Introduction

Commit ID: c24e56f

2023-10-11

# Introduction

In this chapter I introduce the evidence gap I have addressed and the approach I have taken to address it. I begin by describing the prevalence and consequences of poorly reported medical research. I introduce reporting guidelines as promising solutions but with disappointing impact to date, as reporting quality has hardly improved since their publication.

This brings me to my evidence gap: how can we get more authors to adhere to reporting guidelines? I position reporting guidelines as part of a complex behaviour change intervention and I explain how behaviour change theory offers new and fruitful approaches to answer this question. I then outline my aims, objectives, and thesis structure.

## The problem of poor reporting in health research

Facing an uncertain choice during treatment for multiple myeloma, epidemiologist Alessandro Liberati wrote “Why was I forced to make my decision knowing that information was somewhere but not available?”. We may all ask this question at some point in our lives.

When I started my PhD, the world was in the grip of COVID-19. A deluge of research ensued #REF, but because much of it was poorly reported #REF, governments and clinicians had to make decisions without inaccessible information. In the years since, my loved ones have had to make treatment decisions where the evidence is of “low certainty” because key details are missing from research articles.

A selfish silver-lining of these tumultuous years is that my family and friends finally understand the problem my PhD thesis tries to address: when medical researchers inadequately describe what they did or what they found, other people cannot understand, replicate, or use their work. Research costs huge amounts of time, money, and effort, and the written account is typically its sole legacy. When details are omitted they are lost. The remaining gaps are sources of doubt; are they accidental omissions? Oversights? Cover-ups? Whatever their source, the gaps fragment the full picture, and the potential value to patients drains away.

Early concern over reporting quality often came from frustrated reviewers unable to could the data they needed within research reports. For example, in 1963 Glick [1] found many reports of psychiatric therapy used ambiguous descriptions of treatment duration like “at least two months” or “from one to several months”. These descriptions were so vague they were “unsuitable for comparative purposes”. More recently, Dechartes found systematic reviewers could not judge the potential for bias in a third of trials because of poorly described methods, thereby limiting the confidence of conclusions [2].

Reviewers are not the only people affected. When interventions are poorly described, researchers cannot appraise or repeat research. Carp [3] described how a third of 241 brain imaging studies missed information necessary to interpret and repeat them, like the number of examinations, examination duration, and the resolution of images. Doctors and service providers also need clear descriptions to replicate interventions [4]. As Feinstein noted in 1974 [5], it is difficult enough for a clinician to understand the value of unfamiliar procedure, but “it is much more difficult when he is not told what that procedure was”. For example, Davidson et al. [6] reviewed trials describing exercise interventions for chronic back pain and found authors often did not describe their materials, infrastructure, or training, making it impossible for other healthcare providers to offer the interventions themselves.

These are just some of the many studies documenting poor reporting in medical literature. A 2023 systematic review found 148 published between 2020-2022 alone [7]. All investigated reporting quality in different medical research disciplines, and *almost* all concluded reporting was sub-optimal. Hence, poor reporting plagues all disciplines, devalues research, and derails the uptake of new knowledge into clinical practice.

## Reporting guidelines to the rescue?

Concern over reporting quality crescendoed through the eighties and early nineties as systematic reviews became more common. Responding to concerned academics calling for “strategies”, “guides”, and “lists” to help authors prepare their manuscripts, a group of methodologists, trialists, and editors met in 1996 to create the CONsolidated Standards of Reporting Trials (CONSORT) statement [**beggImprovingQualityReporting1996?**]. CONSORT is a set of recommendations detailing what information authors should include in clinical trial reports. It comprised an article describing how it was made, a checklist, flow diagram, and (after an update in 2001) and ‘Explanation and Elaboration’ publication #REF.

CONSORT proved influential, and other groups quickly developed guidelines for different research types. Reporting guidelines are like a theme and variations, where CONSORT forged a path others have followed with varying fidelity (See **?@tbl-rgs**). Most have acronym names. Most were first published as a journal article describing their development. Some, but not all, have checklists and elaboration documents. Some guideline developers publish resources as separate documents, others put them all into a single journal article. Guidelines are developed by different groups, with different composition (possibly including methodologists, editors, clinicians etc) and in different ways (e.g., some by delphi consensus). Although most follow CONSORT’s approach of presenting recommendations focussing on reporting above conduct, guidelines differ in how forceful their recommendations are and whether they also seek to influence design.

| Guideline acronym | Definition | Applicable study type | Publication year | Has Development Article? | Has checklist? | Has explanatory document? | Development process | Focus on design? | Recommendations or requirements? |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| CONSORT | Consolidated Standards of Reporting Trials | Randomised controlled trials | 1996 updated in 2001 and 2010 | Yes | As a Word document | As publication |  |  |  |
| PRISMA |  |  |  |  |  |  |  |  |  |
| STROBE |  |  |  |  |  |  |  |  |  |
| STARD |  |  |  |  |  |  |  |  |  |
| ARRIVE |  |  |  |  |  |  |  |  |  |
| e.t.c. #TODO |  |  |  |  |  |  |  |  |  |

A selection of highly cited reporting guidelines

There are now over 500 reporting guidelines, representing the collective work of thousands of academics. The best-known guidelines are endorsed by large numbers of medical journals and the International Committee of Medical Journal Editors, and are amongst the 1% most highly cited publications indexed by Web of Science [8].

## EQUATOR: uniting the reporting guideline movement

As reporting guidelines grew in number and the problem of poor reporting gained recognition, Doug Altman saw the need to catalogue reporting guidelines and form a community. He united academics from around the world to form The EQUATOR Network, often simply called EQUATOR, standing for Enhancing the QUAlity and Transparency Of health Research. It was the first coordinated attempt to combat poor reporting systematically and on a global scale. One of EQUATOR’s core objectives was to create a database of reporting guidelines, accessible via their website where researchers will also find training and information about developing guidelines.

## Reporting quality has improved but remains sub optimal

Despite their promise, reporting guidelines have not been a silver bullet. Many studies have compared quality of reporting *before* and *after* reporting guidelines were published. Some have found evidence of improvement but all have shown reporting quality remains sub optimal.

In 2023 Kilicoglu et al. [9] analysed 176,620 randomized trial reports published between 1966 and 2018. They used software to classify sentences as to whether they pertained to a CONSORT methodology item. They found reporting quality in clinical trials has improved over time but remains sub optimal. Articles published before 1990 reported 24% of CONSORT items, whereas articles published between 2010-2018 reported 48%.

This result mirrors similar, manual, efforts to monitor reporting quality over time. For example, Dechartres et al. [2] looked at CONSORT items related to bias in 20,920 trial reports. They found information on sequence generation was missing from 69% of articles published between 1986-1990, but only 31% of articles published between 2011-2014. The proportion of articles missing information on allocation concealment fell from 70% to 45% over the same periods.

These two examples focus on clinical trials, but other research types have seen modest improvements too. De Jong et al. [10] reviewed 284 qualitative evidence syntheses. These syntheses appraised whether their primary studies had adhered to COREQ (a reporting guidelines for qualitative research). Studies published before COREQ reported 16 of COREQ’s 32 items, whereas studies published after reported 18. Similarly, when comparing the reporting quality of genetic association studies, Nedovic et al. [11] found reporting quality was better in articles published after journals began endorsing STREGA (63% vs 49% showing full adherence).

However, not all studies have found such a relationship. For example, Howell et al. found no improvement in the reporting of quality improvement studies after the SQUIRE guidelines were published [**howellEffectSQUIREStandards2015?**]. Similarly, Pouwels et al. found no improvement following STROBE’s publication [12].

If reporting *has* improved, is this because of reporting guidelines, or might it reflect general improvements in academic practice? A few studies have tried to isolate the effect of reporting guidelines. For example, in Nedovic’s STREGA study, they only saw improvements amongst STREGA-endorsing journals. There was no improvement amongst articles published in journals not endorsing STREGA. In Kilicoglu’s machine learning study, the fastest improvement occurred in the years immediately after CONSORT was published. However, other studies muddy the picture. For example, in a time series analysis of 456 observational epidemiology [13], although reporting quality improved over time *before* STROBE was published, there was no further improvement *afterwards*.

Some studies have tried to isolate the effect of asking authors to complete reporting checklists. In 2015 PLOS One randomly allocated 1689 incoming submissions reporting *in vivo* animal research manuscripts to either a) request completion the ARRIVE reporting checklist or b) usual practice [14]. No article achieved full compliance with ARRIVE, and only one sub item (details of animal husbandry) showed improvement between groups.

In another study [**barnesImpactOnlineWriting2015?**], 197 authors submitting to 46 participating journals were randomly allocated to receive either a) access to an online tool (called WebCONSORT) to generate customised reporting checklists and flow diagrams based on CONSORT and its extensions or b) a standard CONSORT flow diagram generator without a reporting checklist. There was no difference in reporting quality between groups: authors only reported ~50% of required items.

In 2018, before my PhD, I collaborated with EQUATOR and BMJ Open to compare manuscript versions before and after authors completed a reporting checklist. Authors made few edits to their work. Of 20 included authors, three added information for a single reporting item, one added information for two items, and one added information for six items. The remaining 15 authors made no changes. On average, manuscripts described 57% of necessary reporting items before the author completed the checklist and 60% afterwards. No manuscript fully described all reporting items.

Together, these studies paint a mixed picture. Reporting standards may have improved but, reporting guidelines may have led to modest improvements at best, standards remain low overall, and there is huge room for improvement. In a systematic review of 124 studies assessing adherence to one of eight reporting guidelines, Jin et al [15] found 88% of studies reported suboptimal adherence to reporting guidelines. Similarly, Dal Santo et al [**santoResearchWastePoor2022?**] reviewed 148 studies of reporting quality, and almost all described reporting quality as suboptimal. Ironically, the same review found the studies on reporting quality were *themselves* poorly reported: only 14 (10%) explained how they coded adherence in sufficient detail for replication, and only 49 (33%) provided data for all included studies. The authors of these studies were aware of reporting guidelines and presumably mindful of reporting If these authors still produced incomplete articles, what hope should we have for less motivated authors?

## Evidence gap: how can we make reporting guidelines more impactful?

This brings me to the evidence gap my thesis addresses. Although we know reporting guidelines’ impact has been limited, we do not know why, or how to improve them.

Beyond postulating that “a successful strategy is”a successful strategy is likely to be multi-dimensional” Hair et al. [14] make two concrete suggestion based on evidence from their trial. Because outcome assessors rarely agreed whether a study adhered correctly to ARRIVE, they suggested refining “the content” and “perceived clarity”. They also suggest reducing the number of items to make the guideline quicker to use.

Many articles end their discussion section with a rallying cry for better enforcement.

“We need to promote more active implementation, such as submission of the checklist with the manuscript.” [2]

“Therefore, we suggest that it is not sufficient for journals to simply recommend the use of STREGA to authors in the authors’ instructions; instead, journals should require submission of the STREGA checklist together with the manuscript” [11]

Some articles end with no conjecture on why they saw little effect, or how guidelines could be improved. [**howellEffectSQUIREStandards2015?**]

Most made no attempt to collect feedback from authors or editors.

Some researchers have tried to innovate to make guidelines more usable, but these innovations are not based on evidence or theory. For example, Hopewell et al. created WebCONSORT because they worried combining CONSORT with its extensions may be “cumbersome and difficult” without providing evidence for this claim. They made some guesses for why their intervention failed. Perhaps the custom combined checklists were too long? Perhaps reporting items were not adequately explained? Perhaps intervening at revision stage was too late? But because they did not collect feedback from authors nor editors during their study, these were stabs in the dark.

Their data did, however, reveal one unexpected insight. They had to exclude 39% of manuscripts because editors had incorrectly identified them as randomised trials when they were not, and a quarter of authors selected inappropriate extensions. Perhaps then, identifying study design and matching it to a reporting guideline is a barrier in itself.

I am also guilty of building an intervention to fix a problem without evidence. In 2018 I worked with EQUATOR as a freelance developer to create GoodReports.org, a website where authors could fill out checklists on-line. We made this website to address three problems we believed, anecdotally, to exist. First, some guideline checklists were in unusable formats (e.g. PDFs that could not be filled). Secondly, some guidelines were being paywalls. And thirdly, we worried that some authors might struggle to identify which guideline to use. And so we made a website where checklists were fillable, all guidance was open, and a questionnaire could help authors find the right guideline. Instinctively, I tried to make the website look nice and be easy to use. But those decisions were just default developer behaviour, and I not with any kind of behaviour change in mind. We kept the checklist text and guidance text exactly the same as it is in publications.

### Next steps quoted from articles

#### Active endorsement is a common argument. But does it work and is it practical?

Hair et al. [14] wrote “The editorial resource required to ensure that all accepted publications meet the requirements of the ARRIVE checklist is likely to be considerable” and found that “due to the additional time required for ARRIVE checklist requests, both the number of days manuscripts spent in the PLOS editorial office and the number of days from manuscript submission to AE assignment were found to be significantly longer in the intervention group”. Will journals shoulder increased costs in return for reporting guideline adherence? So far, few have shown any appetite to do so.

## How behaviour change theory can help bridge the evidence gap

One reason its difficult to find a path forward is because reporting guidelines have never really been characterized as an intervention. Hereby lies another irony. Guideline developers often complain that authors don’t report interventions clearly (there’s a whole reporting guideline - TIDIER - specifically for that). Yet if an alien nation came to Earth and wanted to disseminate reporting guidance in a similar same way, they wouldn’t find a single description sufficient for replication. Sure, they would find EQUATOR’s article [16] on how to *develop* reporting guidelines, but that article doesn’t describe how to disseminate guidance in much detail. For example, although it recommends to craft “each [reporting] item into a crisply and unambiguously worded checklist item”, it doesn’t specify what other content should be on that checklist (e.g. an introduction? Instructions on how to use it? The author list?), nor the file format (an editable Word file? A PDF?), or how the checklist should be completed (By pasting text? Entering page numbers? Ticks and crosses?). Similarly, the guidance suggests writing an accompanying Explanatory document, but doesn’t specify what content should go in it.

Another important aspect of defining interventions is describing how it is expected to work. Here, again, EQUATOR’s guidance falls short. For example, it recommends guideline developers disseminate guidance by writing “a short document of about 2,000 words reporting on the rationale for developing the guidance and the development process, including a brief description of the meeting and participants involved”. But why? How is knowing the guideline development process expected to influence behaviour? Why 2000 words? Our alien readers could probably guess, but they shouldn’t have to.

I’m being quite critical of EQUATOR’s guidance for guideline developers. Perhaps unfairly so, as it was not intended to be an intervention definition. It is a guide, written by guideline developers with some experience under the belt, to help other prospective developers. Along the lines of *this is the best way we found to do it, so you might want to do it this way too*.

The guidance for guidelines developers was published in 2010. I get the impression that nobody was really thinking of reporting guidelines as an intervention per say back then. In “A history of the EQUATOR Network”, Doug Altman refers to reporting guidelines as “resources” that “influence” reporting. EQUATOR’s training courses “target” editors and authors, but nowhere does he call them interventions. Other studies call them “tools” or “strategies” (#REF).

At the start of this chapter I illustrated how reporting guidelines came into being. The process was slow, grass-roots, and it was driven by lots of groups of people doing what they felt was sensible. This movement has generated guidance publications, checklists, websites, training; it has influenced editors, peer reviewers, and journal policies. The effectiveness of reporting guidelines rests on more that just the individual sets of guidance. It rests on the entirity of this system. All the different parts.

What do these parts have in common? All of them - the guidelines, the checklists, peer reviewer feedback, editorial checks, websites - they all ultimately seek to influence what and how authors write. The reporting guideline community has created something that in health care would be called a complex behaviour change intervention. That’s why I have chosen to use a behaviour change approach.

So why do I feel confident calling reporting guidelines part of a complex behaviour change intervention? According to the Medical Research Council….

* explain why complex (number of parts, heterogeneity)
* explain why behaviour

It's common for behaviour change initiatives to begin in this organic, messy way.   
  
Give examples  
  
Give examples of how formalization can lead to better effectiveness.

## Summary

In summary, I’ve explained that a lot of medical research is poorly reported, and that this makes it difficult for other researchers to understand, appraise, synthesize, or replicate studies. This, in turn, makes research less useful to patients.

I’ve introduced reporting guidelines, which were created by the research community in the hopes that they would improve reporting quality, and I’ve revealed that, sadly, these guidelines have had only a modest effect on reporting quality, at best.

I’ve argued that reporting guidelines and the system of tools, publications, websites, and people that disseminate them, are a complex behaviour change intervention, with the ultimate goal of altering what authors write.

And I’ve explained how behaviour change theory may help bridge the evidence gap: that although we know reporting guidelines haven’t fixed reporting quality, we don’t know *why* their effect is limited, or *how* they could be improved.

## Aims and Objectives

My aim, therefore, is to understand why reporting guidelines have not had a bigger impact on reporting quality, and to address these issues.

My objectives are:

* To identify factors that may limit reporting guideline impact by synthesising existing research and by evaluating the EQUATOR Network website (addressed in chapters 3)
* To work with key stakeholders to identify intervention changes to address these limiting factors (addressed in chapters 6)
* To implement these changes (described in **?var:chapters.development**)
* To refine the new intervention in response to feedback from authors (addressed in chapter 11)

## Thesis structure

Summary

### Chapter **?var:chapters.reflexivity** - Reflexivity and context

I reflect on my background and my prior held opinions, and those of my supervision team, and how these may have influenced the direction of this thesis.

### Chapter 3 - **?var:titles.synthesis**

The next three chapters pertain to my first objective - to identify possible reasons as to why reporting guidelines have had only a limited impact on reporting quality. This chapter describes a thematic synthesis of studies that qualitatively explored authors’ experiences of using reporting guidelines, where I sought to identify what may influence whether an author successfully adheres to reporting guidance.

### Chapter 4 - **?var:titles.review**

This chapter builds on the previous one by identifying additional possible influences from the content of quantitative survey questions.

### Chapter **?var:chapters.website** - **?var:titles.website**

This chapter describes a service evaluation of the EQUATOR Network website which, although an important piece of the reporting guideline infrastructure, was rarely explored by studies reviewed in the previous two chapters. From this evaluation, I then infer possible barriers that authors may encounter when trying to find and access reporting guidelines from EQUATOR’s website.

### Chapter 6 - **?var:titles.bcw**

The next 4 chapters pertain to my second objective - identifying intervention changes. This chapter introduces the Behaviour Change Wheel, which is a framework for designing and defining behaviour change interventions.

### Chapter 7 - **?var:titles.workshops**

This chapter describes how I lead workshops with members of the UK EQUATOR center to identify intervention options using Behaviour Change Wheel framework.

### Chapter 8 - **?var:titles.focus-groups**

This chapter reports focus groups where I collected ideas on how intervention options could be realised.

### Chapter **?var:chapters.behavioural-analysis** - **?var:titles.behavioural-analysis**

In this chapter, I bring together the outputs of the previous two chapters to create a table of intervention components and logic model.

### Chapter **?var:chapters.development** - **?var:titles.development**

This chapter concerns my third objective, implementing the intervention changes by redesigning a reporting guideline (SRQR) and the EQUATOR Network website’s home page.

### Chapter 11 - **?var:titles.pilot**

In this chapter I address my final objective, which was to refine the intervention in response to feedback from authors. I describe a qualitative study where I used observation, think aloud, structured interviews, and a writing evaluation, to gather feedback from an international sample of authors.

1. Glick BS (1963) [Inadequacies in the reporting of clinical drug research](https://doi.org/10.1007/BF01562195). The Psychiatric Quarterly 37:234–244

2. Dechartres A, Trinquart L, Atal I, Moher D, Dickersin K, Boutron I, Perrodeau E, Altman DG, Ravaud P (2017) [Evolution of poor reporting and inadequate methods over time in 20 920 randomised controlled trials included in Cochrane reviews: Research on research study](https://doi.org/10.1136/bmj.j2490). BMJ 357:j2490

3. Carp J (2012) [The secret lives of experiments: Methods reporting in the fMRI literature](https://doi.org/10.1016/j.neuroimage.2012.07.004). NeuroImage 63:289–300

4. Glasziou P, Meats E, Heneghan C, Shepperd S (2008) [What is missing from descriptions of treatment in trials and reviews?](https://doi.org/10.1136/bmj.39590.732037.47) BMJ 336:1472–1474

5. Feinstein AR (1974) [Clinical biostatistics. XXV. A survey of the statistical procedures in general medical journals](https://doi.org/10.1002/cpt197415197). Clinical Pharmacology and Therapeutics 15:97–107

6. Davidson SRE, Kamper SJ, Haskins R, Robson E, Gleadhill C, da Silva PV, Williams A, Yu Z, Williams CM (2021) [Exercise interventions for low back pain are poorly reported: A systematic review](https://doi.org/10.1016/j.jclinepi.2021.05.020). Journal of Clinical Epidemiology 139:279–286

7. Santo TD, Rice DB, Amiri LSN, Tasleem A, Li K, Boruff JT, Geoffroy M-C, Benedetti A, Thombs BD (2023) [Methods and results of studies on reporting guideline adherence are poorly reported: A meta-research study](https://doi.org/10.1016/j.jclinepi.2023.05.017). Journal of Clinical Epidemiology 159:225–234

8. Caulley L, Cheng W, Catalá-López F, Whelan J, Khoury M, Ferraro J, Husereau D, Altman DG, Moher D (2020) [Citation impact was highly variable for reporting guidelines of health research: A citation analysis](https://doi.org/10.1016/j.jclinepi.2020.07.013). Journal of Clinical Epidemiology 127:96–104

9. Kilicoglu H, Jiang L, Hoang L, Mayo-Wilson E, Vinkers CH, Otte WM (2023) Methodology reporting improved over time in 176,469 randomized controlled trials. Journal of Clinical Epidemiology. <https://doi.org/10.1016/j.jclinepi.2023.08.004>

10. de Jong Y, van der Willik EM, Milders J, Voorend CGN, Morton RL, Dekker FW, Meuleman Y, van Diepen M (2021) [A meta-review demonstrates improved reporting quality of qualitative reviews following the publication of COREQ- and ENTREQ-checklists, regardless of modest uptake](https://doi.org/10.1186/s12874-021-01363-1). BMC Medical Research Methodology 21:184

11. Nedovic D, Panic N, Pastorino R, Ricciardi W, Boccia S (2016) [Evaluation of the Endorsement of the STrengthening the REporting of Genetic Association Studies (STREGA) Statement on the Reporting Quality of Published Genetic Association Studies](https://doi.org/10.2188/jea.JE20150173). Journal of Epidemiology 26:399–404

12. Pouwels KB, Widyakusuma NN, Groenwold RHH, Hak E (2016) [Quality of reporting of confounding remained suboptimal after the STROBE guideline](https://doi.org/10.1016/j.jclinepi.2015.08.009). Journal of Clinical Epidemiology 69:217–224

13. Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, Ferrat E, Roujeau J-C, Richard M-A, Canoui-Poitrine F, European Dermatology Network (EDEN) (2013) [Impact of STROBE statement publication on quality of observational study reporting: Interrupted time series versus before-after analysis](https://doi.org/10.1371/journal.pone.0064733). PloS One 8:e64733

14. Hair K, Macleod MR, Sena ES, et al (2019) [A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus)](https://doi.org/10.1186/s41073-019-0069-3). Research Integrity and Peer Review 4:12

15. Jin Y, Sanger N, Shams I, et al (2018) [Does the medical literature remain inadequately described despite having reporting guidelines for 21 years? A systematic review of reviews: An update](https://doi.org/10.2147/JMDH.S155103). Journal of Multidisciplinary Healthcare 11:495–510

16. Moher D, Schulz KF, Simera I, Altman DG (2010) [Guidance for developers of health research reporting guidelines](https://doi.org/10.1371/journal.pmed.1000217). PLOS Medicine 7:e1000217