Completeness of reporting of quality improvement studies in neonatology is inadequate: a systematic literature survey

Zheng Jing Hu ,[1] Gerhard Fusch,[2] Catherine Hu,[3] Jie Yi Wang,[4] Zoe el Helou,[5] Muhammad Taaha Hassan,[5] Lawrence Mbuagbaw ,[1] Salhab el Helou,[2] Lehana Thabane[1]

ABSTRACT

Introduction Quality improvement (QI) is a growing field of inquiry in healthcare, but the reporting quality of QI studies in neonatology remains unclear. We conducted a systematic survey of the literature to assess the reporting quality of QI studies and factors associated with reporting quality.

Methods We searched Medline for publications of QI studies from 2016 to 16 April 2020. Pairs of reviewers independently screened citations and assessed reporting quality using a 31-item modified Standards for Quality Improvement Reporting Excellence, 2nd edition (SQUIRE 2.0) checklist. We reported the number (percentage) of studies that reported each item and their corresponding 95% CIs. We used Poisson regression to explore factors associated with reporting quality, namely, journal endorsement of SQUIRE 2.0, declaration of funding sources, year of publication and number of authors. The results were reported as incidence rate ratio (IRR) and 95% CI.

Results Of 1921 citations, 336 were eligible; among them, we randomly selected 100 articles to assess reporting quality. The mean (standard deviation) number of SQUIRE 2.0 items adhered to was 22.0 (4.5). Percentage of articles reporting each item varied from 26% to 100%. Journal endorsement of SQUIRE 2.0 (IRR=1.11, 95% CI 1.02 to 1.21, p=0.015), declaration of funding sources and increasing number of authors were significantly associated with better reporting.

Conclusions Reporting quality of QI studies in neonatology is inadequate. Endorsing the SQUIRE 2.0 guideline is a step that journals can implement to enhance the completeness of reporting.

INTRODUCTION

Quality improvement (QI) efforts in healthcare have become an increasingly active field of inquiry. QI efforts have been implemented in various health settings with diverse aims, including reducing medical errors, improving patient safety, providing better satisfaction with care, increasing efficiency of healthcare delivery, or training healthcare practitioners to adhere to evidence-based practices.[1 2] QI projects that are reported clearly and rigorously can provide clear evidence of effective activities for improving the quality of care at local health settings, and thus accelerate the dissemination and adaptation of these effective practices.

The current reporting quality guideline for QI studies is the Standards for Quality Improvement Reporting Excellence, 2nd edition (SQUIRE 2.0), published in 2015. SQUIRE 2.0 is intended for any study that report on systematic, data-driven efforts to improve the quality, safety and value of healthcare. SQUIRE 2.0 was developed as an improvement to SQUIRE 1.0, to provide better guidance for authors in writing more clearly, precisely, completely and transparently about QI studies.[3]

Commonly used reporting guidelines such as the Consolidated Standards of Reporting Trials (CONSORT), The Strengthening the Reporting of Observational Studies in Epidemiology or Preferred Reporting Items for Systematic Reviews and Meta-Analyses are intended for specific study designs (ie, randomised controlled trials (RCTs), observational studies and systematic reviews).[4] In contrast, the applicability of SQUIRE 2.0 depends on a study's objectives, rather than its design; namely, that the study sought to report on a systematic effort to improve the quality, safety and value of healthcare at a systems-level. For example, a study that reported on a systematic effort to reduce nosocomial sepsis qualifies for SQUIRE 2.0 evaluation. However, a study that aimed to assess risk factors of nosocomial sepsis, or investigate a specific treatment regimen, would not qualify for SQUIRE 2.0 evaluation, because a systematic improvement effort was absent. In this context, QI studies undoubtedly encompass a diverse range of study designs and methodologies.

Thus, a significant challenge to achieving clear and consistent reporting of QI studies is the large variation in how QI research is conducted and the definition of QI itself.[5–8] QI studies may vary in intervention methodology, such as iterative Plan-Do-Study-Act cycles, Lean six sigma and Total Quality management. QI studies may also differ in study objectives, such as evaluate the success and feasibility implementing evidence-based practice into a local setting, reduce adverse events, or improve healthcare workers' well-being. Each of these study designs and objectives may have their own reporting practices and cater to different stakeholders' priorities. Consequently, one would anticipate a large variation in reporting, and challenges in achieving clarity and consistency when these studies are reported in the general healthcare improvement context. Woods and Martin describe 'insufficient attention to rigorous evaluation of improvement and to sharing the lessons of successes and failures'[9] as a critical barrier to the effectiveness of QI. Altogether, these realities emphasise the need to develop evidence-based inquiry in this field and assess the completeness of reporting of QI studies as part of this process.

A previous study by Howell *et al* showed that the quality of reporting of QI studies did not demonstrate improvement following the publication of SQUIRE 1.0.[10] Otherwise, to our best knowledge, there have not been any studies that evaluated the completeness of reporting of QI studies using SQUIRE 2.0 in the neonatology literature.

In a Neonatal Intensive Care Unit environment, unsafe care such as medication administration errors or inconsistencies in the quality and key processes of care, can lead to neonates' adverse outcomes that incur long-term developmental consequences.[1] QI efforts have been shown to improve outcomes in neonatal care.[2] Thus, clear reporting of these efforts is paramount to facilitate knowledge translation and accelerate progress on the safety and quality of neonatal care.

This literature survey aims to inform clinicians and researchers in neonatology on the current state of reporting quality of QI studies. The primary objective is to assess the published studies' compliance with SQUIRE 2.0. The secondary objective is to identify the characteristics of published articles associated with their quality of reporting.

METHODS
This study was a systematic survey of the literature. In a systematic survey, the literature review is conducted on a random sample of all eligible articles retrieved from a search strategy; whereby the sampling strategy is determined a-priori. Furthermore, the search strategy aims at retrieving a sufficient sample of articles that reflect the research question being addressed.[11] We searched the Medline database for publications from 2016 to 16 April 2020, as defined by the "Year of Publication" field using the search strategy shown in online supplemental appendix A1. Search terms consisted of Medical Subject Headings

(MeSH) keywords pertaining to quality improvement and neonatology. These terms were determined with a librarian and based on previous studies of QI search strategies.[12] The search aimed to find QI studies published after the release of SQUIRE 2.0 in September 2015.[13]

The primary outcome was the overall quality of reporting, as measured by the number of items that published papers adhered to in a modified SQUIRE 2.0 checklist. In addition, we reported the percentage of studies that adhered to each checklist item along with 95%CIs. We modified the SQUIRE 2.0 checklist for our assessment following a pilot testing of SQUIRE 2.0 with 10 articles and discussion with QI expert SeH. Online supplemental appendix A2 details the corresponding SQUIRE 2.0 statement of these items and reasons for modifying them. Thus, this checklist consists of 31 items.

Herein, we modified the SQUIRE 2.0 checklist for our three main reasons. First, the title was excluded because it was considered as inclusion criteria. Second, the abstract was excluded as well because all included studies had an abstract. Third, some items were not universally applicable to all QI reports, and as such, were excluded from our quality of reporting assessment. Finally, reporting items that expressed similar ideas were combined, while single reporting items that consisted of multiple ideas were split into separate items. We made these decisions based on the details of the items' explanation and elaboration, and after pilot-testing the SQUIRE 2.0 checklist on 10 articles. The developers SQUIRE 2.0 themselves also stated that "…some (SQUIRE 2.0) items may not be relevant for inclusion in a particular manuscript".[3] Therefore, we used pilot testing and clinical rationale to decide which items were universally applicable to all QI reports. See online supplemental appendix A2 for a detailed table of specific items modified, our rationale for modifying them, and the implications of these modifications.

We defined QI publications as studies whose primary objective was to test interventions that lead to better patient outcomes, stronger system performance or enhanced professional development. Our exclusion criteria consisted of literature reviews, study protocols, articles not written in English, editorial commentary and studies whose primary focus was not on healthcare improvement. We implemented a two-stage title/abstract screening process due to documented challenges of defining QI and selecting appropriate studies for a QI literature review.[5]

In stage 1, we included all single studies that consisted of an intervention, primary outcome, and a description of how outcomes changed over time, and all studies that explicitly declared their study as a 'quality improvement' or 'quality initiative' effort in the title or abstract. Two pairs of student reviewers (ZeH and MTH, and CH and JYW) screened the title, abstract, and full text where necessary, and disagreements between two reviewers were resolved by ZJH. Online supplemental appendix A4, box 1 provides details on the decision process for determining the inclusion of articles.
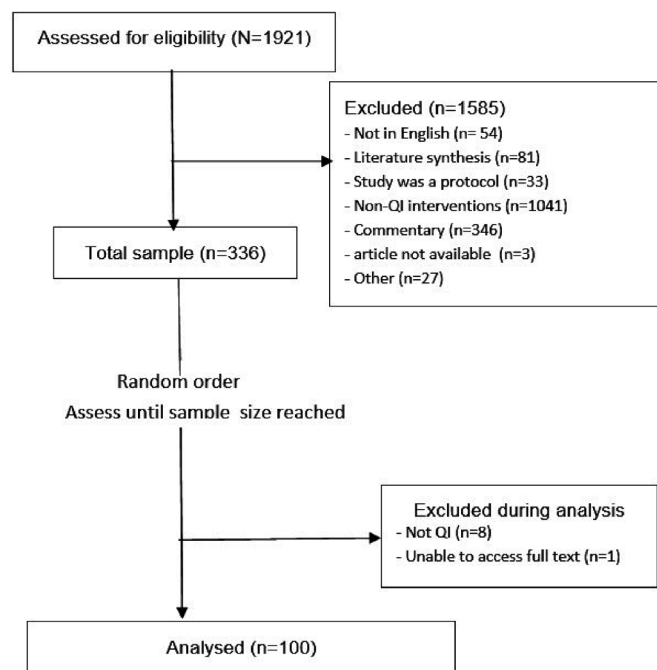
Figure 1 Flow diagram describing the articles' selection process. QI, quality improvement.

Subsequently, the title/abstract of all selected articles were rescreened to determine whether the authors intended to publish their studies as a healthcare improvement effort. Here, we included studies that described intervention(s) to improve a specific process, quality or safety of care. ZJH and QI experts SeH and GF assessed each article's abstract and full-text and reached a consensus on selecting articles for data abstraction. Online supplemental appendix A3 provides a list of attributes that the assessors examined when evaluating a study's eligibility for data abstraction.

We stratified all successfully screened articles by year of publication. Within each stratum, the articles were first sorted randomly using Excel. Subsequently, each article was sampled without replacement in their sorted order, according to the probability that an article belonged to a specific stratum (year). Using this process, we ordered all articles randomly and assessed them in this order until we reached the desired sample size.

We conducted data abstraction on the modified SQUIRE 2.0 checklist. Ten articles were assessed during the initial pilot testing. Each reviewer assessed four articles, and ZJH assessed all 10 articles and compared results. We resolved disagreements through discussion. Subsequent articles were divided between the four student reviewers and assessed independently with ZJH. Disagreements were resolved through discussion, and with a third author (SeH or GF) if needed.

We examined explanatory variables to determine factors associated with the quality of reporting. We determined these factors a-priori based on evidence from previous literature that assessed reporting quality. These factors include: endorsement of SQUIRE 2.0 by the journal, defined as the presence of recommendation or requirement to comply with SQUIRE 2.0 among instructions for authors (ie, Journal endorsement of reporting guideline),[13–16] declaration of funding source, defined as both the presence of a funding section and a declaration of a specific organisation that provided funding,[17 18] year of publication (implying more recent studies have a better quality of reporting),[19–21] and the number of authors listed on a manuscript, excluding organisations or groups as authors.[17 21 22]

Analysis

We summarised the characteristics of the included studies using descriptive statistics with categorical variables reported as frequencies. We summarised continuous variables using median and IQR or mean (SD) where appropriate. We computed the proportion of articles that reported each item, along with their 95% CIs; and we calculated 95% CIs using the Wilson Score method.[23] To explore factors associated with reporting quality, we fitted a Poisson model assuming an identity link to the data for both univariable and multivariable analysis. The incidence rate ratio (IRR) for each factor was reported with 95% CIs and p values. We checked the Poisson model's assumptions by examining the dispersion parameters. We assessed collinearity by examining the correlation matrix between all variables, condition index number and variance inflation factor.

Finally, to explore interrater agreement, we computed Cohen's Kappa[24] for a selected list of SQUIRE 2.0 items. We chose to assess agreement for items that had a more subjective interpretation, where reporting the extent of agreement would be informative. We also chose three items with more objective interpretation as a counterbalance. We excluded data for 26 articles in the agreement assessment, as one reviewer had a particularly low preconsensus agreement. All analyses were conducted using Microsoft Excel and SAS V.9.4.

Sample size

The aim of sample size calculation in this study was to estimate the proportion of articles that reported each item in the modified SQUIRE 2.0 checklist, in line with numerous previous studies that assessed the completeness of reporting.[25–27] Thus, we estimated sample size using the population proportion CI method.[28] The desired margin of error was 10%, and the estimated population proportion was 0.50 based on the conservative approach. Using this approach, we calculated a sample size of 97. As we could assess additional articles, 100 articles were evaluated.

Patient and public involvement

Patients and the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

RESULTS

Our search retrieved 1921 articles. After the selection process, 336 articles remained available for data

abstraction. These articles were then randomly ordered and assessed sequentially until we reached 100 articles. During this process, nine studies were found to be non-QI studies and removed. Figure 1 shows the flow diagram describing the articles' selection process. A description of selected characteristics of publications are shown in table 1.

The proportion of articles that reported each SQUIRE 2.0 item, along with their CIs, are shown in table 2. The mean number of items reported per article was 22.0 (SD=4.5, 95% CI 21.1 to 22.9). The most frequently reported item was the 'Name and significance of the local problem', reported by all articles assessed. The least frequently reported item was 'Details about missing data', reported among only 26% of articles. In general, items in the background section have good reporting and most of the methods section. Reporting was suboptimal on items that described processes of care (ie, QI-specific reporting items), such as contextual elements, the effect of time as a variable and impact of the project on people and systems.

Table 3 shows the IRR in the number of SQUIRE 2.0 items reported, for each factor assessed. Articles published in journals that endorsed SQUIRE 2.0, declared funding sources, and publications with a larger number of authors were all positively associated with a larger number of SQUIRE 2.0 items reported. Articles published more recently had a positive association with the quality of reporting, though it did not achieve statistical significance.

The inter-rater reliability was poor, with Kappa values ranging from 0 to 0.64. Table 4 lists the level of agreement for selected SQUIRE 2.0 items. See online supplemental appendix A4, box 2, for further details on the reasons for preconsensus disagreements between reviewers.

DISCUSSION

The reporting quality of QI studies in neonatology published since the inception of SQUIRE 2.0 is inadequate. The mean (SD) number of items reported was 22.0 (4.5), out of 31 possible items. Factors positively associated with the quality of reporting were the endorsement of SQUIRE 2.0 by the publishing journal, declaration of funding sources and a greater number of authors.

Previous studies that assessed reporting quality of RCTs in various clinical specialties have also found inadequate reporting quality. The overall reporting quality of this study was better compared with previous studies. The current study found a larger average number of reporting items reported per article, compared with other clinical specialties.[22 25 29–32] This variation may be attributed to several factors. First, the publications examined in this literature survey were published more recently compared with previous studies. Second, SQUIRE 2.0 items' wording allows for a broader range of acceptable responses to meet reporting requirements. For instance, SQUIRE 2.0 asks for 'details of the process measures and outcome'[8] when reporting outcome and process measures. In

Table 1 Descriptive characteristics of studies

Study characteristics		Frequency (N=100)
Study design	Continuous QI-iterative improvement	31
	Pre–post; before–after	57
	Pre–post and continuous combined	7
	Non-randomised comparison trial	3
	Randomised controlled trial	1
	Uncategorised	1
Target population	Patients	70
	Caregivers	6
	Healthcare providers	11
	Community	6
	Mother–infant dyad	5
	Other domains	2
Country/region of publication	USA	61
	Americas except USA	4
	Europe	25
	Africa	0
	Asia and Australia	10
Year	2020	10
	2019	30
	2018	26
	2017	19
	2016	15
QI framework used	Model for improvement/ Plan-Do-Study-Act	34
	Other frameworks	12
	None	54
Reporting guideline followed	SQUIRE 2.0	9
	CONSORT	1
	None	90
Sample size	Sample size calculated or justified	14
Journal endorses SQUIRE 2.0—required or recommended in authors' guidelines		51
Declared funding sources (yes)		36
Number of authors, mean (SD)		6.63 (3.1)
Study stated set specific, measurable and time-defined goals		31

CONSORT, Consolidated Standards of Reporting Trials; QI, quality improvement; SQUIRE, Standards for Quality Improvement Reporting Excellence.

comparison, CONSORT required much greater details for reporting outcome measures, including the primary outcome, secondary outcome, effect size, precision and

Table 2 Frequency of reporting for each item in the modified SQUIRE 2.0 checklist

SQUIRE 2.0 item	Frequency	Proportion (95% CI)
Background		
Nature and significance of the local problem	100	1.00 (0.96 to 1.00)
Summary of what is currently known about the problem, including relevant previous studies	98	0.98 (0.93 to 0.99)
Informal or formal frameworks, models, concepts, and/or theories used to explain the problem, any reasons or assumptions that were used to develop the intervention(s), and reasons why the intervention(s) was expected to work	80	0.80 (0.71 to 0.87)
Purpose of the project and of this report	93	0.93 (0.86 to 0.97)
Methods		
Contextual elements considered important at the outset of introducing the intervention(s)	92	0.92 (0.85 to 0.96)
Description of the intervention(s) in sufficient detail that others could reproduce it	96	0.96 (0.90 to 0.98)
Specifics of the team involved in the work	88	0.88 (0.80 to 0.93)
Approach chosen for assessing the impact of the intervention(s)	86	0.86 (0.78 to 0.91)
Measures chosen for studying processes and outcomes of the intervention(s), including rationale for choosing them, their operational definitions, and their validity and reliability (outcome)	82	0.82 (0.73 to 0.88)
Measures chosen for studying processes and outcomes of the intervention(s), including rationale for choosing them, their operational definitions, and their validity and reliability (process)	63	0.63 (0.53 to 0.72)
Description of the approach to the ongoing assessment of contextual elements that contributed to the success, failure, efficiency, and cost	49	0.49 (0.39 to 0.59)
Qualitative and quantitative methods used to draw inferences from the data	69	0.69 (0.59 to 0.77)
Methods for understanding variation within the data, including the effects of time as a variable	48	0.48 (0.38 to 0.58)
Methods employed for assessing completeness and accuracy of data	54	0.54 (0.44 to 0.63)
Ethical aspects of implementing and studying the intervention(s) and how they were addressed, including, but not limited to, formal ethics review and potential conflict(s) of interest	89	0.89 (0.81 to 0.94)
Results		
Initial steps of the intervention(s) and their evolution over time (eg, time-line diagram, flow chart or table), including modifications made to the intervention during the project	56	0.56 (0.46 to 0.65)
Details of the process measures and outcome (outcome)	97	0.97 (0.92 to 0.99)
Details of the process measures and outcome (process)	73	0.73 (0.64 to 0.81)
Contextual elements that interacted with the intervention(s)	29	0.29 (0.21 to 0.39)
Unintended consequences such as unexpected benefits, problems, failures, or costs associated with the intervention(s).	43	0.43 (0.34 to 0.53)
Details about missing data	26	0.26 (0.18 to 0.35)
Discussion and conclusion		
Key findings, including relevance to the rationale and specific aims	97	0.97 (0.92 to 0.99)
Particular strengths of the project	60	0.60 (0.50 to 0.69)
Nature of the association between the intervention(s) and the outcomes	83	0.83 (0.74 to 0.89)
Comparison of results with findings from other publications	80	0.80 (0.71 to 0.87)
Impact of project on people and systems	50	0.50 (0.40 to 0.60)
Reasons for any differences between observed and anticipated outcomes, including the influence of context	59	0.59 (0.49 to 0.68)

Table 2 Continued

SQUIRE 2.0 item	Frequency	Proportion (95% CI)
1. Limits to the generalisability of the work 2. Factors that might have limited internal validity such as confounding, bias, or imprecision in the design, methods, measurement, or analysis	92	0.92 (0.85 to 0.96)
Efforts made to minimise and adjust for limitations	44	0.44 (0.35 to 0.54)
Usefulness of the work	71	0.71 (0.61 to 0.79)
Sustainability	52	0.52 (0.42 to 0.62)

SQUIRE, Standards for Quality Improvement Reporting Excellence.

absolute and relative effect sizes if the outcome measure is binary.[33] Similar 'loose requirements' can be found for numerous items in SQUIRE 2.0.

Thus, a critical methodological limitation of using SQUIRE 2.0 for assessing reporting quality is the absence and subjectivity of information on how much details should be described for specific important items to satisfy the reporting criteria. Consequently, assessment of reporting quality can differ considerably between reviewers. Furthermore, some poorly written articles can still receive a high SQUIRE 2.0 score. A second limitation is the difficulty of assessing the presence of QI expertise in an authorship team. Practitioners who conduct QI projects specialise in diverse academic disciplines, and may lack a comprehensive understanding of QI. Thus, one would anticipate that the presence of a QI expert in the authorship team would improve reporting or ensure that the manuscript is more QI-focused. However, ascertaining whether an author is a QI expert may involve an extensive web search of their institutional affiliations, publication activities and curriculum vitae (if available online). Furthermore, the manuscript may not indicate the extent of involvement of the QI expert in the study. Hence, we could not assess authors' QI expertise as a factor for influencing the quality of reporting.

The subjective description of SQUIRE 2.0 items also impeded inter-rater reliability. In this study, preconsensus agreement between reviewers was poor. The Kappa statistic for various selected items ranged from 0 to 0.64. In comparison, previous RCT reporting quality studies on emergency medicine, brain tumour RCTs, and heart failure showed Kappa ranging from 0.80 to 0.90.[18 26 34] However, even among QI experts, agreement on whether published papers met specific criteria was suboptimal, with a Kappa of 0.52.[35]

We examined post-hoc whether published studies primarily adhered to other reporting guidelines, and if this may have affected their adherence to SQUIRE 2.0. However, the results show that only one study adhered to a guideline besides SQUIRE 2.0. Most studies did not indicate following any reporting guideline and the rest adhered to SQUIRE 2.0 itself. Thus, we were unable to make any conclusions in this regard.

Presently, journals should endorse the SQUIRE 2.0 guideline by recommending its use or mandating its adherence to improve the completeness of reporting of QI studies. Nonetheless, improving published studies' adherence to SQUIRE 2.0 alone is not sufficient for publishing well-written QI reports. Ultimately, SQUIRE 2.0 was intended to provide general reporting guidance for authors who are interested in publishing their QI efforts. As such, SQUIRE 2.0 has limitations for assessing important methodological aspects of reporting regarding specific QI methodologies. Future work can include conducting a critical appraisal of QI publications, assessing the reporting of the interventions themselves, or assessing the association between SQUIRE 2.0 adherence, requirement, and endorsement on the quality of QI evidence. Both assessments would provide valuable insights into the reporting of methodological aspects of QI studies in neonatology.

Table 3 Univariable and multivariable analysis of factors associated with the number of items reported in each article

Attribute	Category	Univariable analysis		Multivariable analysis	
		Incidence rate ratio (95% CI)	P value	Incidence rate ratio (95% CI)	P value
Journal endorses SQUIRE 2.0	Yes	1.09 (1.01 to 1.19)	0.035	1.11 (1.02 to 1.21)	0.015
	No	REF		REF	
Declared funding sources	Yes	1.14 (1.04 to 1.24)	0.003	1.11 (1.02 to 1.21)	0.021
	No	REF		REF	
Year		1.02 (0.99 to 1.06)	0.221	1.02 (0.98 to 1.05)	0.340
Number of authors		1.02 (1.01 to 1.04)	0.001	1.02 (1.01 to 1.03)	0.003

SQUIRE, Standards for Quality Improvement Reporting Excellence; REF, reference level.

Table 4 Assessment of reporting agreement on SQUIRE items, preconsensus

Items	Percent Agreement	Kappa	95% CI (Kappa)
Rationale: What QI framework, model, concept, or theory is being used to address the problem?	0.63	0.02	−0.21 to 0.25
Description of QI team involved	0.79	0.32	0.06 to 0.59
Qualitative and quantitative methods used to draw inferences from the data	0.69	0.31	0.09 to 0.54
Methods for understanding variation within the data, including the effects of time as a variable	0.68	0.36	0.13 to 0.57
Process measures stated	0.67	0.32	0.11 to 0.53
Measures for *assessing contextual factors* (that contributed to the success, failure, efficiency or cost of intervention)	0.54	0.07	−0.14 to 0.29
Initial steps of the intervention(s) and their evolution over time (eg, time-line diagram, flow chart or table), including modifications made to the intervention during the project	0.82	0.64	0.47 to 0.82
Process measures evaluated	0.65	0.21	−0.02 to 0.45
Contextual factors analysed	0.81	0.39	0.13 to 0.64
Limitations to internal and external validity	0.92	0.58	0.28 to 0.88
Impact of project on people and systems	0.61	0.21	−0.02 to 0.43
Usefulness of work	0.63	0.17	−0.02 to 0.35

*26 articles removed due to low agreement with one individual.
QI, quality improvement.

CONCLUSION

Overall, the quality of reporting of QI studies in neonatology is inadequate. Although SQUIRE 2.0 serves as a suitable guideline for reporting QI efforts clearly, its ability to assess reporting of key methodological details is limited. Future studies examining the reporting of QI methodologies, or the relationship between SQUIRE 2.0 adherence and strength of evidence, would inform a better understanding of QI reporting and how it can be improved.

Author affiliations
[1]Department of Health Research Methods Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
[2]Division of Neonatology, Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada
[3]Bachelor of Arts and Science, McMaster University, Hamilton, Ontario, Canada
[4]Bachelor of Medical Sciences, Schulich School of Medicine & Dentistry, University of Western Ontario, London, Ontario, Canada
[5]Bachelor of Health Sciences, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Zheng Jing Hu http://orcid.org/0000-0002-2215-3353
Lawrence Mbuagbaw http://orcid.org/0000-0001-5855-5461

REFERENCES

1 Lachman P, Jayadev A, Rahi M. The case for quality improvement in the neonatal intensive care unit. *Early Hum Dev* 2014;90:719–23.
2 Ellsbury DL, Clark RH. Does quality improvement work in neonatology improve clinical outcomes? *Curr Opin Pediatr* 2017;29:129–34.
3 Goodman D, Ogrinc G, Davies L, *et al*. Explanation and elaboration of the Squire (standards for quality improvement reporting excellence) guidelines, V.2.0: examples of Squire elements in the healthcare improvement literature. *BMJ Qual Saf* 2016;25:e7.
4 Ogrinc G, Davies L, Goodman D, *et al*. Squire 2.0 (standards for quality improvement reporting excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf* 2016;25:986–92.
5 Jones E, Furnival J, Carter W. Identifying and resolving the frustrations of reviewing the improvement literature: the experiences of two improvement researchers. *BMJ Open Qual* 2019;8:e000701.

6   Jones EL, Dixon-Woods M, Martin GP. Why is reporting quality improvement so hard? A qualitative study in perioperative care. *BMJ Open* 2019;9:e030269.

7   Leviton L. Reconciling complexity and classification in quality improvement research. *BMJ Qual Saf* 2011;20(Suppl 1):i28–9.

8   BMJ Quality & Safety. How to study improvement interventions: a brief overview of possible study types. Available: https://qualitysafety-bmj-com.libaccess.lib.mcmaster.ca/content/24/5/325 [Accessed 13 Sep 2020].

9   Dixon-Woods M, Martin GP. Does quality improvement improve quality? *Future Hosp J* 2016;3:191–4.

10 Howell V, Schwartz AE, O'Leary JD, *et al*. The effect of the Squire (standards of quality improvement reporting excellence) guidelines on reporting standards in the quality improvement literature: a before-and-after study. *BMJ Qual Saf* 2015;24:400–6.

11 Mbuagbaw L, Lawson DO, Puljak L, *et al*. A tutorial on methodological studies: the what, when, how and why. *BMC Med Res Methodol* 2020;20:226.

12 Hempel S, Rubenstein LV, Shanman RM, *et al*. Identifying quality improvement intervention publications--a comparison of electronic search strategies. *Implement Sci* 2011;6:85.

13 Stevens A, Shamseer L, Weinstein E, *et al*. Relation of completeness of reporting of health research to journals' endorsement of reporting guidelines: systematic review. *BMJ* 2014;348:g3804.

14 Panic N, Leoncini E, de Belvis G, *et al*. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PLoS One* 2013;8:e83138.

15 Liu LQ, Morris PJ, Pengel LHM. Compliance to the CONSORT statement of randomized controlled trials in solid organ transplantation: a 3-year overview. *Transpl Int* 2013;26:300–6.

16 Ethgen M, Boutron L, Steg PG, *et al*. Quality of reporting internal and external validity data from randomized controlled trials evaluating stents for percutaneous coronary intervention. *BMC Med Res Methodol* 2009;9:24.

17 Balasubramanian SP, Wiener M, Alshameeri Z, *et al*. Standards of reporting of randomized controlled trials in general surgery: can we do better? *Ann Surg* 2006;244:663–7.

18 Borg Debono V, Zhang S, Ye C, *et al*. The quality of reporting of RCTs used within a postoperative pain management meta-analysis, using the CONSORT statement. *BMC Anesthesiol* 2012;12:13.

19 Chowers MY, Gottesman BS, Leibovici L, *et al*. Reporting of adverse events in randomized controlled trials of highly active antiretroviral therapy: systematic review. *J Antimicrob Chemother* 2009;64:239–50.

20 Farrokhyar F, Chu R, Whitlock R, *et al*. A systematic review of the quality of publications reporting coronary artery bypass grafting trials. *Can J Surg* 2007;50:266–77.

21 Zheng SL, Chan FT, Maclean E, *et al*. Reporting trends of randomised controlled trials in heart failure with preserved ejection fraction: a systematic review. *Open Heart* 2016;3:e000449.

22 Yao AC, Khajuria A, Camm CF, *et al*. The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. *Eye* 2014;28:1341–9.

23 7.2.4.1. confidence intervals. Available: https://www.itl.nist.gov/div898/handbook/prc/section2/prc241.htm [Accessed 26 Sep 2020].

24 McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012;22:276–82.

25 Shanthanna H, Kaushal A, Mbuagbaw L, *et al*. A cross-sectional study of the reporting quality of pilot or feasibility trials in high-impact anesthesia journals. *Can J Anaesth* 2018;65:1180–95.

26 Lai R, Chu R, Fraumeni M, *et al*. Quality of randomized controlled trials reporting in the primary treatment of brain tumors. *J Clin Oncol* 2006;24:1136–44.

27 Isiguzo GC, Zunza M, Chirehwa M, *et al*. Quality of pilot trial Abstracts in heart failure is suboptimal: a systematic survey. *Pilot Feasibility Stud* 2018;4:107.

28 5.3.3 - Sample Size Computation | STAT 500. PennState: Statistics Online Courses. Available: https://online.stat.psu.edu/stat500/lesson/5/5.3/5.3.3 [Accessed 29 Mar 2021].

29 Kosa SD, Monize J, Leenus A, *et al*. Reporting quality of pilot clinical trials in chronic kidney disease patients on hemodialysis: a methodological survey. *Pilot Feasibility Stud* 2019;5:53.

30 Thabane L, Thomas T, Ye C, *et al*. Posing the research question: not so simple. *J Can Anesth* 2009;56:71–9.

31 Adie S, Harris IA, Naylor JM, *et al*. Consort compliance in surgical randomized trials: are we there yet? A systematic review. *Ann Surg* 2013;258:872–8.

32 Marshman Z, Farid F. The quality of reporting of randomised controlled trials in dental public health. *Community Dent Health* 2010;27:253–6.

33 Moher D, Hopewell S, Schulz KF, *et al*. Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012;10:28–55.

34 Germini F, Marcucci M, Fedele M, *et al*. Quality of reporting in Abstracts of RCTs published in emergency medicine journals: a systematic survey of the literature suggests we can do better. *Emerg Med J* 2019;660–5.

35 Hempel S, Shekelle PG, Liu JL, *et al*. Development of the quality improvement minimum quality criteria set (QI-MQCS): a tool for critical appraisal of quality improvement intervention publications. *BMJ Qual Saf* 2015;24:796–804.