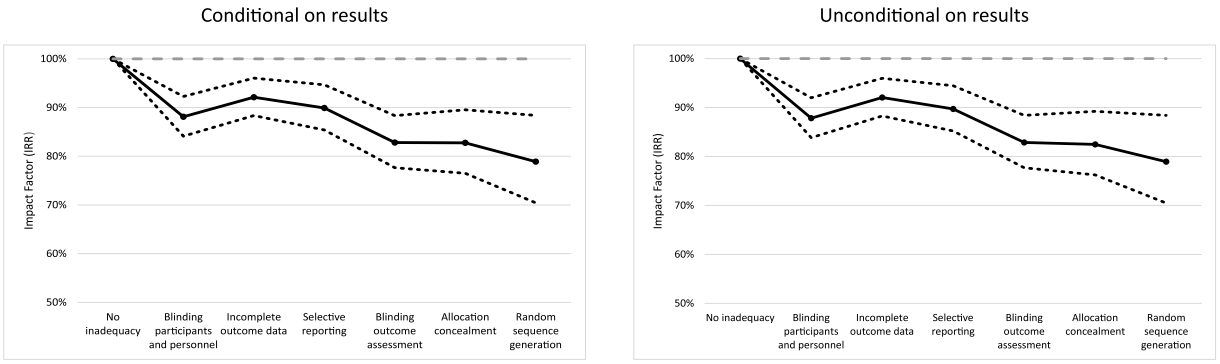
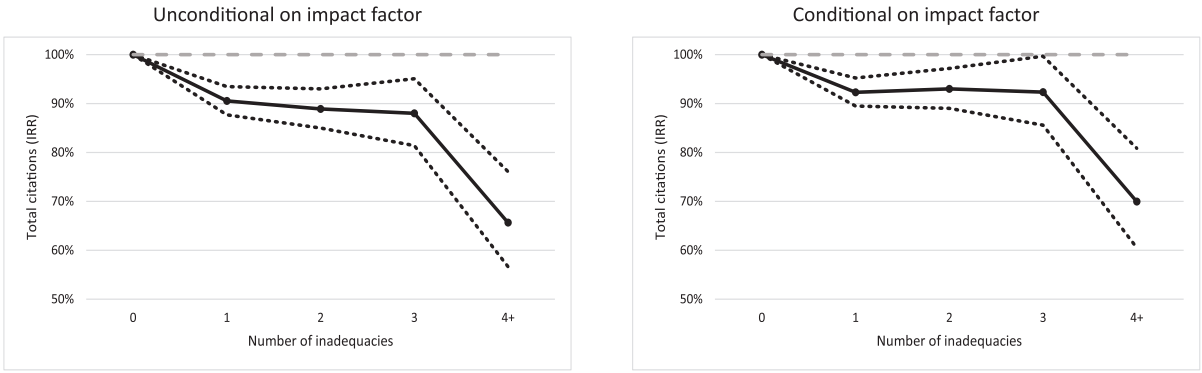
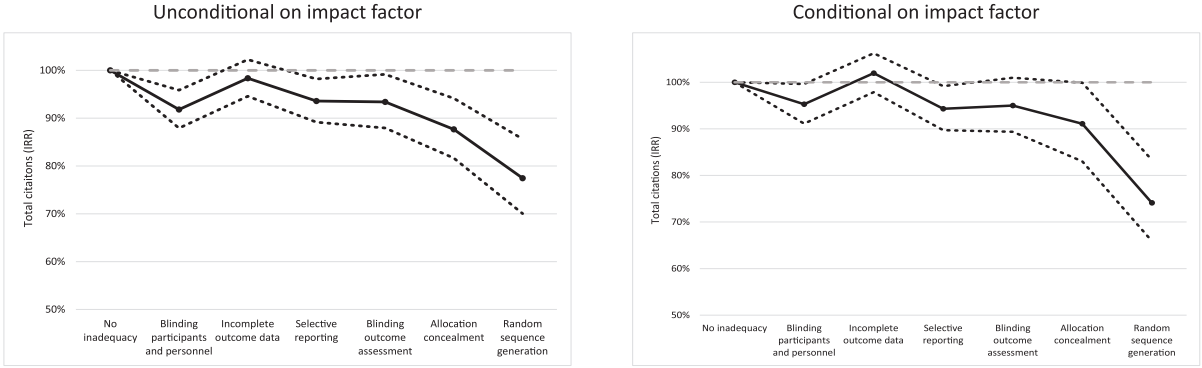


Figure 6: Effect of inadequacies on total citations

(a) Effect of the number of inadequacies on total citations



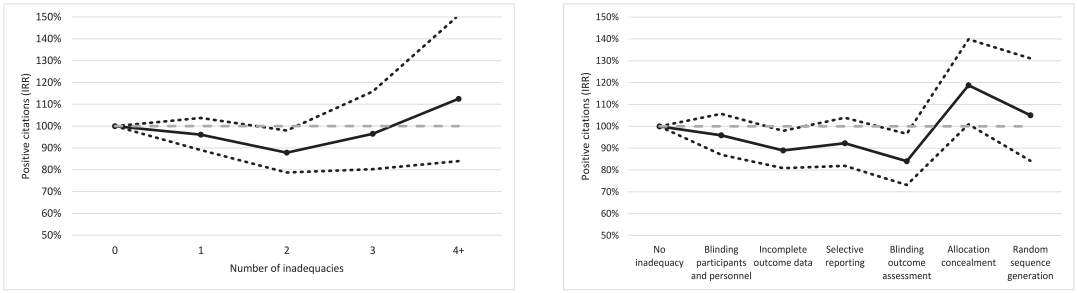
(b) Effect of each inadequacy on total citations



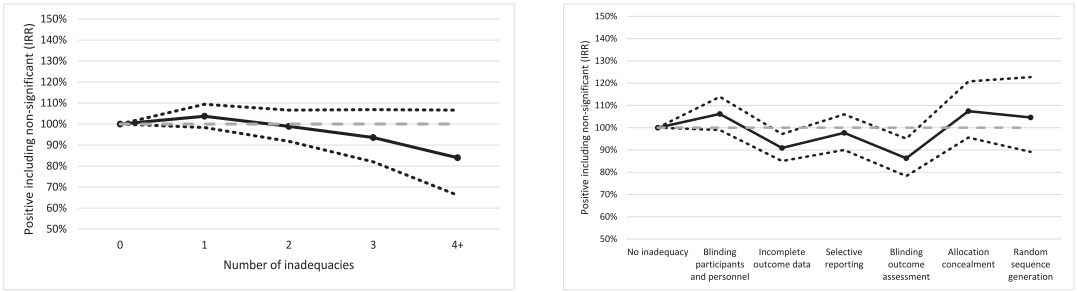
This figure represents the incidence rate ratios from the negatice binomial regressions of total citations on (a) each number of inadequacy, conditional on results with all controls, and (b) each inadequacy, conditional on results, with all controls. In the two figures in the left column, the regressions do not control for journal impact factor. In the two figures on the right column, the regressions control for impact factor (to ask whether inadequacies affect citations when comparing articles published in a similar journal). The coefficients of each of these regressions unconditional on the experimental results of the RCT are very similar (so the figures are not reported here as they are indistinguisable).

Figure 7: Effect of inadequacies on citations to results supporting (or against) the drug

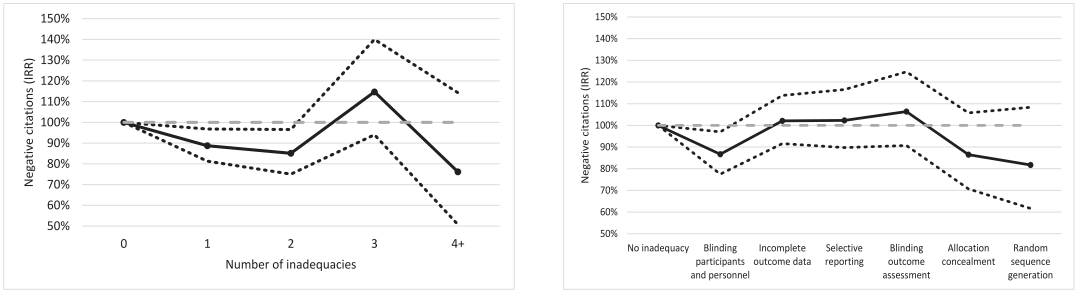
(a) Effect of inadequacies on citations to statistically significant results supporting the drug



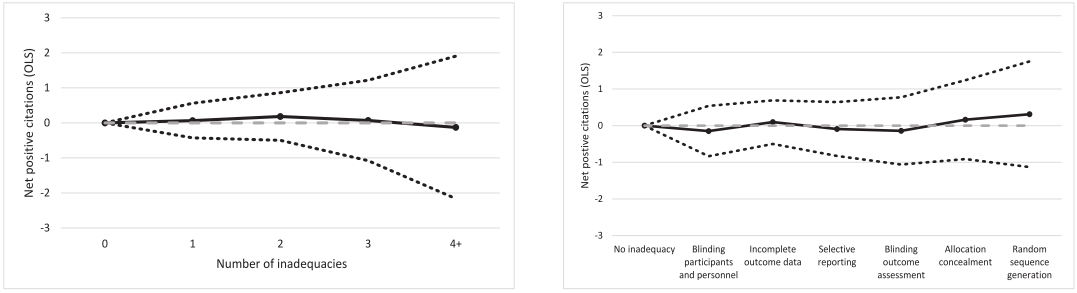
(b) Effect of inadequacies on citations to non-significant results supporting the drug



(c) Effect of inadequacies on citations to statistically significant results against the drug



(d) Effect of inadequacies on net positive citations



Tables

Table 1: Distribution of outcome types

Outcome Type	**Freq.**	**Percent**
Physical Health	56,802	32.03
Adverse Event	44,529	25.11
Death	21,923	12.36
Efficacy (other)	17,955	10.13
Adverse Effect	12,072	6.81
Mental Health	5,922	3.34
Attrition	5,801	3.27
Behavior	5,459	3.08
Utilization	3,568	2.01
Quality Of Life	2,172	1.22
Process	706	0.4
Satisfaction	423	0.24
Total	177,332	100

Table 2: Descriptive statistics

	RCT level ($n = 23,321$)			
	Mean	**Std. dev.**	**Min**	**Max**
Methods				
Adequate*	0.17	0.38	0	1
Inadequate* (ordered by frequency)	0.45	0.50	0	1
Blinding participants and personnel	0.23	0.42	0	1
Incomplete outcome data	0.16	0.36	0	1
Selective reporting	0.11	0.31	0	1
Blinding outcome assessment	0.10	0.30	0	1
Allocation concealment	0.05	0.22	0	1
Random sequence generation	0.03	0.16	0	1
Unclear*	0.37	0.48	0	1
Number of inadequacies	0.68	0.90	0	6
Number of unclear methods	1.74	1.45	0	5
Number inadequate or unclear	2.42	1.71	0	6
Results				
Mean z*	0.16	3.84	−347.8	134.7
Positive	0.15	0.35	0	1
Negative	0.12	0.32	0	1
Null	0.74	0.44	0	1
Scientific impact				
Journal impact factor*	5.34	8.95	0	79.26
Scopus CiteScore (2017)	2.61	2.99	0	15.85
Citations per year	4.78	18.43	0	1047.3
Citations to positive results (annual)	5.38	23.09	0	1047.3
Citations to negative results (annual)	6.22	25.92	0	880.4
Citations to mixed results (annual)	9.31	25.27	0	398
Citations to null results (annual)	3.14	9.19	0	364.9
Affiliation (First author)				
Top University*	0.10	0.30	0	1
Other University*	0.39	0.49	0	1
Top Pharma*	0.01	0.10	0	1
Funding				
NIH*	0.08	0.28	0	1
Other grant*	0.02	0.13	0	1
Industry*	0.05	0.20	0	1
Sample size	349	3,155	4	182,609
Year	2003	7	1990	2018

*Overall Methods are "adequate" if all methods were adequate, "inadequate" if at least one method was inadequate, "unclear" if reviewers were not able to assess from reported information whether methods were adequate or inadequate. All RCT results are standardized, normed and coded so that a higher z-score is better. A binary variable indicates a missing impact factor. Positive cites include citations to articles reporting at least one positive result. Top universities are identified from the AWRU list for 2007. Other universities come from first author affiliation address. Top pharma includes companies with revenue greater than $10 billion since 2011. NIH grant and other grants are retrieved from PubMed. Industry funding is retrieved from the reviews.

Table 3: Effect of number of inadequacies on the probability of positive results

	DV = Significant Positive Result (Logit Model)					
	(1) Main model	**(2) Probabilities**	**(3) Inad. x Sponsor**	**(4) Inad. x post 2005**	**(5) Inad. x post 2007**	**(6) Inad. x post 2012**
Inadequacies						
1	0.17***	0.17***	0.15***	0.15***	0.18***	0.17***
	(0.02)	(0.002)	(0.02)	(0.03)	(0.03)	(0.02)
2	0.10***	0.16***	0.10**	0.12**	0.12**	0.09**
	(0.03)	(0.003)	(0.03)	(0.04)	(0.04)	(0.03)
3	0.26***	0.19***	0.23***	0.08	0.11	0.22**
	(0.05)	(0.005)	(0.05)	(0.06)	(0.06)	(0.05)
4+	0.27***	0.19**	0.32**	0.17	0.17	0.26**
	(0.09)	(0.010)	(0.10)	(0.13)	(0.11)	(0.10)
Sponsor						
NIH	-0.08*		-0.11*	-0.08*	-0.08*	-0.08*
	(0.03)		(0.04)	(0.03)	(0.03)	(0.03)
Pharma	-0.18***		-0.17**	-0.19***	-0.19***	-0.19***
	(0.04)		(0.05)	(0.04)	(0.04)	(0.04)
Post Time				**Post 2005**	**Post 2007**	**Post 2012**
				-0.19***	-0.69*	-0.73*
				(0.05)	(0.31)	(0.32)
Inadequacies x			**NIH**	**Post 2005**	**Post 2007**	**Post 2012**
1			0.08	0.03	-0.06	-0.02
			(0.07)	(0.04)	(0.04)	(0.07)
2			0.06	-0.04	-0.05	0.17
			(0.11)	(0.05)	(0.06)	(0.09)
3			-0.37	0.39***	0.41***	0.31*
			(0.25)	(0.09)	(0.09)	(0.14)
4+			0.10	0.22	0.31	0.02
			(0.70)	(0.18)	(0.19)	(0.41)
Inadequacies x			**Industry**			
1			0.08			
			(0.07)			
2			-0.02			
			(0.09)			
3			0.53			
			(0.17)			
4+			0.10			
			(0.68)			
N Reviews	3,199	3,199	3,199	3,199	3,199	3,199
N Obs.	175,025	175,025	175,025	175,025	175,025	175,025

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Observations are RCT results, compared within review. All models report the logistic regression of a binary indicator for a significant positive result (a result supporting the treatment compared to the comparator) and control for sample size and include fixed effects for outcome type, number of assessments, years and number of comparisons in RCT.

Table 4: Effect of number of inadequacies on journal placement

	DV = Journal Impact Factor						
	(1) Conditional on results	**(2) Unconditional on results**	**(3) Inad. x Sponsor**	**(4) Inad. x Acad. Prestige**	**(5) Inad. x Firm involvement**	**(6) Inad. x Novelty (First)**	**(7) Inad. x Novelty (1st-3rd)**
Inadequacies							
1	0.87***	0.87***	0.87***	0.88***	0.86***	0.88***	0.87***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
2	0.81***	0.81***	0.78***	0.77***	0.79***	0.81***	0.82***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
3	0.71***	0.70***	0.76***	0.71***	0.72***	0.70***	0.72***
	(0.03)	(0.03)	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)
4+	0.64***	0.64***	0.62***	0.56***	0.60***	0.68***	0.63***
	(0.05)	(0.06)	(0.05)	(0.07)	(0.05)	(0.06)	(0.06)
Sponsor							
NIH	1.40***	1.40***	1.44***				
	(0.03)	(0.03)	(0.04)				
Industry	1.17***	1.20***	1.10*				
	(0.04)	(0.04)	(0.05)				
Academic Prestige							
Top Univ.	1.24***	1.24***		1.30***			
	(0.03)	(0.03)		(0.04)			
Other Univ.	1.08***	1.08***		1.08***			
	(0.02)	(0.02)		(0.02)			
Firm involvement (Top pharma first author versus other firm involvement)							
Top First					0.98		
					(0.76)		
Other					1.11*		
					(0.06)		
Novel						**First**	**1st-3rd**
						1.37***	1.34***
						(0.05)	(0.05)
Inadequacies x			**NIH**	**Top Univ.**	**Top P. 1st**	**First**	**1st-3rd**
1			0.92*	1.06*	1.05	0.94*	0.99
			(0.04)	(0.05)	(0.14)	(0.04)	(0.03)
2			0.98	1.19*	1.26*	0.95*	0.94*
			(0.07)	(0.09)	(0.24)	(0.06)	(0.04)
3			0.60*	0.94	0.36*	1.00	0.95
			(0.10)	(0.14)	(0.34)	(0.13)	0.08)
4+			1.83*	1.08	1.00	0.66	1.04
			(0.07)	(0.45)	(.)	(0.14)	(0.17)
Inadequacies x			**Industry**	**Other U.**	**Other F.I.**		
1			1.10*	0.95*	1.10*		
			(0.07)	(0.03)	(0.08)		
2			1.17*	1.02	1.17*		
			(0.10)	(0.05)	(0.11)		
3			0.74*	0.99	0.77*		
			(0.13)	(0.09)	(0.14)		
4+			1.16	1.23*	1.24*		
			(0.27)	(0.20)	(0.29)		
N Reviews	3,103	3,103	3,103	3,103	3,103	3,103	3,103
N Obs.	23,321	23,321	23,321	23,321	23,321	23,321	23,321

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Observations are RCT results, compared within review. The table reports the incidence rate ratios (IRR) from the negative binomial regressions of impact factor on the number of inadequacies, with controls and interactions. The reference for RCTs with zero inadequacies is IRR=1. OLS models predicting the logarithm of impact factor plus one yield similar results (see appendix table A5). All models include all controls as well as outcome, order of entry, and year fixed effects.

Table 5: Effect of number of inadequacies on total citations

	Citations Received Through 2018 (Web of Science)							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Model	Negative Binomial (IRR) DV = Total citations (exposure: age)				OLS DV = log(Citations by year + 1)			
Conditional on Results	Yes	Yes	No	No	Yes	Yes	No	No
Controlling for IF	No	Yes	No	Yes	No	Yes	No	Yes
Reference:								
0 inad	1	1	1	1	1.06	1.06	1.06	1.06
# inadequacies								
1	0.91***	0.92***	0.90***	0.92***	−0.08***	−0.05***	−0.08***	−0.05***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
2	0.89***	0.93***	0.89***	0.93**	−0.11***	−0.07***	−0.12***	−0.07***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)
3	0.88***	0.92*	0.88***	0.93	−0.19***	−0.12***	−0.19***	−0.12***
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
4+	0.66***	0.70***	0.65***	0.70***	−0.30***	−0.24***	−0.31***	−0.24***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Sponsor								
NIH	1.50***	1.32***	1.49***	1.33***	0.40***	0.28**	0.39***	0.28***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)
Industry	1.25***	1.22***	1.28***	1.24***	0.22***	0.20***	0.24***	0.20***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
First Author								
Top University	1.20***	1.14***	1.20***	1.12***	0.17***	0.10***	0.17***	0.10***
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Other University	1.09***	1.08***	1.09***	1.08***	0.04***	0.03**	0.04***	0.03**
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Top Pharma	1.10	1.11	1.12	1.13*	0.09	0.10*	0.10*	0.10*
	(0.06)	(0.06)	(0.07)	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)
N Reviews	3,103	3,103	3,103	3,103	3,103	3,103	3,103	3,103
N Observations	23,308	23,308	23,308	23,308	23,321	23,321	23,321	23,321

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. The regressions predicting citations use article age in 2018 as the time exposure variable and control for order in review, prestige (first author affiliation), log sample size, number of dimensions assessed, missing journal impact factor, outcome type fixed effects, year fixed effects.

Table 6: Effect of the number of inadequacies on citations to results supporting (or against) the drug

	DV = Annual citations						
	Total citations		Positive citations		Negative citations		Net citations
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	NB (IRR)	OLS	NB (IRR)	OLS	NB (IRR)	OLS	OLS
Number of inadequacies							
1	0.90***	−0.08***	0.96	−0.01	0.88**	−0.03**	−0.01
	(0.01)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.01)
2	0.89***	−0.12***	0.88*	−0.03*	0.85*	−0.03**	−0.03
	(0.02)	(0.02)	(0.05)	(0.01)	(0.05)	(0.01)	(0.01)
3	0.88**	−0.19***	0.96	−0.05*	1.14	−0.04	−0.05
	(0.03)	(0.03)	(0.09)	(0.02)	(0.12)	(0.02)	(0.03)
4+	0.65***	−0.31***	1.12	−0.02	0.76	−0.07	−0.01
	(0.05)	(0.05)	(0.17)	(0.04)	(0.16)	(0.04)	(0.04)
Sponsor							
NIH	1.49***	0.39***	0.79***	0.01	1.08	0.08***	0.02
	(0.03)	(0.02)	(0.05)	(0.02)	(0.07)	(0.01)	(0.02)
Industry	1.28***	0.24***	1.20*	0.06*	1.19	0.06*	0.06*
	(0.04)	(0.03)	(0.09)	(0.02)	(0.10)	(0.02)	(0.03)
First Author Affiliation							
Top University	1.20***	0.17***	1.02	0.04**	1.17**	0.07***	0.05**
	(0.03)	(0.02)	(0.06)	(0.02)	(0.07)	(0.01)	(0.02)
Other University	1.09***	0.04***	1.12**	0.04***	1.07	0.02	0.03**
	(0.02)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)	(0.01)
Top Pharma	1.12	0.10*	0.58**	−0.02	1.08	0.00	-0.03
	(0.07)	(0.05)	(0.11)	(0.04)	(0.17)	(0.04)	(0.04)
N Reviews	3,103	23,308	3,103	23,308	3,103	23,308	23,308
N Observations	23,308	3,103	23,308	3,103	23,308	3,103	3,103

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. All models control for log sample size, number of assessments, order in review, missing first author affiliation and include year fixed effects. The models do not control for impact factor (to get the net effect of inadequacies on citations). The exposure variable in the negative binomial model is the age of the RCT in 2018.

Appendix

Appendix Figures

Figure A1: Distribution of results by main outcome category

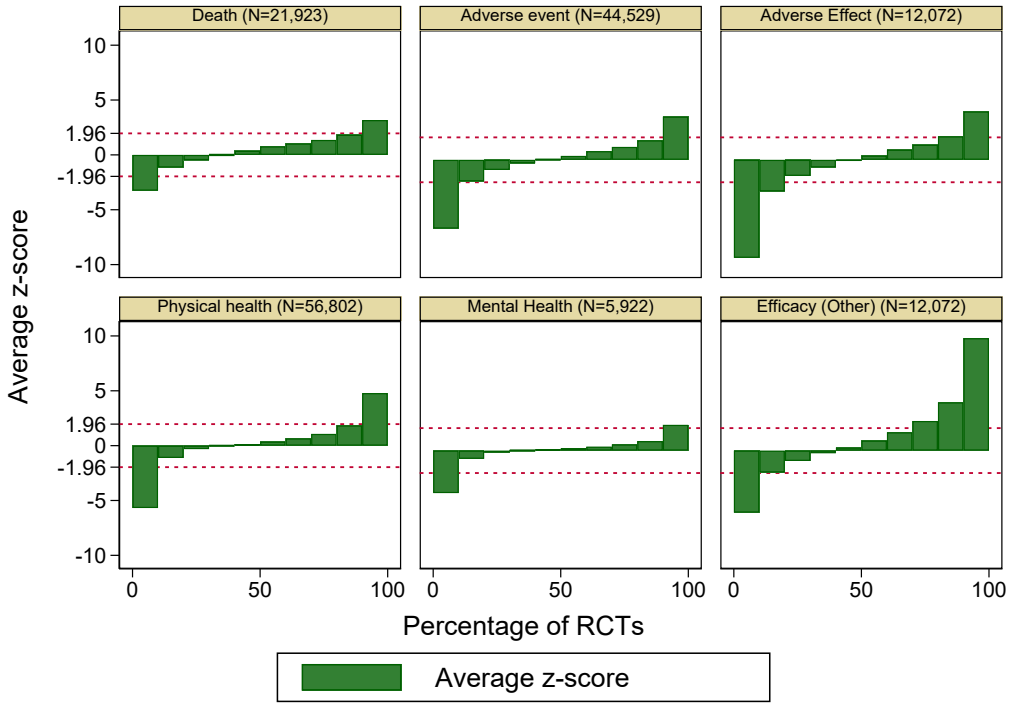
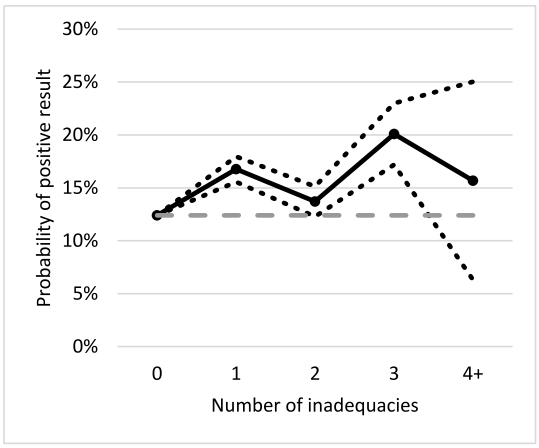
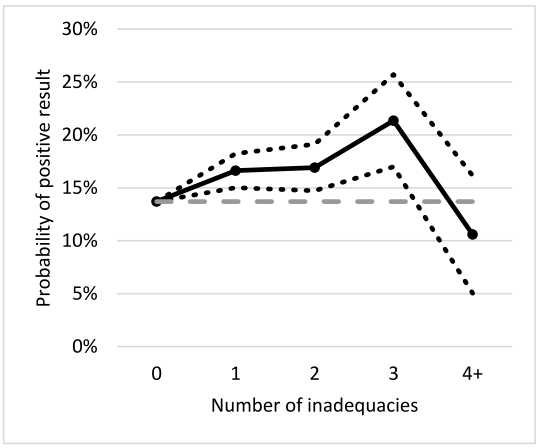


Figure A2: Effect of number of inadequacies on the probability of positive result by outcome type

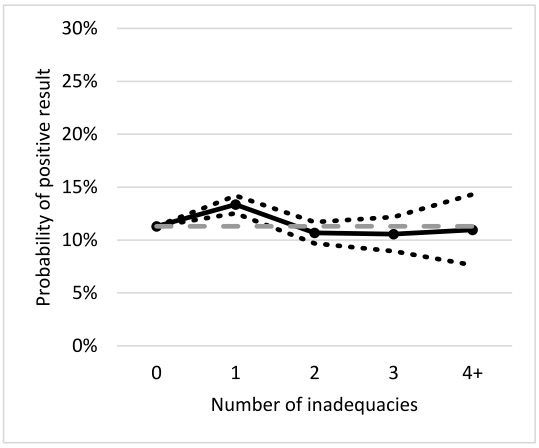
(a) Death



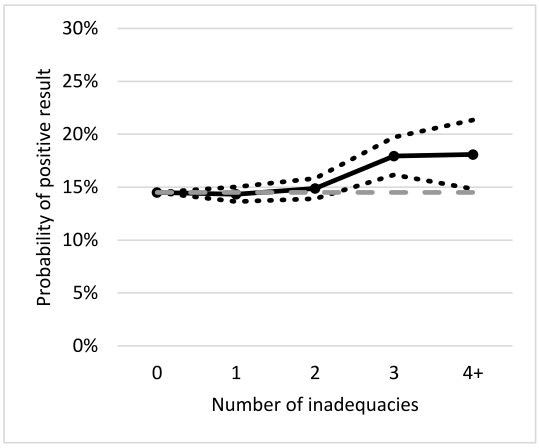
(b) Adverse effect



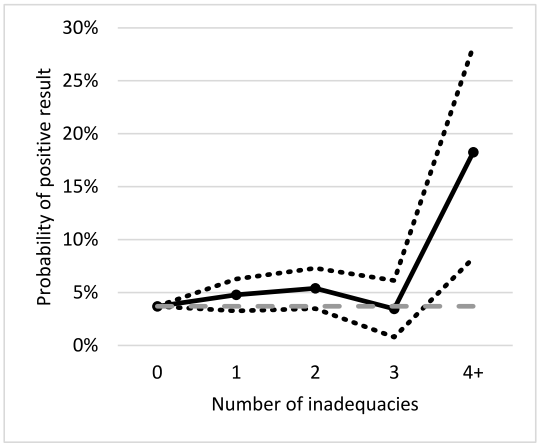
(c) Adverse event



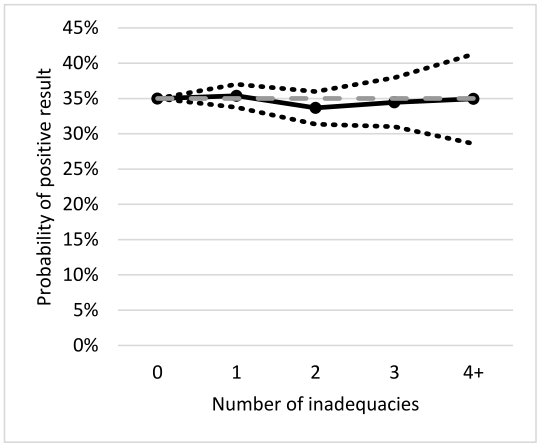
(d) Physical health



(e) Mental Health



(f) Efficacy (other outcomes)



Appendix Tables

Table A1: Definition of clinical trial standards (Cochrane Risk of Bias Assessment Tool)

Bias domain	**Source of bias**	**Support for judgement**	**Review Authors judgement**
Selection bias	Random sequence generation	Describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups.	Selection bias (biased allocation to intervention) due to inadequate generation of a randomized sequence.
Selection bias	Allocation concealment	Describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen before or during enrollment.	Selection bias (biased allocation to interventions) due to inadequate concealment of allocations before assignment.
Performance bias	Blinding of participants and personnel	Describe all measures used, if any, to blind participants and researchers from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.	Performance bias due to knowledge of the allocated intervention by participants and personnel during the study.
Detection bias	Blinding of outcome assessment	Describe all measures used, if any, to blind outcome assessment from knowledge of which intervention a participant received. Provide any information relating to whether the intended blinding was effective.	Performance bias due to knowledge of the allocated intervention by outcome assessment.
Attrition bias	Incomplete outcome data	Describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attritions and exclusions were reported, the numbers in each intervention group (compared with total randomized participants), reasons for attrition or exclusion where reported, and any re-inclusion in the analysis for the review.	Attrition bias due to amount, nature or handling of incomplete outcome data
Reporting bias	Selective reporting	State how selective outcome reporting was examined and what was found.	Reporting bias due to selective outcome reporting.
Other bias	Anything else, pre-specified	State any important concerns about bias not covered in other domains in the tool.	Bias due to problems not covered elsewhere.

Source: Adapted from Higgins et al, 2011

Table A2: Definition of variables

Variable	Definition	Origin, type and values
Methods		
Random sequence generation	Cochrane reviewers' assessment for whether participants were assigned to intervention groups on the basis of a chance (random) process characterized by unpredictability.	Origin: Cochrane reviews Type: categorical variable Values: adequate, inadequate, unclear, missing
Allocation concealment	Cochrane reviewers' assessment for whether appropriate mechanisms were used to prevent foreknowledge of treatment assignment and prevent those who enroll participants from being influenced by this knowledge.	Origin: Cochrane reviews Type: categorical variable Values: adequate, inadequate, unclear, missing
Blinding of participants and personnel	Cochrane reviewers' assessment for whether appropriate mechanisms were used to withhold information about the assigned interventions from participants and personnel.	Origin: Cochrane reviews Type: categorical variable Values: adequate, inadequate, unclear, missing
Blinding of outcome assessment	Cochrane reviewers' assessment for whether appropriate mechanisms were used to withhold information about the assigned interventions from outcome assessors.	Origin: Cochrane reviews Type: categorical variable Values: adequate, inadequate, unclear, missing
Incomplete outcome data	Cochrane reviewers' assessment for whether appropriate measures were taken to prevent missing outcome data, due to attrition (drop-out) during the study or exclusions from the analysis to bias the results.	Origin: Cochrane reviews Type: categorical variable Values: adequate, inadequate, unclear, missing
Selective reporting	Cochrane reviewers' assessment for whether results were selectively reported.	Origin: Cochrane reviews Type: categorical variable Values: adequate, inadequate, unclear, missing
Number of dimensions assessed	Count variable for how many of the six standards were assessed in a Cochrane review.	Origin: Cochrane reviews Type: count variable. Values: 1 to 6.
Number of adequate methods	Count variable for how many of the dimensions were assessed as "adequate".	Origin: Cochrane reviews Type: count variable. Values: 0 to 6.
Number of inadequacies	Count variable for how many of the dimensions were assessed as "inadequate".	Origin: Cochrane reviews Type: count variable. Values: 0 to 6.
Number of unclear assessments	Count variable for how many of the dimensions were assessed as "unclear".	Origin: Cochrane reviews Type: count variable. Values: 0 to 6.
Adequate methods	RCT methods are adequate if all methods assessed are "adequate".	Origin: Cochrane reviews Type: Binary variable. Values: 0 / 1
Inadequate methods	RCT methods are inadequate if at least one method is assessed as "inadequate".	Origin: Cochrane reviews Type: Binary variable. Values: 0 / 1
Unclear methods	RCT methods are unclear if the available information does not allow the classification of RCT methods as "adequate" or "inadequate".	Origin: Cochrane reviews Type: Binary variable. Values: 0 / 1

Variable	Definition	Origin, type and values
Results		
Outcome	Outcome measured in the treatment and control groups, as stated in the statistical file of the meta-analysis.	Origin: Cochrane reviews, statistical file Type: Free text Values: death prior to hospital discharge, stroke, blood pressure, length of stay
Outcome category	Classification of "outcome" variable, based on clustering of similar outcomes (e.g., mortality, death before hospital discharge, maternal mortality, survival) into outcome categories (e.g., death), and coding to identify positive and negative outcomes (e.g., survival versus death).	Origin: Cochrane reviews, statistical file (from "Outcome") Type: Categorical variable Values: death, physical health, adverse event, efficacy (other), adverse effect, mental health, attrition, behavior, utilization, quality of life, process, satisfaction.
Results: comparison level		
z-score	Z-scores for dichotomous and continuous outcomes are calculated using the following formulas. Categorical outcomes: p_1, p_2 the proportion of successes in the treatment and control groups, n_1, n_2 the size of the treatment and control groups, and p the proportion of successes in the two samples combined: $$Z = \frac{p_1 - p2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ Continuous outcomes: μ_1, μ_2 are the mean in the treatment and control groups, σ_1, σ_2 are the standard deviation in the treatment and control groups, n_1, n_2 are the sample sizes in the treatment and control groups. $$Z = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$	Origin: Cochrane reviews, statistical file Type: Continuous variable Values: from -1000 to 1000.
Significant positive result	Indicator for whether $Z > 1.96$	Origin: Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1
Significant negative result	Indicator for whether $Z < -1.96$	Origin: Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1

Variable	Definition	Origin, type and values
Null result	Indicator for whether $-1.96 \leq Z \leq 1.96$	Origin: Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1
Results: RCT level		
Number of results in RCT	Number of treatment, control, outcome, subgroup level comparisons in RCT.	Origin: Cochrane reviews, statistical file Type: Count variable Values: 1,2,3...
Positive RCT	Indicator for whether all RCT results are positive or null.	Origin: Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1
Negative RCT	Indicator for whether all RCT results are negative or null.	Origin: Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1
Mixed RCT	Indicator for whether the RCT includes both significantly positive results and significantly negative results (e.g., the treatment significantly reduces death but also significantly increases side effects).	Origin: Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1
Null RCT	Indicator for whether all RCT results are null.	Origin: Derived, Cochrane reviews, statistical file Type: Binary variable Values: 0 / 1
Scientific impact		
Journal impact factor	Journal impact factor	Origin: WOS (new: Scopus)
Citations	Citations through 2018	Origin: WOS (new: Scopus)
Citations to positive results	Citations * 1[Positive RCT]	Origin: Derived
Citations to negative results	Citations * 1[Negative RCT]	Origin: Derived
Citations to null results	Citations * 1[Null RCT]	Origin: Derived
Citations to mixed results	Citations * 1[Mixed RCT]	Origin: Derived
Citation weighted positive results	Citations * 1[Positive RCT] + Citations * -1[Negative RCT]	Origin: Derived
Bibliometric data		
Publication year	Publication year	Origin: PubMed
First author affiliation	Free text of first author address	Origin: PubMed

Variable	Definition	Origin, type and values
Top university	See list in Appendix	Origin: Derived, PubMed
Other university	University not in top university list	Origin: Derived, PubMed
Top pharma	See list in Appendix. Top pharma defined by revenue threshold. From free text of author affiliation.	Origin: Derived, PubMed
Funding		
NIH	Indicator for NIH grant.	Binary variable. Origin: Derived, PubMed
Other grant	Indicator for other grant.	Binary variable. Origin: Derived, PubMed
Industry	Indicator for industry funding listed in article reported in Cochrane review.	Binary variable. Origin: Derived, Cochrane.
Novelty		
Order of entry in review	Variable representing the chronological rank of the RCT as compared to other RCTs on the same topic based on publication year within review (if the first study on the topic was published in 2000, its order is 1. If the next study was published in 2003, its order is 2...).	Ordinal variable Origin: Derived, Cochrane.
Topic maturity	Age of the first study published on the topic. For instance, if the first study was published in 1990, the topic maturity in 2018 is 28. If the first study was published in 2000, the topic maturity in 2018 is 18.	Integer Origin: Derived, Cochrane.
Sample size	Study sample size including patients in the treatment and control groups	Origin: Cochrane, statistical file.

Table A3: Relationship between inadequacies, pseudo R-squared

	Random Sequence Generation	**Allocation Concealment**	**Blinding Patients & Personnel**	**Blinding Outcome Assessment**	**Incomplete Outcome Data**
Random Sequence Generation	1				
Allocation Concealment	**0.20**	1			
Blinding Patients and Personnel	0.00	0.02	1		
Blinding Outcome Assessment	0.01	0.01	**0.12**	1	
Incomplete Outcome Data	0.00	0.00	0.00	0.00	1
Selective Reporting	0.00	0.00	0.00	0.00	0.01

Each cell reports the pseudo R-squared from the Logit regression of the indicator for one type of inadequacy (e.g., random sequence generation) on the indicator for another type of inadequacy (e.g., allocation concealment).

Table A4: Effect of each inadequacy on experimental results and journal placement

	Positive result		**Journal impact factor**	
	(1)	**(2)**	**(3)**	**(4)**
	Logit	**Probability**	**Negative Binomial**	**OLS**
Inadequacies				
Blinding	0.42***	0.04***	0.88*	−0.09***
Part. & Pers.	(0.04)	(0.004)	(0.02)	(0.02)
Incomplete	0.01	0.00	0.92***	−0.08***
Outcome Data	(0.03)	(0.003)	(0.02)	(0.01)
Selective	0.08	0.01	0.90***	−0.08***
Reporting	(0.04)	(0.003)	(0.02)	(0.02)
Blinding	-0.31***	-0.03***	0.83***	−0.14***
Outcome Assess.	(0.05)	(0.004)	(0.03)	(0.02)
Allocation	0.37***	0.03***	0.82***	−0.17***
Concealment	(0.09)	(0.004)	(0.03)	(0.03)
Random	0.06	0.01	0.79***	−0.18***
Seq. Generation	(0.10)	(0.009)	(0.04)	(0.04)
N Reviews	1,469	1,469	3,103	3,103
N observations	94,535	94,535	23,321	23,321

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Observations are compared within reviews. In columns 1 and 2, an observation is an RCT result, and the regression includes all RCTs assessed on all dimensions. The regression controls for sample size and includes year, outcome, and number of comparisons fixed effects. In columns 3 and 4, and observation is an RCT. The models control for sample size, number of assessments, missing assessment, order of entry, academic prestige (first author affiliated with top university or other university), top pharma affiliation, sponsorship and include year and order of entry fixed effects.

Table A5: Effect of number of inadequacies on journal placement (OLS)

	DV = log (Journal Impact Factor + 1)						
	(1) Conditional on results	**(2) Unconditional on results**	**(3) Inad. x Sponsor**	**(4) Inad. x Acad. Prestige**	**(5) Inad. x Firm involvement**	**(6) Inad. x Novelty (First)**	**(7) Inad. x Novelty (1st-3rd)**
Inadequacies							
1	-0.05***	-0.06***	-0.10***	-0.09***	-0.12***	-0.10***	-0.11***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)
2	-0.07***	-0.07***	-0.17***	-0.15***	-0.19***	-0.16***	-0.14***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
3	-0.12***	-0.13***	-0.22***	-0.22***	-0.26***	-0.26***	-0.24***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)
4+	-0.24***	-0.24***	-0.38***	-0.39***	-0.41***	-0.33***	-0.40***
	(0.05)	(0.05)	(0.05)	(0.07)	(0.06)	(0.06)	(0.07)
Sponsor							
NIH	0.28***	0.27***	0.44***				
	(0.02)	(0.02)	(0.02)				
Industry	0.20***	0.21***	0.09*				
	(0.03)	(0.03)	(0.04)				
Academic Prestige							
Top Univ.	0.10***	0.10***		0.29***			
	(0.02)	(0.02)		(0.02)			
Other Univ.	0.03**	0.03**		0.08***			
	(0.01)	(0.01)		(0.02)			
Firm involvement (Top pharma first author versus other)							
Top First					-0.23		
					(0.07)		
Other					0.11*		
					(0.04)		
Novel						**First**	**1st-3rd**
						0.39***	0.38***
						(0.03)	(0.03)
Inadequacies x			**NIH**	**Top Univ.**	**Top P. 1st**	**First**	**1st-3rd**
1			-0.12*	-0.06	-0.02	-0.05*	-0.01
			(0.04)	(0.04)	(0.01)	(0.03)	(0.02)
2			-0.09*	0.03	0.20*	-0.06*	-0.08*
			(0.06)	(0.06)	(0.17)	(0.05)	(0.03)
3			-0.56***	-0.16*	-0.76*	-0.02	-0.08*
			(0.12)	(0.11)	(0.45)	(0.08)	(0.06)
4+			0.50*	-0.12	0	-0.26*	0.04
			(0.35)	(0.27)	(.)	(0.13)	(0.10)
Inadequacies x			**Industry**	**Other U.**	**Other F.I.**		
1			0.05	-0.06*	0.04*		
			(0.05)	(0.02)	(0.06)		
2			0.11	-0.07*	0.09*		
			(0.07)	(0.03)	(0.08)		
3			-0.28*	-0.12*	-0.25*		
			(0.12)	(0.06)	(0.13)		
4+			0.04	0.01	0.00		
			(0.17)	(0.10)	(0.17)		
N Reviews	3,103	3,103	3,103	3,103	3,103	3,103	3,103
N Obs.	23,321	23,321	23,321	23,321	23,321	23,321	23,321

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Observations are RCT results, compared within review. The table report the OLS coefficients from the regressions of impact factor on the number of inadequacies, with controls and interactions. All models include all controls (including those non reported), as well as outcome, order of entry, and year fixed effects.

Table A6: Effect of each inadequacy on citations (no control for impact factor)

	Total citations		**Positive citations**		**Negative citations**		**Net citations**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	NB (IRR)	**OLS**	**NB (IRR)**	**OLS**	**NB (IRR)**	**OLS**	**OLS**
Inadequacies							
Blinding	0.91***	−0.08***	0.97	−0.01	0.86**	−0.03	−0.00
Part. & Pers.	(0.02)	(0.02)	(0.05)	(0.01)	(0.05)	(0.01)	(0.02)
Incomplete	0.98	−0.04**	0.89*	−0.03*	1.02	0.01	−0.02
Outcome Data	(0.02)	(0.01)	(0.04)	(0.01)	(0.06)	(0.01)	(0.01)
Selective	0.94**	−0.11***	0.93	−0.02	1.02	−0.03	−0.02
Reporting	(0.02)	(0.02)	(0.06)	(0.02)	(0.07)	(0.01)	(0.01)
Blinding	0.93*	−0.05*	0.83**	−0.04*	1.07	−0.01	−0.05∗
Outcome Assess.	(0.03)	(0.02)	(0.06)	(0.02)	(0.09)	(0.01)	(0.02)
Allocation	0.87***	−0.13***	1.19*	−0.00	0.87	−0.03	0.00
Concealment	(0.03)	(0.03)	(0.10)	(0.02)	(0.09)	(0.02)	(0.02)
Random	0.78***	−0.14***	1.06	−0.00	0.82	−0.07*	0.00
Seq. Generation	(0.04)	(0.03)	(0.12)	(0.03)	(0.12)	(0.03)	(0.03)
N Reviews	3,103	3,103	3,103	3,103	3,103	3,103	3,103
N observations	23,308	23,308	23,308	23,308	23,308	23,308	23,308

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. This table presents the net effect of each inadequacy on total citations unconditional on results, and not controlling for impact factor. An observation is an RCT. RCTs are compared within reviews. All models control for log sample size, sponsor and fist author affiliation (top university, other university, top pharma or missing), and include fixed effects for novelty (order of entry in review). The exposure variable in the negative binomial models (NB) is the age of the article in 2018. Columns 1, 3 and 5 report the incidence rate rations (IRR) from the negative binomial regressions of total citations on all inadequacies and controls. Columns 2, 4, 6 and 7 reports the OLS coefficients of the regression of the log of citations plus one on all inadequacies and controls.

List of top 25 universities in clinical medicine and pharmacy (AWRU, 2007)

- Harvard University
- University of California, San Francisco
- University of Washington
- The Johns Hopkins University
- Columbia University
- University of California, Los Angeles
- The University of Texas Southwestern Medical Center at Dallas
- University of Michigan – Ann Arbor
- Karolinska Institute
- University of Pittsburgh
- Stanford University
- Mayo Medical School
- University of Oxford
- University of Minnesota, Twin Cities
- University of Cambridge
- Yale University
- University College London
- The University of Texas M. D. Anderson Cancer Center
- University of Wisconsin – Madison
- Vanderbilt University
- University of Pennsylvania
- Duke University
- University of California, San Diego
- Tufts University
- The Imperial College of Science, Technology and Medicine

List of top pharmaceutical companies used in this study (by revenue)

- Johnson & Johnson
- Roche
- Pfizer
- Novartis
- Bayer
- Merck & Co
- GlaxoSmithKline
- Sanofi
- Abbvie
- Abbott Laboratories
- Eli Lilly & Co
- Amgen
- Bristol-Myers Squibb
- Gilead Sciences
- AstraZeneca
- Teva Pharmaceutical Industries
- Boehringer Ingelheim
- Merck Group
- Novo Nordisk
- Takeda Pharmaceutical
- Allergan plc
- Shire
- Celgene
- Biogen