OPEN ACCESS

Findings from a novel approach to publication guideline revision: user road testing of a draft version of SQUIRE 2.0

Louise Davies,[1,2] Kyla Z Donnelly,[2] Daisy J Goodman,[3] Greg Ogrinc[3]

▶ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/bmjqs-2015-004117).

[1]VA Outcomes Group, White River Junction VA Medical Center, White River Junction, Vermont, USA
[2]The Dartmouth Institute for Health Policy & Clinical Practice, Geisel School of Medicine, Lebanon, New Hampshire, USA
[3]Department of Education, Department of Veterans Affairs Medical Center, White River Junction, Vermont, USA

Correspondence to
Dr Louise Davies, VA Outcomes Group, White River Junction VA Medical Center, Mailstop 111B, 215 North Main Street, White River Junction, VT 05009, USA; louise.davies@dartmouth.edu

Received 25 February 2015
Revised 10 July 2015
Accepted 12 July 2015
Published Online First 11 August 2015

To cite: Davies L, Donnelly KZ, Goodman DJ, *et al*. *BMJ Qual Saf* 2016;**25**:265–272.

ABSTRACT

Background The Standards for Quality Improvement Reporting Excellence (SQUIRE) Guideline was published in 2008 (SQUIRE 1.0) and was the first publication guideline specifically designed to advance the science of healthcare improvement. Advances in the discipline of improvement prompted us to revise it. We adopted a novel approach to the revision by asking end-users to 'road test' a draft version of SQUIRE 2.0. The aim was to determine whether they understood and implemented the guidelines as intended by the developers.

Methods Forty-four participants were assigned a manuscript section (ie, introduction, methods, results, discussion) and asked to use the draft Guidelines to guide their writing process. They indicated the text that corresponded to each SQUIRE item used and submitted it along with a confidential survey. The survey examined usability of the Guidelines using Likert-scaled questions and participants' interpretation of key concepts in SQUIRE using open-ended questions. On the submitted text, we evaluated concordance between participants' item usage/interpretation and the developers' intended application. For the survey, the Likert-scaled responses were summarised using descriptive statistics and the open-ended questions were analysed by content analysis.

Results Consistent with the SQUIRE Guidelines' recommendation that not every item be included, less than one-third (n=14) of participants applied every item in their section in full. Of the 85 instances when an item was partially used or was omitted, only 7 (8.2%) of these instances were due to participants not understanding the item. Usage of Guideline items was highest for items most similar to standard scientific reporting (ie, 'Specific aim of the improvement' (introduction), 'Description of the improvement' (methods) and 'Implications for further studies' (discussion)) and lowest (<20% of the time) for those unique to healthcare improvement (ie, 'Assessment methods for context factors that contributed to success or failure' and 'Costs and strategic trade-offs'). Items unique to healthcare improvement, specifically 'Evolution of the improvement', 'Context elements that influenced the improvement', 'The logic on which the improvement was based', 'Process and outcome measures', demonstrated poor concordance between participants' interpretation and developers' intended application.

Conclusions User testing of a draft version of SQUIRE 2.0 revealed which items have poor concordance between developer intent and author usage, which will inform final editing of the Guideline and development of supporting supplementary materials. It also identified the items that require special attention when teaching about scholarly writing in healthcare improvement.

INTRODUCTION

In 2008, the Standards for Quality Improvement Reporting Excellence (SQUIRE) Guidelines (SQUIRE 1.0) were published.[1] Designed to help advance the science of improvement, they were the product of a detailed 3-year development process.[2] They had been preceded in 1999 by the publication of the Quality Improvement Report Guidelines, designed for reporting the results of local quality improvement work. These Guidelines used a format that departed from traditional medical IMRaD reporting sections (Introduction, Methods, Results and Discussion).[3] SQUIRE 1.0 explicitly retained the IMRaD format both because it captured the inherent logic of all scientific inquiry

and because it was the traditional reporting framework in many scientific journals. SQUIRE 1.0 hewed to a scientific approach by, for example, including items that asked authors to describe methods for assessing whether the changes observed during an intervention were due to the intervention or some other outside factor. But SQUIRE also forged new ground by calling for reflection upon the experiential learning of improvement efforts.[4]

SQUIRE 1.0 was one of several important milestones in the recognition of improvement as a separate field of study. Another was a change to medical training so that it now includes improvement as a major competency.[5 6] In addition to changes in medical training, one need go no further than an internet search to find several pages of results when one types in 'fellowships and training in quality improvement'. But the growth has not been limited to education—the field is also still defining itself. [7 8] This rapid evolution has created a happy problem: SQUIRE 1.0 needs revision and updating.

Thus it was that in 2012, in the setting of this growth and evolution, we initiated the development of 'SQUIRE 2.0'. We began the process with an evaluation by users and an advisory group of thinkers (reported separately),[9] followed by a consensus conference to evaluate an interim draft. These feedback opportunities revealed controversy around the meaning and use of theory in improvement work, disagreement about what should be reported in improvement work and confusion about what specific items meant and who the SQUIRE Guidelines were intended for, among other things.

Based on the results of the evaluation and the consensus conferences, we decided that success in disseminating SQUIRE 2.0 required a detailed understanding of what items needed to be clarified and/or better explained. We created a draft, 'SQUIRE 1.6', with the specific intent of gaining this information from likely users. Conducting formal testing of draft Guidelines by users is a novel idea and is an expansion over what is generally called for in guideline development.[10] We are not aware of any other publications reporting results of a formal 'road testing' process for a publication guideline. The aim of having users test the Guideline was to determine whether SQUIRE 1.6 was understood and implemented as intended by the developers. The results were used to maximise the effectiveness of education and dissemination efforts, and to finalise SQUIRE 2.0, released in autumn 2015.

METHODS

Potential study participants (n=427) were identified from the records of the SQUIRE Development team. They comprised a list of graduates, faculty and directors of healthcare improvement fellowships or programmes (eg, The Institute for Healthcare Improvement, Cambridge, Massachusetts, USA; the UK Health Foundation, London, England; VA Quality Scholars Programs around the USA; the Dartmouth Leadership and Preventive Medicine Residency, Lebanon, New Hampshire, USA, the Healthcare Improvement Fellowship at Jönköping University, Sweden, among others), people involved in medical education about healthcare improvement, people who had been involved in the evaluation of SQUIRE 1.0 and other professional contacts in the field. Invitations were sent out to all 427 people, with two follow-up invitations if there was no response.

The invitation requested that participants do two things within a 3-month timeline. First, they were to use the SQUIRE 1.6 Guidelines (see online supplementary appendix table 1) to write a section of a manuscript they were working on or to rewrite one they had recently finished. They were also asked to annotate their manuscript section using the 'track changes' function of Microsoft Word (Microsoft Corp., Redmond, Washington, USA) to show which SQUIRE 1.6 items they had used and to which text they felt the Guideline item applied. Then, they completed a confidential survey about their item usage and their interpretation of key concepts in the SQUIRE Guidelines.

People who responded that they would participate in the study were allocated through blocked randomisation to a manuscript section: introduction, methods, results or discussion. As participants accrued, some came from the same programme or geographic location, though not consecutively. In this case, the randomisation was stratified to ensure participants from the same location did not share the same assignment (eg, three people from the same location given the methods section).

The survey data and manuscript section were collected electronically using Qualtrics Survey Software (Qualtrics, LLC, Provo, Utah, USA). The survey contained open-ended questions on key concepts in SQUIRE and areas of controversy, and Likert-scaled questions to assess item usage. Quantitative data from the survey were transferred into Stata V.13 (Statacorp, College Station, Texas, USA) and descriptive statistics were calculated and interpreted in the context of the qualitative data and submitted manuscript sections. Qualitative data from the survey were transferred into Hyperresearch V.3.7.1 (Researchware, Randolph, Massachusetts, USA) and analysed using a content analysis approach.[11] The manuscript sections were evaluated for concordance between the item usage as identified by the respondent and the intended application of the item by the developers.

RESULTS

Eighty-three people responded that they would be willing to participate as writers. Of those, 44 completed the survey and 41 submitted the writing sample. Physicians, nurses and other allied health personnel

participated. There was a wide distribution of experience with scientific writing. More detailed characteristics of those who completed the survey are shown in table 1.

A follow-up survey was sent to the 39 non-completers of the writing sample and survey, and 21 responded. Among those who did not finish the tasks, the reasons stated were that they ran out of time (n=12), felt their work did not fit the SQUIRE guidelines (n=3), felt the guidelines were hard to use (n=2) or had technical issues uploading their documents and filling out the survey (n=4).

Of the 41 manuscript sections submitted, 4 submitted an entire manuscript instead of just the section they were assigned, and so for the qualitative analysis, we took advantage of this to increase the numbers of discussion sections analysed (by chance, only 5 of the 20 people assigned to submit a discussion section did so). Thus, we performed a qualitative analysis of 10 introduction sections, 10 methods sections, 12 results sections and 9 discussion sections.

Table 1 Demographic characteristics of participants (n=44)

	Frequency (%)
Practice setting	
Academic/university	29 (65.9%)
Community/private practice	1 (2.3%)
Business	1 (2.3%)
Government/publically funded	9 (20.5%)
Other	4 (9.1%)
Highest educational attainment (n=43)	
Bachelor's degree	2 (4.7%)
Doctorate of philosophy	10 (23.3%)
Master's degree	4 (9.3%)
Medical degree	22 (51.2%)
Other	5 (11.6%)
Field of training	
Allied health	1 (2.3%)
Nursing	7 (15.9%)
Medicine	30 (68.2%)
Other	6 (13.6%)
Career publications	
0	3 (6.8%)
1–5	18 (40.9%)
6–10	5 (11.4%)
11–15	6 (13.6%)
≥16	12 (27.3%)
Total publications in healthcare improvement (n=39)	
0	15 (38.5%)
1–5	17 (43.6%)
6–10	4 (10.3%)
11–15	1 (2.6%)
≥16	2 (5.1%)
Ever-used SQUIRE Guidelines	
Yes	26 (59.1%)
No	18 (40.9%)
Ever-used other Guidelines	
Yes	25 (56.8%)
No	19 (43.2%)

Frequencies may not sum to the total due to missing data.
SQUIRE, Standards for Quality Improvement Reporting Excellence.

SQUIRE 1.6 applicability and usability

Applicability of Guideline items

We asked participants to rate the extent to which they used each item in their assigned section, the options were in full, in part or not at all (see online supplementary appendix table 2). Consistent with the recommendation in the SQUIRE Guidelines that not every item be included in a manuscript, less than one-third (n=14) of people stated that they applied every item in their section in full. Virtually all (43/44) respondents indicated that all the items contained in the Guidelines were relevant to healthcare improvement and appropriate for inclusion.

Stated understandability of Guideline items

Respondents stated that they understood most items. Of the 85 instances when an item was only partially used or was omitted, only 7 (8.2%) of these instances were due to respondents stating it was because they did not understand the item. Only one item was identified as hard to understand by more than one respondent: 'methods employed to ensure completeness of data', which two participants said they left out because of difficulty in comprehending the item.

SQUIRE 1.6 item usage and interpretation

Respondent reports of item usage

On average, the three items in the introduction section were included in full nearly all the time. There were no instances where an item from the introduction was not used at all. The 8 items in the methods section, 5 items in the results and the 15 items in the discussion section, respectively, were used less consistently. A pattern was noted in item usage overall—items that were reported as used fully were most similar to scientific writing in general. An exception was the discussion item 'Relation of the key findings to the original logic and the mechanisms by which the study was expected to work'. This item, on average, was reported as used fully, but the qualitative analysis revealed that its usage was frequently inconsistent with the intention of the developers. Omitted items were most often those relatively unique to healthcare improvement—concerning context or opportunity costs of an intervention, for example (table 2).

Qualitative analysis of item usage

Respondents were asked to identify the text in their submitted manuscript section that fulfilled each of the items they used from the SQUIRE 1.6 Guidelines. We examined their classifications to determine whether the items were applied as intended by the developers

Table 2 Key qualitative and quantitative findings related to item usage, and their interpretation, from the analysis of manuscript sections submitted by testers of SQUIRE 1.6.

Finding	Interpretation	Implication
Participant reported item usage (Quantitative) **Respondents reported using only a few items in full:** 1of 3 introduction items *"Specific aim of the improvement…"* 1 of 8 methods items *'Description of the improvement…'* 0 of 5 results items *4 of 15* discussion items *"Key findings…"* *"Relation of key findings to… logic… mechanism…"* *''Particular strengths of the work"* *'Implications …for further studies'.* **The largest number of omitted items came from the following:** 9 of 10 methods section respondents reported omitting: "*Assessment methods for context factors that contributed to the success, failure, efficiency, and cost of the improvement"* 4 out of 5 discussion section respondents reported omitting: *"Costs and strategic trade-offs, including opportunity costs"*	Most of the items utilized in full are shared with medical scientific writing more generally. The omitted items are more unique to healthcare improvement work, are outside the domain of typical health services research reporting, or represent less well-developed areas of the science of healthcare improvement (i.e., assessment methods for context factors…)	There is work to be done in at least three areas: - Make SQUIRE 2.0 items that are unique to healthcare improvement work and/or are evolving with the science in the field as clear as possible to promote interpretation and use as intended by the guideline developers. - Develop frameworks and methods of measurement for underdeveloped areas of the science of healthcare improvement that are specified as important by SQUIRE, such as measurement of context. - Help people gain comfort and skill in using the less familiar SQUIRE items when they plan, do, and report work to improve the quality, safety, and value of healthcare.
Analysis of Item interpretation (Qualitative) **The following items were commonly applied in ways not intended by the developers of the Guidelines:** Results (n=10 evaluated): 'Evolution of the improvement' *- Intended use: descriptions of alteration of the intervention or steps of the implementation over time* *- Observed use: reporting baseline outcome data from before the intervention* Methods (n=12 evaluated): 'Context elements that influenced the improvement…' *- Intended use: reports of context that would affect the intervention, e.g. that changed over time or that created external pressures to change* *- Observed use: describing only general setting of the work (e.g., size and type of facility)* 'The logic on which the improvement was based, including mechanism by which improvement expected to work' *- Intended use: description of the theory or rationale that underpinned the work* *- Observed use: describing the method (e.g., lean) used for improvement, or the medical evidence on which the intervention was based* Process and outcome measures used for the improvement, including rationale for choice…' *- Intended use: description of measures that would capture fidelity to processes of the intervention, and measures of outcomes* *- Observed use: describing outcome measures but not process measures* Discussion (n=9 evaluated): "Relation of the key findings to the original logic and the mechanisms by which the study was expected to work." *- Intended use: description of whether theory or rationale underlying work was sound based on study findings.* *- Observed use: reporting primary outcome, no reflection on whether the theory was supported by findings. Also, reporting data on the intervention but not why the intervention was expected to work*	Items unique to the evolving science of healthcare improvement are the items most frequently applied in ways not intended by the developers of the Guidelines	

of the Guideline. A pattern emerged that was similar to the usage pattern reported above: items unique to the field of healthcare improvement were frequently applied in ways that were different from that intended by the developers of the SQUIRE Guidelines (table 2).

For example, the item 'Evolution of the improvement' was intended for the reporting of steps of an intervention or changes that were made to an intervention during the course of an improvement project. Respondents commonly identified the reporting of

baseline data from before the intervention as fulfilling this item. While baseline data are necessary, they do not reveal changes to or steps of an intervention over time.

Another example was the application of the item 'The logic on which the improvement was based, including mechanism by which improvement expected to work'. This item was intended for reporting the theory or rationale on which the intervention was based. In more than one instance, respondents labelled the method used for the improvement (eg, lean) as the logic, or mechanism. Another interpretation was to list the medical evidence on which the intervention was based. Medical evidence of effectiveness might be part of a mechanism for why an improvement would be expected to work, but it is not enough by itself to explain why a particular intervention would be effective. Here is an example of correct use of the item for a hypothetical study aimed at improving rates of hand washing: "…we provided individualized feedback to study subjects on their own and their aggregated peer hand washing rates, because it is known that competition and social pressure lead to greater compliance: people will tend to improve if they receive information about their own performance." If provided, a description of the fact that hand washing is associated with lower infection rates would provide the medical evidence for the intervention, but that information alone would not explain how the hand washing intervention was expected to decrease infection rates—the example statement provided above would be needed to fulfil that Guideline item.

Deviations from the Guidelines: moving items to other places in a manuscript

When we first began the process of revising SQUIRE 1.0, we noted that there was disagreement about whether certain items belonged in the methods or results section of a manuscript. A principal reason for this is that some felt iterations of an intervention were a method, and others felt it was a result. The consequence was that in using the Guidelines people would move items to the results section that the developers had intended to be reported in the methods section, and vice versa.

In SQUIRE 1.6, the practice of moving items from one section to another in the writing process continued. Nineteen (43.2%) respondents included items from another section(s) in their writing sample, most commonly items were drawn from the methods (n=12) and introduction (n=8) sections. Among the introduction items 'Nature and severity of the local problem, and its context' (n=6) and 'Specific aim of the improvement and intent of this report' (n=5) were moved most often. Among the methods section, items 'Qualitative and quantitative methods used to draw inferences from the data on efficacy and understand the variation' (n=6) and 'The logic on which the improvement was based, including the mechanism by which the improvement was expected to work' (n=5) were moved most often.

In general, items were moved most frequently to the results section. The item related to the nature of the problem was most commonly moved to the results section (three out of six times) followed by the methods (two out of six times). The item related to the qualitative and quantitative methods was most commonly moved to the results section (five out of six times). The item related to logic was also most commonly included in the results section (three out of five times). The item related to specific aim was most often included in the methods section (four out of five times). The most common reasons that an item was moved from another section was because it was the writer's 'usual place', followed by 'it made sense to the writing flow' and, lastly, for some 'other' reason (which we did not require they specify).

Respondents' interpretation of key SQUIRE concepts

The concept of 'context'

Forty-two people answered the open-ended survey item "Write a sentence that describes your interpretation of what is being asked for in the methods section item: describe the context elements that influenced the improvement, and the reasons these elements were considered important". 'Context' was intended by the developers to encompass the factors thought likely to affect the development of the intervention, iterations of the intervention and its success or failure. Among respondents, the meaning of the word context, as it applies to SQUIRE and the intention of the Guidelines, was generally congruent with the intention of the developers, with some variation in depth of response. The meaning of short or long answers should be interpreted with caution since the question was open-ended with no minimum or maximum response required.

The simplest recorded answer given for the survey item was 'key factors' by one respondent. The most complex list included 10 features of context that ranged from 'type of business or activity' to 'competencies' to 'traditions'.

The concept of 'the logic on which the improvement was based'

Forty-two people answered the open-ended survey item that asked respondents to provide their interpretation of the following Guideline item 'logic upon which the improvement was based, mechanism by which it was expected to work'. The intention was for authors to describe the theory or rationale that underpinned the work.

Consistent with the usage data reported above, the concept of 'the logic on which the improvement was based' was not always interpreted as intended by the developers of SQUIRE. The developers intended this item to specify a rationale, theoretical framework,

logic model, mechanism or other explanation that served as a 'reason giving' for why it was thought the intervention would work.[12]

Five respondents interpreted the item to mean the improvement method or improvement theory used for the work (ie, lean, model for improvement), which was not the intended meaning of the item. A few others stated it should be the theory of change used to guide the work. Other answers included "the theoretical and empirical basis for the improvement…", "key drivers to accomplish the project's aim…", "sequence of steps that are expected to lead to the outcomes of interest or the basis on which we think the improvement will be effective", and "…your rationale for choosing to do an improvement project to make this change… your hypothesis about what would be the most impactful drivers of change". All of these latter answers capture some (but not all) of the qualities of what was intended by the developers.

DISCUSSION

In a novel approach to publication guideline revision, 44 people road tested a draft version of the SQUIRE Guidelines (called SQUIRE 1.6) and answered survey questions designed to assess the state of the field. The results revealed the areas most challenging and confusing in writing about improvement work. The data revealed areas of SQUIRE requiring further development, clarification and educational support. The findings were incorporated into the final form of the Guidelines, called SQUIRE 2.0, to be released in the fall of 2015.

Findings and interpretation

We found that even among this engaged group of authors there was lower usage of, and more variable interpretation of, SQUIRE 1.6 items that are relatively specific to healthcare improvement work. The first instance was found in the 'Evolution of the improvement' item, meant to be fulfilled by providing descriptions of how interventions changed or were implemented over time. Most manuscript results and methods sections did not include descriptions of any iterations of the intervention. This is not surprising given that the evaluation of SQUIRE 1.0 revealed debate about whether iterations of improvement interventions should be described.

In the second instance, while participants interpreted the open-ended question about the meaning of the word 'context' similarly to the developers of the Guideline, they did not employ the context item as hoped. Participants described the setting of their work but not often the elements that influenced its development, iterations/implementation steps, success or failure—the item intent.

The last instance comes from the use of the two items 'the logic on which the improvement was based, including mechanism by which improvement expected to work' and 'process and outcome measures used for the improvement, including rationale for choice…'. The open-ended question about the logic on which the improvement was based elicited a range of answers; some were inconsistent with the developers' intent, while others were not complete. Neither the logic item nor the process and outcome measures items—which are somewhat interrelated—were generally used as intended. These findings are also consistent with the findings in the evaluation of SQUIRE 1.0; participants in that study identified these as difficult concepts. The above-noted items will need explicit explanation and instruction in SQUIRE 2.0

We hypothesise that issues with usage and interpretation of items were primarily a result of the ongoing evolution of the field, as it draws from and expands upon multiple areas of knowledge.[12–17] SQUIRE 2.0 could help by explicitly pointing people to the developing resources. However, it is also true that the concepts associated with some of the SQUIRE items are difficult and necessitate wide-ranging knowledge that will be new for many who are joining the field from either clinical practice or health administration. Some require in-depth understanding of issues of establishing internal validity in research design (items 10a–10e in the SQUIRE 1.0 Guidelines; and the items under the section labelled 'improvement' in SQUIRE 1.6), which may not be familiar to people not well grounded in a clinical research background. Other concepts come from the social sciences (eg, the 'logic and mechanisms' item in SQUIRE 1.6) and would likely be unfamiliar to people without a firm background in implementation or public health intervention development.

Significance

Our findings show that writing scholarly healthcare improvement work requires a specific knowledge base, and this knowledge is not universally held. We now know some of the specific gaps that should be addressed to help SQUIRE 2.0 reach its goal of improving the reporting of improvement work. The findings should be helpful not just for the development of SQUIRE 2.0 but also in the education of the next generation, for whom exposure to improvement work is now becoming standard.

These data are also useful because we are learning that the mere presence of publication guidelines to support reporting of scientific studies does not necessarily by itself result in improved reporting. Focusing on the effects of publication guidelines for trials (the Consolidated Standards of Reporting Trials[18]) and for observational trials (Strengthening the Reporting of Observational studies in Epidemiology[19]), the effects of the Guidelines on the quality of reporting have been mixed. Papers from the early to mid-2000s reported that the introduction of publication guidelines increased the quality of the published literature,[20 21] and that the adoption of guidelines by a

journal increased quality of reporting.[22 23] However, a more recent paper suggests that secular trends in the quality of reporting overall could be confounding these findings, and more controlled trials would be required to determine the true effect of the Guidelines.[24] Further, some others have not noted a change in quality of reporting when journals adopted publication guidelines.[25] In addition, a follow-up study to one of the earlier papers that described improvements in randomised trial reporting between 1975 and 2000 noted no further improvements in the quality of reporting between 2001 and 2010.[26] From these findings, we infer that publication guidelines probably help improve reporting, but their dissemination period is lengthy and possibly prone to stalling out or failing if people working in the field lack the knowledge to implement them as intended. SQUIRE 2.0 would do well to provide many pathways for education and dissemination both at the time of release and in the future.

Limitations

Publication guidelines for the healthcare field are most often developed by consensus panels of experts in the methodology for which the guideline is intended, such as that for meta-analyses and cohort studies. An advantage to this is that presumably the most knowledgeable and methodologically advanced thinkers are guiding the field. A potential limitation to our study is the utility of data we obtained during this phase of our development process—the sample is from a self-selected group and the respondents had various levels of expertise in healthcare improvement. Can these data really be used to help formulate a document intended to help pull the field forward? We think the qualities of the respondents make our results quite useful. If people interested enough in the work and writing of healthcare improvement to participate in a study such as this reveal results such as shown here, then the findings in the rest of the field are only likely to be more marked. Asking potential users of a revised Guideline to trial it is unique, to our knowledge, and vital in a field like this that is still maturing. The road testing and survey have provided valuable input that could not have been obtained from experts alone. Understanding the areas of confusion and controversy are crucial if we are to have a publication guideline that achieves its stated goal of improving scientific rigour in reporting. In the evaluation of SQUIRE 1.0, which informed the development of SQUIRE 1.6, we involved thinkers and experts from across the field of healthcare improvement, so we feel that the advantages of both approaches are well represented in the overall evaluation and revision process.

CONCLUSIONS

A key purpose of the SQUIRE Guidelines—both now and when they were first developed—is to advance the science of healthcare improvement and increase the breadth, frequency and quality of published reports of healthcare improvement by encouraging and guiding authors.[27] Respondents who road tested SQUIRE 1.6 showed us what work will lie ahead. To bridge the gap between the various sciences that are contributing to the scholarly work of improving healthcare, SQUIRE 2.0 will need to provide explanations that facilitate comfort and skill in using the less familiar and still evolving SQUIRE items. SQUIRE 2.0 items that are unique to healthcare improvement work will likely require clarification and explanation so that they are interpreted and applied as intended by the guideline developers. Lastly, as evidenced by the growth in the field since SQUIRE 1.0 was released, SQUIRE 2.0 will need to leave room for the further development of the science of improvement.

Twitter Follow Louise Davies at @louisedaviesmd

Acknowledgements The SQUIRE development team is deeply grateful to the authors who tested SQUIRE 1.6 for this study. The time they took to participate was an extra layer of work on top of the already hard work of writing for scientific publication. The data they provided were invaluable to the revision process and will inform future efforts to support authors writing about healthcare improvement work.

Contributors LD: conception and design, data acquisition, analysis, drafting of the manuscript and obtaining funding. KZD: design, analysis, interpretation of data, drafting of the manuscript and critical revision of manuscript. DJG: design, interpretation of data and critical revision of manuscript. GO: conception and design, critical revision of manuscript and obtaining funding.

Funding The Robert Wood Johnson Foundation (grant number 70024), The UK Health Foundation (grant number 7099).

Competing interests None declared.

Ethics approval This research study was reviewed and approved as 'human subjects research, exempt from further review' by the Committee for the Protection of Human Subjects at Dartmouth College.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement We can provide de-identified raw data to persons who request it.

REFERENCES

1 Davidoff F, Batalden P, Stevens D, *et al*. Publication guidelines for quality improvement in health care: evolution of the SQUIRE project. *Qual Saf Health Care* 2008;17(Suppl 1):i3–9.
2 Davidoff F, Batalden P. Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 2005;14:319–25.
3 Moss F, Thompson R. A new structure for quality improvement reports. *Qual Health Care* 1999;8:76.
4 Stevens DP, Thomson R. SQUIRE arrives—with a plan for its own improvement. *Qual Saf Health Care* 2008;17(Suppl 1):i1–2.

5 Batalden P, Leach D, Swing S, *et al*. General competencies and accreditation in graduate medical education. *Health Aff (Millwood)* 2002;21:103–11.
6 World federation for medical education global standards for quality improvement in medical education, European specifications for basic and post graduate medical education and continuing professional development. Copenhagen, Denmark: University of Copenhagen, 2007.
7 Marshall M, Pronovost P, Dixon-Woods M. Promotion of improvement as a science. *Lancet* 2013;381:419–21.
8 Wensing M, Grimshaw JM, Eccles MP. Does the world need a scientific society for research on how to improve healthcare? *Implement Sci* 2012;7:10.
9 Davies L, Batalden P, Davidoff F, *et al*. The SQUIRE Guidelines: an evaluation from the field, five years post release. *BMJ Qual Saf* 2015;24:769–75.
10 Moher D, Schulz KF, Simera I, *et al*. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.
11 Krippendorff K. *Content analysis an introduction to its methodology*. 3rd edn. Thousand Oaks, CA: Sage, 2013.
12 Davidoff F, Dixon-Woods M, Leviton L, *et al*. Demystifying theory and its use in improvement. *BMJ Qual Saf* 2015;24:228–38.
13 Kaplan HC, Provost LP, Froehle CM, *et al*. The Model for Understanding Success in Quality (MUSIQ): building a theory of context in healthcare quality improvement. *BMJ Qual Saf* 2012;21:13–20.
14 Taylor SL, Dy S, Foy R, *et al*. What context features might be important determinants of the effectiveness of patient safety practice interventions? *BMJ Qual Saf* 2011;20:611–17.
15 Hoffmann TC, Glasziou PP, Boutron I, *et al*. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;348:g1687.
16 Portela MC, Pronovost PJ, Woodcock T, *et al*. How to study improvement interventions: a brief overview of possible study types. *BMJ Qual Saf* 2015;24:325–36.
17 Batalden P, Davidoff F, Marshall M, *et al*. So what? Now what? Exploring, understanding and using the epistemologies that inform the improvement of healthcare. *BMJ Qual Saf* 2011;20(Suppl 1):i99–105.
18 Schulz KF, Altman DG, Moher D, *et al*. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010;63:834–40.
19 von Elm E, Altman DG, Egger M, *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology* 2007;18:800–4.
20 Moher D, Jones A, Lepage L, *et al*. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA* 2001;285:1992–5.
21 Latronico N, Botteri M, Minelli C, *et al*. Quality of reporting of randomised controlled trials in the intensive care literature. A systematic analysis of papers published in Intensive Care Medicine over 26 years. *Intensive Care Med* 2002;28:1316–23.
22 Smith BA, Lee HJ, Lee JH, *et al*. Quality of reporting randomized controlled trials (RCTs) in the nursing literature: application of the consolidated standards of reporting trials (CONSORT). *Nurs Outlook* 2008;56:31–7.
23 Plint AC, Moher D, Morrison A, *et al*. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust* 2006;185:263–7.
24 Bastuji-Garin S, Sbidian E, Gaudy-Marqueste C, *et al*. Impact of STROBE statement publication on quality of observational study reporting: interrupted time series versus before-after analysis. *PLoS ONE* 2013;8:e64733.
25 Sinha S, Sinha S, Ashby E, *et al*. Quality of reporting in randomized trials published in high-quality surgical journals. *J Am Coll Surg* 2009;209:565–71 e1.
26 Latronico N, Metelli M, Turin M, *et al*. Quality of reporting of randomized controlled trials published in Intensive Care Medicine from 2001 to 2010. *Intensive Care Med* 2013;39:1386–95.
27 Huber AJ, Ogrinc G, Davidoff F. SQUIRE (Standards for Quality Improvement Reporting Excellence). In: Moher D, Altman D, Schulz K, *et al*., ed. *Guidelines for reporting health research: a user's manual*. London, England: Wiley Blackwell, 2014:227–40.