Universiteit
Leiden
The Netherlands

Standardization in science: effects and issues of guidelines for biomedical reporting

Schniedermann, A.

Citation
Schniedermann, A. (2026, June 11). *Standardization in science: effects and issues of guidelines for biomedical reporting*. Retrieved from https://hdl.handle.net/1887/4304878

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4304878

Note: To cite this publication please use the final published version (if applicable).

Look heah, now, I've got the wuhks of all the old mastahs - the gweat ahchaeologists of the past. I wigh them against each othah - balance the disagweements - analyze the conflicting statements - decide which is pwobably cowwect - and come to a conclusion. That is the scientific method- At least '-patronizingly-' as I see it. How insuffewable cwude it would be to go to Ahctuwus, oah Sol, foah instance, and blundah about, when the old mastahs have covahed the gwound so much moah effectually than we could possibly hope to do.

Lord Dorwin, Chancellor of the galactic empire (in decline).
From Isaac Asimov's Foundation.

Introduction

The greatest scandal in modern science?

Medical research is fundamentally broken, and it has been since somewhere between the 1990s and the 2010s. In 1994, Doug Altman, one of the most prolific medical researchers at that time, proclaimed "the scandal of poor medical research" (Altman, 1994, p. 283). In his short letter to the editor of the famous British Medical Journal (BMJ), he publicly criticized how the pressure to publish has fueled the rise of scientific fraud and misconduct. In his words, "[w]e should be appalled" (Altman, 1994, p. 283) about the misinterpretation of results, selective reporting, unjustified conclusions, and much more. During the following years, there were no signs of improvement. Instead, other researchers chimed in, revealing more and more problems in the way research is done and how findings are communicated. By combining their efforts under the term *metascience*, they began a concerted attempt to gain a systematic understanding of what had happened to the medical sciences and elsewhere (Hardwicke et al., 2020; Penders, 2024; Peterson & Panofsky, 2023). Around 10 years and an impressive record of such studies later, John Ioannidis published his famous essay titled "Why most published research findings are false" (Ioannidis, 2005). In this essay, which was referenced in over 13,000 other publications and which even got a dedicated Wikipedia page,[1] he explains how the problems prevail. He persuasively argues that the biomedical research community is lead astray by a set of fundamental misconceptions about what valuable research results are. In recalling on Doug Altman's early criticisms, other experts conclude that medical research is "still a scandal" (R. Smith, 2014), even decades later (Glasziou & Chalmers, 2018; Sauerbrei et al., 2021). So, what has happened to medical research, which has been credited with some of the greatest advances in improving the quality of life among mankind?

Many have followed the critics' arguments and added their own accusations. They argue that an overfocus on publishable research results, rather than rigorous execution of the right methods, misleads the entire medical research system. Today, research is not

[1] Viz. https://en.wikipedia.org/w/index.php?title=Why_Most_Published_Research_Findings_Are_False&oldid=1230006292 [retrieved on 29 August 2024 14:51 UTC].

performed purely out of curiosity. Having the right results means getting promotions and research grants because modern academia has become so disciplined by research evaluation and performance-based funding, leading to a culture that is often described as "publish or perish" (Sugimoto & Larivière, 2018; see also Moskovkin, 2024). But getting out papers with *the right results* also provides medical companies with access to billion-dollar markets.

Taken together, these contexts create incentives for researchers and research managers alike to conduct and report medical research in a way that fashions study results and emphasizes the positive outcomes while ignoring or downplaying the negative ones (Bekelman et al., 2003; Bothwell et al., 2016; I. Chalmers, 1990; I. Chalmers et al., 2002; I. Chalmers, 2006; Fister, 2005; Head et al., 2015; Hutchison et al., 1995; McGauran et al., 2010; Sismondo, 2008). As a result, medical study reports often fail to accurately reflect the actual outcomes of medical trials and their findings. Rather, the metascientific evidence suggests that the whole medical literature suffers from a reporting bias.

Reporting biases range from the selective reporting of outcomes to the complete non-publication of results.[2] Selective reporting means that a medical publication lacks in crucial details about the study design, its performance, taken decisions, data analysis strategies, or information about potential conflicts of interest (Altman & Moher, 2014; McGauran et al., 2010). For example, an early criticism was that randomized controlled trials fail to report details about how randomization was performed, i.e. if only the whole patient sample was drawn randomly or if also the allocation into treatment or control groups was done randomly. Furthermore, it has also been argued how publications of randomized controlled trials often overestimate efficacy of the interventions they try to test, whilst underestimating the safety risks, especially if those risks could discourage the prescription of a drug, hence drug sales (I. Chalmers, 2006; McGauran et al., 2010; Rennie, 1995; Schulz et al., 1995).

Given that formal scientific communication is often depicted as the backbone of a modern, globally interconnected science system, one might ask with legitimate cause if biased reporting is a fundamental problem in biomedical research. Contemporary scholars have argued that selective reporting can lead to incorrect treatment decisions because doctors cannot accurately assess the relevance or quality of the studies they

[2] In the case of the complete non-publication of results, typically a negative finding (e.g. the treatment has no benefit over a placebo) is kept in the file drawer because journals won't publish it, or companies fear a competitive disadvantage when showing others which of their investments failed.

read. Moreover, healthcare policies may be based on biased evidence. For this reason, biased reporting became a central topic to metascientists and science reformers.

> *Biased under-reporting of research harms and sometimes kills patients, quite apart from the waste of resources that results from this form of scientific and ethical misconduct.*
>
> *(I. Chalmers, 2006, p. 339)*

But even in non-clinical research, biased reporting is a waste of research money and skews the scientific discourse because future research is built on previous research. In that manner, selectively reported trials do not add to a progressing stock of knowledge. Rather, researchers become unable to replicate each other's findings or have to rerun experiments that were performed before but remained unpublished. Therefore, experts concluded how "underreporting research is scientific misconduct" (I. Chalmers, 1990, p. 1405), and that "bad reporting of research breaches moral and ethical standards" (Altman & Moher, 2014, p. 7; see also I. Chalmers, 2006).

So, should we conclude that biomedical research authors are artistic taletellers or even fraudsters that have ruined biomedical sciences and its literature, only to be published, funded, and promoted? Maybe. But all the criticisms and accusations seem to originate from a position that assumes there are the right tools for the job—medical research methods and procedures that can be executed in a proper way. The rigor and value of scientific practice mandate following the right procedures, and whatever the results are, they should be accepted as valid medical knowledge (Peterson & Panofsky, 2023; see also Douglas, 2004). Indeed, the high level of standardization in the clinical sciences has fueled this perception. For example, randomized controlled trials have become the "gold standard" of proving medical efficacy and building medical knowledge (Epstein, 2007; Timmermans & Berg, 2003). Its level of standardization has turned the question of interventional efficacy into a standard problem that can be, and must be, solved before licensing new drugs. The resulting "clinical trial industry" (Meldrum, 2000, p. 755) even replaced academia as the leading producer of randomized controlled trials in the 1990s (Bothwell et al., 2016). Furthermore, meta-analyses and systematic reviews, aggregations and combinations of the vast amount of published research that is currently available, were placed at the top of the hierarchy of evidence. This hierarchy is a clear demonstration that it is rigorous methods and standardized procedures that determine epistemic value (Goldenberg, 2009; Hammersley, 2001; Stegenga, 2011).

One might acquire the impression that such a depiction of scientific value is overly idealistic, overemphasizing the production of research. But then, what is happening in the trial centers around the world? Since the second half of the 20th century, scholars from different disciplines and motivations have entered the laboratories, trial centers, and medical guideline boards to understand the inner workings of these incubators of medical knowledge. By shedding light on the day-to-day practices of researchers, they uncovered the many aspects that are non-standard and uncertain—moments dominated by tinkering rather than planned action (Knorr-Cetina, 1981; Kontopodis et al., 2011; Law, 2017; Suchman, 1985). For example, there are so many methodological decisions during a meta-analysis that lie outside the standard procedure that any idea of a homogeneous practice must be the exception, rather than the norm (Stegenga, 2011). By asking what the actual effects and practical functions of standards are, sociologists found that it is exactly the interplay between standards and contingent procedures that leads to valuable results (e.g., Fujimura, 1987, 1988; Timmermans & Epstein, 2010). Local deviations and adaptations make standards valuable and effective, rather than any formulaic, machine-like form of application (Field & Derksen, 2021; Suchman, 1985). This interplay allows for a pluralistic and diverse epistemic environment in which researchers can go beyond the average and become creative, innovative, and breakthrough discoverers.

Thereby, we might be justified in thinking that many of the depictions of the reporting crisis and the suggested solutions stem from the idealization of research practices. Instead of categorizing research as simply "the good, the bad, and the ugly" (Yuan & Hunt, 2009, p. 1086), it is important to recognize the complexity involved in conducting and reporting research. How could there even be homogeneity if we consider that research projects, albeit similar in design, are performed in different places around the world with different resources and research cultures? In addition, researchers have different backgrounds and types of expertise. These can make them value different aspects and neglect others so that there might be a more mundane idea of proper and improper behaviors, which are more distant from the dramatic descriptions of misconduct (De Vries et al., 2006). If strict science recipes are idealizations, what can we expect from scientific publications? Who considers which details as crucial when reading a study? How can an author even know which details are considered crucial, let alone write appropriately?

Finding understandable descriptions and weaving them into an understandable story is not just a technical task. Even when a technical community like medical researchers has established a standardized vocabulary about patients, diagnoses, and treatments, their language remains contextual and contingent (Bazerman, 1988; Bazerman & Paradis,

1991; Golinski, 1998; Keller, 2011). Texts and their formulations can have a spin that inadvertently assigns prominence to certain study characteristics or findings. Many methodological decisions and epistemic assumptions are made unconsciously but nevertheless shape the scientific language. Debates on causality and causal language are illustrative in this regard (e.g., Han et al., 2022).[3] But contextuality also means that a bit of information important to the author is not important to the reader and vice versa. Study authors further must choose information and level of detail in order to balance understandability with readability (Devereaux et al., 2001; K. C. Elliott, 2020; Rennie, 1995). This perspective rather renders any common understanding as an improbable phenomenon that is worth studying.[4]

However, the medical research community has already taken a different direction. With the 1990s on the rise, researchers, publishing experts, and editors of prolific journals developed and issued formal guidelines to reinstate proper ways of scientific reporting. Reporting guidelines provide rules and checklists for authors about what to include in medical research articles (Altman & Simera, 2016). They seem to refuel traditional pictures and promises of (re)standardization, where reporting bias is an error that can be fixed. While selective reporting has created a gap between the performance of medical trials and descriptions thereof, reporting guidelines can close this gap by demanding authors provide predefined bits of information. Thereby, they promise to make research more transparent and may help in solving any form of reporting crisis (Altman & Moher, 2014; Vazire, 2017).

But with all that is known about the diversity and plurality of medical research and reporting, how is it even possible to find and define a shared set of rules for writing? Reporting guidelines may reiterate idealized depictions of scientific practices, but this time with regard to writing. But what lies below the idealization, and how do reporting guidelines impact the day-to-day writing practices of scientific authors? After all, even the history of the highly standardized design of randomized controlled trials quickly reveals its dynamicity and contextuality (Bothwell et al., 2016; Marks, 1997). So, what is the case for standardization (again)? Or in other words, what does the type of a standard

[3] "Repeat after me, correlation is not causation, correlation is not causation, correlation is not causation … 'Correlation is not causation' is a statistics mantra", says not a scientific article but one in The Guardian, illustrating the publicity for this problem (N. Green, 2012).

[4] For example, the idea that social order requires constant practices of ordering became a basic tenet of social constructivism. In this regard, mutual understanding rather than misunderstanding becomes the important phenomenon that needs to be studied and theorized (Keller, 2011).

and its level of formalization tell us about how the medical research community frames epistemic problems and solutions, and organizes collective action around those?

During the 1990s, reporting checklists for almost every major study type in medicine were developed and distributed (Andrew et al., 1994; see also Rennie, 1995). What followed the first successful attempts was a bull run in the development of ever-new guidelines and their concentration in the "**E**nhancing the **QUA**lity and **T**ransparency **O**f health **R**esearch" network (EQUATOR) in 2008 (Simera et al., 2010). Today, this network lists over 600 different reporting guidelines. The "**CON**solidated **S**tandards **O**f **R**eporting **T**rials" (CONSORT) have been endorsed by hundreds of journals, and some even turned formal compliance into a mandatory requirement before publishing (Begg et al., 1996; Turner et al., 2012). Similarly, the "**P**referred **R**eporting **I**tems for **S**ystematic reviews and **M**eta-Analyses" (PRISMA) (Moher et al., 2009e), which will be the empirical object of my research, has accumulated over 100,000 citations from publications that apply the guideline.

Although this portrayal demonstrates remarkable achievement, it fuels the question of how a research community can maintain its epistemic plurality and creativity if formal standards become pervasive at this level. How can a research system that has often been characterized as highly competitive (Castel, 2009; S. Moore et al., 2017; Vessuri et al., 2014), find its way to agree on rather precise standards across different fields and disciplines? Why did researchers accept these standards? How were reporting guidelines disseminated, distributed and marketed?

Reporting guidelines undoubtedly call for a new chapter in the story of scientific writing and formal scientific communication. They seem to be a culmination point of debates on what counts as valuable science and what the qualities and goals are that scientists should strive for. Understanding how they came about, why they were adopted, and which role they play in solving practical problems is crucial in understanding scientific practices and central elements of scientific communication. Addressing the many questions that emerge around reporting guidelines, we can learn how the biomedical research community defines and solves current and future problems in research. The ways guidelines are formulated and disseminated provide insights into the epistemological self-concept of the biomedical field, which has already transformed traditional epistemic boundaries, such as struggles between qualitative and quantitative methods or barriers between applied and basic research (Roth, 2022). Further, it may help us to project future developments in other scientific disciplines to which biomedicine is a "Leitwissenschaft" (Müller & de Rijcke, 2017, p. 158) in which "new things happen first" (S. Fuller, 2022) and then spill over to other fields, such as STS where

formal scholarly communication increasingly resembles that of the life science (Kaltenbrunner et al., 2022).

I will treat reporting guidelines as a social phenomenon in the biomedical sciences throughout this research. It is based on the already mentioned PRISMA guideline, which it takes as a case that can be illuminated from different perspectives. My research takes several analytical dimensions and conceptual strands from the science studies into account. For instance, traditional assessments of forms of (formal) scientific communication that focus on the tacit and codified forms of communication between researchers, including how scientific publications are produced (Bazerman, 1988; Knorr & Knorr, 1978; Latour, 1987; Shapin & Schaffer, 1985/2011). Other accounts are concerned with scientific work practices and especially with the standardization and proceduralization of research tasks (Fujimura, 1987; Gaudillière & Löwy, 1998; Timmermans & Epstein, 2010; Whitley, 1985/2000), as well as more recent perspectives on the scientific reward system and the effect of current evaluative regimes (Biagioli & Lippman, 2020; Dahler-Larsen, 2014; de Rijcke et al., 2016). Other systemic perspectives involve the proliferation of certain study designs in biomedicine, i.e., medical trials (Marks, 1997), or the configuration and characterization of biomedicine as a distinct research sector (A. E. Clarke et al., 2010; Collins & Pinch, 2005; Roth, 2022).

These perspectives are addressed and combined with the help of a methodological package that combines an analysis of the guideline documents, interviews with their developers and users, and a bibliometric analysis of the use and dissemination of PRISMA. Further, my research does not just add the sociological lens to a new phenomenon in the medical sciences. I would rather attempt to get closer to the metascientific and reform discourses, which bore reporting guidelines and other techniques and novel standardizations of scientific practice. At the same time, the sociological perspective preserves some distance from the normative assumptions of that community (see also Bastian, 2021).

This book offers a deep look into how reporting standards came about and how they transformed the writing and publishing of systematic reviewing. The individual chapters will each provide a different perspective on the transparency crisis. I will provide a lengthy review on scientific writing and how it evolved into a state of crisis during the recent history of medical research in chapter 1. Having set the scene, I will provide reflections on the methodological and theoretical stances that shaped my research in chapter 2. From chapter 3 onwards, we will delve into the case of the PRISMA guidelines by loosely following the sequence of its emergence, dissemination, and use. In chapter 3 we will take a detailed look at the contents and characteristics of the reporting

guideline. Insights from qualitative interviews with PRISMA's developers will shed light on how the guidelines came about and how the community around it professionalized its craft. In chapter 4, I will take a closer look at the role of transparency as an epistemic value and discuss the guideline's justification in the field. This perspective gives us an idea of how the rules inscribed in standards are related to the more abstract goals and promises of broader science reforms. In chapter 5, I make use of the thousands of citations of PRISMA and analyze in which fields and countries the guideline is used. Afterwards I will take a closer look at the users of PRISMA. First, by quantitatively analyzing the publishing careers of review authors who use PRISMA in chapter 6. Finally, by turning towards individual use cases in chapter 5, which helps us to understand how standards are integrated into the day-to-day practice of scientific authors. I will end this book with a broader and more interpretative reflection on my research in chapter 8.