Discussion

Table of contents

## Chapter Overview

## Summary of findings

In this thesis I aimed to identify and address barriers authors face when using reporting guidelines. I chose this aim because I believe eliminating barriers will increase the proportion of authors adhering to reporting guidelines. This will make research articles easier for other researchers, clinicians, and patients to understand, synthesise, replicate, and use, leading ultimately to better patient outcomes.

In chapters 3 - I explored barriers through a qualitative evidence synthesis, a review of survey questions, and a service evaluation of the EQUATOR website. I identified affecting how authors discover, find, understand, and apply reporting guidelines.

In chapter I described how I used a framework for designing behaviour change interventions called the Behaviour Change Wheel #REF to prioritise intervention options. Over a series of workshops, EQUATOR UK staff and I decided to prioritise interventions using education, training, persuasion, modelling (demonstrating using reporting guidelines), and restructuring the environment (both physical and digital environments). Conversely, we saw restriction and incentivization as inequitable. When considering policy categories, we prioritized communication, guidelines, and service provision.

In chapter I described leading focus groups with stakeholders to generate 28 ideas to address barriers, and in chapters and I describe how I turned some of these ideas into 58 intervention components and brought them to life by redesigning a reporting guideline and the EQUATOR Network home page.

Finally, in chapter 11 I described how I evaluated the redesigned reporting guideline and home page with a diverse group of authors. The 53 deficiencies I identified can be addressed in future iterations.

## Overview of outputs

### The website

The main output of my thesis is the redesigned reporting guideline and EQUATOR Network home page. Although I have only redesigned one reporting guideline, SRQR, my approach will easily scale to others. I have built the website so guideline developers can upload and edit their own content. Developers (or EQUATOR staff, or I) simply need to upload content as plain text files: a file for each reporting item, a file for meta data (like the guideline’s scope, authors, publication, version e.t.c.), and a glossary. The platform will then automatically generate a fully functional wep page for the guideline, with hover definitions, discussion pages, collapsible content e.t.c. Tracking analytics to monitor authors behaviour is baked in.

Hence although I’ve described my output as a website, it is more like a platform. I believe this will have a big impact on guideline development groups, as few have sufficient funding nor expertise to develop or maintain websites, even when using DIY “no code” tools like Wordpress. One guideline developer I spoke to spent weeks creating a simple website with little functionality. The website I’ve built is comparatively far more feature-full, and requires zero technical work from guideline developers. What previously took guideline developers weeks is now achievable in a matter of hours. Even guideline groups with no website budgets can use my platform to turn plain text into online resources that mirror the redesigned SRQR guideline.

Two prominent reporting guideline websites (CONSORT and PRISMA) went down during my PhD. With no software expertise within in the guideline groups, and no budget to hire a developer, the websites stayed offline for many months. Providing a single platform like mine means development groups need not worry about maintaining their own systems. The platform itself uses simple, reliable, globally used gold-standard infrastructure familiar to many PhD students, so future maintenance will be easy and cheap.

### Conferences and publications

I presented chapters at 3 conferences and won 3 awards. I won 1st prize for Excellence in Doctoral Research, and 2nd prize for my oral presentation in the Early Career Researcher category at the 2022 World Conference for Research Integrity, where I presented an overview of chapters 3 - 8, covering my approach, current limitations and ideas to address them. A few months later I presented a poster covering the results of my focus groups (chapter 8) at the Reproducibility, Replicability and Trust in Science 2022 organised by the Wellcome Trust. I won 3rd prize amongst departmental final year DPhil students in 2023 at the Botnar Institute Student Symposium, for my presentation showcasing my redesigned reporting guideline (chapters 9 and 10).

I intend to publish 5 articles originating from this thesis. These are: 1) my qualitative evidence synthesis (chapter 3), 2) my review of survey content (chapter 4), 3) the workshops and focus groups (chapters 7 and 8), 4) intervention refinement (chapter 11 updated after more design iterations), 5) a finalised intervention description (an update of chapters 9 and 10 once the design is finalised).

## Contributions and transferability to the meta-research community

Beyond the immediate impact of the website, I believe my work will bring three other benefits to the reporting guideline community and other meta-researchers by 1) providing possible explanations for previous research findings 2) opening new lines of enquiry and funding options and 3) as a model for other grass-root academic movements.

In my introduction chapter I summarised previous research evaluating the impact of reporting guidelines. Few studies employed qualitative methods and process evaluations were scant and shallow. Consequently, although these studies offered a depressing survey of poor reporting standards, they did not explain *why* reporting guidelines had little effect or *how* they could be improved. Researchers were operating in the dark and could not see how to move forward. My work turns the light on and illuminates hypotheses to explain and address past failures.

I believe EQUATOR staff had a lightbulb moment of their own. My thesis was never meant to be a behaviour change project and my initial supervisory team had no experience in behaviour change nor qualitative methods. They were firmly in the quantitative camp, familiar with statistics, systematic reviews, and randomised trials. They repeatedly warned me against creating software, as they felt it fell firmly within the domain of *development* and not *research*. They typically shoehorned meta-research into clinically-focussed grant applications and maintained their website on a shoestring. I have demonstrated how development requires a great deal of research, and thus can be packaged as a passable thesis project and placed centre stage in funding applications. Framing reporting guidelines as a behavioural intervention means EQUATOR is no longer restricted to medical research funders, and can access new funding sources like the Economic and Social Research Council who have recently announced plans to radically expand UK behavioural research capacity [1]. Framing reporting guidelines as an *online* intervention might release funding to support EQUATOR’s website. Therefore, pursuing my approach will bestow EQUATOR with new research directions, collaborations, and funding options.

These new avenues are open to other meta-researchers too, and I hope my work inspires guideline developers and grass roots movements further afield. Some of my results may be directly applicable to others. For instance, one of my intervention components was to use language to convey confidence and benefits instead of judgement and fear. Sadly, negative language pervades discussions on research integrity and vilifies researchers, accusing them of “waste”, “questionable practices”, “failing”, or “lacking integrity”. This may alienate “average”, well-meaning researchers. Shifting the narrative towards “efficiency”, “ease”, “confidence”, or “strengthening integrity” makes conversations more welcoming and attractive.

However, it’s not my *results* that I think are most transferable, it’s my *approach*. Many of my intervention components are too tethered to reporting guidelines to be of interest to other fields. But the approach that I took and the methods I used could be useful to most meta-researchers seeking to drive change. Although commonly used in medical research, I have not found any meta-research groups using behaviour change frameworks to improve the scholarly system. I think they are missing a trick. Without a framework, change-drivers risk getting hung-up on their favourite intervention types (in my experience, most often regulation or education) to the neglect of others. For example, in 2022 I attended a workshop to brainstorm strategies to increase equity and diversity funding applications. One participant gave a rich account of how their research support department re-designed their systems, shared case studies, praised examples of best practice, ran training courses, and held people to account when necessary. The facilitator only noted the word ‘training’. Had he been more familiar with behaviour change, perhaps he would have recognized the participant’s examples of environmental restructuring, persuasion, education, incentivization, and coercion.

## Strengths

Beyond using a behaviour change framework, meta researchers could benefit from integrating the other strengths of my work, including the use of systematic methods, qualitative methods to solicit rich information, and diverse recruitment.

I used systematic methods throughout my thesis. My literature search for chapters 3 and 4 was systematic. The behaviour change wheel and APEASE criteria that I used in chapters 3 - 9 were themselves made systematically, and require users to consider and prioritise options from 360 degrees. In all of my data analysis, from collecting barriers, ideas, and deficiencies, I sought to code and collate *all* available information: just as systematic search seeks to identify all relevant literature, my coding strategy sought to identify all themes. Similarly, when moving from one stage to another - from barriers to ideas, from ideas to intervention components, from components to deficiencies - I considered and linked items fastidiously, thereby drawing threads through my thesis from start to end.

Another strength is my use of qualitative methods to elicit rich description from participants. I’ve already extolled the benefit of using a qualitative approach, but within the world of qualitative research one must still be judicious when selecting methods. For example, in our 2019 study with BMJ Open we chose poorly. We were seeking to identify deficiencies in a different website, and we thought adding a free text question to an online survey would help. Over 21 months, we contacted 11,000 authors, of whom 93 answered the question “How could we make [the website](#the-website) more useful?”, mostly with very short answers. We only identified 6 themes. In this thesis, by stark comparison, I identified 53 deficiencies in a fraction of the time and by recruiting only 11 participants, by using more appropriate qualitative methods.

This tale reveals a warning to guideline development groups: although a qualitative approach may seem accessible, doing it *well* requires expertise. Guideline development groups would be wise to include qualitative experts, preferably those with experience in behaviour change interventions and refining text.

My diverse recruitment was another strength. I wanted diversity because I wanted to understand the perspectives of all stakeholders and because believed it would lead to more discoveries: more barriers, more ideas, more deficiencies. I achieved diversity in my stakeholder focus groups (academics, publishers, and guideline developers) and when evaluating the website (authors of varying demographic, disciplines, and experience). Whereas my qualitative synthesis found little geographic diversity amongst participants, my previous chapter included participants from South America, Africa, Asia, Europe, and Australia. I feel proud to have addressed this need, but it was not easy. Twitter proved useless. I relied largely on Penelope.ai - the manuscript checker I created - and I was fortunate to have budget to pay participants. Other development groups may not have access to such luxuries.

## Limitations

### In choosing a framework, I neglected others

Although using a behaviour change framework was a strength, in choosing it I decided *against* using others, and some researchers would consider this a limitation. In chapter 6 I explained why I chose the Behaviour Change Wheel above two other popular frameworks - the Theoretical Domains Framework and the Person Based Approach. There are plenty of others out there, and behaviouralists may grumble about my decision not to use their preferred framework. I chose not to do a formal comparison as others have already, and shown frameworks vary in their focus, evidence base, and ease of use #REF. In choosing a framework I did not look for *the best* but rather the one that was *best for me*. It had to be based on evidence - that was a given - but beyond that it had to be the framework-of-least-resistance. I was already leading my (original) supervisors down new avenues, so my framework had to be easy to understand and not too far from familiar epistemological ground.

Research groups leaning towards other frameworks may avoid elements of friction I encountered. For example, in applying the Behaviour Change Wheel to redesigning a reporting guideline and creating web pages, some intervention functions and policy categories did not obviously generalise. For instance, I labelled many components as examples of “environmental restructuring”, because I saw the website as a digital environment, and so (for example) adding metaphorical ‘signposts’ to other web content felt the same as installing physical signposts in a hospital. Readers may feel uneasy seeing me compare a digital environment with a physical one, especially if they read the Behaviour Change Wheel’s definition of *environmental restructuring: Changing the physical or social context*. They should be reassured, as I was, after reading the example given straight after: *“Providing on-screen prompts for GPs to ask about smoking behaviour”*. The Behaviour Change Wheel developers’ example of environmental restructuring is itself a digital one ( [2]).

Translating the Behaviour Change Wheel to a purely digital environment was one challenge. Another, was that in some instances the framework did not go far enough. For instance, although the framework helped me identify that information should be easy to find, or that design should look simple and professional, the Behaviour Change Wheel does not tell you *how* to do that. It is not a web development framework. A user experience expert might grumble that I did not use a user experience checklist or information architecture heuristics. They would be justified. Future work would do draw on these domains, although doing so is often harder than one may expect, and is generally outsourced to experts #REF.

### My intervention components are granular

Some intervention designers may be more used to interventions having more “offline” elements: conversations, leaflets, physical objects, places, in-person services. Such designers may consider “a leaflet” or “a website” to be a single element. When I had coffee with one such researcher, he suggested I label my redesign as a single “enablement” component. I did not follow his advice. Instead, I have tried to describe and justify the redesign as the sum of many components. Others may think my interpretation is a distortion of the Behaviour Change Wheel’s intention, or may feel that I stretched the framework too far by applying it with such granularity.

When the behaviour change wheel authors’ applied it to their own digital intervention their components were much larger than mine. Each component in their *Drink Less* mobile phone application was a complete module #REF. These modules included interactivity: in the normative feedback module a widget displays how the user’s drinking compares to others in the UK. The self-monitoring module allows users to track consumption, and so on. In contrast, many of my intervention components are subtler. For instance, I have counted tone of voice and visual design as persuasive components, and I’ve counted the user journey from home page through the reporting guideline (2 pages now, compared to 4 or more previously). Drink Less’ creators will have made decisions about tone, design, and layout, yet do not consider them as components.

Although I consider my breadth of components a strength (and a testament to my diligent approach to identifying barriers and ideas), having many small, intermingled components makes it difficult to isolate the efficacy of individual components. In developing Drink Less, the creators used a factorial screening trail “to identify the individual components, or combinations of components, within the multi-component intervention that affect change and to screen out the ineffective ones” [**garnettRefiningContentDesign2021?**]. This was possible because Drink Less only had six, separated components. Applying such an analysis to my website would be difficult and, probably, not useful.

### My logic model is rudimentary

I found another drawback of having so many components when trying to succinctly describe the website to other people. Whereas I can describe the core components of Drink Less quite easily, my table of intervention components 9 is unwieldy. When communicating my work to guideline developers, I will explore how best to present components. In their topology of logic models, Mills et al [3] propose a way of moving from a rudimentary list of components like mine, which they define as a *type 1* logic model, towards a model that concisely captures complexity and context. As my intervention matures beyond planning and refinement, I could explore depicting it as relationships between resources, activities, outputs, outcomes, impact, and domains (what Mills et al refer to as *type 2* and *type 3* models).

### I neglected the behaviour of other stakeholders

Such a model should also include the behaviour of editors, peer reviewers, and other stakeholders. This was another limitation of my approach - I focussed exclusively on authors’ behaviour. Future research could explore barriers and solutions faced by others and incorporate them into intervention design and logic models.

In addition to considering the behaviour of all stakeholders, future research should also consider differences between reporting guidelines. In my introduction I described reporting guidelines as variations on a theme. Because they shared so many commonalities, I justified treating them all the same. However, in practice, some variation may matter. For example, guidelines that cater to writing protocols may be best delivered in a different format, by different stakeholders (funders? registries?) and aimed at researchers at early stages of work. Therefore, subsets of reporting guidelines may deserve their own dissemination strategies and logic models.

### My context did not reflect the real world

Logic models should also reflect real-world context as far as possible. Although I fought to mimic parts of real life in my interviews with authors (chapter {{<var chapters.pilot >}}) by allowing naïve authors to explore the website similar to how they would in real life, the context was still far from real. Future research should explore how authors use the redesigned reporting guidelines in a real-world context, and how it impacts their writing.

In my introduction I described a real-world study we carried out in collaboration with BMJ Open before my PhD, where we were evaluating an older version of reporting guidelines. I can imagine performing a similar study with the redesigned guidelines but with an important difference: future real-world evaluations should be based upon a logic model. If I were to repeat the BMJ Open study today, my logic model would suggest alternative outcome measures. In the workshops (chapter 7) EQUATOR and I decided we want authors to use reporting guidelines *as early as possible* in their research journey. My evidence synthesis (3) revealed that for many authors, receiving a reporting checklist at the time of journal submission was a bad time to give advice, as authors lacked the time and motivation to act on the guidance. In the workshops, EQUATOR and I began to think differently about the role of journal endorsement: instead of seeing article submission as the moment where authors should be applying reporting guidelines, we realised that journal endorsement is merely a good way to make authors *aware* of reporting guidelines, and that we not should expect authors to fully apply them there and then, but rather we would hope authors come back to the website to use a reporting guideline *earlier* in their *next* project. We could adjust the BMJ Open study protocol to reflect this shift in logic model. In our original BMJ Open study we used reporting adherence as our primary outcome, and we tracked manuscripts through journal submission to look for evidence that authors improved their manuscripts after completing a checklist. We found no such evidence. According to our new logic model, we would not expect to find such immediate changes. Instead, we would hope to see the same authors *returning* to our website in the future (after a few weeks or months), and we believe that authors using our redesigned guidelines will be more likely to return than authors using the old version. To test this refined logic model, we could choose *return rate* as our primary outcome measure, and then compare reporting adherence within those returning authors, comparing adherence in their second manuscripts to their first. Tracking authors over time will be possible using web analytics and by following authors that repeatedly cite my new platform.

### Lack of quantification

I made little use of quantitative data in this thesis. I sought to understand, identify, and ideate, not to count, measure, or compare. Consequently, whilst my thesis has generated may hypotheses, it tests none of them.

As intervention development moves beyond planning, designing, refinement, and towards evaluation, new questions will require a quantitative approach. Do authors prefer the redesigned guidelines or the original ones? How many authors come back to use guidelines again? Of the remaining barriers that are difficult or expensive to address, which occur most frequently? And of course the ultimate question: Which version of the guidelines - the original or the redesigned - leads to better adherence? All of these questions will require a quantitative approach. They all too have an implicit follow-up question - *why?* - and so any quantitative approach should have a qualitative accompaniment.

In summary, my thesis had a number of limitations. In choosing a framework I neglected to consider others which may have shaped my work differently, especially my approach to digital design. My detailed approach to identifying barriers and solutions led to a large number of intervention components, some of which are small or subtle. This may complicate communicating my logic model or identifying the effectiveness of individual components. In focussing on authors’ behaviour I have neglected to consider editors, peer reviewers, or other stakeholders. By considering reporting guidelines as a homogenous group I have not accounted for the differences between them or their users. I prioritized qualitative questions above quantitative, and so whilst my thesis raises many hypotheses it tests none of them. The context in which authors gave me feedback did not reflect real life. Future studies addressing these questions should refine my rudimentary logic model, and make it specific to the context being evaluated.

## Recommendations for future research

In addition to the future work to extend the website and to address the limitations permeating my thesis, my work opens up new lines of exploration that could be developed into research strands. This work fell into three major groups: training, networks, and engaging new stakeholders. My focus groups and author interviews identified a need for training on how to use reporting guidelines and on how to write in general. Focus group participants felt that a network of “reporting champions”, similar to UKRN’s network model, could be a useful way to both advertise reporting guidelines and offer assistance in using them. Focus group participants also felt that getting funders, registries, and institutions to endorse or enforce reporting guidelines would help get authors using them earlier in their research. Future research projects could inform the development of these opportunities, explore how best to deliver them, and evaluate their feasibility and efficacy.

My work also offers new directions for guideline developers. Having identified in my qualitative synthesis that few reporting guidelines have undergone meaningful user testing despite exhibiting many barriers, developers can use my work to justify funding applications to support such user testing. I hope that future developers will consider my findings and designs when creating their own resources. EQUATOR could facilitate this by updating their existing guidance for guideline developers to address the barriers and solutions I identified that must be considered when formulating guidance, creating checklists, and writing example and elaboration documents.

Hence future work following from this thesis may include creating new guidance, training, or networks of stakeholders and champions.

## Implications for policy

My work touches on policy in two ways: reporting guideline policies held by journals and other stakeholders, and funders’ policies towards funding meta-research and grass-roots services.

In my introduction I mentioned that guideline developers have long been calling for journals to better enforce reporting guidelines. I argued that pointing the finger solely at journals was unfair and unrealistic. Nevertheless, journal endorsement and enforcement *are* an important piece of the puzzle. In making reporting guidelines easier to use and understand, my work will make such policies easier to enact. The smaller the hurdle, the more likely journal editors will be to lay it in front of their authors. And if editors can better understand reporting guidelines, it will be easier for them to check adherence. Similarly, reducing friction will make it easier for funders to begin requiring reporting guidelines, especially if new guidelines are developed for protocols.

Funders looking to support changing the scholarly system should allocate money for meta-research, behaviour change, intervention development and maintenance. They should accommodate software development costs and value the importance of thorough user testing. Once a resource becomes established, they should fund its evaluation, monitoring, and periodic updating, and if necessary they should provision for long term maintenance far beyond the terminus of a traditional grant life cycle. Maintenance costs may be a fraction of the initial research costs, but they need to be reliable and persist for years if not decades. For grass-roots movements to successfully change the scholarly system, academics need their digital tools to be adopted by private sector stakeholders, most of whom will value stability and sustainability.

## Conclusion

I have demonstrated how I identified and addressed barriers preventing authors from adhering to reporting guidelines. I identified barriers through a qualitative evidence synthesis, survey review, and website service evaluation. I identified solutions through workshops and focus groups with stakeholders, applied a subset of these ideas by redesigning reporting guidelines, and refined the redesign by interviewing authors.

My redesigned reporting guideline and EQUATOR Network home page could be adopted and extended to other reporting guidelines. The EQAUTOR Network and guideline developers can use my work to inform and justify future funding applications and develop impactful resources.

My work could be extended by adding functionality and reporting guidelines to the website. Future research should address the limitations of my work by exploring the utility of alternative frameworks, developing logic models that include the behaviour of other stakeholders, testing these logic models in real world contexts, exploring authors preferences and their reporting quality.

I hope other researchers will draw inspiration from my pragmatic approach that made heavy use of qualitative methods and an established behaviour change framework. However, for academic-lead movements to develop digital tools that successfully change the scholarly system, funders will need to reconsider how they fund such endeavours.

1. (2023) ESRC to radically expand UK behavioural research capacity.

2. Michie S, van Stralen MM, West R (2011) [The behaviour change wheel: A new method for characterising and designing behaviour change interventions](https://doi.org/10.1186/1748-5908-6-42). Implementation Science 6:42

3. Mills T, Lawton R, Sheard L (2019) [Advancing complexity science in healthcare research: The logic of logic models](https://doi.org/10.1186/s12874-019-0701-4). BMC Medical Research Methodology 19:55